

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

Парилина Елена Михайловна

**РЕШЕНИЯ КООПЕРАТИВНЫХ СТОХАСТИЧЕСКИХ
ИГР С ТРАНСФЕРАБЕЛЬНЫМИ ВЫИГРЫШАМИ**

Том 1

Специальность 01.01.09 —

Дискретная математика и математическая кибернетика

Диссертация

на соискание ученой степени

доктора физико-математических наук

Научный консультант:

доктор физико-математических наук,

профессор Петросян Леон Аганесович

Санкт-Петербург

2018

Оглавление

Введение	7
Глава 1. Кооперативные стохастические игры конечной продолжительности	29
§ 1.1 Некооперативные стохастические игры конечной продолжительности	29
§ 1.2 Основные функциональные уравнения для стохастических игр конечной продолжительности	32
§ 1.3 Определение кооперативной стохастической игры конечной продолжительности	34
§ 1.4 Вектор Шепли, s -ядро и n -ядро	39
§ 1.5 Процедура распределения дележа в стохастических играх конечной продолжительности	41
§ 1.6 Позиционная состоятельность решения кооперативной стохастической игры конечной продолжительности	43
§ 1.7 Неотрицательность компонент процедуры распределения дележа в стохастических играх конечной продолжительности. Регуляризация дележей	47
§ 1.8 Регуляризация вектора Шепли и s -ядра в стохастических играх конечной продолжительности	51
§ 1.9 Сильная позиционная состоятельность s -ядра в стохастических играх конечной продолжительности	66

Глава 2. Кооперативные стохастические игры бесконечной продолжительности	77
§ 2.1 Некооперативные стохастические игры бесконечной продолжительности	77
§ 2.2 Основные функциональные уравнения для стохастических игр бесконечной продолжительности в стационарных стратегиях	81
§ 2.3 Определение кооперативной стохастической игры бесконечной продолжительности	82
§ 2.4 Принципы устойчивой кооперации в стохастических играх бесконечной продолжительности	86
§ 2.4.1 Позиционная состоятельность принципа оптимальности в стохастических играх бесконечной продолжительности	87
§ 2.4.2 Стратегическая устойчивость кооперативного решения в стохастических играх бесконечной продолжительности	92
§ 2.4.3 Защита от иррационального поведения кооперативного решения в стохастических играх бесконечной продолжительности ...	99
§ 2.5 Существование кооперативного решения в стохастических играх бесконечной продолжительности, удовлетворяющего принципам устойчивой кооперации	102
§ 2.6 Сильное трансферабельное равновесие в стохастических играх бесконечной продолжительности	109
§ 2.7 Стохастическая игра с одним поглощающим состоянием	111
§ 2.8 Сильная позиционная состоятельность s -ядра в стохастических играх бесконечной продолжительности	116
Глава 3. Динамические игры, разыгрываемые на деревьях событий	120
§ 3.1 Определение динамической игры, разыгрываемой на дереве событий	120
§ 3.2 Позиционная состоятельность s -ядра в динамических играх, разыгрываемых на деревьях событий	125

§ 3.3	Позиционно состоятельное s -ядро в игре совместного управления загрязнением окружающей среды	131
§ 3.4	Абсолютное ε -равновесие в динамических играх, разыгрываемых на деревьях событий.....	137
§ 3.5	Построение абсолютного ε -равновесия в игре совместного управления загрязнением окружающей среды	143
§ 3.6	Условие стратегической устойчивости кооперативного решения в динамических играх, разыгрываемых на деревьях событий.....	149
§ 3.7	Абсолютное ε -равновесия с использованием процедуры распределения дележа в игре управления загрязнением окружающей среды	155
§ 3.8	Динамические игры, разыгрываемые на бинарных деревьях событий.....	158
§ 3.8.1	Динамические игры с линейной динамикой состояния.....	158
§ 3.8.2	Динамические игры, разыгрываемые на бинарных деревьях событий, с симметричными игроками.....	161
§ 3.8.3	S -адаптивное равновесие по Нэшу в динамических играх, разыгрываемых на деревьях событий.....	162
§ 3.8.4	Кооперативное решение динамической игры, разыгрываемой на бинарном дереве событий	167
§ 3.8.5	Цена анархии в динамической игре, разыгрываемой на бинарном дереве событий	173
§ 3.9	Цена анархии в одной игре охраны окружающей среды	174
§ 3.9.1	Модель	174
§ 3.9.2	Основные результаты.....	176
§ 3.9.3	Интерпретация результатов	183
§ 3.10	Позиционно состоятельный вектор Шепли в игре, разыгрываемой на дереве событий со случайным временем окончания.....	185
§ 3.10.1	Описание игры	185
§ 3.10.2	Позиционно состоятельный вектор Шепли в динамических играх на деревьях событий.....	190

§ 3.10.3 Необходимые условия существования S -адаптивного равновесия по Нэшу	196
§ 3.10.4 Теоретико-игровая модель охраны окружающей среды со случайным временем окончания	198
Глава 4. Приложения стохастических игр	203
§ 4.1 Модели передачи данных в беспроводных сетях	203
§ 4.1.1 Игра «Дилемма пересылки»	203
§ 4.1.2 Игра «Совместная пересылка пакета»	211
§ 4.1.3 Игра «Множественный доступ»	214
§ 4.1.4 Игра «Множественный доступ» в случае неполной информации	219
§ 4.1.5 Игра «Посылай и пересылай»	226
§ 4.1.6 Стохастическая игра передачи данных при наличии буферов конечной емкости	237
§ 4.2 Устойчивые в динамике коалиционные структуры	244
§ 4.2.1 Постановка задачи. Определение вспомогательной стохастической игры	244
§ 4.2.2 D-устойчивые коалиционные структуры	250
§ 4.2.3 D-устойчивые коалиционные структуры для игры двух лиц ..	252
§ 4.2.4 Примеры нахождения устойчивых в динамике коалиционных структур в играх трех лиц	255
§ 4.3 Стохастическая игра «Дилемма заключенного» с неполной информацией о дисконтирующих факторах	262
§ 4.3.1 Модель	262
§ 4.3.2 Случай с известными дисконтирующими факторами	263
§ 4.3.3 Двухфазная игра с неизвестными дисконтирующими факторами игроков	272
§ 4.3.4 Двухфазовая игра с кооперативным и полukoоперативным равновесиями	277
§ 4.3.5 Двухфазная игра с обучающей фазой непредписанной продолжительности	290

§ 4.3.6 Некооперативные равновесия	296
§ 4.3.7 Численный пример	298
Заключение	303
Литература	306

Введение

Актуальность темы диссертации

Диссертационная работа посвящена изучению методов построения кооперативного поведения в стохастических конфликтно-управляемых динамических системах. Раздел теории динамических игр, учитывающий неопределенность и случайность, является актуальным при моделировании реальных конфликтных процессов. Поскольку кооперация дает игрокам не меньший суммарный выигрыш нежели индивидуально рациональное поведение, то важной задачей является построение кооперативного варианта изначально заданной некооперативной стохастической игры. При решении кооперативных игр, разыгрываемых в динамике, необходимо учитывать особенности реализации заранее выбранных кооперативных решений, а именно учитывать изменение «окружающей среды» и изменения в поведении участников конфликта, ввиду смены их интересов, что не может не повлиять на характер совместных договоренностей и привести к подрыву кооперативного соглашения. В связи с этим возникает актуальная задача разработки принципов устойчивой кооперации в конфликтно-управляемых динамических системах конечной и бесконечной продолжительности, в том числе, в системах, учитывающих стохастический характер динамики.

При моделировании возможных сценариев развития событий в технологических системах разной сложности, а также бизнес-процессов часто используется анализ деревьев событий, которые представляют стохастический процесс с изначально заданными вероятностями, на которые не влияют действия игроков. Построение позиционно состоятельных кооперативных решений в конфликтно-управляемых стохастических системах, заданных на деревьях событий, является важной задачей для вышеупомянутых областей. Особенно в таких систе-

мах практический интерес представляет задача реализации стратегической поддержки кооперативных решений, позволяющей защитить кооперативное соглашение от выхода участников конфликта, желающих увеличить свой выигрыш.

Другой актуальной задачей работы является создание моделей передачи данных в сетях различных конфигураций с использованием теории стохастических игр. Решение стохастических игр как некооперативных, так и кооперативных, позволяет оценивать необходимость координации действий элементов—участников сети. В подавляющем большинстве случаев координация действий участников сети приводит к значительному увеличению её пропускной способности, и тем самым, к уменьшению суммарных издержек её участников. В работе ставится важная задача нахождения оценки стоимости отказа от кооперации и приводятся вычисления этой оценки в сетях различной конфигурации.

Актуально также применение теории стохастических игр к задаче определения устойчивых коалиционных структур в динамике, которые представляют собой разбиение множества игроков на непересекающиеся подмножества. Этот подход позволяет учитывать развитие процесса перехода игроков из одной структуры в другую и найти ту коалиционную структуру, от которой ни один из игроков не будет отклоняться в индивидуальном порядке. Устойчивая в динамике или d -устойчивая коалиционная структура находится в результате решения стохастической игры.

Степень разработанности проблемы в литературе

Началом развития теории стохастических игр можно считать публикацию L. Shapley [210], в которой он ввел понятие стохастической игры и доказал существование значения антагонистической стохастической игры бесконечной продолжительности. Стоит также отметить, что несколькими годами ранее в работах Н. W. Kuhn было введено понятие позиционной игры [120, 121]. Основные уравнения, используемые при решении стохастических игр являются прямым обобщением уравнения Беллмана [3, 75] для игровых динамических задач. Обобщение результатов Шепли на случай стохастической игры n лиц было получено в работах [91, 217], где доказано существование равновесия по Нэшу в стационарных стратегиях в стохастических играх с компактным мно-

жеством стратегий и конечным множеством состояний. Впоследствии большое количество работ было посвящено доказательству существования равновесия по Нэшу в различных классах стратегий, изучению стохастических игр с неполной информацией, асимметричными игроками, стохастических игр специальной структуры, среди которых хотелось бы отметить вклад следующих ученых в развитие теории стохастических игр: N. Vieille [214, 220], J.-F. Mertens [141], A. Neyman [141, 147, 148], A. S. Nowak [152, 153, 154], E. Solan [106, 212, 214], A. Jászkiewicz [110]. Стоит также упомянуть книги [140, 149] и обзоры результатов в области стохастических игр [213, 216, 221]. Вычислительные задачи нахождения равновесий в стохастических играх обсуждаются и решаются в работах [77, 78, 105].

Метод построения кооперативного варианта стохастической игры, реализуемой на конечном древовидном графе, впервые предложен Л. А. Петросяном в работе [178], где была сформулирована проблема позиционной состоятельности вектора Шепли и решена задача регуляризации позиционно несостоятельного вектора Шепли для данного класса игр. После этого метод построения кооперативного варианта игры был обобщен на класс стохастических игр бесконечной продолжительности в работе Е. М. Парилиной и Л. А. Петросяна [72]. Кооперативные стохастические игры бесконечной продолжительности с конечным множеством стратегий позднее были изучены в статьях [115, 163, 170]. В диссертационной работе формулируются принципы устойчивой кооперации или поддержки кооперативных решений стохастических игр. Эти принципы сформулированы Л. А. Петросяном и Н. А. Зенкевичем в статье [44] для динамических и дифференциальных игр.

Первый принцип или свойство динамической устойчивости (позиционной состоятельности) кооперативных решений сформулирован Л. А. Петросяном [35] для дифференциальных игр. Механизм определения выплат игрокам для регуляризации динамически неустойчивых кооперативных решений посредством так называемой процедуры распределения дележа предложен Л. А. Петросяном и В. В. Даниловым в статье [43]. Проблема построения динамически устойчивых кооперативных решений также изучена в работах Е. В. Шевкопляс для диф-

ференциальных игр со случайной продолжительностью [49], Д. В. К. Янгом и Л. А. Петросяном для динамических игр со случайной продолжительностью [225], Л. А. Петросяном и А. А. Седаковым для сетевых игр [186, 188], В. В. Захаровым [16] и Е. А. Корниенко [20] для иерархических игр, М. В. Марковским для линейно-квадратичных дифференциальных игр [25], А. В. Тур для линейно-квадратичных динамических игр [54], Д. В. Кузютиным для динамических игр с неполной информацией [21], Н. В. Козловской и Н. А. Зенкевичем [118] и Л. А. Петросяном и G. Zassoug [182], В. В. Захаровым и Л. А. Петросяном [17] в задачах охраны окружающей среды, В. В. Мазаловым и А. В. Реттиевой в задачах управления биоресурсами [137, 138]. Проблема устойчивой кооперации в динамических играх маршрутизации транспорта изучается в статье [18]. Подробное описание условия динамической устойчивости решений можно найти в книге [224] и различных подходов к этой проблеме в обзоре [227].

Второй принцип устойчивой кооперации в динамических и дифференциальных играх, стратегическая устойчивость или стратегическая поддержка кооперативного решения, сформулирован в работе Л. А. Петросяна [38]. Этот принцип оказался актуальным и позднее адаптирован для различных классов дифференциальных и динамических игр [55, 180, 185, 187].

Третий принцип устойчивой кооперации, защита от иррационального поведения, сформулирован в работе Д. В. К. Янга [223] и позднее применен в линейно-квадратичных играх в работах [53, 133]. Условия сохранения кооперации в марковских процессах, допускающих кооперацию игроков, включая условие защиты от иррационального поведения, сформулированы в статье [65].

Условие динамической устойчивости в случае, когда кооперативное решение является множеством, было усилено Л. А. Петросяном в работе [37] и получило название сильной динамической устойчивости. В последние годы это условие исследовано в различных классах игр (см., например, [15, 52, 83]).

В диссертационной работе среди прочих исследуется класс динамических игр, разыгрываемых на деревьях событий, в которых стратегии игроков не влияют на вероятности переходов из одних вершин дерева в другие. Анализ деревьев событий получил широкое распространение при моделировании воз-

можных инцидентов на ядерных реакторах [193]. Класс игр, разыгрываемых на деревьях событий, впервые представлен в работах G. Zassour [226] и A. Haurie et al. [104] для изучения некооперативных равновесий на европейском рынке природного газа. Подробное описание этого класса игр можно найти также в книге Haurie et al. [103]. Стохастические игры на деревьях событий используются при моделировании нерегулируемых рынков электроэнергии и прогнозирования равновесных инвестиций в технологии разных поколений [96, 97, 189, 190]. В статье [194] построен позиционно состоятельный вектор Шепли для данного класса игр. В диссертационной работе строится позиционно состоятельное s -ядро в играх, разыгрываемых на деревьях событий. Динамической устойчивости s -ядра в детерминированных динамических играх посвящены работы [81, 89, 98, 122, 228], тогда как в [191, 222] изучается устойчивость s -ядра в стохастических динамических играх специального вида. Идея построения ε -равновесия, используемая в диссертационной работе, предложена в [192] для конечно повторяющихся игр. Существование ε -равновесия в динамических и дифференциальных играх доказано в работах [23, 24]. Метод построения ε -оптимальных стратегий в дифференциальных играх предложен в [56]. Сравнение ε -равновесия и абсолютного ε -равновесия в повторяющихся играх описаны в [111]. Среди работ о построении ε -равновесий стоит отметить [92, 93, 215].

Для динамических игр, разыгрываемых на деревьях событий, в работе вычисляется цена анархии (*price of anarchy* или PoA), которая впервые предложена в [117]. Она является мерой различия выигрышей игроков при координации их действий и при ее отсутствии. В телекоммуникационных сетях со сложной конфигурацией или большой размерности бывает непросто вычислить точное значение цены анархии, поэтому часто ставится задача нахождения её границ [57, 82, 84, 196]. Мера, аналогичная цене анархии, называемая ценой устойчивости (*price of stability* или PoS), определена в [64]. В последнее время большое количество работ Т. Roughgarden и других авторов посвящено теоретико-игровому моделированию передачи данных, где вычисляется цена анархии сети [85, 86, 201, 202, 203, 204]. Результат вычисления цены анархии в дифференциальных играх представлен в работах [73, 229]. Теоретико-игровое

моделирование передачи данных с помощью стохастических игр применяется как для беспроводных, так и для других видов телекоммуникационных сетей [74, 80, 112, 150, 151, 205]. В этой области отметим работы К. Е. Авраченко, А. Ю. Гарнаева, Е. Altman и других [59, 60, 61, 124, 127].

В качестве приложения стохастических игр в работе представлена модель нахождения устойчивой в динамике коалиционной структуры. Обобщение кооперативных решений на случай наличия коалиционной структуры получено в статье [68], и далее в [155]. Различные статические подходы к определению устойчивых коалиционных структур были предложены в статьях [76, 102]. В диссертационной работе моделируется динамический подход, близкий к изложенному в [116], проводящий аналогию между поглощающими состояниями марковских цепей и устойчивыми коалиционными структурами. Отметим работы по динамическим играм, в которых допускается возможность формирования игроками коалиционной структуры, а не только максимальной коалиции [48, 181, 199].

Теория стохастических игр позволяет создавать адекватные реальности модели в области налогообложения [197], охраны окружающей среды [175] и экономики [63]. Теоретико-игровые прикладные модели можно найти в [9, 45]. Часто для приложений используется упрощенная модель стохастической игры — повторяющаяся игра. В частности, широко исследуется проблема стимулирования кооперации и оптимальное поведение при неполной или асимметричной информации [95, 101, 107, 108, 119, 129, 177].

Объектом исследования диссертационной работы являются стохастические конфликтные системы конечной и бесконечной продолжительности, а также конфликтные системы, развитие которых происходит по так называемым деревьям событий, в которых стохастический процесс не зависит от поведения участников конфликта. **Предметом исследования** являются методы построения устойчивых в широком смысле кооперативных решений в стохастических конфликтно-управляемых системах.

Цель работы

Целью диссертационной работы является построение теории кооперативных

стохастических игр конечной и бесконечной продолжительности с дисконтированными выигрышами, а также развитие теории динамических игр, разыгрываемых на деревьях событий, в которых стохастический процесс не зависит от стратегий игроков. Также целью работы является проверка принципов устойчивой кооперации, а именно, позиционной состоятельности, сильной позиционной состоятельности, стратегической устойчивости и защиты от иррационального поведения классических кооперативных решений в стохастических играх, разработка методов регуляризации решений с целью получения позиционно состоятельных решений на основе предлагаемых в работе процедур распределения дележа. Целью работы также является построение прикладных теоретико-игровых моделей с использованием стохастических игр.

Основные задачи

Одной из задач работы является разработка метода построения кооперативных вариантов стохастических игр в случае конечной и бесконечной продолжительности с конечным множеством состояний игры. Также в работе ставится задача построения метода регуляризации кооперативных решений для таких стохастических игр на основе процедуры распределения дележа с целью получения позиционно состоятельных кооперативных решений. Для случая неотрицательных функций выигрыша игроков высказывается предположение о неотрицательности выплат игрокам в реализованных вершинах (состояниях). В связи с этим ставится задача построения неотрицательной процедуры распределения дележа, удовлетворяющей свойству позиционной состоятельности решения для стохастических игр конечной продолжительности, заданных на графах. Также для класса стохастических игр конечной продолжительности ставится задача построения сильно позиционно состоятельного s -ядра.

При изучении стохастических игр бесконечной продолжительности одной из задач является регуляризация кооперативных решений путем построения позиционно состоятельной процедуры распределения дележей, принадлежащих этим решениям. Реализация кооперативного решения в динамике требует проверки выполнения принципа его стратегической устойчивости, который гарантирует существование равновесия по Нэшу (или в некоторых случаях — абсо-

лютного равновесия по Нэшу) с выигрышами, равными компонентам дележа. Для того, чтобы «застраховать» реализацию кооперативного решения в динамике от отклонения, ставится задача нахождения условий, при которых любой игрок получит не меньший суммарный выигрыш по сравнению с гарантированным выигрышем, даже если на каком-то шаге кооперативное соглашение будет расторгнуто, и в оставшейся части игры игроки будут играть индивидуально (защита от иррационального поведения). Одной из задач, формулируемых для класса стохастических игр бесконечной продолжительности, которые ставятся в работе, является построение сильного трансферабельного равновесия с выигрышами игроков, равными соответствующим компонентам дележа, а также — построение сильно позиционно состоятельного s -ядра.

Еще одной задачей диссертационной работы является построение позиционно состоятельного s -ядра на основе процедуры распределения дележа в кооперативной стохастической игре, разыгрываемой на дереве событий. Так как стохастические игры, заданные на деревьях событий, имеют конечную продолжительность, то гарантировать существование равновесия по Нэшу с кооперативными выигрышами в подавляющем большинстве игр невозможно. В связи с этим для стохастических игр такого вида ставится задача построения приближенного равновесия или абсолютного ε -равновесия с кооперативными выигрышами. При этом находится минимальное значение ε в зависимости от параметров игры. В работе ставится задача построения абсолютного ε -равновесия в регуляризованной стохастической игре, при котором выигрыши игроков равны соответствующим компонентам дележа, на основе которого была проведена регуляризация.

При моделировании реальных динамических процессов с помощью деревьев событий часто используются бинарные деревья, в которых из каждой нетерминальной вершины выходит два ребра. В диссертационной работе ставится задача изучения стохастических игр, разыгрываемых на бинарных деревьях событий с симметричными игроками. В этом случае находится равновесие по Нэшу и кооперативное решение, максимизирующее суммарный ожидаемый выигрыш игроков и выводится формула цены анархии для стохастических игр

данного вида, которая позволяет оценить, во сколько раз суммарный выигрыш игроков при кооперации больше их суммарного выигрыша в равновесии по Нэшу.

Одной из задач диссертационной работы является разработка теоретико-игровых моделей передачи данных в беспроводных сетях различных конфигураций с использованием стохастических игр, а также нахождение равновесий по Нэшу и кооперативных решений, вычисление стоимости отказа от кооперации с целью обоснования необходимости координации действий участников сети. Для одной модели передачи данных ставится задача проверки кооперативного решения на позиционную состоятельность и стратегическую устойчивость и проведения регуляризации позиционно несостоятельных решений.

Одной из задач работы также является построение теоретико-игровой модели нахождения «устойчивых во времени» или d -устойчивых коалиционных структур с применением теории стохастических игр, в рамках решения которой предлагается метод построения стохастической игры, в ней находятся равновесия по Нэшу, удовлетворяющие условиям, которые позволяют определить d -устойчивые коалиционные структуры. Ставится задача провести численное моделирование решения игры трех лиц и найти все d -устойчивые коалиционные структуры.

Также одной из задач работы является построение двухфазной стохастической игры двух лиц «Дилемма заключенного» с неполной информацией, в которой имеется два состояния. Неполнота информации заключается в отсутствии информации о значении дисконтирующего фактора другого игрока. Ставится задача нахождения байесовских равновесий в этой игре.

Научная новизна

В диссертационной работе предложен метод построения кооперативного варианта стохастической игры по изначально заданной некооперативной стохастической игре конечной и бесконечной продолжительности. В работе подробно изучены механизмы построения кооперативных решений, удовлетворяющих принципам устойчивой кооперации в вышеупомянутых классах стохастических игр, а также в играх, разыгрываемых на деревьях событий. Для построения по-

зиционно состоятельных решений применена процедура распределения дележа. Для кооперативных стохастических игр конечной и бесконечной продолжительностей с конечным множеством состояний получены достаточные условия сильной позиционной состоятельности s -ядра. Все эти методы и механизмы впервые применяются к вышеупомянутым классам игр. Для стохастических игр бесконечной продолжительности получены условия существования кооперативных решений, удовлетворяющих принципу стратегической устойчивости и защиты от иррационального поведения.

Также подробно изучаются методы поддержки кооперации в динамических играх, разыгрываемых на деревьях событий. В результате получены условия позиционной состоятельности s -ядра, условия существования абсолютного ε -равновесия с кооперативными выигрышами игроков, когда перераспределение выигрышей не проводилось, а также условие стратегической устойчивости кооперативного решения, когда перераспределение выигрышей между игроками было произведено. Все эти результаты впервые получены в работах автора. Стоит отметить новый подход к сравнению выигрышей игроков при кооперативном и некооперативном поведении с использованием цены анархии, который был заимствован из моделей передачи данных и впервые был применен к динамическим играм, разыгрываемым на деревьях событий. Для динамических игр, разыгрываемых на бинарных деревьях событий с линейными уравнениями динамики состояний и симметричными игроками, в явном виде вычислена цена анархии.

Методы моделирования передачи данных в беспроводных сетях некоторых конфигураций уже рассматривались в литературе. Нами впервые предложен подход, опирающийся на кооперативный вариант игры, моделирующий передачу данных с помощью стохастической игры, найдены кооперативные решения и проведена их регуляризация в случае их позиционной несостоятельности. Регуляризация позволяет изменить схему выплат участникам сети с целью увеличения их выигрышей и повышения пропускной способности сети. В работе предложен оригинальный метод нахождения d -устойчивых коалиционных структур (структур, устойчивых в динамике), который требует решения стохастической

игры, построенной специальным образом по изначально заданным значениям характеристической функции. В работе предложен подход к нахождению оптимального поведения участников стохастической игры «Дилемма заключенного» с двумя состояниями в случае, когда неизвестен дисконтирующий фактор другого игрока. В этом случае построена двухфазная игра, в первой фазе которой игроки пытаются узнать тип другого игрока, а во второй фазе — реализуют полностью кооперативное (кооперация в обоих состояниях), частично кооперативное (кооперация только в одном состоянии) или некооперативное равновесие.

Методы исследования

В диссертационной работе используются методы теории динамических игр (построение многошаговых игр, подыгр, принципов оптимальности, динамическая устойчивость или позиционная состоятельность, процедура распределения дележа), теории кооперативных игр (дележи, коалиционные структуры, устойчивость коалиционных структур), теории оптимального управления (метод динамического программирования, принцип максимума Понтрягина), оптимизации и теории вероятностей (распределения случайных величин, стохастические процессы, марковские цепи).

Теоретическая и практическая значимости

Полученные в диссертационной работе теоретические результаты относятся к области динамических игр. Их значимость заключается в построении теории кооперативных стохастических игр для различных способов задания динамики игры и стохастического процесса, а также в построении устойчивых механизмов кооперации, в разработке методов построения позиционно состоятельных и сильно позиционно состоятельных кооперативных решений, а также их стратегической устойчивости для изучаемых классов игр.

Практическая ценность работы определяется областью применения стохастических игр: при математическом моделировании экономических, социальных конфликтно-управляемых процессов, при решении задач страхования и охраны окружающей среды. Поэтому сферу применения полученных результатов можно оценить описанной достаточно широкой областью применения сто-

частических игр, в которых кооперация игроков имеет содержательный смысл. Теоретические результаты главы 2 продемонстрированы на экономическом примере, а именно, решена задача построения устойчивого кооперативного решения в условиях конкуренции. Теоретические результаты главы 3 продемонстрированы на примерах задач охраны окружающей среды, когда несколько стран-участниц конфликта производят выбросы и загрязняют общую территорию. Как показали результаты, кооперативное поведение стран приводит к совместному уменьшению издержек и уменьшению общего загрязнения окружающей среды. Глава 4 диссертационной работы посвящена построению прикладных моделей с использованием аппарата стохастических игр. В частности, в рамках этой главы построено несколько моделей динамической передачи данных в беспроводных сетях различных конфигураций, для каждой из которых делается вывод о целесообразности кооперации или совместного управления передачей данных для увеличения пропускных способностей сетей.

Исследования, проведенные в рамках диссертационной работы, были поддержаны следующими грантами: РФФИ № 06-01-39005 «Математический анализ конфликта и кооперации» (2006–2007), № 09-01-00334 «Оптимизация в многоагентных беспроводных сетях» (2009–2010), № 10-01-91160 «Динамические сетевые игры» (2010–2011), № 13-01-91160 «Кооперация в сетевых играх» (2013–2014), № 14-01-31141 «Устойчивость коалиционных соглашений» (2014–2015), № 16-01-00713 «Кооперация при структурных и информационных ограничениях» (2016–2017), тревел-грантом № 08-01-09236 (2008); грантами СПбГУ № 9.0.189.2010 «Математическое моделирование многоагентных управляемых динамических систем» (2010–2012), № 9.38.245.2014 «Принципы оптимальности в динамических и дифференциальных играх с фиксированной и изменяющейся коалиционной структурой» (2014–2016), мероприятие 5 СПбГУ (2011, 2012, 2013, 2014, 2015, 2018); грантом РФФИ № 17-11-01079 «Оптимальное поведение в конфликтно-управляемых системах» (2017–наст.вр.); финансовой поддержкой исследовательского центра GERAD (2013, 2015, 2018).

Результаты диссертационной работы использовались и используются в учебном процессе Санкт-Петербургского государственного университета при прове-

дении занятий по курсам «Теория игр и исследование операций», «Методы и модели исследования операций», «Динамические игры», «Управление рисками в проектах, финансах и логистике» и др.

Достоверность полученных в диссертационной работе результатов обусловлена строгостью математических доказательств.

Краткое описание работы

Диссертация состоит из введения, четырех глав и заключения, включает 329 страниц, 42 таблицы и 24 рисунка. Список литературы содержит 229 наименований.

Первая глава диссертационной работы посвящена изучению кооперативных стохастических игр, разыгрываемых на конечных древовидных графах. Основными целями этой главы является построение кооперативной игры на основе некооперативной стохастической игры, разработка методов проверки кооперативных решений игры на позиционную состоятельность и сильную позиционную состоятельность, а также регуляризация позиционно несостоятельных решений. В § 1.1 вводится определение некооперативной стохастической игры и приводится описание динамики игры, а также определяется класс стратегий, в которых находится решение. Основные функциональные уравнения для вычисления математического ожидания выигрышей игроков в игре и ее подыграх приведены в § 1.2. Кооперативный вариант стохастической игры в форме характеристической функции строится в § 1.3. Также в этом разделе проводится обсуждение подходов к определению характеристических функций. Выбор делается в пользу так называемой α -характеристической функции. Записываются уравнения Беллмана для определения значений этой функций для всех возможных коалиций. Дается определение дележа в стохастической игре и любой ее подыгре, а также вводится понятие решения кооперативной стохастической игры и подыгры. Параграф 1.4 содержит определения используемых в работе решений кооперативных игр, включая вектор Шепли, s -ядро, n -ядро. Определение процедуры распределения дележа для стохастической игры конечной продолжительности приводится в § 1.5, который также содержит функциональные уравнения Беллмана для вычисления ожидаемой суммы выплат игрокам

в соответствии с процедурой распределения дележа и доказательство леммы, позволяющей построить процедуру распределения дележа. В § 1.6 описывается проблема позиционной несостоятельности решений кооперативной стохастической игры и дается определение позиционно состоятельного дележа, а также приводится способ построения позиционно состоятельного кооперативного решения. В случае неотрицательности функции выигрыша изначально заданной игры, а также при требовании к неотрицательности выплат, производимых игрокам согласно процедуре распределения дележа, в § 1.7 предлагается способ построения новой процедуры распределения дележа, определяется новый дележ в кооперативной игре с новой характеристической функцией, удовлетворяющий свойствам позиционной состоятельности и неотрицательности соответствующей ему процедуры распределения. Теоретические результаты конкретизируются на случай выбора вектора Шепли и c -ядра в качестве решений кооперативной стохастической игры в § 1.8, а именно, проводится их регуляризация таким образом, что регуляризованные вектор Шепли и c -ядро обладают свойством позиционной состоятельности. Два примера нахождения позиционно состоятельных решений (в первом случае — вектора Шепли, во втором — c -ядра) приводятся в § 1.8. В случае, когда решением игры является не единственный дележ, а их множество, актуальным становится вопрос построения сильно позиционного состоятельного решения. Проблема сильной позиционной состоятельности c -ядра для кооперативной стохастической игры конечной продолжительности сформулирована в § 1.9. Определены и доказаны достаточные условия сильной позиционной состоятельности c -ядра, а также приведен численный пример построения сильно позиционно состоятельной процедуры распределения дележа из c -ядра.

Вторая глава посвящена изучению принципов устойчивой кооперации в стохастических играх бесконечной продолжительности и конечным множеством состояний. За основу построения кооперативного варианта была взята стохастическая игра, впервые представленная в работе [210]. В § 2.1 описываются модель некооперативной стохастической игры, классы стратегий игроков и способы вычисления их выигрышей. В случае, когда стохастическая игра решается

в классе стационарных стратегий, выигрыши игроков можно вычислить в явном виде при заданных стратегиях игроков, эти формулы можно найти в § 2.2. Кооперативный вариант стохастической игры бесконечной продолжительности и конечным множеством состояний строится в § 2.3. Значение характеристической функции для заданной коалиции игроков определяется как значение антагонистической игры двух лиц, где в качестве максимизирующего игрока выступает рассматриваемая коалиция, а в качестве антагониста — коалиция, состоящая из игроков, не вошедших в первую коалицию. Существование значения такой игры гарантировано теоремой Шепли [210]. В § 2.3 представлены формулы для вычисления значений характеристической функции игры и любой ее подыгры для всех возможных коалиций. В § 2.4 определены основные принципы устойчивой кооперации в стохастических играх рассматриваемого типа, которые в совокупности впервые сформулированы в работе [44]. Принцип позиционной состоятельности, который является обобщением понятия динамической устойчивости, представлен в § 2.4.1, где введено понятие процедуры распределения дележа и определено свойство позиционной состоятельности дележа и соответствующей ей процедуре распределения. Получена формула для вычисления компонент позиционно состоятельной процедуры распределения дележа в явном виде, которая приводится в § 2.4.1, а также предлагается способ регуляризации изначально заданной некооперативной игры, в которой кооперативное решение (заранее выбранный дележ) уже является позиционно состоятельным. Принцип стратегической устойчивости кооперативного решения формулируется в § 2.4.2, а именно, строится ситуация в стратегиях поведения с двумя «режимами»: кооперативным и наказания, позволяющим наказывать отклоняющегося от кооперативного поведения игрока. Доказывается теорема о существовании ситуации равновесия по Нэшу с выигрышами игроков, равными компонентам заранее выбранного дележа. Также в этом параграфе доказывается существование абсолютного равновесия по Нэшу с этими же выигрышами в стохастической игре. Принцип защиты от иррационального поведения приводится в § 2.4.3, где получено достаточное условие выполнения этого принципа. Вопрос существования устойчивого кооперативного решения, удовлетворяющего всем трем

принципам, обсуждается в § 2.5. Там же предлагается стохастическая динамическая модель конкуренции между асимметричными фирмами, находится кооперативное решение стохастической игры и проводится проверка выполнения всех принципов устойчивой кооперации. В § 2.6 показывается «усиление» принципа стратегической поддержки, а именно, доказываемое существование сильного трансферабельного равновесия в регуляризованной игре с кооперативными выигрышами. Принцип сильной позиционной состоятельности для стохастических игр бесконечной продолжительности с конечным множеством состояний формулируется в § 2.8. Параграф 2.8 содержит пример построения кооперативного решения в стохастической игре с одним поглощающим состоянием, а также проверку принципов устойчивой кооперации для такой игры. В качестве итога сформулированы ограничения на дисконтирующий фактор игроков, позволяющие обеспечить устойчивую кооперацию в рассматриваемой игре.

В третьей главе изучаются игры, разыгрываемые на деревьях событий, в которых стохастический процесс не зависит от стратегий игроков. Для данного класса игр задано уравнение динамики состояния системы, строится кооперативный вариант заданной изначально некооперативной игры. Теоретические результаты Главы 3 продемонстрированы на нескольких примерах из области охраны окружающей среды. В § 3.1 описывается модель стохастической игры, разыгрываемой на дереве событий, определяется класс S -адаптивных стратегий, приводятся основные уравнения для нахождения равновесия по Нэшу и кооперативной ситуации, максимизирующей суммарный ожидаемый выигрыш игроков. В § 3.2 определяется характеристическая функция, используя γ -подход, при котором значение γ -характеристической функции для коалиции S представляет собой суммарный выигрыш игроков этой коалиции в ситуации равновесия по Нэшу, когда игроки коалиции S максимизируют свой суммарный выигрыш, а все остальные игроки максимизируют свои индивидуальные выигрыши. В качестве решения стохастической игры выбирается s -ядро. Формулируется проблема позиционной состоятельности s -ядра, и предлагается способ перераспределения выплат игрокам для того, чтобы гарантировать позиционную состоятельность выбранного дележа из s -ядра в случае, когда s -ядра всех

подыгр кооперативной траектории непусты. Пример построения позиционно состоятельного s -ядра в одной игре охраны окружающей среды рассматривается в § 3.3. Для игры с тремя периодами находится позиционно состоятельное s -ядро на основе n -ядер. Абсолютное ε -равновесие строится для нерегуляризованной стохастической игры, разыгрываемой на дереве событий, в § 3.4. В построенном приближенном равновесии выигрыши игроков равны соответствующим выигрышам в ситуации, когда игроки реализуют кооперативные стратегии. Примеры построения абсолютного ε -равновесия в играх охраны окружающей среды с четырьмя и одиннадцатью периодами и бинарными деревьями приводятся в § 3.5. Для игры с одиннадцатью периодами времени проводится численный анализ минимального значения ε как функции дисконтирующего фактора. В предположении, что игроки максимизируют суммарный выигрыш и проводят регуляризацию выигрышей с соответствии с позиционно состоятельной процедурой распределения дележа, в § 3.6 строится абсолютное ε -равновесие в игре и находится минимальное значение ε . Численное сравнение минимальных значений ε , используемых при построении абсолютных ε -равновесий в кооперативной игре с 6 периодами без регуляризации и с регуляризацией выигрышей игроков проводится в § 3.7. Случай бинарного дерева событий с линейно-квадратичными функциями выигрышей подробно изучен в § 3.8. Также в этом параграфе предполагается симметричность игроков. Для данной стохастической игры в явном виде находятся стратегии игроков, образующие S -адаптивное равновесие по Нэшу (см. § 3.8.3), а также стратегии, образующие кооперативное решение (см. § 3.8.4). В § 3.8.5 выводится формула для вычисления цены анархии в описанной игре. Теоретические результаты § 3.8 демонстрируются на одной игре охраны окружающей среды. В § 3.10 рассматривается стохастическая игра на дереве событий случайной продолжительности, для которой предлагается способ построения позиционно состоятельного вектора Шепли, а также приводятся необходимые условия существования S -адаптивного равновесия по Нэшу. Теоретические результаты § 3.10 демонстрируются на примере в § 3.10.4.

Глава 4 посвящена теоретико-игровому моделированию с использованием теории стохастических игр, а также решению прикладных стохастических игр.

В § 4.1 предлагаются четыре модели передачи данных в беспроводных сетях. В игре «Дилемма пересылки» (см. § 4.1.1) описывается конфликтная система, состоящая из четырех вершин, две из которых представляют игроков, принимающих решения о пересылке пакета, появляющегося в данной вершине. Динамическая модель передачи данных моделируется стохастической игрой с четырьмя состояниями и решается в стационарных стратегиях, для которой найдены равновесия по Нэшу и кооперативное решение. Сравнение выигрышей игроков при наличии кооперации и при ее отсутствии делается с помощью предложенной «стоимости отказа от кооперации». Для численного примера игры «Дилемма пересылки» делается вывод о необходимости координации действий игроков с целью увеличения пропускной способности системы. Аналогичные вычисления проводятся для игр «Совместная пересылка пакета» (см. § 4.1.2) и «Множественный доступ» (см. § 4.1.3). Для игры «Совместная пересылка пакета» результаты получены в общем виде, а для игры «Множественный доступ» — при равномерном начальном распределении на множестве состояний. В § 4.1.5 строится модель передачи данных в сети специальной конфигурации, для которой определяется кооперативная игра, находится вектор Шепли в качестве решения игры, строится процедура распределения дележа, гарантирующая позиционную состоятельность вектора Шепли, а также проверяется выполнение принципа стратегической устойчивости вектора Шепли. Модель передачи данных в сети, в которой вершины имеют буферы конечных емкостей, представлена в § 4.1.6 В § 4.2 предлагается динамическая модель нахождения устойчивых коалиционных структур как решений стохастической игры специального вида. Метод построения вспомогательной стохастической игры приводится в § 4.2.1. Численное решение игры трех лиц для двух случаев: с супераддитивной и несупераддитивной характеристическими функциями дано в § 4.2.4. В § 4.3 рассматривается стохастическая байесовская модель бесконечно повторяющейся игры «Дилемма заключенного» двух лиц с двумя возможными состояниями. В § 4.3.1 описывается модель и находятся условия, при которых «кооперативная» ситуация (кооперация в обоих состояниях), «полукооперативная» ситуация (кооперация только в одном состоянии) являются абсолютными равновесиями по Нэшу.

В зависимости от значений дисконтирующих факторов выделяется три класса игроков: «терпеливые», «умеренные» и «нетерпеливые». В § 4.3.3 предлагается модель игры с неизвестными дисконтирующими факторами, с двумя фазами: обучающей и кооперативной, описываются стратегии игроков в обеих фазах и вводится понятие частично марковского абсолютного байесовского равновесия. В § 4.3.4 доказываются существование равновесий, при которых во второй фазе игры стратегии игроков образуют кооперативное или полukoоперативное равновесия. Обучающая фаза в этом случае длится один период. Также здесь производится сравнение выигрышей игроков в равновесиях. Случай, когда обучающая фаза имеет непредписанную продолжительность, рассмотрен в § 4.3.5, где получены условия существования равновесий с кооперативными и полukoоперативными исходами. Теоретические результаты § 4.3 продемонстрированы на численном примере в § 4.3.7.

Заключение содержит краткое описание полученных результатов.

Положения, выносимые на защиту

1. Метод построения кооперативной стохастической игры по заданной изначально некооперативной стохастической игре конечной продолжительности, а также бесконечной продолжительности при конечном множестве состояний.
2. Метод построения позиционно состоятельного решения на основе процедуры распределения дележа в кооперативных стохастических играх конечной и бесконечной продолжительностей, а также в динамических играх, разыгрываемых на деревьях событий.
3. Регуляризация кооперативных решений стохастической игры конечной продолжительности и неотрицательными функциями выигрышей игроков, а также построение «новой» позиционно состоятельной процедуры распределения дележа с неотрицательными компонентами и условия позиционной состоятельности решения для данного класса игр.
4. Метод построения сильно позиционно состоятельного s -ядра в кооперативных стохастических играх конечной и бесконечной продолжительности.

5. Условия выполнения принципов устойчивой кооперации, включая позиционную состоятельность, стратегическую устойчивость и защиту от иррационального поведения кооперативного решения, для стохастических игр бесконечной продолжительности с конечным множеством состояний.
6. Условие существования сильного трансферабельного равновесия с выигрышами, равными соответствующим компонентам дележа, в стохастических играх бесконечной продолжительности с конечным множеством состояний.
7. Условия существования абсолютного ε -равновесия в динамических играх, разыгрываемых на деревьях событий, с выигрышами игроков, равными их выигрышам при реализации кооперативного решения, а также условие стратегической устойчивости кооперативного решения для данного класса игр.
8. Построение равновесия по Нэшу и кооперативного решения в динамических играх, разыгрываемых на бинарных деревьях событий с линейными уравнениями состояний и симметричными игроками, а также метод вычисления цены анархии для данного класса игр.
9. Метод построения кооперативного решения при динамической передаче данных в беспроводных сетях различных конфигураций, метод регуляризации одной стохастической игры передачи данных с целью получения позиционно состоятельного кооперативного решения.
10. Метод нахождения d -устойчивых (устойчивых в динамике) коалиционных структур с помощью решения специальным образом построенной стохастической игры в случае отсутствия ограничений на множество допустимых коалиций.
11. Условия существования байесовских равновесий в двухфазной стохастической игре «Дилемма заключенного» с двумя состояниями, приводящие к полной или частичной кооперации игроков во второй фазе игры.

Апробация результатов

Основные результаты диссертационной работы докладывались на Международных конференциях «Теория игр и менеджмент» (Санкт-Петербург, 2007, 2010, 2011, 2012, 2013, 2014, 2016, 2018), Международном рабочем совещании «Задачи оптимальной остановки и стохастического управления» (Петрозаводск, 2005), Международных конференциях «Устойчивость и процессы управления» (Санкт-Петербург, 2005, 2010, 2015), Международных конференциях «Конструктивный негладкий анализ и смежные вопросы» (Санкт-Петербург, 2012), Международной конференции «Теория управления и теория обобщенных решений уравнений Гамильтона-Якоби» (Екатеринбург, 2005), Российской «Воронежской весенней математической школе «Понтрягинские чтения — XV» (Воронеж, 2004), семинаре российско-финской школы аспирантов «Динамические игры и многокритериальная оптимизация» (Петрозаводск, 2006), Международной конференции «Нелинейный динамический анализ» (Санкт-Петербург, 2007), Международной конференции «Теория игр и приложения» (Циндао, 2007), Международной конференции «Испано-итальяно-голландская конференция по теории игр» SING–2008 (Вроцлав, 2008), Семинаре аспирантов «Ссылки как показатели исследований — кто читает наши статьи?» (Стокгольм, 2007), Международном семинаре по теории игр (Тель-Авив, 2011), 16-ой конференции по динамике, экономическому росту и международной торговле DEGIT-XVI (Санкт-Петербург, 2011), Четвертом международном семинаре по динамическим играм в менеджменте (Падуя, 2012), Международной конференции «Математика, экономика и менеджмент: 100 лет со дня рождения Л. В. Канторовича» (Санкт-Петербург, 2012), Международном семинаре «Сетевые игры и менеджмент» (Петрозаводск, 2012), 15-ом, 16-ом и 18-ом Международных симпозиумах по динамическим играм и приложениям (Бышице, 2012, Амстердам, 2014, Гренобль, 2018), 9-ом и 10-ом Международных семинарах Международного общества динамических игр (Барселона, 2013, Глазго, 2015), Международной конференции «SING11-GTM2015» (Санкт-Петербург, 2015), XIV Международной конференции «Устойчивость и колебания нелинейных систем управления» (Конференция Пятницкого), а также на семинарах Кафедры математической теории игр и статистических решений СПбГУ, Института прикладных математических ис-

следований Карельского научного центра (2006, 2018), Отдела динамических систем Института математики и механики им. Н. Н. Красовского Уральского отделения РАН (2018), Института проблем управления им. В. А. Трапезникова РАН (2018), Университета Падуи (2013), Университета Циндао (2018), Исследовательского центра GERAD (Монреаль, 2013, 2015, 2018).

Публикации

Автором опубликовано более 85 научных работ, из них 39 научных работ по теме диссертационного исследования, в т.ч. одна глава в монографии и 15 статей, опубликованных в ведущих рецензируемых научных журналах (Математическая теория игр и её приложения, Управление большими системами, Труды Института математики и механики УрО РАН, European Journal of Operational Research, Automatica, Operations Research Letters, Theory and Decision, International Game Theory Review, Automation and Remote Control) из списка, рекомендованного ВАК, и приравненных к ним.

Все основные результаты, выносимые на защиту, получены автором самостоятельно.

Благодарности

Автор выражает искреннюю благодарность и глубокую признательность доктору физико-математических наук, профессору Леону Аганесовичу Петросяну за предложенную тему исследования, формирование научного интереса к математической теории игр, многолетнее и плодотворное сотрудничество, внимание к работе и всестороннюю помощь. Также автор выражает благодарность профессору Джоржу Заккуру за поддержку, доброе отношение, интересные дискуссии и совместную работу. Автор благодарит своих родителей и семью за терпеливое отношение к научной работе, помощь и поддержку в любых ситуациях.

Глава 1

Кооперативные стохастические игры конечной продолжительности

§ 1.1 Некооперативные стохастические игры конечной продолжительности

Определим конечную стохастическую игру, разыгрываемую на графе. Пусть задан конечный древовидный граф $\Psi = (Z, L)$ (см. [4]), где Z — множество вершин графа, а $L : Z \rightarrow Z$ — точечно-множественное отображение, заданное на множестве Z , со значениями во множестве подмножеств множества Z . Вершина z_0 — начальная вершина древовидного графа Ψ . Обозначим через $Z^T \subset Z$ терминальные вершины графа Ψ , то есть те вершины z , для которых $L(z) = \emptyset$. Конечный древовидный граф с начальной вершиной z_0 будем обозначать через $\Psi(z_0)$.

Пусть в каждой вершине $z \in Z$ графа $\Psi(z_0)$ задана одновременная игра n лиц в нормальной форме

$$\Gamma(z) = \langle N, A_1^z, \dots, A_n^z, K_1^z, \dots, K_n^z \rangle,$$

где $N = \{1, 2, \dots, n\}$ — конечное множество игроков, одинаковое для всех вершин $z \in Z$; A_i^z — конечное множество стратегий игрока $i \in N$, $K_i^z(a_1^z, \dots, a_n^z) : \prod_{j \in N} A_j^z \rightarrow \mathbb{R}$ — функция выигрыша игрока i , $a_i^z \in A_i^z$. Набор стратегий $a^z = (a_1^z, \dots, a_n^z)$, $a_i^z \in A_i^z$, $i \in N$, будем называть ситуацией в игре $\Gamma(z)$. При этом, $a^z \in A^z = \prod_{i \in N} A_i^z$, A^z — множество ситуаций в игре $\Gamma(z)$.

Для каждой вершины $z \in Z$ определены вероятности перехода в вершины

$y \in L(z)$ графа $\Psi(z_0)$, следующие за вершиной z . Эти вероятности зависят от ситуации a^z , реализовавшейся в игре $\Gamma(z)$. Таким образом, для каждой вершины $z \in Z$ задана функция $p(\cdot|z, a^z) : A^z \rightarrow \Delta(L(z))$, где $\Delta(L(z))$ — вероятностное распределение на множестве $L(z)$. То есть,

$$\begin{aligned} p(y|z, a^z) &\geq 0, \\ \sum_{y \in L(z)} p(y|z, a^z) &= 1 \end{aligned}$$

для любой ситуации $a^z \in A^z$. Величина $p(y|z, a^z)$ — вероятность реализации одновременной игры $\Gamma(y)$, $y \in L(z)$ после игры $\Gamma(z)$, в которой реализовалась ситуация $a^z = (a_1^z, \dots, a_n^z)$.

Предположим также, что продолжительность игры является случайной величиной, которая принимает значения от 0 до l , где l — длина игры (длинной игры будем называть число шагов в максимальной партии игры). Считаем заданными вероятности q_k окончания игры на шаге k . Причем, $0 \leq q_k \leq 1$, $k = 0, \dots, l-1$, $q_l = 1$, и шаг k в вершине $z \in Z$ в стохастической игре со случайной продолжительностью определяется из условия $z \in (L(z_0))^k$.

Замечание 1.1. Следует отметить, что вероятности q_k , $k = 0, \dots, l$, являются условными вероятностями, и по этим вероятностям можно определить распределение случайной величины, представляющей собой продолжительность игры. В случае, когда все пути в графе $\Psi(z_0)$ имеют одинаковую длину l , дискретное распределение продолжительности игры, определенное по условным вероятностям q_k , представлено в таблице 1.1, в которой P_k — вероятность того, что игра закончится на шаге k .

Таблица 1.1: Распределение случайной величины, равной продолжительности игры.

k	P_k
0	q_0
1	$(1 - q_0)q_1$
2	$(1 - q_0)(1 - q_1)q_2$
\vdots	\vdots
l	$(1 - q_0)(1 - q_1) \cdot \dots \cdot (1 - q_{l-1})$

Определение 1.1. *Стохастической игрой со случайной продолжительностью $G(z_0)$, где z_0 — начальная вершина древовидного графа $\Psi(z_0)$, будем называть набор*

$$G(z_0) = \langle N, \Psi(z_0), \{\Gamma(z)\}_{z \in Z}, \{q_k\}_{k=0}^l, \{p(\cdot|z, a^z)\}_{z \in Z \setminus Z^T, a^z \in A^z} \rangle. \quad (1.1.1)$$

Из определения стохастической игры со случайной продолжительностью понятно, что переходы из одних вершин графа $\Psi(z_0)$ в другие, а также момент окончания игры являются случайными.

Стохастическая игра со случайной продолжительностью $G(z_0)$ происходит следующим образом:

1. В вершине z_0 графа $\Psi(z_0)$ осуществляется одновременная игра $\Gamma(z_0)$. Пусть в ней реализуется некоторая ситуация $a^{z_0} \in A^{z_0}$. Каждый игрок $i \in N$ получает выигрыш $K_i^{z_0}(a^{z_0})$. Стохастическая игра $G(z_0)$ либо прекращается с вероятностью q_0 , $0 \leq q_0 \leq 1$, либо с вероятностью $1 - q_0$ продолжается и происходит переход в вершину $y \in L(z_0)$ графа $\Psi(z_0)$ с вероятностью $p(y|z_0, a^{z_0})$, зависящей от ситуации a^{z_0} , реализовавшейся в игре $\Gamma(z_0)$.
2. Предположим, что на k -ом шаге игровой процесс находится в вершине $z_k \in Z$, где задана одновременная игра $\Gamma(z_k)$. Пусть в этой игре реализуется ситуация $a^{z_k} \in A^{z_k}$. Каждый игрок $i \in N$ получает выигрыш $K_i^{z_k}(a^{z_k})$. Стохастическая игра либо заканчивается с вероятностью q_k , $0 \leq q_k \leq 1$, либо с вероятностью $1 - q_k$ продолжается и происходит переход в вершину графа $z_{k+1} \in L(z_k)$ с вероятностью $p(z_{k+1}|z_k, a^{z_k})$, зависящей от ситуации a^{z_k} , реализовавшейся в игре $\Gamma(z_k)$. В случае, если множество $L(z_k)$ пусто, игра заканчивается в вершине z_k с вероятностью 1.
3. Стохастическая игра продолжается, пока не будет достигнута терминальная вершина, или она не закончится в соответствии с реализациями вероятностей q_0, \dots, q_l .

Обозначим через $G(z_k)$ подыгру (см. [120, 121]) игры $G(z_0)$, берущую начало в вершине $z_k \in Z$ графа $\Psi(z_0)$ (с игры $\Gamma(z_k)$), которая также является стохастической игрой со случайной продолжительностью. Подыгра $G(z_k)$ определяется

на подграфе $\Psi(z_k)$ с множеством вершин $Z(z_k) \subset Z$ и задается набором

$$G(z_k) = \langle N, \Psi(z_k), \{\Gamma(z)\}_{z \in Z(z_k)}, \{q_s\}_{s=k}^l, \{p(\cdot|z, a^z)\}_{z \in Z(z_k) \setminus Z^T, a^z \in A^z} \rangle.$$

Для решения стохастической игры (1.1.1) необходимо определить стратегии игроков. Обозначим через $\varphi_i : Z \times H \rightarrow \prod_{z \in Z} \Delta(A_i^z)$ стратегию поведения игрока i в игре $G(z_0)$, где H — множество историй вершин дерева (под историей вершины понимается набор ситуаций, реализованных на отрезке пути от z_0 до этой вершины), $\Delta(A_i^z)$ — множество смешанных стратегий игрока i в вершине $z \in Z$. Ситуацией в стохастической игре $G(z_0)$ называется набор стратегий игроков $\varphi = (\varphi_i : i \in N)$. Обозначим через Σ_i множество стратегий поведения игрока i в стохастической игре $G(z_0)$, тогда $\Sigma = \prod_{i \in N} \Sigma_i$ — множество ситуаций в игре $G(z_0)$. Очевидно, что сужение стратегии φ_i на подграф $\Psi(z_k)$ графа $\Psi(z_0)$ является стратегией в подыгре $G(z_k)$. Это сужение обозначим через $\varphi_i^{z_k}$.

§ 1.2 Основные функциональные уравнения для стохастических игр конечной продолжительности

Предположим, что в стохастической игре $G(z_0)$ игроки реализуют стратегии φ_i , $i \in N$. Определим выигрыш игрока i как математическое ожидание его выигрыша относительно случайной величины, равной продолжительности игры, и например, для реализовавшегося пути $z_0, z_1 \in L(z_0), z_2 \in L(z_1), \dots, z_l \in L(z_{l-1}), L(z_l) = \emptyset$, получаем

$$E_i(z_0, \varphi) = \sum_{k=0}^l P_k \sum_{j=0}^k K_i^{z_j} (a^{z_j}) = \sum_{k=0}^l q_k \left(\prod_{j=0}^{k-1} (1 - q_j) \right) \left(\sum_{m=0}^k K_i^{z_m} (a^{z_m}) \right),$$

где $a^{z_0} \in A^{z_0}, a^{z_1} \in A^{z_1}, \dots, a^{z_l} \in A^{z_l}$ — последовательность реализовавшихся ситуаций в соответствующих вершинах при стратегиях игроков $(\varphi_i : i \in N)$.

Так как переходы из вершин в последующие вершины являются случайными, то в качестве выигрыша игрока в стохастической игре будем рассматривать математическое ожидание выигрыша также относительно случайных переходов из вершин в следующие за ними вершины. Математическое ожидание $E_i(z_0, \varphi)$

выигрыша игрока i в игре удовлетворяет функциональному уравнению

$$\begin{aligned}
 E_i(z_0, \varphi) &= q_0 K_i^{z_0}(a^{z_0}) + (1 - q_0) \left(K_i^{z_0}(a^{z_0}) + \sum_{y \in L(z_0)} p(y|z_0, a^{z_0}) E_i(y, \varphi^y) \right) = \\
 &= K_i^{z_0}(a^{z_0}) + (1 - q_0) \sum_{y \in L(z_0)} p(y|z_0, a^{z_0}) E_i(y, \varphi^y),
 \end{aligned} \tag{1.2.1}$$

где $E_i(y, \varphi^y)$ — математическое ожидание выигрыша игрока i в подыгре $G(y)$, начинающейся в вершине $y \in L(z_0)$ графа $\Psi(z_0)$.

Пусть $z \in (L(z_0))^k$, то есть игровой процесс попадает в вершину $z \in Z$ на k -м шаге, тогда для математического ожидания выигрыша i -го игрока в подыгре $G(z)$ справедливо функциональное уравнение

$$\begin{aligned}
 E_i(z, \varphi^z) &= q_k K_i^z(a^z) + (1 - q_k) \left(K_i^z(a^z) + \sum_{y \in L(z)} p(y|z, a^z) E_i(y, \varphi^y) \right) = \\
 &= K_i^z(a^z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, a^z) E_i(y, \varphi^y).
 \end{aligned}$$

Для построения кооперативной игры необходимо определить набор кооперативных стратегий, реализация которых определяет так называемый «кооперативный путь» (один из кооперативных путей, если их несколько), то есть путь, который максимизирует суммарный выигрыш игроков. В случае стохастических игр — это поддерево с заданными вероятностями перехода, на которых достигается максимум математического ожидания суммарного выигрыша игроков во всей игре. Однако максимум математического ожидания суммарного выигрыша игроков в классе смешанных стратегий равен максимуму математического ожидания суммарного выигрыша игроков в классе чистых стратегий, поэтому для нахождения кооперативных стратегий в стохастической игре можно ограничиться классом чистых стратегий.

§ 1.3 Определение кооперативной стохастической игры конечной продолжительности

Обозначим через $\bar{\varphi} = (\bar{\varphi}_1, \dots, \bar{\varphi}_n)$ ситуацию в стохастической игре $G(z_0)$ в чистых стратегиях, которая максимизирует сумму математических ожиданий выигрышей игроков:

$$V(N, z_0) = \max_{\varphi \in \Sigma} \left[\sum_{i \in N} E_i(z_0, \varphi) \right] = \sum_{i \in N} E_i(z_0, \bar{\varphi}).$$

Назовем такую ситуацию *кооперативной* или *набором кооперативных стратегий*. Пусть ситуация $\bar{\varphi}$ такова, что $\bar{\varphi}_i(z) = \bar{a}_i^z$, $i \in N$, $z \in Z$. Можем определить набор кооперативных стратегий для любой подыгры $G(z)$, $z \in Z$, начинающейся с игры $\Gamma(z)$.

Построим кооперативный вариант стохастической игры на основе некооперативной стохастической игры со случайной продолжительностью $G(z_0)$, описанной выше. Для этого необходимо определить значения характеристической функции для каждого подмножества S (*коалиции*) множества игроков N . Характеристическую функцию, вычисленную для подыгры $G(z)$, $z \in Z$, обозначим через $V(S, z)$, где $S \subset N$.

Характеристическая функция $V(S, z)$ показывает «силу» коалиции S или суммарный выигрыш, который могут получить игроки, объединяясь в коалицию S . Существуют разные подходы к определению характеристической функции, определяющей кооперативную игру, на основе некооперативной игры. Выделим несколько подходов:

1. α -подход. Значение $V(S, z)$ — нижнее значение игры G_S с нулевой суммой между коалициями S и $N \setminus S$ (максимин). Причем, максимин находится в чистых стратегиях коалиции S . Этот подход можно считать «пессимистичным», поскольку $V(S, z)$ равно максимальному суммарному выигрышу коалиции S , который коалиция S может получить при наихудшем поведении коалиции $N \setminus S$. Этот подход был предложен в книге Неймана и Моргенштерна [146].

2. β -подход. Значение $V(S, z)$ — верхнее значение игры G_S с нулевой суммой между коалициями S и $N \setminus S$ (минимакс). Причем, минимакс находится в чистых стратегиях. Этот подход можно считать «оптимистичным». Сравнение α - и β -подходов можно найти в работе [69].
3. Значение игры G_S . В этом случае значение $V(S, z)$ равно значению игры G_S с нулевой суммой между коалициями S и $N \setminus S$. Причем, значение всегда существует в смешанных стратегиях, при этом оно равно максимуму и минимуму игры G_S в смешанных стратегиях. То есть в случае, когда минимакс и максимин находятся в классе смешанных стратегий, значения α - и β -характеристических функций совпадают.
4. γ -подход. Значение $V(S, z)$ равно выигрышу коалиции S в ситуации равновесия по Нэшу, когда игроки из коалиции S действуют как один игрок, максимизируя суммарный выигрыш, а все игроки, не входящие в S , играют индивидуально. Этот подход предложен в статье [81].
5. δ -подход. Значение $V(S, z)$ равно максимальному выигрышу коалиции S в ситуации, когда игроки, не входящие в коалицию S , играют равновесные по Нэшу стратегии, найденные в игре n лиц, когда все игроки играют индивидуально. Этот подход предложен в работе [182] и далее исследован в статье [195].
6. ζ -подход. Значение $V(S, z)$ равно выигрышу коалиции S в ситуации, когда игроки из коалиции S используют стратегии, максимизирующие суммарный выигрыш коалиции N , а игроки, не входящие в коалицию S , минимизируют суммарный выигрыш игроков из коалиции S (см. [14]).

В этой главе будем использовать α -подход и считать, что «сила» коалиции S равна нижнему значению стохастической игры двух лиц с нулевой суммой G_S между коалициями S и $N \setminus S$. Именно такой подход был использован в работе [178], в которой впервые предложен способ построения кооперативной стохастической игры на основе некооперативной и сформулирована проблема позиционной несостоятельности вектора Шепли.

Определим значения характеристической функции. Сначала рассмотрим случай, когда $S = N$, и найдем максимум суммарного выигрыша коалиции N в стохастической игре $G(z_0)$. С этой целью выпишем уравнение Беллмана (см. [75] или [3]) для максимума суммы математических ожиданий выигрышей игроков:

$$\begin{aligned} V(N, z_0) &= \max_{\substack{a_i^{z_0} \in A_i^{z_0} \\ i \in N}} \left[\sum_{i \in N} K_i^{z_0}(a^{z_0}) + (1 - q_0) \sum_{y \in L(z_0)} p(y|z_0, a^{z_0}) V(N, y) \right] = \quad (1.3.1) \\ &= \sum_{i \in N} K_i^{z_0}(\bar{a}^{z_0}) + (1 - q_0) \sum_{y \in L(z_0)} p(y|z_0, \bar{a}^{z_0}) V(N, y) \end{aligned}$$

с граничным условием

$$V(N, z) = \max_{\substack{a_i^z \in A_i^z \\ i \in N}} \sum_{i \in N} K_i^z(a^z), \quad z \in \{z : L(z) = \emptyset\}. \quad (1.3.2)$$

В дальнейшем, в этой главе, для определенности будем предполагать, что $z \in (L(z_0))^k$.

Для подыгры $G(z)$, $z \in Z$, уравнение (1.3.1) с граничным условием (1.3.2) принимает вид:

$$\begin{aligned} V(N, z) &= \max_{\substack{a_i^z \in A_i^z \\ i \in N}} \left[\sum_{i \in N} K_i^z(a^z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, a^z) V(N, y) \right] = \quad (1.3.3) \\ &= \sum_{i \in N} K_i^z(\bar{a}^z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(N, y) \end{aligned}$$

с граничным условием

$$V(N, z) = \max_{\substack{a_i^z \in A_i^z \\ i \in N}} \sum_{i \in N} K_i^z(a^z), \quad z \in \{z : L(z) = \emptyset\}. \quad (1.3.4)$$

Ситуация $(\bar{\varphi}_i : i \in N)$ в стохастической игре $G(z_0)$ порождает вероятностные распределения на множестве Z вершин графа $\Psi(z_0)$.

Определение 1.2. Подграф графа $\Psi(z_0)$, который состоит из вершин $z \in Z$ графа $\Psi(z_0)$, имеющих положительную вероятность реализации, порожденную кооперативной ситуацией $\bar{\varphi}$, назовем кооперативным поддеревом и обозначим через $\bar{\Psi}(z_0)$.

Очевидно, что подграф $\bar{\Psi}(z_0)$ является конечным древовидным графом. Множество вершин в графе $\bar{\Psi}(z_0)$ обозначим через $CZ \subseteq Z$.

Пусть теперь $S \subset N$, $S \neq N$. Для каждой вершины $z \in CZ$ определим вспомогательную игру с нулевой суммой, которую обозначим через $G_S(z)$. Это антагонистическая игра или игра с нулевой суммой между коалицией $S \subset N$, выступающей в качестве максимизирующего игрока, и коалицией $N \setminus S$, выступающей в качестве минимизирующего игрока. При этом, выигрыш коалиции S определяется как сумма выигрышей игроков, входящих в коалицию S . Тогда значение характеристической функции $V(S, z)$ зададим как нижнее значение антагонистической игры $G_S(z)$ в чистых стратегиях (аналогично нижнему значению матричной игры)¹.

Функция $V(S, z)$, $z \in CZ$, удовлетворяет функциональному уравнению

$$V(S, z) = \max_{a_S^z \in A_S^z} \min_{a_{N \setminus S}^z \in A_{N \setminus S}^z} \left[\sum_{i \in S} K_i^z(a_S^z, a_{N \setminus S}^z) + \right. \\ \left. + (1 - q_k) \sum_{y \in L(z)} p(y|z, (a_S^z, a_{N \setminus S}^z)) V(S, y) \right] \quad (1.3.5)$$

с граничным условием

$$V(S, z) = \max_{a_S^z \in A_S^z} \min_{a_{N \setminus S}^z \in A_{N \setminus S}^z} \sum_{i \in S} K_i^z(a_S^z, a_{N \setminus S}^z), \quad z \in \{z : L(z) = \emptyset\}, \quad (1.3.6)$$

где $a_S^z = (a_i^z : i \in S)$ — стратегия коалиции S , а $A_S^z = \prod_{i \in S} A_i^z$ — множество стратегий коалиции S , и $a_{N \setminus S}^z = (a_j^z : j \in N \setminus S)$ — стратегия коалиции $N \setminus S$, а $A_{N \setminus S}^z = \prod_{j \in N \setminus S} A_j^z$ — множество стратегий коалиции $N \setminus S$.

Для всех $z \in CZ$ естественно предположить, что

$$V(\emptyset, z) = 0. \quad (1.3.7)$$

Таким образом, для каждой подыгры $G(z)$, $z \in CZ$, мы определили характеристическую функцию $V(S, z)$, $S \subset N$. Характеристическая функция $V(S, z)$ определяется уравнением (1.3.3) с граничным условием (1.3.4), а также уравнением (1.3.5) с граничным условием (1.3.6) и равенством (1.3.7).

Характеристическая функция $V(S, z)$, определяемая формулами (1.3.3) — (1.3.7), супераддитивна по S , то есть для любой вершины $z \in CZ$ и любых

¹В этой главе мы используем α -подход для построения характеристической функции, предложенный Нейманом и Моргенштерном [146].

коалиций $S, P \subset N$, $S \cap P = \emptyset$, имеет место неравенство

$$V(S \cup P, z) \geq V(S, z) + V(P, z).$$

Доказательство этого утверждения можно найти, например, в книге [51].

Определение 1.3. Кооперативной стохастической игрой со случайной продолжительностью $\bar{G}(z_0)$, построенной по некооперативной стохастической игре $G(z_0)$, назовем пару $\langle N, V(S, z_0) \rangle$, где $V(S, z_0)$ — характеристическая функция, определенная по формуле (1.3.3) с граничным условием (1.3.4) для коалиции N , по формуле (1.3.5) с граничным условием (1.3.6) для коалиции $S \subset N$, $S \neq N$, $S \neq \emptyset$, и по формуле (1.3.7) для коалиции $S = \emptyset$.

Определение 1.4. Дележом в кооперативной стохастической игре $\bar{G}(z_0)$ будем называть вектор $\xi(z_0) = (\xi_1(z_0), \dots, \xi_n(z_0))$, удовлетворяющий двум свойствам:

1. Коллективной рациональности: $\sum_{i \in N} \xi_i(z_0) = V(N, z_0)$;
2. Индивидуальной рациональности: $\xi_i(z_0) \geq V(\{i\}, z_0)$ для всех $i \in N$.

Множество дележей (см. [10, 11, 12, 13, 51]) в кооперативной стохастической игре $\bar{G}(z_0)$ обозначим через $I(z_0)$.

Определение 1.5. Решением кооперативной стохастической игры $\bar{G}(z_0)$ будем называть подмножество $C(z_0)$ множества дележей $I(z_0)$.

Решения кооперативной игры можно условно разделить на одноточечные и множественные. К широко известным одноточечным решениям можно отнести вектор Шепли [211], N -ядро [206]. Самое известное множественное решение — C -ядро [99]. Предположим, что решение $C(z_0)$ кооперативной стохастической игры $\bar{G}(z_0)$ является непустым.

Определение 1.6. Кооперативной подыгрой $\bar{G}(z)$, $z \in Z$, игры $\bar{G}(z_0)$, построенной по некооперативной стохастической подыгре $G(z)$, будем называть пару $\langle N, V(S, z) \rangle$, где $V(S, z)$ — характеристическая функция, определенная по формуле (1.3.3) с граничным условием (1.3.4) для коалиции N , по формуле (1.3.5) с граничным условием (1.3.6) для коалиции $S \subset N$, $S \neq N$, $S \neq \emptyset$, и по формуле (1.3.7) для коалиции $S = \emptyset$.

Определим дележ, множество дележей и решение для произвольной кооперативной подыгры $\bar{G}(z)$, $z \in Z \setminus z_0$.

Определение 1.7. Дележом в кооперативной подыгре $\bar{G}(z)$ будем называть вектор $\xi(z) = (\xi_1(z), \dots, \xi_n(z))$, удовлетворяющий двум свойствам:

1. $\sum_{i \in N} \xi_i(z) = V(N, z)$,
2. $\xi_i(z) \geq V(\{i\}, z)$ для всех $i \in N$.

Множество дележей в кооперативной подыгре $\bar{G}(z)$ обозначим через $I(z)$.

Определение 1.8. Решением кооперативной подыгры $\bar{G}(z)$ будем называть любое фиксированное подмножество $C(z)$ множества дележей $I(z)$.

Предположим, что решение $C(z)$ любой кооперативной подыгры $\bar{G}(z)$ является непустым для всех $z \in CZ$.

§ 1.4 Вектор Шепли, s -ядро и n -ядро

В этом разделе дадим определения решений, которые будут использованы далее в работе. Вектором Шепли кооперативной стохастической игры или подыгры $\bar{G}(z)$, $z \in CZ$, будем называть вектор $Sh(z) = (Sh_1(z), \dots, Sh_n(z))$, элемент $Sh_i(z)$, $i \in N$, которого вычисляется по формуле

$$Sh_i(z) = \sum_{\substack{S \subset N \\ S \ni i}} \frac{(|S| - 1)! (n - |S|)!}{n!} (V(S, z) - V(S \setminus \{i\}, z)),$$

где $|S|$ — мощность множества S . Определение вектора Шепли сформулировано в статье Шепли [211].

S -ядром кооперативной стохастической игры или подыгры $\bar{G}(z)$, $z \in CZ$, которое обозначим через $CO(z)$, будем называть множество

$$CO(z) = \left\{ \xi(z) \in I(z) : \sum_{i \in S} \xi_i(z) \geq V(S, z) \text{ для любого } S \subset N \right\}.$$

Для кооперативной стохастической игры или подыгры $\bar{G}(z)$ и любого вектора $\xi(z) \in I(z)$ обозначим через $\theta(\xi(z))$ вектор значений эксцессов $e(S, \xi(z)) = V(S, z) - \sum_{i \in S} \xi_i(z)$, расположенных в порядке убывания:

$$\theta(\xi(z)) = (e(S_1, \xi(z)), e(S_2, \xi(z)), \dots, e(S_{2^n-1}, \xi(z))),$$

где коалиции перенумерованы таким образом, что $e(S_1, \xi(z)) \geq e(S_2, \xi(z)) \geq \dots \geq e(S_{2^n-1}, \xi(z))$.

На множестве эксцессов $\{\theta(\xi(z)) : \xi(z) \in I(z)\}$ рассмотрим отношение лексикографического упорядочения \succ_{lex} :

$$\theta(\xi(z)) \succ_{lex} \theta(\psi(z)) \iff \exists l \in \{1, \dots, 2^n\},$$

такое что

$$\begin{cases} \theta_k(\xi(z)) = \theta_k(\psi(z)), & \text{для всех } k = 1, \dots, l-1; \\ \theta_l(\xi(z)) > \theta_l(\psi(z)), \end{cases}$$

где $\psi(z) \in I(z)$.

N -ядром кооперативной стохастической игры или подыгры $\bar{G}(z)$, $z \in CZ$, называется подмножество множества дележей, на котором достигается $\min_{\substack{\theta(\xi(z)) \\ \xi(z) \in I(z)}} \succ_{lex}$. Определение n -ядра впервые представлено Шмайдлером [206].

Определение и некоторые полезные теоремы и леммы о свойствах n -ядра можно найти в [51, 87, 113]. Работы [114, 144] посвящены нахождению n -ядра. N -ядро кооперативной игры состоит из единственного вектора.

Если $C(z_0)$ — это решение кооперативной стохастической игры $\bar{G}(z_0)$, то далее в работе под решением $C(z)$ кооперативной подыгры $\bar{G}(z)$ будем понимать решение, построенное по тем же правилам, что и $C(z_0)$. Например, если $C(z_0)$ — это вектор Шепли в кооперативной игре $\bar{G}(z_0)$, то $C(z)$ — это вектор Шепли в кооперативной подыгре $\bar{G}(z)$, $z \in CZ$.

Здесь предполагается, что игроки выбирают для себя какое-то фиксированное подмножество множества дележей, обладающее «оптимальными» для них свойствами, то есть игроки, объединившись в коалицию N , собираются следовать некоторому правилу распределения выигрышей в течение всего игрового процесса. Множество $C(z)$ может состоять из единственного дележа, если, например, игроки решили использовать вектор Шепли или n -ядро, или c -ядро состоит из единственной точки. Случай, когда множество дележей пусто (например, пустым может оказаться c -ядро), мы не рассматриваем в данной работе. Решением игры или подыгры $\bar{G}(z)$ могут быть и любые другие решения из классической «статической» кооперативной теории, такие как NM -решение,

k -ядро, \mathcal{M} -устойчивые множества (см. [51]).

Далее в работе будем предполагать, что $C(z)$ является непустым подмножеством множества $I(z)$ для любого $z \in CZ$, то есть для каждой вершины $z \in CZ$ существует по крайней мере один дележ

$$\xi(z) = (\xi_1(z), \dots, \xi_n(z)) \in C(z) \subset I(z).$$

§ 1.5 Процедура распределения дележа в стохастических играх конечной продолжительности

Введем понятие процедуры распределения дележа [43] из решения кооперативной игры, выбранного игроками, которая определяет выплаты игрокам в каждой вершине кооперативного поддерева $\bar{\Psi}(z_0)$.

Обозначим через a^{z_0, \dots, z_l} набор ситуаций $a^{z_0}, a^{z_1}, \dots, a^{z_l}$, реализовавшихся на пути z_0, \dots, z_l , где a^{z_i} — это ситуация в игре $\Gamma(z_i)$, $z_i \in L(z_{i-1})$, $i = 1, \dots, l$.

Рассмотрим произвольную вершину $z \in CZ$, $z \in (L(z_0))^k$, кооперативного поддерева. Игрок получает некоторые выплаты при реализации вектора кооперативных стратегий ². Пусть в вершине $z \in CZ$ выплата игроку $i \in N$ равна $\beta_i(z)$. В любой кооперативной подыгре $\bar{G}(z)$ игрок может посчитать эти выплаты при \bar{a}^{z, \dots, z_l} на отрезке пути z, \dots, z_l . Причем сумма выплат в стохастической игре является случайной величиной. Обозначим через $B_i(z)$ математическое ожидание сумм таких выплат в кооперативной подыгре $\bar{G}(z)$. Величина $B_i(z)$ удовлетворяют следующему функциональному уравнению:

$$B_i(z) = \beta_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) B_i(y) \quad (1.5.1)$$

с граничным условием

$$B_i(z) = \beta_i(z) \text{ для } z \in \{z : L(z) = \emptyset\}. \quad (1.5.2)$$

Теперь определим процедуру распределения дележа из кооперативного решения $C(z_0)$, выбранного игроками в начале игры.

²Очевидно, что все $z, \dots, z_l \in CZ$, так как CZ — множество вершин кооперативного поддерева, и ситуация $\bar{\varphi}$ фиксирована.

Определение 1.9. Пусть $\xi(z_0) = (\xi_1(z_0), \dots, \xi_n(z_0)) \in C(z_0)$. Набор векторов $\{\beta(z) = (\beta_1(z), \dots, \beta_n(z)) : z \in CZ\}$ назовем процедурой распределения дележа (ПРД) $\xi(z_0)$, если выполнены условия:

1. Для любой вершины $z \in CZ$:

$$\sum_{i \in N} \beta_i(z) = \sum_{i \in N} K_i^z(\bar{a}^z).$$

2. Компоненты $\xi_i(z_0)$, $i \in N$, дележа $\xi(z_0)$ совпадают с математическим ожиданием сумм соответствующих компонент процедуры распределения дележа относительно вероятностного распределения переходов и окончания игры, т. е. $\xi_i(z_0) = B_i(z_0)$, где $B_i(z_0)$ удовлетворяет функциональному уравнению (1.5.1) с граничным условием (1.5.2).

Для каждой кооперативной подыгры $\bar{G}(z)$, $z \in CZ$, запишем функциональное уравнение для компонент $\xi_i(z)$ дележа $\xi(z) \in C(z) \subset I(z)$ типа (1.5.1) и определим величины $\gamma_i(z)$ из уравнения

$$\xi_i(z) = \gamma_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \xi_i(y), \quad (1.5.3)$$

где $\xi(y) = (\xi_i(y) : i \in N)$ — некоторый дележ, принадлежащий решению $C(y)$ кооперативной подыгры $\bar{G}(y)$. Граничное условие для $\gamma_i(z)$ будет следующим:

$$\gamma_i(z) = \xi_i(z) \text{ для } z \in \{z : L(z) = \emptyset\}. \quad (1.5.4)$$

Лемма 1.1. Вектор $\gamma(z) = (\gamma_i(z) : i \in N)$, определяемый формулой (1.5.3) с граничным условием (1.5.4), является процедурой распределения дележа.

Доказательство. Очевидно, что для терминальных вершин и вершин, в которых вероятность окончания равна единице, имеет место равенство (1.5.4), условия 1 и 2 определения 1.9 выполнены.

Теперь покажем выполнение этих условий для остальных вершин кооперативного поддерева. Из (1.5.3) выразим величины $\gamma_i(z)$ и суммируем их по $i \in N$, получаем

$$\sum_{i \in N} \gamma_i(z) = \sum_{i \in N} \xi_i(z) - (1 - q_k) \sum_{i \in N} \left(\sum_{y \in L(z)} p(y|z, \bar{a}^z) \xi_i(y) \right). \quad (1.5.5)$$

Так как

$$\begin{aligned}\xi(z) &= (\xi_i(z) : i \in N) \in C(z) \subset I(z), \\ \xi(y) &= (\xi_i(y) : i \in N) \in C(y) \subset I(y),\end{aligned}$$

то из (1.5.5) получаем

$$\sum_{i \in N} \gamma_i(z) = V(N, z) - (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(N, y). \quad (1.5.6)$$

Из (1.5.6) и (1.3.3) следует, что $\sum_{i \in N} \gamma_i(z) = \sum_{i \in N} K_i^z(\bar{a}^z)$ для ситуации $\bar{a}^z = (\bar{a}_i^z : i \in N)$, которая реализовалась в игре $\Gamma(z)$ при использовании игроками вектора кооперативных стратегий $\bar{\varphi}$. Получается, что $\gamma_i(z)$ удовлетворяет условию 1 определения 1.9.

Теперь проверим выполнение условия 2 определения 1.9, а именно, найдем математическое ожидание сумм $\gamma_i(z)$, определяемых формулой (1.5.3), вдоль вершин кооперативного поддерева. Для вершин $z \in \{z : L(z) = \emptyset \text{ или } q_k = 1\}$ условие 2 выполнено. Перейдем к вершинам кооперативного поддерева, из которых за один шаг можно попасть в вершины, упомянутые выше. Для них получаем равенство:

$$\begin{aligned}B_i(z_l) &= \xi_i(z_l) - (1 - q_l) \sum_{y \in L(z_l)} p(y|z_l, \bar{a}^{z_l}) \xi_i(y) + (1 - q_l) \sum_{y \in L(z_l)} p(y|z_l, \bar{a}^{z_l}) \gamma_i(y) = \\ &= \xi_i(z_l),\end{aligned}$$

поскольку $\xi_i(y) = \gamma_i(y)$. Следуя от терминальных вершин к начальной последовательно доказываем, что условие 2 определения 1.9 выполнено. Лемма доказана. \square

§ 1.6 Позиционная состоятельность решения кооперативной стохастической игры конечной продолжительности

Игроки перед началом игры приходят к соглашению о кооперации, то есть договариваются максимизировать математическое ожидание суммарного выигрыша и рассчитывают получить дележ $\xi(z_0) \in C(z_0)$. Развитию игры во времени соответствует движение вдоль вершин кооперативного поддерева $\bar{\Psi}(z_0)$.

Однако поскольку стохастическая структура игры подразумевает неоднозначность в реализации вершин кооперативного поддерева, то движение вдоль вершин кооперативного поддерева ещё не обеспечивает сохранение кооперации. Действительно, при движении вдоль пути игроки попадают в кооперативные подыгры с текущими начальными состояниями, в которых один и тот же игрок имеет различные возможности. Со временем изменяются условия конфликта и возможности участвующих в нем сторон. Но было бы разумным потребовать сохранения принципа оптимальности или «подхода» в выборе решений кооперативных подыгр. Условие сохранения принципа оптимальности нарушается, если в некоторый момент времени, в вершине $z \in CZ$, сумма оставшихся выплат для игрока i не равняется i -ой компоненте дележа из решения $C(z)$ кооперативной подыгры $\bar{G}(z)$. Следовательно, в вершине $z \in CZ$ перед игроком i может встать вопрос о том, стоит ли ему далее придерживаться намеченного перед началом игры соглашения действовать «совместно оптимально» или же игроку i стоит отклониться от кооперативного решения. Если такое отклонение будет выгодно хотя бы для одного из игроков, то это и будет означать позиционную несостоятельность дележа $\xi(z_0) \in C(z_0)$.

Определение 1.10. Дележ $\xi(z_0) \in C(z_0)$ называется позиционно состоятельным в кооперативной стохастической игре $\bar{G}(z_0)$, если для каждой вершины $z \in CZ \cap (L(z_0))^k$ существует процедура распределения дележа $\beta(z) = (\beta_i(z) : i \in N)$ такая, что

$$\xi_i(z) = \beta_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \xi_i(y), \quad (1.6.1)$$

и

$$\xi_i(z) = \beta_i(z), \quad z \in \{z : L(z) = \emptyset\}, \quad (1.6.2)$$

где $\xi(y) = (\xi_i(y) : i \in N)$ — некоторый дележ, принадлежащий решению $C(y)$ кооперативной подыгры $\bar{G}(y)$.

Замечание 1.2. Если $C(z_0)$ состоит более чем из одной точки, то в выборе дележа $\xi(z_0)$ есть неоднозначность. Выбрав определенный дележ $\xi(z_0) \in C(z_0)$ и решив проверить его на позиционную состоятельность, необходимо проверить условие (1.6.1) сначала для вершины z_0 , то есть проверить существует

ли ПРД $\beta(z_0) = (\beta_i(z_0) : i \in N)$, удовлетворяющая условию (1.6.1) для некоторого дележа $\xi(y) \in C(y)$, где $y \in L(z_0)$. Очевидно существует неоднозначность в выборе дележа $\xi(y) \in C(y)$, который в свою очередь должен быть также позиционно состоятелен в кооперативной подыгре $\bar{G}(y)$, то есть для него также должно выполняться условие (1.6.1), что уточняется в определении, когда говорится, что условие выполняется для всех z из множества вершин кооперативного поддерева.

Определение 1.11. Будем говорить, что кооперативная стохастическая игра $\bar{G}(z_0)$ имеет позиционно состоятельное решение $C(z_0)$, если все дележи $\xi(z_0) \in C(z_0)$ являются позиционно состоятельными.

Очевидно, что если производить выплаты игрокам в вершинах кооперативного поддерева в соответствии с изначально заданными функциями выигрышей, невозможно в общем случае добиться позиционной состоятельности кооперативного решения. Это может привести к подрыву кооперативного соглашения. В связи с этим естественной становится задача нахождения схемы или процедуры выплат игрокам в вершинах кооперативного поддерева для удовлетворения свойства позиционной состоятельности выбранного кооперативного решения. Для этого следует найти такую ПРД $(\beta_i : i \in N)$ для всех $z \in CZ$, для которой условия (1.6.1) и (1.6.2) будут выполнены.

Теорема 1.1. Пусть в кооперативной стохастической игре $\bar{G}(z_0)$ и любой ее подыгре кооперативные решения $C(z_0)$ и $C(z)$, $z \in CZ \setminus z_0$, непусты. Если для каждого $\xi(z) = (\xi_i(z) : i \in N) \in C(z)$ процедура распределения дележа определена по формуле

$$\beta_i(z) = \xi_i(z) - (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \xi_i(y), \quad (1.6.3)$$

для всех $z \in CZ$, $z \notin \{z : L(z) = \emptyset\}$, где $\xi(y) = (\xi_i(y) : i \in N) \in C(y)$, и по формуле (1.6.2) для всех $z \in \{z : L(z) = \emptyset\}$, то кооперативное решение $C(z_0)$ является позиционно состоятельным.

Доказательство. Для доказательства позиционной состоятельности кооперативного решения $C(z_0)$ требуется доказать, что для каждого вектора $\xi(z_0) \in C(z_0)$ выполняются условия (1.6.1) и (1.6.2).

Из леммы 1.1 следует, что выплаты, определенные формулами (1.6.3) и (1.6.2), являются ПРД. Условие (1.6.1) следует из (1.6.3), учитывая, что $\xi(y) = (\xi_i(y) : i \in N)$ — элемент кооперативного решения подыгры $\bar{G}(y)$. \square

Предлагаемый способ реализации дележа обладает важным свойством: в каждой вершине кооперативного пути игроки ориентируются на один и тот же «принцип оптимальности» (свойство позиционной состоятельности) и, в этом смысле, не имеют оснований для нарушения ранее принятого кооперативного соглашения и отклонения от кооперативного решения. Также сумма выплат игрокам в каждой вершине кооперативного поддерева равна сумме выигрышей, полученных игроками в этой вершине (условие 1 определения 1.9 процедуры распределения дележа). Последнее условие можно назвать условием достижимости выплат, поскольку игроки перераспределяют сумму, которую «зарабатывают» и не берут средства извне.

Заметим, что в определении 1.10 не требовалась неотрицательность функций $\beta_i(z)$, где $z \in CZ$. Все дележи, принадлежащие решению $C(z)$, будут являться позиционно состоятельными, если решение $C(z) \neq \emptyset$ для всех вершин $z \in CZ$ при условии, что выплаты игрокам производятся не в соответствии с их функциями выигрышей в играх, по которым проходит кооперативный путь, а в соответствии с кооперативной процедурой распределения дележа $\beta(z) = (\beta_1(z), \dots, \beta_n(z))$, определенной формулами (1.6.1), (1.6.2) для всех $z \in CZ$, где $\beta_i(z)$ — это выплата i -му игроку в вершине $z \in CZ$. При этом математическое ожидание всех выплат i -му игроку будет совпадать с математическим ожиданием i -ой компоненты выбранного игроками дележа из решения, что следует из Теоремы 1.1. Таким образом, игроки могут пойти и на получение в каких-то вершинах отрицательных выплат, чтобы гарантировать сохранение кооперации на протяжении всей игры для того, чтобы во всей игре гарантировать себе получение компонент заранее выбранного дележа $\xi(z_0)$, принадлежащего решению $C(z_0)$ кооперативной стохастической игры $\bar{G}(z_0)$.

§ 1.7 Неотрицательность компонент процедуры распределения дележа в стохастических играх конечной продолжительности. Регуляризация дележей

В этом разделе рассмотрим случай, когда функция выигрыша любого игрока $i \in N$ неотрицательна: $K_i^z(a^z) \geq 0$ для всех вершин $z \in CZ$. Предположим, что игроки заинтересованы в получении неотрицательных выплат в каждой вершине кооперативного поддерева и при этом хотят гарантировать позиционную состоятельность кооперативного решения. В случае, когда нельзя гарантировать неотрицательность $\beta_i(z)$ для всех вершин $z \in CZ$, используя процедуру распределения дележа, описанную в предыдущем параграфе, можно пойти по пути построения нового позиционно состоятельного решения на основе заранее выбранного игроками решения из множества $C(z_0)$. Покажем, как это делается, когда в качестве решения рассматривается множество $C(z_0) \subset I(z_0)$. Заметим, что данная процедура может быть применена для дележей, известных в классической статической кооперативной теории (C -ядро, N -ядро, вектор Шепли).

Для всех вершин $z \in CZ$ определим новую ПРД по формуле

$$\beta_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}_1^z, \dots, \bar{a}_n^z)}{V(N, z)} \xi_i(z), \quad (1.7.1)$$

где $\xi(z) = (\xi_1(z), \dots, \xi_n(z)) \in C(z)$, а $\bar{a}^z = (\bar{a}_1^z, \dots, \bar{a}_n^z)$ — ситуация в игре $\Gamma(z)$, полученная при реализации вектора кооперативных стратегий $\bar{\varphi} = (\bar{\varphi}_1(\cdot), \dots, \bar{\varphi}_n(\cdot))$ в вершине $z \in CZ$, $V(N, z)$ — значение характеристической функции для коалиции N , вычисленное для кооперативной подыгры $\bar{G}(z)$.

Поскольку $K_i^z(a^z) \geq 0$ для всех вершин $z \in CZ$ и всех $i \in N$, то $\beta_i(z) \geq 0$ для любой вершины $z \in CZ$. Из (1.7.1) и того, что $\sum_{i \in N} \xi_i(z) = V(N, z)$, следует также, что текущая выплата $\beta_i(z)$ игроку i в игре $\Gamma(z)$ должна быть пропорциональна i -ой компоненте дележа $\xi(z) \in C(z)$ в кооперативной подыгре $\bar{G}(z)$ стохастической игры $\bar{G}(z_0)$.

Определим новый дележ в кооперативной подыгре $\bar{G}(z)$, $z \in CZ$, и $z \in (L(z_0))^k$ на основе «старого» дележа $\xi(z)$ как решение функционального урав-

нения

$$\hat{\xi}_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{\xi}_i(y) \quad (1.7.2)$$

с граничным условием

$$\hat{\xi}_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \xi_i(z) = \xi_i(z) \quad (1.7.3)$$

для $z \in \{z : L(z) = \emptyset\}$.

Построим новую характеристическую функцию $\hat{V}(S, z)$ для каждой кооперативной подыгры $\bar{G}(z)$ для всех $z \in CZ$, используя функциональное уравнение

$$\hat{V}(S, z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} V(S, z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{V}(S, y) \quad (1.7.4)$$

с граничным условием

$$\hat{V}(S, z) = V(S, z) \text{ для } z \in \{z : L(z) = \emptyset\}. \quad (1.7.5)$$

Функции $\hat{V}(S, z)$ и $V(S, z)$ супераддитивны, и $\hat{V}(N, z) = V(N, z)$, поскольку $\hat{V}(N, z)$ и $V(N, z)$ удовлетворяют функциональному уравнению (1.3.3) с начальным условием (1.3.4).

Для всех вершин $z \in CZ$ вычислим дележи $\hat{\xi}(z)$ и найдем множество решений $\hat{C}(z)$ следующим образом:

$$\hat{C}(z) = \left\{ \hat{\xi}(z) : \hat{\xi}_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{\xi}_i(y), \right. \quad (1.7.6)$$

$$\left. \hat{\xi}_i(z) = \xi_i(z) \text{ для } z \in \{z : L(z) = \emptyset\}, \xi(z) \in C(z) \right\}.$$

Определение 1.12. Множество $\hat{C}(z_0)$, определенное формулой (1.7.6), назовем регуляризованным решением кооперативной стохастической игры $\bar{G}(z_0)$.

Таким образом, у игроков есть возможность регуляризовать выбранное в начале игры решение, чтобы в каждой вершине стохастической игры $\bar{G}(z_0)$ «новое» решение $\hat{C}(z_0)$ было позиционно состоятельным, и выплаты игрокам были бы неотрицательными. Но дележ, принадлежащий новому регуляризованному

решению $\hat{C}(z_0)$, вообще говоря, не будет являться дележом для кооперативной игры с характеристической функцией $V(S, z_0)$, определенной по формулам (1.3.5), (1.3.6), а будет являться дележом для кооперативной стохастической игры с новой характеристической функцией $\hat{V}(S, z_0)$, построенной по формулам (1.7.4), (1.7.5).

Теорема 1.2. Дележ $\hat{\xi}(z) = (\hat{\xi}_1(z), \dots, \hat{\xi}_n(z))$, определенный формулой (1.7.2) с граничным условием (1.7.3), является позиционно состоятельным дележом кооперативной игры $\langle N, \hat{V} \rangle$, где характеристическая функция $\hat{V}(S, z)$ задана функциональным уравнением (1.7.4) с граничным условием (1.7.5). При этом процедура распределения дележа, определенная формулой (1.7.1), неотрицательна для любых $i \in N$ и $z \in CZ$.

Доказательство. Позиционная состоятельность следует из построения «нового дележа» $\hat{\xi}(z)$. Неотрицательность компоненты

$$\frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \xi_i(z),$$

очевидна, поскольку

$$K_i^z(a_1^z, \dots, a_n^z) \geq 0$$

для всех $z \in Z$ и любого игрока $i \in N$.

Теперь покажем, что $\hat{\xi}(z) = (\hat{\xi}_1(z), \dots, \hat{\xi}_n(z))$ обладает свойствами дележа для кооперативной игры с характеристической функцией $\hat{V}(S, z)$, которая задана функциональным уравнением (1.7.4) с граничным условием (1.7.5). Для этого для любого игрока $i \in N$ и любой вершины $z \in CZ$ необходимо доказать выполнение двух свойств:

1. $\sum_{i \in N} \hat{\xi}_i(z) = \hat{V}(N, z)$,
2. $\hat{\xi}_i(z) \geq \hat{V}(\{i\}, z)$.

Первое свойство очевидным образом выполняется для вершин $z \in \{z : L(z) = \emptyset\}$ и таких, что $z \in CZ$. Теперь докажем его для вершин

$z \in \{z : L(z) \ni y \text{ и } L(y) = \emptyset\}$ и таких, что $z \in CZ$:

$$\begin{aligned} \sum_{i \in N} \hat{\xi}_i(z) &= \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \sum_{i \in N} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} \left(p(y|z, \bar{a}^z) \sum_{i \in N} \hat{\xi}_i(y) \right) = \\ &= \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} V(N, z) + (1 - q_k) \sum_{y \in L(z)} \left(p(y|z, \bar{a}^z) \hat{V}(N, z) \right) = \\ &= V(N, z) = \hat{V}(N, z), \end{aligned}$$

так как вершина $y \in \{y : L(y) = \emptyset\}$.

Второе свойство также очевидно выполняется для вершин $z \in \{z : L(z) = \emptyset \text{ или } q_k = 1\}$. Покажем, что $\hat{\xi}_i(z) - \hat{V}(\{i\}, z) \geq 0$ для вершин $z \in \{z : L(z) \ni y \text{ и } L(y) = \emptyset\}$, используя формулы (1.7.2) и (1.7.4):

$$\begin{aligned} \hat{\xi}_i(z) - \hat{V}(\{i\}, z) &= \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{\xi}_i(y) - \\ &- \left\{ \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} V(\{i\}, z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{V}(\{i\}, y) \right\} = \\ &= \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} (\xi_i(z) - V(\{i\}, z)) + \\ &+ (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \left(\hat{\xi}_i(y) - \hat{V}(\{i\}, y) \right) \geq 0. \end{aligned}$$

Первое слагаемое неотрицательно, так как $\xi(z)$ является дележом кооперативной подыгры $\bar{G}(z)$, а второе слагаемое неотрицательно, потому что вершина $y \in \{y : L(y) = \emptyset\}$. Рекуррентно доказываем для предыдущих вершин и таких что $z \in CZ$ и до вершины z_0 . \square

Важно, в каком соотношении находится множество $\hat{C}(z)$ — регуляризованное решение, определенное по формуле (1.7.6), и множество $\tilde{C}(z)$ — решение, найденное для кооперативной подыгры $\bar{G}(z)$ с характеристической функцией $\hat{V}(S, z)$ (то есть, построенное по тем же правилам, что и решение $C(z) \subset I(z)$ для кооперативной подыгры $\bar{G}(z)$). Найдем множества $\tilde{C}(z)$ и $\hat{C}(z)$ для кооперативной стохастической подыгры $\bar{G}(z)$, когда в качестве решений стохастической игры $\bar{G}(z_0)$ выбраны вектор Шепли и s -ядро.

§ 1.8 Регуляризация вектора Шепли и s -ядра в стохастических играх конечной продолжительности

Сначала рассмотрим случай, когда в качестве решения игроки выбрали одноточечный принцип оптимальности — вектор Шепли. Вектор Шепли, вычисленный для стохастической игры $\bar{G}(z_0)$, обозначим через $Sh(z_0) = (Sh_i(z_0) : i \in N)$, а для кооперативной подыгры $\bar{G}(z)$, где $z \in CZ$, через $Sh(z) = (Sh_i(z) : i \in N)$.

Определим регуляризованный вектор Шепли в кооперативной подыгре $\bar{G}(z)$, где $z \in CZ$, и $z \in (L(z_0))^k$ на основе вектора Шепли, посчитанного для изначально заданной игры, как решение функционального уравнения

$$\hat{Sh}_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} Sh_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{Sh}_i(y) \quad (1.8.1)$$

с граничным условием

$$\hat{Sh}_i(z) = Sh_i(z) \quad (1.8.2)$$

для $z \in \{z : L(z) = \emptyset\}$.

Справедлива следующая теорема.

Теорема 1.3. *Вектор, удовлетворяющий функциональному уравнению (1.8.1) с граничным условием (1.8.2), позиционно состоятелен и является вектором Шепли для кооперативной подыгры $\langle N, \hat{V}(\cdot, z) \rangle$, $z \in CZ$ стохастической игры $\langle N, \hat{V}(\cdot, z_0) \rangle$, где значения характеристической функции $\hat{V}(\cdot, z)$ вычисляются по формулам (1.7.4) и (1.7.5).*

Замечание 1.3. *Теорема 1.3 дает соотношение между множествами $\tilde{C}(z)$ и $\hat{C}(z)$, о которых шла речь в конце предыдущего параграфа. Если в качестве решения в стохастической игре $\bar{G}(z_0)$ выбран вектор Шепли, то $\tilde{C}(z) = \hat{C}(z)$ для любого $z \in CZ$. Поэтому можно привести эквивалентную формулировку Теоремы 1.2.*

Теорема 1.4. *Вектор, удовлетворяющий функциональному уравнению (1.8.1) с граничным условием (1.8.2), позиционно состоятелен, и $\hat{Sh}(z_0) = \hat{C}(z_0) = \tilde{C}(z_0)$, где $\hat{C}(z_0)$ — регуляризованное решение, удовлетворяющее (1.7.6), $\tilde{C}(z_0)$*

— это вектор Шепли, вычисленный для кооперативной стохастической игры $\langle N, \hat{V}(\cdot, z_0) \rangle$ с характеристической функцией, определенной формулами (1.7.4), (1.7.5).

Доказательство. То, что вектор, удовлетворяющий функциональному уравнению (1.8.1) с начальным условием (1.8.2), позиционно состоятелен, следует из Теоремы 1.1, которая сформулирована для произвольного решения $C(z)$.

Вычислим вектор Шепли для кооперативной стохастической подыгры $\langle N, \hat{V}(\cdot, z) \rangle$, $z \in CZ$, с регуляризованной характеристической функцией, определенной формулами (1.7.4), (1.7.5):

$$\hat{S}h_i(z) = \sum_{\substack{S \subset N \\ S \ni i}} \frac{(|S| - 1)! (n - |S|)!}{n!} \left(\hat{V}(S, z) - \hat{V}(S \setminus \{i\}, z) \right).$$

Перепишем (1.7.4) для коалиции $S \setminus \{i\}$, получаем

$$\begin{aligned} \hat{V}(S \setminus \{i\}, z) = & \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} V(S \setminus \{i\}, z) + \\ & + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{V}(S \setminus \{i\}, y). \end{aligned} \quad (1.8.3)$$

Вычитая (1.8.3) из (1.7.4), умножая на $\frac{(|S|-1)!(n-|S|)!}{n!}$, суммируя по всем возможным коалициям $S \subset N$, таким, что $S \ni i$, получаем

$$\begin{aligned} & \sum_{\substack{S \subset N \\ S \ni i}} \frac{(|S| - 1)! (n - |S|)!}{n!} \left[\hat{V}(S, z) - \hat{V}(S \setminus \{i\}, z) \right] = \quad (1.8.4) \\ & = \left\{ \sum_{\substack{S \subset N \\ S \ni i}} \frac{(|S| - 1)! (n - |S|)!}{n!} [V(S, z) - V(S \setminus \{i\}, z)] \right\} \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} + \\ & + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \times \\ & \times \left\{ \sum_{\substack{S \subset N \\ S \ni i}} \frac{(|S| - 1)! (n - |S|)!}{n!} \left[\hat{V}(S, y) - \hat{V}(S \setminus \{i\}, y) \right] \right\} = \\ & = Sh_i(z) \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{S}h_i(y). \end{aligned}$$

Из (1.8.4) и (1.8.1) следует утверждение теоремы. \square

Теперь предположим, что в качестве решения кооперативной стохастической игры $\bar{G}(z_0)$ выбрано c -ядро. Также как и ранее в работе, будем предполагать, что $CO(z) \neq \emptyset$ для любой вершины $z \in CZ$. Предположим также, что $CO(z_0)$ не является позиционно состоятельным, то есть по крайней мере для одного дележа $\xi(z_0) \in CO(z_0)$ нарушается условие позиционной состоятельности.

Определение 1.13. *Регуляризованным c -ядром стохастической игры $\bar{G}(z_0)$ называется множество $\widehat{CO}(z_0)$:*

$$\widehat{CO}(z_0) = \left\{ \hat{\xi}(z_0) : \hat{\xi}_i(z_0) = \frac{\sum_{i \in N} K_i^{z_0}(\bar{a}^{z_0})}{V(N, z_0)} \xi_i(z_0) + (1 - q_0) \sum_{y \in L(z_0)} (p(y|z_0, \bar{a}^{z_0}) \hat{\xi}_i(y)), \right. \\ \left. \hat{\xi}_i(z_0) = \xi_i(z_0), z_0 \in \{z : L(z) = \emptyset\}, \xi(z_0) \in CO(z_0) \right\}. \quad (1.8.5)$$

Определение 1.14. *Регуляризованным c -ядром кооперативной подыгры $\bar{G}(z)$ называется множество $\widehat{CO}(z)$:*

$$\widehat{CO}(z) = \left\{ \hat{\xi}(z) : \hat{\xi}_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{\xi}_i(y) \right. \\ \left. \hat{\xi}_i(z) = \xi_i(z) \text{ для } z \in \{z : L(z) = \emptyset\}, \xi(z) \in CO(z) \right\}. \quad (1.8.6)$$

Обозначим через $\widetilde{CO}(z)$ — c -ядро, вычисленное для кооперативной подыгры $\langle N, \hat{V}(\cdot, z) \rangle$, $z \in CZ$, с характеристической функцией $\hat{V}(S, z)$, определенной формулами (1.7.4), (1.7.5). Сформулируем теорему, показывающую связь между $\widetilde{CO}(z)$ и $\widehat{CO}(z)$.

Теорема 1.5. *Регуляризованное c -ядро, определяемое формулой (1.8.5), — позиционно состоятельное решение. При этом, $\widehat{CO}(z_0) \subset \widetilde{CO}(z_0)$, где $\widetilde{CO}(z_0)$ — c -ядро, вычисленное для кооперативной стохастической игры $\langle N, \hat{V}(\cdot, z) \rangle$ с характеристической функцией, определенной по формулам (1.7.4), (1.7.5).*

Доказательство. Позиционная состоятельность c -ядра следует из Теоремы 1.1. Для доказательства включения $\widehat{CO}(z_0) \subset \widetilde{CO}(z_0)$ необходимо показать, что каждый дележ $\hat{\xi}(z_0) \in \widehat{CO}(z_0)$ принадлежит множеству $\widetilde{CO}(z_0)$, что равносильно следующему: для любых $\hat{\xi}(z) \in \widehat{CO}(z)$, $z \in CZ$ и $S \subset N$ выполняется

неравенство

$$\sum_{i \in S} \hat{\xi}_i(z) \geq \hat{V}(S, z). \quad (1.8.7)$$

Для вершин $z \in \{z : L(z) = \emptyset\}$ это утверждение очевидно. Докажем выполнение этого неравенства для вершин $z \in \{z : L(z) \ni y \text{ и } L(y) = \emptyset\}$:

$$\begin{aligned} \sum_{i \in S} \hat{\xi}_i(z) &= \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \sum_{i \in S} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} \left(p(y|z, \bar{a}^z) \sum_{i \in S} \hat{\xi}_i(y) \right) \geq \\ &\geq \hat{V}(S, z), \end{aligned}$$

это следует из того, что $y \in \{z : L(z) = \emptyset\}$ и $\sum_{i \in S} \xi_i(z) \geq V(S, z)$, так как $\xi(z)$ — дележ из c -ядра $CO(z)$. Дальнейшее доказательство проводится для последующих вершин вплоть до начальной вершины z_0 так же, как в теореме 1.1. \square

Теперь рассмотрим примеры построения и регуляризации решений в кооперативных стохастических играх, задаваемых на графах.

Пример 1. [см. [42]] Рассмотрим стохастическую игру $G(z_0)$, заданную на графе $\Psi(z_0)$, который имеет вид, представленный на Рис. 1.1.

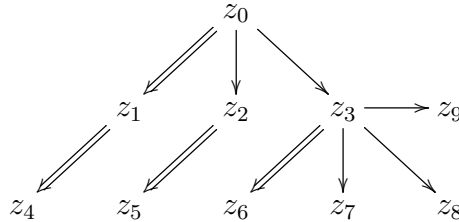


Рис. 1.1: Граф для примера 1.

Множество вершин графа $\Psi(z_0)$ есть $Z = \{z_0, \dots, z_9\}$. Множество игроков $N = \{1, 2\}$. В каждой вершине графа $\Psi(z_0)$ задана игра двух лиц в нормальной форме $\Gamma(z)$, $z \in Z$, выигрыши в них представлены матрицами:

$$\begin{aligned} \Gamma(z_0) &: \begin{pmatrix} (5, 5) & (0, 8) \\ (8, 0) & (1, 1) \end{pmatrix}, & \Gamma(z_2) &: \begin{pmatrix} (3, 0) & (6, 4) \\ (5, 6) & (2, 2) \end{pmatrix}, \\ \Gamma(z_3) &: \begin{pmatrix} (1, 11) & (4, 2) \\ (1, 3) & (1, 1) \end{pmatrix}, & \Gamma(z_7) &: \begin{pmatrix} (1, 1) & (0, 2) \\ (2, 0) & (1, 2) \end{pmatrix}, \end{aligned}$$

$$\Gamma(z_8) : \begin{pmatrix} (5, 5) & (6, 1) \\ (1, 6) & (6, 6) \end{pmatrix}, \quad \Gamma(z_9) : \begin{pmatrix} (4, 2) & (3, 4) \\ (5, 6) & (1, 5) \end{pmatrix},$$

$$\Gamma(z_1), \Gamma(z_4), \Gamma(z_5), \Gamma(z_6) : \begin{pmatrix} (0, 0) & (1, 0) \\ (1, 0) & (0, 1) \end{pmatrix}.$$

Для задания некооперативной стохастической игры $G(z_0)$ определим вероятности перехода и вероятности окончания игры. Определим сначала вероятности перехода из одних вершин древовидного графа в другие. Если в игре $\Gamma(z_0)$ реализуется ситуация $(2, 2)$, то происходит переход в вершину z_2 с вероятностью $1/3$ и в вершину z_3 с вероятностью $2/3$. Если же реализуется ситуация, отличная от $(2, 2)$ (стрелка \implies означает детерминированный переход), то происходит переход в вершину z_1 . В вершинах z_1, z_2 при реализации любой ситуации происходит переход в вершины z_4 и z_5 соответственно. Если в одновременной игре $\Gamma(z_3)$ реализуется ситуация $(2, 2)$, то происходит переход в вершины z_8 и z_9 с равными вероятностями $1/2$, а если в игре $\Gamma(z_3)$ реализуется ситуация $(2, 1)$, то с вероятностью 1 происходит переход в вершину z_7 . Из остальных ситуаций совершается переход в вершину z_6 с вероятностью 1 .

Заданы вероятности q_k окончания игры на k -м шаге:

$$q_1 = \frac{1}{8}, \quad q_2 = 0, \quad q_3 = 1.$$

Пусть в качестве решения игроки выбрали вектор Шепли, который вычисляется для игры двух лиц по формуле:

$$Sh_1(z) = V(\{1\}, z) + \frac{V(\{1, 2\}, z) - V(\{1\}, z) - V(\{2\}, z)}{2},$$

$$Sh_2(z) = V(\{2\}, z) + \frac{V(\{1, 2\}, z) - V(\{1\}, z) - V(\{2\}, z)}{2},$$

где $V(\{1\}, z)$ и $V(\{2\}, z)$ — значения характеристической функции в вершине z для коалиций $\{1\}$ и $\{2\}$ соответственно.

Все описанные выше множества и величины определяют стохастическую игру со случайной продолжительностью $G(z_0)$ (см. (1.1.1)).

Решение кооперативной игры начинаем искать с терминальных вершин графа, т.е. с тех вершин, из которых нельзя попасть ни в какие другие вершины древовидного графа. Сначала для игры $\Gamma(z_9)$ по формуле (1.3.6) вычислим

$V(\{1\}, z_9)$ и $V(\{2\}, z_9)$ как максимальные гарантированные выигрыши игроков:

$$V(\{1\}, z_9) = 3, \quad V(\{2\}, z_9) = 4, \quad V(\{1, 2\}, z_9) = 11.$$

Тогда можно вычислить вектор Шепли в подыгре $\bar{G}(z_9)$ игры $\bar{G}(z_0)$, начинающейся с игры $\Gamma(z_9)$:

$$Sh_1(z_9) = 5, \quad Sh_2(z_9) = 6.$$

По формуле (1.3.6) проводим аналогичные вычисления для подыгр, начинающихся с игр $\Gamma(z_4)$, $\Gamma(z_5)$, $\Gamma(z_6)$, $\Gamma(z_7)$ и $\Gamma(z_8)$, так как они реализуются в вершинах, принадлежащих множеству $\{z : L(z) = \emptyset\}$. Значения характеристических функций для этих подыгр и вектора Шепли представлены в табл. 1.2.

Таблица 1.2: Характеристические функции и вектора Шепли для подыгр $\bar{G}(z)$, $z \in \{z_4, z_5, z_6, z_7, z_8, z_9\}$.

Вершина z	$V(\{1\}, z)$	$V(\{2\}, z)$	$V(\{1, 2\}, z)$	$Sh_1(z)$	$Sh_2(z)$
z_4	0	0	1	1/2	1/2
z_5	0	0	1	1/2	1/2
z_6	0	0	1	1/2	1/2
z_7	1	2	3	1	2
z_8	5	5	12	6	6
z_9	3	4	11	5	6

Перейдем к вершинам из множества $\{z : (L(z))^2 = \emptyset\}$. Начнем с вершины z_3 . Так как игровой процесс может перейти в другие вершины графа, то тогда для того, чтобы найти вектор Шепли в кооперативной подыгре $\bar{G}(z_3)$, надо преобразовать матрицу игры. В ситуации (2,2) математические ожидания выигрышей игроков будем вычислять следующим образом:

- для первого игрока:

$$1 + (1 - q_2) \left(\frac{1}{2}V(\{1\}, z_8) + \frac{1}{2}V(\{1\}, z_9) \right) = 5,$$

- для второго игрока:

$$1 + (1 - q_2) \left(\frac{1}{2}V(\{2\}, z_8) + \frac{1}{2}V(\{2\}, z_9) \right) = 5.5.$$

В ситуации (2,1) они будут равны

- для первого игрока:

$$1 + (1 - q_2)V(\{1\}, z_7) = 2,$$

- для второго игрока:

$$1 + (1 - q_2)V(\{2\}, z_7) = 3.$$

Аналогично для ситуации (1,1) математические ожидания выигрышей будут равны

- для первого игрока:

$$1 + (1 - q_2)V(\{1\}, z_6) = 1,$$

- для второго игрока:

$$11 + (1 - q_2)V(\{2\}, z_6) = 11;$$

и для ситуации (1,2) математические ожидания выигрышей будут равны

- для первого игрока:

$$4 + (1 - q_2)V(\{1\}, z_6) = 4,$$

- для второго игрока:

$$2 + (1 - q_2)V(\{2\}, z_6) = 2.$$

Тогда биматричная игра, сформированная для вычисления значений характеристических функций $V(\{1\}, z_3)$ и $V(\{2\}, z_3)$, будет выглядеть следующим образом:

$$\begin{pmatrix} (1, 11) & (4, 2) \\ (2, 5) & (5, 5.5) \end{pmatrix}.$$

Значения характеристической функции для кооперативной подыгры $\bar{G}(z_3)$ стохастической игры $\bar{G}(z_0)$ для коалиций $\{1\}$, $\{2\}$ равны

$$V(\{1\}, z_3) = 2, \quad V(\{2\}, z_3) = 5.$$

Для вычисления $V(\{1, 2\}, z_3)$ используем формулу (1.3.5) и получаем матрицу игры:

$$\begin{pmatrix} 12 + (1 - q_2)V(\{1, 2\}, z_6) & 6 + (1 - q_2)V(\{1, 2\}, z_6) \\ 4 + (1 - q_2)V(\{1, 2\}, z_7) & 2 + (1 - q_2)(0.5V(\{1, 2\}, z_8) + 0.5V(\{1, 2\}, z_9)) \end{pmatrix}$$

или в численном выражении:

$$\begin{pmatrix} 13 & 7 \\ 7 & 13.5 \end{pmatrix}.$$

Тогда

$$\begin{aligned} V(\{1, 2\}, z_3) &= 13, 5, \\ Sh_1(z_3) &= 5.25, \quad Sh_2(z_3) = 8.25. \end{aligned}$$

Аналогичные вычисления проводим для кооперативной подыгры $\bar{G}(z_1)$:

$$\begin{aligned} V(\{1\}, z_1) &= 0, \quad V(\{2\}, z_1) = 0, \quad V(\{1, 2\}, z_1) = 2, \\ Sh_1(z_1) &= Sh_2(z_1) = 1, \end{aligned}$$

и подыгры $\bar{G}(z_2)$:

$$\begin{aligned} V(\{1\}, z_2) &= 3, \quad V(\{2\}, z_2) = 2, \quad V(\{1, 2\}, z_2) = 12, \\ Sh_1(z_2) &= 6.5, \quad Sh_2(z_2) = 5.5. \end{aligned}$$

Для кооперативной стохастической игры $\bar{G}(z_0)$, матричная игра для вычисления значения характеристической функции для коалиций $\{1\}$, $\{2\}$ будет вычислена по формуле (1.3.5). Для ситуации (2,2) математические ожидания выигрышей игроков:

- для первого игрока:

$$1 + (1 - q_1) \left(\frac{1}{3}V(\{1\}, z_2) + \frac{2}{3}V(\{1\}, z_3) \right) = 3\frac{1}{24},$$

- для второго игрока:

$$1 + (1 - q_1) \left(\frac{1}{3}V(\{2\}, z_2) + \frac{2}{3}V(\{2\}, z_3) \right) = 4\frac{1}{2}.$$

В ситуации (2,1) математические ожидания выигрышей равны

- для первого игрока:

$$8 + (1 - q_1)V(\{1\}, z_1) = 8,$$

- для второго игрока:

$$0 + (1 - q_1)V(\{2\}, z_1) = 0.$$

Аналогично, для ситуации (1,1) математические ожидания выигрышей будут равны

- для первого игрока:

$$5 + (1 - q_1)V(\{1\}, z_1) = 5,$$

- для второго игрока:

$$5 + (1 - q_1)V(\{2\}, z_1) = 5.$$

Для ситуации (1,2) математические ожидания выигрышей будут равны

- для первого игрока:

$$0 + (1 - q_1)V(\{1\}, z_1) = 0,$$

- для второго игрока:

$$8 + (1 - q_1)V(\{2\}, z_1) = 8.$$

Получаем матрицу игры:

$$\begin{pmatrix} (5, 5) & (0, 8) \\ (8, 0) & (3\frac{1}{24}, 4\frac{1}{2}) \end{pmatrix},$$

$$V(\{1\}, z_0) = 3\frac{1}{24}, V(\{2\}, z_0) = 4\frac{1}{2}.$$

Для вычисления $V(\{1, 2\}, z_0)$ составим матричную игру по формуле (1.3.5):

$$\begin{pmatrix} 10 + (1 - q_1)V(\{1, 2\}, z_1) & 8 + (1 - q_1)V(\{1, 2\}, z_1) \\ 8 + (1 - q_1)V(\{1, 2\}, z_1) & 2 + (1 - q_1)(\frac{1}{3}V(\{1, 2\}, z_2) + \frac{2}{3}V(\{1, 2\}, z_3)) \end{pmatrix}$$

или в численном выражении:

$$\begin{pmatrix} 11\frac{3}{4} & 9\frac{3}{4} \\ 9\frac{3}{4} & 13\frac{3}{8} \end{pmatrix}.$$

Вычисляем $V(\{1, 2\}, z_0)$ и $Sh_1(z_0)$, $Sh_2(z_0)$, получаем:

$$V(\{1, 2\}, z_0) = 13\frac{3}{8}, \quad Sh_1(z_0) = 5\frac{23}{24}, \quad Sh_2(z_0) = 7\frac{5}{12}.$$

Множество вершин, образующих кооперативное поддерево, состоит из вершин $z_0, z_2, z_3, z_5, z_8, z_9$.

Проверим, является ли ПРД неотрицательной. Вычисления показывают, что ПРД отрицательна уже в вершине z_3 . Это следует из уравнения (1.6.1), где в качестве вершины выбрана z_3 :

$$Sh_1(z_3) = \beta_1(z_3) + (1 - q_1) \left[\frac{1}{2} \cdot Sh_1(z_8) + \frac{1}{2} \cdot Sh_1(z_9) \right],$$

$$5.25 = \beta_1(z_3) + (1 - 0) \left[\frac{1}{2} \cdot 5 + \frac{1}{2} \cdot 6 \right],$$

$$\beta_1(z_3) = -0.25.$$

Так как $\beta_1(z_3)$ отрицательна, то проведем регуляризацию вектора Шепли с целью построения «нового» неотрицательного вектора Шепли.

Строим новый вектор Шепли для вершин кооперативного поддерева с вершинами $z_0, z_2, z_3, z_5, z_8, z_9$ по формулам (1.8.1) и (1.8.2):

$$\hat{Sh}_1(z_5) = 0.5, \hat{Sh}_1(z_8) = 6, \hat{Sh}_1(z_9) = 5,$$

$$\hat{Sh}_2(z_5) = 0.5, \hat{Sh}_2(z_8) = 6, \hat{Sh}_2(z_9) = 6,$$

$$\hat{Sh}_1(z_2) = \frac{11}{12} \cdot 6.5 + \frac{1}{2} = 6\frac{11}{24},$$

$$\hat{Sh}_2(z_2) = \frac{11}{12} \cdot 5.5 + \frac{1}{2} = 5\frac{13}{24},$$

$$\hat{Sh}_1(z_3) = \frac{2}{13.5} \cdot 5.25 + \left[\frac{1}{2} \cdot 5 + \frac{1}{2} \cdot 6 \right] = 6\frac{5}{18},$$

$$\hat{Sh}_2(z_3) = \frac{2}{13.5} \cdot 8.25 + \left[\frac{1}{2} \cdot 6 + \frac{1}{2} \cdot 6 \right] = 7\frac{2}{9},$$

$$\hat{Sh}_1(z_0) = \frac{2}{13\frac{3}{8}} \cdot 5\frac{23}{24} + \left(1 - \frac{1}{8}\right) \left[\frac{1}{3} \cdot 6\frac{11}{24} + \frac{2}{3} \cdot 6\frac{5}{18} \right] = 6\frac{80741}{184896} \approx 6.437,$$

$$\hat{Sh}_2(z_0) = \frac{2}{13\frac{3}{8}} \cdot 7\frac{5}{12} + \left(1 - \frac{1}{8}\right) \left[\frac{1}{3} \cdot 5\frac{13}{24} + \frac{2}{3} \cdot 7\frac{2}{9} \right] = 6\frac{173491}{184896} \approx 6.938.$$

Этот «новый» вектор является вектором Шепли для кооперативной игры с характеристической функцией, определенной формулами (1.7.4), (1.7.5). Он будет позиционно состоятельным, что следует из Теоремы 1.4. При этом компоненты ПРД неотрицательны.

Для подыгр, начинающихся в вершинах кооперативного поддерева, новые характеристические функции представлены в табл. 1.3.

Таблица 1.3: «Новые» характеристические функции.

Вершина z	$\hat{V}(\{1\}, z)$	$\hat{V}(\{2\}, z)$	$\hat{V}(\{1, 2\}, z)$
z_0	3.763	4.265	13.375
z_2	2.750	1.833	12.000
z_3	4.296	5.574	13.500
z_5	0.000	0.000	1.000
z_8	5.000	5.000	12.000
z_9	3.000	4.000	11.000

Замечание 1.4. В качестве решения можно выбрать n -ядро (см. [206]). Заметим, что n -ядро состоит из единственного вектора, поэтому не будет возникать вопроса о выборе единственного дележа из множества. В случае, когда s -ядро непусто, n -ядро принадлежит s -ядру.

Пример 1.2. Рассмотрим стохастическую игру $G(z_0)$, заданную на графе $\Psi(z_0)$, который представлен на Рис. 1.2. Множество вершин графа $\Psi(z_0)$ есть

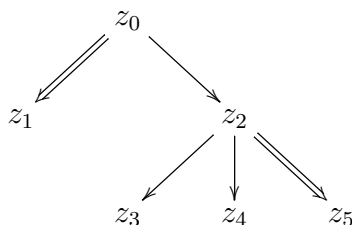


Рис. 1.2: Граф для Примера 1.2.

$Z = \{z_0, \dots, z_5\}$. Множество игроков $N = \{1, 2, 3\}$. В каждой вершине графа $G(z_0)$ задана одновременная игра трех лиц $\Gamma(z)$, $z \in Z$, с матрицами выигры-

шей:

$$\begin{aligned} \Gamma(z_0), \Gamma(z_2) &: \left(\begin{pmatrix} (1, 1, 1) & (2, 2, 0) \\ (2, 2, 0) & (0, 0, 3) \end{pmatrix} \begin{pmatrix} (1, 1, 2) & (2, 2, 1) \\ (1, 3, 1) & (3, 0, 1) \end{pmatrix} \right), \\ \Gamma(z_1), \Gamma(z_5) &: \left(\begin{pmatrix} (2, 0, 1) & (1, 0, 1) \\ (3, 1, 2) & (2, 2, 2) \end{pmatrix} \begin{pmatrix} (2, 2, 1) & (1, 1, 3) \\ (2, 1, 1) & (2, 1, 2) \end{pmatrix} \right), \\ \Gamma(z_3) &: \left(\begin{pmatrix} (1, 1, 1) & (2, 2, 2) \\ (3, 2, 0) & (1, 4, 1) \end{pmatrix} \begin{pmatrix} (2, 0, 1) & (2, 1, 1) \\ (4, 0, 1) & (0, 4, 1) \end{pmatrix} \right), \\ \Gamma(z_4) &: \left(\begin{pmatrix} (2, 1, 0) & (2, 1, 3) \\ (3, 1, 2) & (3, 6, 4) \end{pmatrix} \begin{pmatrix} (4, 5, 0) & (0, 5, 4) \\ (2, 8, 0) & (0, 8, 2) \end{pmatrix} \right). \end{aligned}$$

В каждой игре, определенной выше, первый игрок выбирает строки, второй игрок — столбцы, а третий — матрицы.

Определим сначала вероятности перехода. Если в игре $\Gamma(z_0)$ реализуется ситуация $(1,1,1)$, то совершается переход в вершину z_1 с вероятностью $1/3$ и в вершину z_2 с вероятностью $2/3$. Если же реализуется ситуация, отличная от $(1,1,1)$ (стрелка \implies означает детерминированный переход), то совершается переход в вершину z_1 . В вершине z_2 при реализации ситуации $(2,1,2)$ происходит переход в вершины z_3 и z_4 с вероятностями $1/3$, $2/3$ соответственно, из остальных ситуаций совершается переход в вершину z_5 с вероятностью 1.

Также заданы вероятности q_k окончания игры $G(z_0)$ на k -м шаге:

$$q_1 = 0.5, \quad q_2 = 0, \quad q_3 = 1.$$

Сначала предположим, что игроки в качестве решения выбрали вектор Шепли. Начинаем решение игры с вершин графа, которые принадлежат множеству: $\{z : L(z) = \emptyset\}$. Для кооперативной подыгры $\bar{G}(z_3)$ вычисляем значения характеристической функции и вектор Шепли. Аналогичные вычисления проводим для вершин z_1 , z_5 , z_4 , и далее для вершин z_2 и z_0 по формуле (1.3.5). Результаты представлены в табл. 1.4 и 1.5.

Множество вершин кооперативного поддерева есть $CZ = \{z_0, z_1, z_2, z_3, z_4\}$. Проведем регуляризацию вектора Шепли

$$Sh(z_0) = \left(3\frac{31}{54}, 2\frac{89}{108}, 2\frac{89}{108} \right)$$

Таблица 1.4: Характеристические функции для подыгр $\bar{G}(z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

z	$V(\{1\}, z)$	$V(\{2\}, z)$	$V(\{3\}, z)$	$V(\{1, 2\}, z)$	$V(\{2, 3\}, z)$	$V(\{1, 3\}, z)$	$V(\{1, 2, 3\}, z)$
z_0	2	1	3/2	11/2	9/2	2	83/9
z_1	2	0	1	3	4	3	6
z_2	3	1	4/3	7	6	7	47/3
z_3	1	1	1	4	4	3	6
z_4	0	1	0	8	9	5	13
z_5	2	0	1	3	4	3	6

Таблица 1.5: Вектора Шепли для подыгр $\bar{G}(z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

z	$Sh_1(z)$	$Sh_2(z)$	$Sh_3(z)$
z_0	193/54	305/108	305/108
z_1	8/3	7/6	13/6
z_2	37/6	14/3	29/6
z_3	11/6	7/3	11/6
z_4	10/3	35/6	23/6
z_5	8/3	7/6	13/6

и проверим неотрицательность ПРД. Для этого вычислим величины $\beta_i(z)$ в вершинах $z_0 \in CZ$ и $z_2 \in CZ$ по формуле (1.6.3) и проверим, неотрицательна ли ПРД $\beta_i(z)$:

$$\beta_i(z_2) = Sh_i(z_2) - (1 - q_2) \left(\frac{1}{3} Sh_i(z_3) + \frac{2}{3} Sh_i(z_4) \right),$$

$$\beta_i(z_0) = Sh_i(z_0) - (1 - q_1) \left(\frac{1}{3} Sh_i(z_1) + \frac{2}{3} Sh_i(z_2) \right),$$

получаем

$$\beta_1(z_2) = 3\frac{1}{3}, \quad \beta_2(z_2) = 0, \quad \beta_3(z_2) = 1\frac{2}{3};$$

$$\beta_1(z_0) = 1\frac{2}{27}, \quad \beta_2(z_0) = 1\frac{2}{27}, \quad \beta_3(z_0) = \frac{23}{27}.$$

Для $z \in \{z_0, z_2\}$ выполнены условия: $\beta_i(z) \geq 0$ и $\sum_{i \in N} \beta_i(z_2) = 5$, а также $\sum_{i \in N} \beta_i(z_0) = 3$.

Во всех вершинах кооперативного поддерева выполнено условие позиционной состоятельности и неотрицательности ПРД. Можно утверждать, что вектор Шепли $Sh(z_0)$ является позиционно состоятельным дележом в игре $\bar{G}(z_0)$, если выплаты производятся в соответствии с ПРД.

Проведем теперь все рассуждения, предполагая, что игроки в качестве решения в игре $\bar{G}(z_0)$ выбрали n -ядро. N -ядро на множестве дележей впервые введено Шмайдлером [206]. Для вычисления n -ядра может быть использован математический пакет Matlab [136] и программа TUGlab [218], написанная для вычислений, производимых в кооперативной теории игр, или пакет Mathematica [126] и программа TUGames [139], написанная для тех же целей.

Характеристическая функция была вычислена ранее. N -ядро, вычисленное для подыгры $\bar{G}(z)$, $z \in CZ$, будем обозначать $n(z) = (n_i(z) : i \in N)$.

Для всех подыгр игры $\bar{G}(z_0)$ вычислим n -ядра. Результат представлен в табл. 1.6.

Таблица 1.6: N -ядра в подыграх $\bar{G}(z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

z	$n_1(z)$	$n_2(z)$	$n_3(z)$
z_0	$3\frac{5}{9}$	$3\frac{1}{18}$	$2\frac{11}{18}$
z_1	$2\frac{1}{2}$	$1\frac{1}{4}$	$2\frac{1}{4}$
z_2	$6\frac{1}{3}$	$4\frac{1}{2}$	$4\frac{2}{3}$
z_3	$1\frac{2}{3}$	$2\frac{2}{3}$	$1\frac{2}{3}$
z_4	$2\frac{2}{3}$	$6\frac{2}{3}$	$3\frac{2}{3}$
z_5	$2\frac{1}{2}$	$1\frac{1}{4}$	$2\frac{1}{4}$

Проверим позиционную состоятельность n -ядра, используя формулу (1.6.1). Для вершины z_2 вычислим $\beta_i(z_2)$ из уравнения:

$$n_i(z_2) = \beta_i(z_2) + (1 - q_2)(p(z_3|z_2, \bar{a}^{z_2})n_i(z_3) + p(z_4|z_2, \bar{a}^{z_2})n_i(z_4)).$$

Получаем

$$\beta_1(z_2) = 4, \quad \beta_2(z_2) = -\frac{5}{6}, \quad \beta_3(z_2) = 1\frac{5}{6}.$$

N -ядро кооперативной стохастической игры $\bar{G}(z_0)$ является позиционно состоятельным, если выплаты игрокам производить в соответствии с ПРД, но ПРД не является неотрицательной, так как $\beta_2(z_2) < 0$. Для последующих вершин не будем проверять выполнение условия существования неотрицательной ПРД (1.6.1), а перейдем к регуляризации n -ядра.

По формуле (1.7.2) с начальным условием (1.7.3) вычислим «новое» n -ядро

для каждой вершины $z \in CZ$. Для вершины z_2 используем формулу:

$$\hat{n}_i(z_2) = \frac{\sum_{i \in N} K_i^{z_2}(\bar{a}^{z_2})}{V(N, z_2)} n_i(z_2) + (1 - q_2) \sum_{y \in L(z_2)} p(y|z_2, \bar{a}^{z_2}) \hat{n}_i(y),$$

для вершины z_0 используем формулу:

$$\hat{n}_i(z_0) = \frac{\sum_{i \in N} K_i^{z_0}(\bar{a}^{z_0})}{V(N, z_0)} n_i(z_0) + (1 - q_1) \sum_{y \in L(z_0)} p(y|z_0, \bar{a}^{z_0}) \hat{n}_i(y).$$

«Новые» n -ядра для вершин из множества CZ представлены в табл. 1.7.

Таблица 1.7: N -ядра в подыграх $\bar{G}(z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

z	$\hat{n}_1(z)$	$\hat{n}_2(z)$	$\hat{n}_3(z)$
z_0	$3 \frac{3487}{140436}$	$3 \frac{128867}{280872}$	$2 \frac{69149}{93624}$
z_1	$2 \frac{1}{2}$	$1 \frac{1}{4}$	$2 \frac{1}{4}$
z_2	$4 \frac{50}{141}$	$6 \frac{217}{282}$	$4 \frac{51}{94}$
z_3	$1 \frac{2}{3}$	$2 \frac{2}{3}$	$1 \frac{2}{3}$
z_4	$2 \frac{2}{3}$	$6 \frac{2}{3}$	$3 \frac{2}{3}$
z_5	$2 \frac{1}{2}$	$1 \frac{1}{4}$	$2 \frac{1}{4}$

Вычислим характеристическую функцию $\hat{V}(S, z)$ для всех $z \in CZ$ по формулам (1.7.4) и (1.7.5). При этом, $\hat{V}(S, z_3) = V(S, z_3)$, $\hat{V}(S, z_4) = V(S, z_4)$, $\hat{V}(S, z_1) = \hat{V}(S, z_5) = V(S, z_1) = V(S, z_5)$. Для вычисления $\hat{V}(S, z_2)$ используем формулу:

$$\hat{V}(S, z_2) = \frac{\sum_{i \in N} K_i^{z_2}(\bar{a}^{z_2})}{V(N, z_2)} V(S, z_2) + (1 - q_2) \sum_{y \in L(z_2)} p(y|z_2, \bar{a}^{z_2}) \hat{V}(S, y),$$

и для $\hat{V}(S, z_0)$:

$$\begin{aligned} \hat{V}(S, z_0) = & \frac{\sum_{i \in N} K_i^{z_0}(\bar{a}^{z_0})}{V(N, z_0)} V(S, z_0) + \\ & + (1 - q_1) \left(p(z_1|z_0, \bar{a}^{z_0}) \hat{V}(S, z_1) + p(z_2|z_0, \bar{a}^{z_0}) \hat{V}(S, z_2) \right). \end{aligned}$$

Функции $\hat{V}(S, z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$, представлены в табл. 1.8.

Заметим, что «новое» n -ядро $\hat{n}(z_2)$, вычисленное для подыгры $\bar{G}(z_2)$, принадлежит множеству дележей кооперативной игры с характеристической функцией $\hat{V}(S, z_2)$ ($\hat{n}(z_2)$ также принадлежит множеству дележей $I(z_2)$, что не является обязательным), но не является n -ядром изначально заданной кооперативной

Таблица 1.8: Характеристические функции $\hat{V}(S, z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

z	$\hat{V}(\{1\}, z)$	$\hat{V}(\{2\}, z)$	$\hat{V}(\{3\}, z)$	$\hat{V}(\{1, 2\}, z)$	$\hat{V}(\{2, 3\}, z)$	$\hat{V}(\{1, 3\}, z)$	$\hat{V}(\{1, 2, 3\}, z)$
z_0	$1 \frac{245}{249}$	$\frac{164}{249}$	$1 \frac{74}{747}$	$4 \frac{155}{249}$	$3 \frac{80}{83}$	$4 \frac{52}{83}$	$9 \frac{2}{9}$
z_1	2	0	1	3	4	3	6
z_2	$1 \frac{41}{141}$	$1 \frac{15}{47}$	$\frac{107}{141}$	$8 \frac{127}{141}$	$9 \frac{35}{141}$	$6 \frac{80}{141}$	$15 \frac{2}{3}$
z_3	1	1	1	4	4	3	6
z_4	0	1	0	8	9	5	13
z_5	2	0	1	3	4	3	6

игры. N -ядро кооперативной игры с характеристической функцией $\hat{V}(S, z_2)$, определенной выше, обозначим через $\tilde{n}(z_2) = (\tilde{n}_1(z_2), \tilde{n}_2(z_2), \tilde{n}_3(z_2))$, оно будет следующим (результаты приведены округленными до третьего знака после запятой):

$$\tilde{n}(z_2) \approx (4.213, 6.894, 4.560) \neq \hat{n}(z_2).$$

«Новое» n -ядро $\hat{n}(z_0)$, вычисленное для подыгры $\bar{G}(z_0)$, принадлежит множеству дележей кооперативной игры с характеристической функцией $\hat{V}(S, z_0)$ ($\hat{n}(z_0)$ также принадлежит множеству дележей $I(z_0)$), но не является n -ядром изначально заданной кооперативной игры. N -ядро кооперативной игры с характеристической функцией $\hat{V}(S, z_0)$, определенной выше, обозначим через $\tilde{n}(z_0)$, оно будет следующим:

$$\tilde{n}(z_0) \approx (3.621, 2.720, 2.881) \neq \hat{n}(z_0).$$

§ 1.9 Сильная позиционная состоятельность s -ядра в стохастических играх конечной продолжительности

В этом параграфе мы рассмотрим случай, когда решение кооперативной стохастической игры представляет собой множество, а не точку из множества дележей. В качестве примера такого решения будем использовать s -ядро. Мы сформулируем проблему сильной позиционной состоятельности s -ядра и предложим достаточные условия его сильной позиционной состоятельности. Впервые эта проблема сформулирована для дифференциальных игр [36], а далее — для многокритериальных задач оптимального управления [37].

Предположим, что c -ядра стохастической игры $\bar{G}(z_0)$ и любой подыгры $\bar{G}(z)$, $z \in CZ \setminus z_0$, непусты. Кооперируясь, игроки договариваются о совместной реализации набора кооперативных стратегий $\bar{\varphi}$ и рассчитывают получить компоненты дележа, принадлежащего c -ядру $CO(z_0)$. Попадая в промежуточную вершину $z \in CZ \setminus \{z_0\}$ кооперативного поддерева, игрок $i \in N$ выбирает стратегию \bar{a}_i^z в соответствии с кооперативной стратегией $\bar{\varphi}_i$ и получает выигрыш $K_i^z(\bar{a}^z)$. Если игроки произведут пересчет решения, т.е. найдут решение кооперативной подыгры, начинающейся в вершине z , то текущим решением будет c -ядро $CO(z)$. Было бы разумным потребовать, чтобы выигрыш, полученный игроком в вершине z , суммированный с ожидаемой суммой любых дележей из решений $CO(y)$, $y \in L(z)$, игр кооперативного поддерева, следующих за игрой $\Gamma(z)$, являлся бы дележом решения $CO(z)$. Если это свойство выполнено для любой вершины z кооперативного поддерева, то c -ядро кооперативной стохастической игры $\bar{G}(z_0)$ является сильно позиционно состоятельным.

Для того, чтобы дать строгое определение сильно позиционно состоятельного c -ядра, необходимо определить так называемое ожидаемое c -ядро. Для любой нетерминальной вершины кооперативного поддерева определим множество ожидаемых дележей, принадлежащих c -ядрам, которые являются решениями подыгр, следующих за рассматриваемой вершиной. Для каждой вершины $z \in CZ$, $L(z) \neq \emptyset$, определим *ожидаемое c -ядро*:

$$EC(L(z)) = \left\{ \alpha(L(z)) = \sum_{y \in L(z)} p(y|z, \bar{a}^z) \alpha(y) \mid \alpha(y) \in CO(y) \right\}. \quad (1.9.1)$$

Множеству $EC(L(z))$ принадлежат вектора $\alpha(L(z))$, которые являются математическими ожиданиями всевозможных наборов дележей из c -ядер подыгр, начинающихся из следующих за z вершин, относительно вероятностного распределения $\{p(y|z, \bar{a}^z), y \in L(z)\}$.

Также определим процедуру перераспределения выигрышей игроков в вершинах кооперативного поддерева. Вспомним определение 1.9 процедуры распределения дележа. Первое условие в Определении 1.9 можно назвать условием «достижимости процедуры распределения дележа», поскольку оно позволяет гарантировать, что в любой вершине кооперативного поддерева сумма выплат

игрокам равна сумме полученных ими выигрышей при реализации кооперативных стратегий. Второе условие гарантирует игрокам получение компонент заранее выбранного дележа из c -ядра кооперативной игры $\bar{G}(z_0)$ в смысле математического ожидания, если выплаты игрокам на протяжении всей игры будут производиться в соответствии с процедурой распределения $\{\beta(z) : z \in CZ\}$.

Теперь нам необходимо определить процедуру распределения дележа $\alpha(z_0)$ из c -ядра $CO(z_0)$ таким образом, чтобы c -ядро было сильно позиционно состоятельно.

Определение 1.15. Назовем c -ядро $CO(z_0)$ кооперативной стохастической игры $\bar{G}(z_0)$ сильно позиционно состоятельным, если для любого дележа из c -ядра $CO(z_0)$ существует процедура распределения $\{\beta(z)\}_{z \in CZ}$ дележа такая, что для любой вершины $z \in CZ$ имеют место включения

$$\beta(z) \oplus (1 - q_k)EC(L(z)) \subset CO(z), \quad (1.9.2)$$

$$B(z_0) \in CO(z_0), \quad (1.9.3)$$

где

$$\beta(z) \oplus (1 - q_k)EC(L(z)) = \left\{ \beta(z) + (1 - q_k)\alpha(L(z)) : \alpha(L(z)) \in EC(L(z)) \right\}.$$

При этом процедура распределения $\{\beta(z)\}_{z \in CZ}$ также называется сильно позиционно состоятельной³.

Условие (1.9.2) означает, что множество векторов, равных сумме компонент процедуры распределения дележа в вершине z и дележа из ожидаемого c -ядра для вершины z , содержится в c -ядре подыгры, начинающейся в вершине z . Это условие накладывает ограничения на выплаты игрокам в одновременных играх, определенных в вершинах, и чаще всего не выполняется для произвольной игры, если выплаты игрокам производятся в соответствии с изначально заданными функциями выигрышей.

Наложим дополнительные ограничения на характеристические функции подыгр, начинающихся из вершин кооперативного поддерева, чтобы получить достаточные условия сильной позиционной состоятельности c -ядра. Обозначим

³Сумма, обозначаемая знаком \oplus , называется суммой Минковского (см. [207], где доказаны некоторые свойства этой операции).

через $EV(S, L(z))$ ожидаемое значение характеристической функции, вычисленное для коалиции $S \subseteq N$ в вершинах, следующих за вершиной z :

$$EV(S, L(z)) = \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(S, y).$$

Обозначим через

$$\Delta V(S, z) = V(S, z) - (1 - q_k) EV(S, L(z))$$

разность между значениями характеристической функции в вершине z и ожидаемым значением характеристической функции при условии, что игра не закончится в вершине z . Обозначим через $\Delta CO(z)$ — аналог c -ядра, построенного по функции $\Delta V(S, z)$. Сформулируем достаточное условие сильной позиционной состоятельности процедуры распределения дележа из c -ядра $CO(z_0)$.

Теорема 1.6. Пусть для любой вершины $z \in CZ$ c -ядро $CO(z)$ и множество $\Delta CO(z)$ непусты. Если для любой вершины $z \in CZ$ процедура распределения $\{\beta(z) : z \in CZ\}$ дележа $\alpha(z_0)$ из c -ядра $CO(z_0)$ удовлетворяет условиям:

$$\beta(z) \in \Delta CO(z), \quad (1.9.4)$$

$$B(z_0) \in CO(z_0). \quad (1.9.5)$$

то дележ $\alpha(z_0) \in CO(z_0)$ и процедура $\{\beta(z) : z \in CZ\}$ сильно позиционно состоятельны.

Доказательство. Покажем, что произвольный вектор $\beta(z) \in \Delta CO(z)$, удовлетворяющий условиям (1.9.4) и (1.9.5), является сильно позиционно состоятельной процедурой распределения некоторого дележа $\alpha(z_0) \in CO(z_0)$, то есть, что условия (1.9.2) и (1.9.3) из Определения 1.15 выполнены. Условие (1.9.5) совпадает с (1.9.3), поэтому осталось показать, что включение (1.9.2) выполняется для любой вершины $z \in CZ$. Для вершины z возьмем произвольный вектор $\alpha(L(z)) \in EC(L(z))$ и найдем сумму $\beta(z) + (1 - q_k)\alpha(L(z))$. Проверим, принадлежит ли последний вектор c -ядру $CO(z)$. Сначала вычислим сумму всех компонент этого вектора:

$$\sum_{i \in N} \beta_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \sum_{i \in N} \alpha_i(y) =$$

$$\begin{aligned}
&= V(N, z) - (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(N, y) + \\
&+ (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \sum_{i \in N} \alpha_i(y) = V(N, z),
\end{aligned}$$

что означает выполнение свойства коллективной рациональности.

Теперь пусть $S \subset N$:

$$\begin{aligned}
&\sum_{i \in S} \beta_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \sum_{i \in S} \alpha_i(y) \geq \\
&\geq V(S, z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(S, y) - \\
&- (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(S, y) = V(S, z).
\end{aligned}$$

В силу произвольности выбора вершины $z \in CZ$, можно сделать вывод о том, что дележ $\alpha(z_0)$ из c -ядра кооперативной игры $\bar{G}(z_0)$ и ПРД $\{\beta(z) : z \in CZ\}$ сильно позиционно состоятельны. \square

В случае непустоты аналогов c -ядер $\Delta CO(z)$ для любой вершины z кооперативного поддерева Теорема 1.6 дает метод построения сильно позиционно состоятельной процедуры распределения дележей из c -ядра, равных $B_i(z_0)$ из условия (1.9.5). Стоит отметить, что в общем случае не все дележи из c -ядра могут быть реализованы с помощью определенной выше процедуры распределения $\{\beta(z) : z \in CZ\}$.

Пример 3 Рассмотрим стохастическую игру $G(z_0)$, заданную на графе $\Psi(z_0)$, который представлен на рис. 1.3.

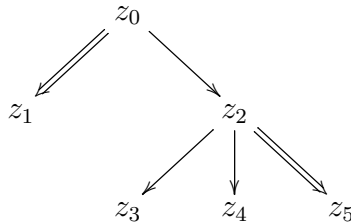


Рис. 1.3: Дерево $\Psi(z_0)$.

Множество вершин графа $\Psi(z_0)$ есть $Z = \{z_0, \dots, z_5\}$. Множество игроков $N = \{1, 2, 3\}$. В каждой вершине графа $G(z_0)$ задана одновременная игра трех

лиц $\Gamma(z)$, $z \in Z$, с матрицами выигрышей:

$$\begin{aligned} \Gamma(z_0) &: \left(\begin{pmatrix} (2, 2, 2) & (2, 2, 0) \\ (2, 2, 0) & (0, 0, 3) \end{pmatrix} \begin{pmatrix} (1, 1, 2) & (2, 2, 1) \\ (1, 3, 1) & (3, 0, 1) \end{pmatrix} \right), \\ \Gamma(z_2) &: \left(\begin{pmatrix} (1, 1, 1) & (2, 2, 0) \\ (2, 2, 0) & (0, 0, 3) \end{pmatrix} \begin{pmatrix} (1, 1, 2) & (2, 2, 1) \\ (1, 3, 1) & (3, 0, 1) \end{pmatrix} \right), \\ \Gamma(z_1), \Gamma(z_5) &: \left(\begin{pmatrix} (2, 0, 1) & (1, 0, 1) \\ (3, 1, 2) & (2, 2, 2) \end{pmatrix} \begin{pmatrix} (2, 2, 1) & (1, 1, 3) \\ (2, 1, 1) & (2, 1, 2) \end{pmatrix} \right), \\ \Gamma(z_3) &: \left(\begin{pmatrix} (1, 1, 1) & (2, 2, 2) \\ (3, 2, 0) & (1, 4, 1) \end{pmatrix} \begin{pmatrix} (2, 0, 1) & (2, 1, 1) \\ (4, 0, 1) & (0, 4, 1) \end{pmatrix} \right), \\ \Gamma(z_4) &: \left(\begin{pmatrix} (2, 1, 0) & (2, 1, 3) \\ (3, 1, 2) & (3, 6, 4) \end{pmatrix} \begin{pmatrix} (4, 5, 0) & (0, 5, 4) \\ (2, 8, 0) & (0, 8, 2) \end{pmatrix} \right). \end{aligned}$$

В каждой игре первый игрок выбирает строку, второй игрок — столбец, а третий — матрицу. Множество стратегий игрока $i \in N$ в игре $\Gamma(z)$ есть $A_i^z = \{1, 2\}$.

Определим вероятности перехода из одних вершин графа в другие. Если в игре $\Gamma(z_0)$ реализуется ситуация $(1, 1, 1)$, то игровой процесс переходит в вершину z_1 с вероятностью $1/3$ и в вершину z_2 с вероятностью $2/3$, если же реализуется ситуация, отличная от $(1, 1, 1)$ (стрелка \implies означает детерминированный переход), то игровой процесс переходит в вершину z_1 . В вершине z_2 при реализации ситуации $(2, 1, 2)$ осуществляется переход в вершины z_3 и z_4 с вероятностями $1/3$, $2/3$ соответственно, из остальных ситуаций этой игры, осуществляется переход в вершину z_5 с вероятностью 1.

Также заданы вероятности q_k окончания игры $G(z_0)$ на k -м шаге:

$$q_1 = 0.5, \quad q_2 = 0, \quad q_3 = 1.$$

Для построения кооперативного варианта стохастической игры сначала найдем набор кооперативных стратегий $\bar{\varphi}$. Ситуация $\bar{\varphi}$ предписывает в вершине z_0 разыгрывать ситуацию $(1, 1, 1)$. Игра закончится в вершине z_0 с вероятностью 0.5 или осуществляется переход на следующий шаг также с вероятностью 0.5. При продолжении игровой процесс переходит в вершину z_1 с вероятностью $1/3$, в которой игроки должны реализовать любую из ситуаций $(2, 1, 1)$ или $(2, 2, 1)$,

и с вероятностью $2/3$ в вершину z_2 , в которой игроки должны разыграть ситуацию $(2, 1, 2)$. В вершине z_2 игра не заканчивается, поскольку $q_2 = 0$, и происходит переход в вершины z_3 и z_4 с вероятностями $1/3$ и $2/3$ соответственно. В вершинах z_3 и z_4 игра заканчивается. Таким образом, множество вершин кооперативного поддерева, изображенного на рис. 1.4, есть $\bar{\Psi}(z_0) = \{z_0, z_1, z_2, z_3, z_4\}$.

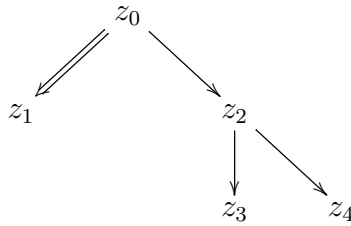


Рис. 1.4: Кооперативное поддерево $\bar{\Psi}(z_0)$ игры $\bar{G}(z_0)$.

Вычисляем значения характеристической функции по формулам (1.3.3) с граничным условием (1.3.4) для $S = N$, (1.3.5) с граничным условием (1.3.6) для $S \subset N$ и (1.3.7) для $S = \emptyset$. Значения функций представлены в табл. 1.9. Для дальнейших вычислений используем пакет **TUGlab** в программе Matlab [16].

Таблица 1.9: Характеристические функции $V(S, z)$ для $\bar{G}(z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

$z \setminus S$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1, 2\}$	$\{2, 3\}$	$\{1, 3\}$	$\{1, 2, 3\}$
z_0	2	1	1.5	5.5	4.5	6	110/9
z_1	2	0	1	3	4	3	6
z_2	3	1	4/3	7	6	7	47/3
z_3	1	1	1	4	4	3	6
z_4	0	1	0	8	9	5	13
z_5	2	0	1	3	4	3	6

Определим s -ядра подыгр, начинающихся из вершин кооперативного поддерева $\bar{\Psi}(z_0)$. Удостоверимся, что все они непусты, чтобы можно было использовать этот принцип оптимальности в качестве решения кооперативной стохастической игры. Системы линейных неравенств и равенства, определяющие s -ядра, а также их графическое представление даны в табл. 1.10 и 1.11. На рисунках множества дележей изображены светло-серым цветом, темно-серые подмножества множеств дележей являются s -ядрами. Заметим, что для вершин z_1 и z_5

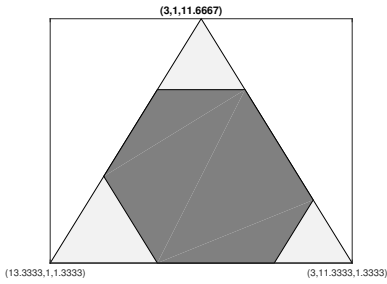
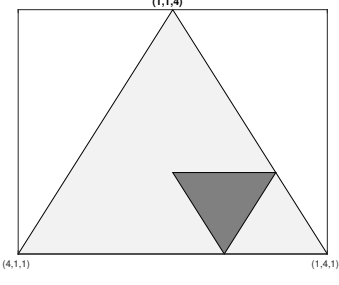
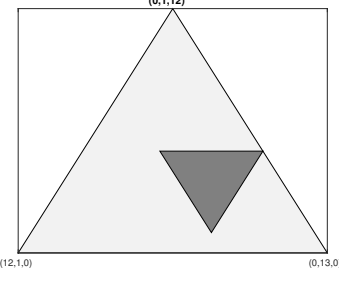
для всех элементов c -ядра: $\alpha_1 = 2$. C -ядром является отрезок, соединяющий точки $(2, 1, 3)$ и $(2, 3, 1)$.

Таблица 1.10: C -ядро для вершины $z_0, z_1, z_5 \in CZ$.

z	c -ядро	Граф. представление c -ядра
z_0	$\begin{cases} \alpha_1 \geq 2 \\ \alpha_2 \geq 1 \\ \alpha_3 \geq 1.5 \\ \alpha_1 + \alpha_2 \geq 5.5 \\ \alpha_1 + \alpha_3 \geq 6 \\ \alpha_2 + \alpha_3 \geq 4.5 \\ \alpha_1 + \alpha_2 + \alpha_3 = 110/9 \end{cases}$	
z_1, z_5	$\begin{cases} \alpha_1 \geq 2 \\ \alpha_2 \geq 0 \\ \alpha_3 \geq 1 \\ \alpha_1 + \alpha_2 \geq 3 \\ \alpha_1 + \alpha_3 \geq 3 \\ \alpha_2 + \alpha_3 \geq 4 \\ \alpha_1 + \alpha_2 + \alpha_3 = 6 \end{cases}$	

Для всех вершин кооперативного поддерева $\bar{\Psi}(z_0)$ определим аналоги c -ядер — множества $\Delta CO(z)$. Напомним, что для терминальных вершин z_1, z_3, z_4 множества $\Delta CO(\cdot)$ совпадают с ядрами $CO(\cdot)$. Системы линейных неравенств и равенства, определяющие множества $\Delta CO(z_0)$ и $\Delta CO(z_2)$, а также их графическое представление можно найти в табл. 1.12. Заметим, что аналоги c -ядер $\Delta CO(\cdot)$ непусты для всех вершин кооперативного поддерева. Сначала проверим, является ли c -ядро сильно позиционно состоятельным, если выплаты игрокам производятся в соответствии с изначально заданными функциями выигрышей, то есть проверим, принадлежат ли вектора выигрышей соответствующим множествам $\Delta CO(\cdot)$ в вершинах кооперативного поддерева при реализации ко-

Таблица 1.11: C -ядро для вершин $z_2, z_3, z_4 \in CZ$.

z_2	$\begin{cases} \alpha_1 \geq 3 \\ \alpha_2 \geq 1 \\ \alpha_3 \geq 4/3 \\ \alpha_1 + \alpha_2 \geq 7 \\ \alpha_1 + \alpha_3 \geq 7 \\ \alpha_2 + \alpha_3 \geq 6 \\ \alpha_1 + \alpha_2 + \alpha_3 = 47/3 \end{cases}$	
z_3	$\begin{cases} \alpha_1 \geq 1 \\ \alpha_2 \geq 1 \\ \alpha_3 \geq 1 \\ \alpha_1 + \alpha_2 \geq 4 \\ \alpha_1 + \alpha_3 \geq 3 \\ \alpha_2 + \alpha_3 \geq 4 \\ \alpha_1 + \alpha_2 + \alpha_3 = 6 \end{cases}$	
z_4	$\begin{cases} \alpha_1 \geq 0 \\ \alpha_2 \geq 1 \\ \alpha_3 \geq 0 \\ \alpha_1 + \alpha_2 \geq 8 \\ \alpha_1 + \alpha_3 \geq 5 \\ \alpha_2 + \alpha_3 \geq 9 \\ \alpha_1 + \alpha_2 + \alpha_3 = 13 \end{cases}$	

оперативного решения:

$$K^{z_0}(1, 1, 1) = (2, 2, 2) \in \Delta CO(z_0),$$

$$K^{z_1}(2, 2, 1) = (2, 2, 2) \in CO(z_1) = \Delta CO(z_1),$$

$$K^{z_2}(2, 1, 2) = (1, 3, 1) \notin \Delta CO(z_2),$$

$$K^{z_3}(1, 2, 1) = (2, 2, 2) \in CO(z_3) = \Delta CO(z_3),$$

$$K^{z_4}(2, 2, 1) = (3, 6, 4) \in CO(z_4) = \Delta CO(z_4).$$

Как мы видим, в вершине z_2 условие включения нарушается, поэтому мы не можем гарантировать сильную позиционную состоятельность дележа из c -ядра,

Таблица 1.12: Множества $\Delta CO(z)$ для вершин z_0 и z_2 .

z	$\Delta C(z)$	Граф. представление $\Delta CO(z)$
z_0	$\begin{cases} \alpha_1 \geq 2/3 \\ \alpha_2 \geq 2/3 \\ \alpha_3 \geq 8/9 \\ \alpha_1 + \alpha_2 \geq 8/3 \\ \alpha_1 + \alpha_3 \geq 19/6 \\ \alpha_2 + \alpha_3 \geq 11/6 \\ \alpha_1 + \alpha_2 + \alpha_3 = 6 \end{cases}$	
z_2	$\begin{cases} \alpha_1 \geq 8/3 \\ \alpha_2 \geq 0 \\ \alpha_3 \geq 1 \\ \alpha_1 + \alpha_2 \geq 1/3 \\ \alpha_1 + \alpha_3 \geq 8/3 \\ \alpha_2 + \alpha_3 \geq -4/3 \\ \alpha_1 + \alpha_2 + \alpha_3 = 5 \end{cases}$	

если игрокам производить выплаты в вершинах в соответствии с изначально заданными функциями выигрышей.

Покажем, что условие (1.9.2) нарушается в вершине z_2 . Согласно Определению 1.15 игроки могут выбрать любой дележ из ожидаемого s -ядра для вершины z . Пусть, ими будут выбраны дележи $(1.5, 3, 1.5) \in CO(z_3)$ и $(0, 8, 5) \in CO(z_4)$, тогда сумма в левой части включения (1.9.2) будет иметь вид:

$$(1, 3, 1) + \frac{1}{3}(1.5, 3, 1.5) + \frac{2}{3}(0, 8, 5) = \left(\frac{3}{2}, \frac{28}{3}, \frac{29}{6} \right),$$

причем этот вектор не принадлежит s -ядру $CO(z_2)$, что говорит о нарушении условия (1.9.2) сильной позиционной состоятельности s -ядра.

Согласно Теореме 1.6, набор векторов $\beta(z)$ из $\Delta CO(z)$, $z \in CZ$, является процедурой распределения дележа из s -ядра $CO(z_0)$ исходной игры. Из Теоремы 1.6 также следует, что такой набор $\{\beta(z) : z \in CZ\}$ является сильно позиционно состоятельным. Например, возьмем следующие элементы из $\Delta C(z)$, $z \in CZ$: $\beta(z_0) = (4, 1, 1)$, $\beta(z_1) = (2, 2, 2)$, $\beta(z_2) = (3, 1, 1)$, $\beta(z_3) = (2, 2, 2)$, $\beta(z_4) = (3, 6, 4)$. Посчитаем математическое ожидание выигрышей игроков, ес-

ли в вершинах кооперативного поддерева им производятся выплаты в соответствии с этой процедурой распределения $\{\beta(\cdot)\}$:

$$\begin{aligned} B(z_0) &= (4, 1, 1) + 0.5 \left\{ \frac{1}{3}(2, 2, 2) + \frac{2}{3} \left((3, 1, 1) + \frac{1}{3}(2, 2, 2) + \frac{2}{3}(3, 6, 4) \right) \right\} = \\ &= \left(\frac{56}{9}, \frac{29}{9}, \frac{25}{9} \right). \end{aligned}$$

Очевидно, что $B(z_0) \in CO(z_0)$.

Таким образом, мы показали способ построения сильно позиционно состоятельной процедуры распределения дележа, когда в качестве решения игры выбирается множественный принцип оптимальности — s -ядро.

Глава 2

Кооперативные стохастические игры бесконечной продолжительности

§ 2.1 Некооперативные стохастические игры бесконечной продолжительности

В этой главе рассмотрим стохастические игры бесконечной продолжительности, определение которых ввел Л. Шепли в работе [210]. Основные классические результаты по некооперативным стохастическим играм изложены в книгах [90, 149]. Как и в предыдущей главе, игра происходит в дискретном времени. Существенное отличие этой стохастической игры от рассмотренной ранее заключается в том, что игра имеет бесконечную продолжительность, а множество возможных состояний игры на каждом шаге конечно и остается неизменным на протяжении всей игры. Дадим определение стохастической игры, опишем множество стратегий и функции выигрышей игроков. Отметим, что обозначения этой главы, которые являются широко используемыми в современных публикациях по стохастическим играм, немного отличаются от обозначений главы 1.

Рассмотрим стохастическую игру G , определяемую следующими элементами:

1. $N = \{1, \dots, n\}$ — конечное множество игроков;
2. $\Omega = \{\omega_1, \dots, \bar{\omega}\}$ — конечное непустое множество состояний;
3. A_i^ω — конечное непустое множество стратегий игрока $i \in N$ в состоянии $\omega \in \Omega$, элемент которого $a_i^\omega \in A_i^\omega$ называется стратегией игрока $i \in N$ в

этом состоянии. Ситуацией в состоянии $\omega \in \Omega$ называется набор стратегий игроков: $a^\omega = (a_i^\omega : i \in N)$. Множество ситуаций в состоянии ω есть $A^\omega = A_1^\omega \times \dots \times A_n^\omega$.

4. $K_i^\omega : \prod_{k \in N} A_k^\omega \rightarrow \mathbb{R}$ — конечная функция выигрыша игрока $i \in N$ в состоянии $\omega \in \Omega$.
5. $p(\cdot | \omega, a^\omega) : \Omega \times A^\omega \rightarrow \Delta(\Omega)$ — функция перехода из состояния $\omega \in \Omega$ и ситуации $a^\omega \in \prod_{i \in N} A_i^\omega$. Здесь $\Delta(\Omega)$ — вероятностное распределение на множестве Ω .
6. $\pi_0 = (\pi_0^{\omega_1}, \dots, \pi_0^{\omega_{\bar{\omega}}})$ — начальное состояние игры, определенное вероятностным распределением, где π_0^ω — вероятность того, что состояние ω реализуется в первый период времени, $\sum_{\omega \in \Omega} \pi_0^\omega = 1$.

Игра происходит в дискретном времени t и длится бесконечное количество периодов. Стохастическая игра G разыгрывается следующим образом:

1. Перед началом игры «случай» выбирает начальное состояние из множества Ω в соответствии с вероятностным распределением π_0 , т. е. с вероятностью π_0^ω стохастическая игра начинается из состояния ω .
2. В первый период времени реализуется состояние ω . Игроки одновременно выбирают свои стратегии. Игрок $i \in N$ выбирает стратегию $a_i^\omega \in A_i^\omega$. Таким образом, в первый период времени реализуется ситуация $a^\omega = (a_i^\omega : i \in N) \in A_1^\omega \times \dots \times A_n^\omega$. При этом, игрок i получает выигрыш $K_i^\omega(a^\omega)$. Как только ситуация a^ω становится известна игрокам, игра переходит на следующий шаг в состояние $\omega' \in \Omega$ с вероятностью $p(\omega' | \omega, a^\omega)$.
3. Во второй период реализуется состояние $\omega' \in \Omega$ с вероятностью $p(\omega' | \omega, a^\omega)$. Игрок $i \in N$ выбирает стратегию $a_i^{\omega'} \in A_i^{\omega'}$. Таким образом, во второй период времени реализуется ситуация $a^{\omega'} = (a_i^{\omega'} : i \in N) \in A_1^{\omega'} \times \dots \times A_n^{\omega'}$. При этом игрок i получает выигрыш $K_i^{\omega'}(a^{\omega'})$.
4. Далее игра продолжается по вышеописанной схеме.

Пусть $\hat{a}_i^\omega \in \Delta(A_i^\omega)$ — смешанная стратегия игрока i , где $\Delta(A_i^\omega)$ — множество смешанных стратегий игрока i в состоянии ω .

Определение 2.1. *Стохастической игрой G называется набор*

$$G = \left\langle N, \Omega, \{A_i^\omega\}_{\substack{i \in N \\ \omega \in \Omega}}, \{K_i^\omega\}_{\substack{i \in N \\ \omega \in \Omega}}, \pi_0, \left\{ p(\omega'' | \omega', a^{\omega'}) \right\}_{\substack{\omega', \omega'' \in \Omega \\ a^{\omega'} \in \prod_{i \in N} A_i^{\omega'}}}, \delta \right\rangle, \quad (2.1.1)$$

где $\delta \in (0, 1)$ — дисконтирующий фактор, одинаковый для всех игроков.

Каждое состояние ω определяется игрой n лиц в нормальной форме

$$\langle N, \{A_i^\omega\}_{i \in N}, \{K_i^\omega\}_{i \in N} \rangle.$$

Переход из одного состояния в другое можно интерпретировать как (положительный или отрицательный) шок, воздействующий на систему и на выигрыши игроков.

Подыгру некооперативной стохастической игры G , начинающуюся с периода k , обозначим через $G(k)$.

Для того, чтобы решить стохастическую игру, необходимо определить множество стратегий игроков и способ вычисления выигрышей игроков во всей игре. Сначала определим стратегии игроков, выделив два класса стратегий:

- Стратегия поведения игрока $i \in N$ — это функция $\varphi_i = \{\varphi_i(k)\}_{k=1}^\infty$ и $\varphi_i(k) : h(k) \times \Omega \mapsto \Delta(A_i^\omega)$, где $h(k)$ — история шага k , которая представляет собой набор пар состояний и ситуаций, которые реализовались до шага k : $((\omega(1), a(1)), (\omega(2), a(2)), \dots, (\omega(k-1), a(k-1)))$. Обозначим множество стратегий поведения игрока i через Φ_i и ситуацию в стохастической игре в стратегиях поведения через $\varphi = (\varphi_i : i \in N)$, $\varphi_i \in \Phi_i$.
- Также рассмотрим подмножество множества стратегий поведения — множество стационарных стратегий. Стационарная стратегия предписывает игроку выбрать одну и ту же стратегию в одном и том же состоянии независимо от истории периода. Так как стационарная стратегия является стратегией поведения, то можно использовать то же обозначение, как и для стратегии поведения. Но мы введем обозначение для стационарной стратегии, чтобы в дальнейшем различать стационарную стратегию и стратегию

поведения (не обязательно стационарную) в общем понимании. Обозначим через $\eta_i = \{\eta_i(k)\}_{k=1}^{\infty}$, $\eta_i(k) : \Omega \mapsto \Delta(A_i^{\omega})$ стационарную стратегию игрока i . Обозначим ситуацию в стационарных стратегиях в стохастической игре через $\eta = (\eta_i : i \in N)$, а множество стационарных стратегий игрока i обозначим через H_i , причем, $H_i \subset \Phi_i$.

Теперь перейдем к способу определения выигрышей игроков в стохастической игре вида (2.1.1):

- Для конечного числа периодов t выигрыш игрока i в стохастической игре определяется как математическое ожидание:

$$E_i(\varphi) = E^{\omega(1), \varphi} \frac{1}{t} \sum_{k=1}^t K_i^{\omega(k)}(a(k)),$$

т. е. математическое ожидание выигрыша относительно начального состояния $\omega(1)$ и ситуации в игре φ . При этом, $K_i^{\omega(k)}(a(k))$ — выигрыш игрока i в состоянии $\omega(k)$, которое реализуется на шаге k , $a(k)$ — ситуация в состоянии $\omega(k)$, реализованном на шаге k в соответствии со стратегией φ .

- В случае бесконечной игры G дисконтированный выигрыш игрока i определяется по формуле:

$$E_i(\varphi) = E^{\omega(1), \varphi} \sum_{k=1}^{\infty} \delta^{k-1} K_i^{\omega(k)}(a(k)) \quad (2.1.2)$$

как математическое ожидание выигрыша относительно начального состояния $\omega(1)$ и ситуации φ .

Сформулируем основные результаты о существовании значений стохастических игр с нулевой суммой и равновесия по Нэшу в играх двух и более лиц, которые будут использованы далее в работе.

Теорема 2.1. [210] *Любая стохастическая игра двух лиц с нулевой суммой с дисконт-фактором $\delta \in (0, 1)$ имеет значение для любого начального состояния. При этом у игроков существуют оптимальные стационарные стратегии.*

Этот результат был обобщен на случай игры с ненулевой суммой с числом игроков, большим двух, в работах Финка и Такахаши в 1964 году.

Теорема 2.2. [91, 217] *В любой стохастической игре многих лиц с ненулевой суммой с дисконт-фактором $\delta \in (0, 1)$ с конечными множествами состояний и стратегий существует равновесие по Нэшу для любого начального состояния. При этом, равновесные стратегии игроков являются стационарными.*

§ 2.2 Основные функциональные уравнения для стохастических игр бесконечной продолжительности в стационарных стратегиях

В этом разделе покажем, как вычисляются выигрыши игроков в стохастической игре, если игроки используют стационарные стратегии. Так как множество состояний Ω конечно, то число различных возможных подыгр игры G равно числу состояний во множестве Ω , поскольку стационарные стратегии предписывают одинаковое поведение игрокам в одних и тех же состояниях при различных историях текущего периода. Таким образом, если игра решается в стационарных стратегиях, то число подыгр равно $|\Omega|$. Обозначим подыгру игры G , начинающуюся из состояния $\omega \in \Omega$, через G^ω .

Функцию перехода $p(\omega'|\omega, a^\omega)$ в случае стационарных стратегий можно записать в виде матрицы размерности $\bar{\omega} \times \bar{\omega}$:

$$\Pi(\eta) = \begin{pmatrix} p(\omega_1|\omega_1, a^{\omega_1}) & \dots & p(\bar{\omega}|\omega_1, a^{\omega_1}) \\ p(\omega_1|\omega_2, a^{\omega_2}) & \dots & p(\bar{\omega}|\omega_2, a^{\omega_2}) \\ \dots & \dots & \dots \\ p(\omega_1|\bar{\omega}, a^{\bar{\omega}}) & \dots & p(\bar{\omega}|\bar{\omega}, a^{\bar{\omega}}) \end{pmatrix} \quad (2.2.1)$$

как функцию ситуации в стационарных стратегиях $\eta = (\eta_i : i \in N)$ такой, что $\eta_i(\omega) = a_i^\omega \in \Delta(A_i^\omega)$, $\omega \in \Omega$, $i \in N$, и $a^\omega = (a_1^\omega, \dots, a_n^\omega)$ для любого состояния $\omega \in \Omega$. Элемент матрицы (2.2.1), стоящий на пересечении строки j и столбца j' — вероятность перехода из состояния j в состояние j' при условии реализации ситуации $\eta = (\eta_i : i \in N)$.

Упростим формулу (2.1.2) для вычисления выигрыша игрока i , а именно, вычислим математическое ожидание его выигрыша в явном виде. Пусть $E_i^\omega(\eta)$ — выигрыш игрока i в подыгре G^ω в случае, когда реализуется ситуация в стационарных стратегиях $\eta = (\eta_1, \dots, \eta_n)$. Запишем выигрыш игрока i в векторной форме: $E_i(\eta) = (E_i^{\omega_1}(\eta), \dots, E_i^{\bar{\omega}}(\eta))^T$.

Функция выигрыша игрока i в подыгре G^ω удовлетворяет рекуррентному уравнению:

$$E_i^\omega(\eta) = K_i^\omega(a^\omega) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^\omega) E_i^{\omega'}(\eta). \quad (2.2.2)$$

Учитывая, что вероятности перехода записаны в виде матрицы (2.2.1), перепишем уравнение (2.2.2) в матричной форме следующим образом:

$$E_i(\eta) = K_i(a) + \delta \Pi(\eta) E_i(\eta), \quad (2.2.3)$$

где $K_i(a) = (K_i^{\omega_1}(a^{\omega_1}), \dots, K_i^{\bar{\omega}}(a^{\bar{\omega}}))^T$. Уравнение (2.2.3) эквивалентно уравнению

$$E_i(\eta) = (\mathbb{I} - \delta \Pi(\eta))^{-1} K_i(a),$$

где \mathbb{I} — единичная матрица размерности $\bar{\omega} \times \bar{\omega}$. Матрица $(\mathbb{I} - \delta \Pi(\eta))^{-1}$ всегда существует для дисконт фактора $\delta \in (0, 1)$. При этом выигрыш игрока i в игре G с учетом выбора начального состояния в соответствии с распределением π_0 в ситуации η равен

$$\bar{E}_i(\eta) = \pi_0 E_i(\eta) = \pi_0 (\mathbb{I} - \delta \Pi(\eta))^{-1} K_i(a). \quad (2.2.4)$$

§ 2.3 Определение кооперативной стохастической игры бесконечной продолжительности

В этом параграфе построим кооперативный вариант стохастической игры G , который обозначим через G_c . Предположим, что игроки решили кооперироваться, образуя максимальную коалицию N и максимизировать суммарный выигрыш игроков коалиции. Существование максимума дисконтированного суммарного выигрыша доказано в статье [210], причем ситуация, доставляющая максимум, является ситуацией в стационарных стратегиях. Обозначим чистую стратегию

игрока i через $\eta_i \in H_i$, где $H_i \subset \Phi_i$, а смешанную стационарную стратегию игрока i через $\hat{\eta}_i \in \hat{H}_i$, где $H_i \subset \hat{H}_i$.¹

Вектор кооперативных стратегий, при котором достигается максимум суммарного выигрыша игроков в игре G , обозначим через $\eta^* = (\eta_1^*, \dots, \eta_n^*)$, где²

$$\max_{\eta \in \prod_{i \in N} H_i} \sum_{i \in N} \bar{E}_i(\eta) = \sum_{i \in N} \bar{E}_i(\eta^*). \quad (2.3.1)$$

Определив вектор кооперативных стратегий, определим значения характеристической функции для любой коалиции $S \subseteq N$. Эта функция показывает «силу» коалиции, которую игроки формируют. Обозначим характеристическую функцию через $V(S) = (V^{\omega_1}(S), \dots, V^{\omega}(S))$. Как и в [115], пусть $V(S)$ — минимаксное значение игры с нулевой суммой G_S между коалицией S и коалицией $N \setminus S$.³ Обозначим чистые стационарные стратегии коалиций S и $N \setminus S$ через $\eta_S \in H_S = \prod_{i \in S} H_i$ и $\eta_{N \setminus S} \in H_{N \setminus S} = \prod_{i \in N \setminus S} H_i$ соответственно.

Замечание 2.1. При определении характеристической функции $V(S)$, $S \subseteq N$, мы предполагаем, что игроки из коалиции S играют в интересах коалиции S . Следовательно, стратегии игроков из S коррелированы [67].

В состоянии $\omega \in \Omega$, коррелированные стратегии игроков из S есть $\hat{a}_S^\omega \in \Delta(A_S^\omega)$, где $A_S^\omega = \prod_{i \in S} A_i^\omega$. Коррелированные стратегии игроков из коалиции S и $N \setminus S$ есть $\hat{\eta}_S(\omega) \in \Delta(A_S^\omega)$ и $\hat{\eta}_{N \setminus S}(\omega) \in \Delta(A_{N \setminus S}^\omega)$, соответственно. Пусть множество коррелированных стратегий коалиций S и $N \setminus S$ есть \hat{H}_S и $\hat{H}_{N \setminus S}$ соответственно.

Начнем с определения характеристической функции для максимальной коалиции $S = N$. Запишем уравнение Беллмана для характеристической функции $V(N)$:

$$V(N) = \max_{\eta \in \prod_{i \in N} H_i} \sum_{i \in N} \bar{E}_i(\eta) = \sum_{i \in N} K_i(a^*) + \delta \Pi(\eta^*) V(N), \quad (2.3.2)$$

¹Начиная с этого момента и далее мы будем использовать обозначение η_i , если игрок i использует стационарную стратегию в игре. Если игрок i использует стратегию поведения (не обязательно стационарную), будем использовать обозначение φ_i .

²Без потери общности, максимум в левой части (2.3.1) ищется на множестве чистых стратегий коалиции N .

³Существование минимаксного значения дисконтированной стохастической игры двух лиц доказано Шепли [210].

где η^* — кооперативное решение, удовлетворяющее условию (2.3.1), и $\eta^*(\omega) = a^{\omega^*}$, $\omega \in \Omega$, а также $K_i(a^*) = (K_i^{\omega_1}(a^{\omega_1^*}), \dots, K_i^{\bar{\omega}}(a^{\bar{\omega}^*}))$. Из (2.3.2) получим матричную форму для $V(N)$:

$$V(N) = (\mathbb{I} - \delta\Pi(\eta^*))^{-1} \sum_{i \in N} K_i(a^*), \quad (2.3.3)$$

где \mathbb{I} — единичная матрица размерности $\bar{\omega} \times \bar{\omega}$ и $\Pi(\eta^*)$ — матрица размерности $\bar{\omega} \times \bar{\omega}$ вероятностей перехода в игре G , когда игроки реализуют ситуацию η^* . Матрица $\Pi(\eta^*)$ приведена выше формулой (2.2.1).

Далее определим характеристическую функцию $V^\omega(S)$ для значений $S \subset N$ как минимаксный выигрыш в подыгре G_S^ω , начинающейся из состояния ω :⁴

$$V^\omega(S) = \min_{\hat{\eta}_{N \setminus S}} \max_{\eta_S} \sum_{i \in S} E_i^\omega(\eta_S, \hat{\eta}_{N \setminus S}) = \max_{\hat{\eta}_S} \min_{\eta_{N \setminus S}} \sum_{i \in S} E_i^\omega(\hat{\eta}_S, \eta_{N \setminus S}). \quad (2.3.4)$$

Уравнение Беллмана для значения характеристической функции $V^\omega(S)$ принимает вид:

$$\begin{aligned} V^\omega(S) &= \min_{\hat{\eta}_{N \setminus S} \in \hat{H}_{N \setminus S}} \max_{\eta_S \in H_S} \sum_{i \in S} E_i^\omega(\eta_S, \hat{\eta}_{N \setminus S}) = \sum_{i \in S} E_i^\omega(\eta_S, \hat{\eta}_{N \setminus S}) \\ &= \sum_{i \in S} K_i^\omega(\mathbf{a}_S^\omega, \hat{\mathbf{a}}_{N \setminus S}^\omega) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, (\mathbf{a}_S^\omega, \hat{\mathbf{a}}_{N \setminus S}^\omega)) V^{\omega'}(S), \end{aligned} \quad (2.3.5)$$

где $(\mathbf{a}_S^\omega, \hat{\mathbf{a}}_{N \setminus S}^\omega)$ — ситуация в коррелированных стратегиях в состоянии $\omega \in \Omega$ такая, что $\eta_S(\omega) = \mathbf{a}_S^\omega$, $\hat{\eta}_{N \setminus S}(\omega) = \hat{\mathbf{a}}_{N \setminus S}^\omega$, и $K_i(\mathbf{a}_S^\omega, \hat{\mathbf{a}}_{N \setminus S}^\omega) = (K_i^{\omega_1}(\mathbf{a}_S^{\omega_1}, \hat{\mathbf{a}}_{N \setminus S}^{\omega_1}), \dots, K_i^{\bar{\omega}}(\mathbf{a}_S^{\bar{\omega}}, \hat{\mathbf{a}}_{N \setminus S}^{\bar{\omega}}))$. Теперь перепишем уравнение (2.3.5) в матричной форме:

$$V(S) = (\mathbb{I} - \delta\Pi(\eta_S, \hat{\eta}_{N \setminus S}))^{-1} \sum_{i \in N} K_i(\mathbf{a}_S^\omega, \hat{\mathbf{a}}_{N \setminus S}^\omega). \quad (2.3.6)$$

Наконец определим характеристическую функцию $\bar{V}(S)$ для стохастической игры G :

$$\bar{V}(S) = \pi_0 V(S), \quad (2.3.7)$$

для любой коалиции $S \subseteq N$, где $V(S) = (V^{\omega_1}(S), \dots, V^{\bar{\omega}}(S))$, и $V^\omega(S)$ — значение характеристической функции для подыгры G^ω для коалиции S .

⁴В уравнении (2.3.4), максимум в $\min_{\hat{\eta}_{N \setminus S}} \max_{\eta_S} \sum_{i \in S} E_i^\omega(\eta_S, \eta_{N \setminus S})$ находится на множестве чистых стратегий коалиции S , также как и минимум в $\max_{\hat{\eta}_S} \min_{\eta_{N \setminus S}} \sum_{i \in S} E_i^\omega(\eta_S, \eta_{N \setminus S})$ находится на множестве чистых стратегий коалиции $N \setminus S$.

Также предположим, что для любого состояния $\omega \in \Omega$:

$$V^\omega(\emptyset) = 0. \quad (2.3.8)$$

Характеристические функции $\bar{V}(S)$ и $V^\omega(S)$, $\omega \in \Omega$, определенные формулами (2.3.7) и (2.3.3)-(2.3.8), являются супераддитивными [69]. Это означает, что для любых непересекающихся коалиций $S, T \subset N$, $S \cap T = \emptyset$, справедливо неравенство $V(S) + V(T) \leq V(S \cup T)$. Если это неравенство не выполнено для коалиции $S \cup T$, то коалицию $S \cup T$ не выгодно формировать⁵.

Таким образом мы построили кооперативный вариант стохастической игры из определения 2.1, а также для всех её подыгр.

Определение 2.2. Кооперативной стохастической игрой G_c , соответствующей стохастической игре G , называется пара $\langle N, \bar{V} \rangle$, где N — множество игроков, $\bar{V} : 2^N \rightarrow \mathbb{R}$ — характеристическая функция, определенная формулой (2.3.7). Кооперативную стохастическую подыгру, начинающуюся из состояния ω , обозначим через G_c^ω , она задается парой $\langle N, V^\omega \rangle$, где $V^\omega : 2^N \rightarrow \mathbb{R}$ — характеристическая функция, определяемая формулами (2.3.3), (2.3.4) и (2.3.8).

Когда формируется максимальная коалиция, игроки принимают решение не только о совместном выборе стратегий в игре с целью максимизации суммарного выигрыша максимальной коалиции, но и о том, каким образом будет распределяться этот выигрыш между игроками. Дадим определение дележа в подыгре G_c^ω и игре G_c . Для определения дележа максимального суммарного выигрыша (2.3.1) чаще всего нам необходимо знать значения характеристической функции для любой коалиции $S \subset N$.

Определение 2.3. Дележ в подыгре G_c^ω , $\omega \in \Omega$, — это вектор $\sigma^\omega = (\sigma_1^\omega, \dots, \sigma_n^\omega)$, удовлетворяющий условиям: (1) $\sum_{i \in N} \sigma_i^\omega = V^\omega(N)$, и (2) $\sigma_i^\omega \geq V^\omega(\{i\})$ для любого игрока $i \in N$. Множество дележей в подыгре G_c^ω обозначим через Σ^ω .

⁵Свойство супераддитивности не считается обязательным в современной теории кооперативных игр и часто опускается, поскольку в реальной жизни могут существовать как прибыльные, так и неприбыльные коалиции. Как было замечено в [68] (см. стр. 233), что доводы в пользу супераддитивности в кооперативных играх весьма убедительны, но добиться супераддитивности в некоторых экономических задачах бывает зачастую проблематично.

Определение 2.4. Дележ в игре G_c — вектор $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n)$, где $\bar{\sigma}_i = \pi_0 \sigma_i$, $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\omega_n})^T$, и $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$, $\omega \in \Omega$. Множество дележей в игре G_c обозначим через $\bar{\Sigma}$.

Согласно определению 2.3 дележ удовлетворяет следующим свойствам: (1) любой игрок должен получить в результате кооперации не меньше, чем он мог бы получить, играя индивидуально (*условие индивидуальной рациональности*), (2) сумма компонент дележа равняется значению характеристической функции максимальной коалиции (*условие коллективной рациональности*). Множество дележей непусто в любой подыгре G_c^ω , $\omega \in \Omega$ и в кооперативной стохастической игре G_c . Это следует из способа определения характеристической функции, а именно, из формул (2.3.2)-(2.3.8).

В следующем параграфе будут сформулированы и изучены принципы устойчивой кооперации для рассматриваемого класса игр.

§ 2.4 Принципы устойчивой кооперации в стохастических играх бесконечной продолжительности

В кооперативных играх решение игры определяется принципом оптимальности. Под принципом оптимальности понимается подмножество множества дележей. Таким образом, принцип оптимальности содержит один или более дележей, или может быть пустым. Вектор Шепли как принцип оптимальности всегда существует и содержит единственный дележ, то есть решение кооперативной игры в соответствии с этим принципом оптимальности существует и единственно. Под решением кооперативной стохастической игры будем понимать дележ⁶. Чтобы не рассматривать вопрос о том, каким образом выбирается дележ из множества, предположим, что принцип оптимальности содержит единственный дележ. Примерами одноточечных принципов оптимальности могут быть вектор Шепли [211] и n -ядро [206]. Реализация решения в кооперативной стохастической игре требует выполнения некоторых принципов, которые могут служить основой *устойчивой кооперации*. Следуя [44, 183], выделим основные принци-

⁶Ниже будет рассмотрен случай, когда в качестве решения кооперативной стохастической игры рассматривается множество дележей.

пы устойчивой кооперации, включающие позиционную состоятельность, стратегическую устойчивость (или стратегическую поддержку) и защиту от иррационального поведения кооперативного решения стохастической игры. Дадим определения и исследуем условия выполнения каждого принципа устойчивой кооперации.

§ 2.4.1 Позиционная состоятельность принципа оптимальности в стохастических играх бесконечной продолжительности

Принцип позиционной состоятельности или динамической устойчивости заключается в следующем. Игроки в начале игры выбирают принцип оптимальности для определения дележа совместного выигрыша. Для реализации кооперативного решения игроки придерживаются кооперативного поведения или кооперативных стратегий η_i^* , $i \in N$. Если на каком-либо шаге игроки решат рассчитать свои ожидаемые выигрыши в подыгре в соответствии с изначально заданными функциями выигрышей, то чаще всего эти ожидаемые дисконтированные выигрыши не будут совпадать с кооперативным решением, рассчитанным в соответствии с изначально выбранным принципом оптимальности. Это означает позиционную несостоятельность кооперативного решения (или принципа оптимальности). Если для любой подыгры дисконтированные выигрыши игроков совпадают с кооперативными решениями, рассчитанными в соответствии с начальным принципом оптимальности, то будем говорить о позиционной состоятельности кооперативного решения или принципа оптимальности (см. [35]). Для того, чтобы сделать кооперативное решение позиционно состоятельным, можно предложить механизм трансфертных выплат или так называемую *процедуру распределения дележа*⁷ (ПРД), которая изначально предложена Л. А. Петросяном для дифференциальных игр [43].

Применение процедуры распределения дележа приводит к модификации функций выигрышей игроков. Назовем модифицированную игру σ -регуляризацией, где σ — дележ, выбранный изначально в качестве кооперативного решения в игре G_c . Несмотря на то, что модифицированная игра изменяет функции выигры-

⁷Процедура распределения дележа была адаптирована для класса дисконтированных стохастических игр в работе [72].

шей игроков, она имеет несколько положительных свойств. Во-первых, кооперативное решение σ является позиционно состоятельным в модифицированной игре с «новыми» функциями выигрышей, то есть суммы ожидаемых выплат в σ -регуляризации начальной игры совпадают с компонентами решения σ . Во-вторых, сумма выплат игрокам на каждом шаге в σ -регуляризации совпадает с суммой выигрышей игроков на этом шаге в игре G_c . Например, если игроки выбрали вектор Шепли в качестве кооперативного решения в начальной игре, то условие позиционной состоятельности означает, что в любой ее подыгре G_c^ω вектор, состоящий из сумм ожидаемых выплат в этой подыгре, совпадает с вектором Шепли, рассчитанным для этой подыгры.

Пусть игроки реализуют кооперативное решение в стохастической игре и для любой подыгры G_c^ω выбирается дележ $\sigma^\omega = (\sigma_1^\omega, \dots, \sigma_n^\omega) \in \Sigma^\omega$ в соответствии с одним и тем же принципом оптимальности. Задача состоит в том, чтобы определить выплаты игрокам, гарантирующие игроку i получение σ_i^ω в подыгре G_c^ω в качестве ожидаемого выигрыша.

Когда игроки реализуют кооперативные стационарные стратегии в стохастической игре (2.1.1), то число различных подыгр стохастической игр равно числу состояний в игре. Поэтому для определения ПРД необходимо задать вектор выплат $\beta_i = (\beta_i^{\omega_1}, \dots, \beta_i^{\omega_n})^T$, где β_i^ω — выплата игроку $i \in N$ в состоянии $\omega \in \Omega$.

Определение 2.5. Множество выплат $\{\beta_i\}_{i \in N}$ называется процедурой распределения дележа $\bar{\sigma}$ (ПРД), если выполнены условия:

1. В каждом состоянии $\omega \in \Omega$ сумма выплат игрокам равна сумме выигрышей игроков при кооперативном решении $\bar{\eta}$:

$$\sum_{i \in N} \beta_i^\omega = \sum_{i \in N} K_i^\omega(a^{\omega*}). \quad (2.4.1)$$

2. Ожидаемая сумма выплат игроку $i \in N$, во всей игре \bar{G} равна i -ой компоненте дележа $\bar{\sigma}$.

Определим позиционно состоятельное кооперативное решение и соответствующую ему ПРД.

Определение 2.6. Дележ $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n)$ и соответствующая ему ПРД $\{\beta_i\}_{i \in N}$ называются позиционной состоятельными, если ожидаемая сумма выплат игроку i в подыгре G^ω равна i -ой компоненте дележа, рассчитанного для подыгры G_c^ω в соответствии с изначально выбранным принципом оптимальности (в соответствии с которым был рассчитан дележ $\bar{\sigma}$ для всей игры).

Следующее утверждение позволяет построить динамически устойчивую ПРД для дележа $\bar{\sigma}$.

Лемма 2.1. Пусть дележ $\bar{\sigma}$ такой, что $(\bar{\sigma}_1, \dots, \bar{\sigma}_n) \in \bar{\Sigma}$, где $\bar{\sigma}_i = \pi_0 \sigma_i$, $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\omega_n})^T$ и $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$, $\omega \in \Omega$. Тогда набор $\{\beta_i\}_{i \in N}$, где вектор β_i вычисляется по формуле

$$\beta_i = (\mathbb{I} - \delta \Pi(\eta^*)) \sigma_i, \quad (2.4.2)$$

есть процедура распределения дележа⁸ $\bar{\sigma}$ в игре G .

Доказательство. Проверим первое условие ПРД:

$$\sum_{i \in N} \beta_i^\omega = \sum_{i \in N} K_i^\omega(a^{\omega*}),$$

где $a^{\omega*}$ — ситуация в состоянии ω , являющаяся реализацией ситуации η^* во всей игре. Нетрудно показать, что β_i из (2.4.2) удовлетворяет (2.4.1). Так как $\sum_{i \in N} \beta_i$ равна $(\mathbb{I} - \delta \Pi(\eta^*)) \sum_{i \in N} \sigma_i = (\mathbb{I} - \delta \Pi(\eta^*)) V(N)$, и $V(N)$ определена формулой (2.3.3), то равенство (2.4.1) выполнено.

Второе условие ПРД также выполнено, поскольку ожидаемый выигрыш игрока i , обозначенный через B_i , в игре с трансферными выплатами β_i^ω в состоянии $\omega \in \Omega$, удовлетворяет рекуррентному уравнению:

$$B_i^\omega = \beta_i^\omega + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) B_i^{\omega'},$$

⁸ПРД определяется единственным образом по формуле (2.4.2), если принцип оптимальности дает единственное кооперативное решение $\bar{\sigma}$ (например, в случае n -ядра, вектора Шепли или любого другого одноточечного принципа оптимальности). В случае, когда кооперативное решение — это множество дележей, состоящее более чем из одной точки, то метод построения ПРД требует модификации, которая, например, предложена в работе [171].

или в векторной форме:

$$B_i = \beta_i + \delta\Pi(\eta^*)B_i, \quad (2.4.3)$$

где $B_i = (B_i^{\omega_1}, \dots, B_i^{\bar{\omega}})^T$. Уравнение (2.4.3) эквивалентно следующему:

$$B_i = (\mathbb{I} - \delta\Pi(\eta^*))^{-1} \beta_i. \quad (2.4.4)$$

Из второго условия ПРД и уравнения (2.4.4) получаем

$$\sigma_i = (\mathbb{I} - \delta\Pi(\eta^*))^{-1} \beta_i, \quad (2.4.5)$$

где $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\bar{\omega}})^T$, $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$. Уравнение (2.4.5) может быть записано в эквивалентной форме:

$$\beta_i = (\mathbb{I} - \delta\Pi(\eta^*))\sigma_i. \quad (2.4.6)$$

В итоге, уравнение (2.4.2) эквивалентно следующему уравнению:

$$\sigma_i = \beta_i + \delta\Pi(\eta^*)\sigma_i. \quad (2.4.7)$$

Второе слагаемое в правой части уравнения (2.4.7) — ожидаемая сумма выплат, посчитанная для подыгры, начинающейся со следующего шага. Предположим, что дележ в каждой подыгре выбирается в соответствии с одним и тем же принципом оптимальности, что и в начале игры. Очевидно, что, кооперируясь, игроки реализуют вектор кооперативных стратегий η^* , тогда ожидаемый выигрыш игрока i при условии реализации выплат в соответствии в ПРД равен соответствующей компоненте дележа $\bar{\sigma}$ в кооперативной стохастической игре G_c . \square

По определению 2.5 для каждого $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n) \in \bar{\Sigma}$, где $\bar{\sigma}_i = \pi_0\sigma_i$, $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\bar{\omega}})^T$, $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$, мы можем провести регуляризацию игры G следующим образом.

Определение 2.7. Назовем σ -регуляризацией стохастической игры G (подыгры G^ω , $\omega \in \Omega$) некооперативную стохастическую игру G_σ (подыгру G_σ^ω), если для любого $i \in N$ в состоянии ω функция выигрыша $K_i^{\sigma, \omega}(a^\omega)$ определена следующим образом:

$$K_i^{\sigma, \omega}(a^\omega) = \begin{cases} \beta_i^\omega, & \text{если } a^\omega = a^{\omega*}, \\ K_i^\omega(a^\omega), & \text{если } a^\omega \neq a^{\omega*}, \end{cases} \quad (2.4.8)$$

где β_i^ω — компонента ПРД игрока i , определяемая формулой (2.4.2), и $a^{\omega*} = \eta^*(\omega)$.

Функция выигрыша вида (2.4.8) является модифицированной функцией выигрыша игры G .

Замечание 2.2. σ -регуляризация меняет функции выигрышей в каждом состоянии $\omega \in \Omega$ только при условии, что реализуется кооперативная ситуация $a^{\omega*} = \eta^*(\omega)$. Можно ожидать, что игроки согласятся на изменение начально заданных функций выигрышей для того, чтобы выбранное ими кооперативное решение удовлетворяло принципу позиционной состоятельности.

Следующая теорема показывает, что в σ -регуляризации начальной стохастической игры G дележ $\bar{\sigma}$ удовлетворяет принципу позиционной состоятельности.

Теорема 2.3. Пусть $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n) \in \bar{\Sigma}$ — дележ в игре G , где $\bar{\sigma}_i = \pi_0 \sigma_i$, $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\bar{\omega}})^T$, $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$, $\omega \in \Omega$, тогда σ -регуляризация стохастической игры G удовлетворяет принципу динамической устойчивости, то есть $\bar{\sigma}$ является позиционно состоятельным решением в игре G_σ .

Доказательство. Пусть в начале игры в качестве решения выбран дележ $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n) \in \bar{\Sigma}$, где $\bar{\sigma}_i = \pi_0 \sigma_i$, $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\bar{\omega}})^T$, $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$. Рассмотрим σ -регуляризацию игры G , которая произведена в соответствии с определением 2.7, при этом трансфертные выплаты $\{\beta_i\}_{i \in N}$, вычисляемые по формуле (2.4.2), являются ПРД, что следует из леммы 2.1. Для того чтобы доказать, что σ -регуляризация игры G удовлетворяет принципу позиционной состоятельности, необходимо посчитать дисконтированные выигрыши игроков в каждой подыгре игры G_σ в случае, когда игроки реализуют ситуацию η^* . Рассмотрим подыгру G_σ^ω , начинающуюся из состояния $\omega \in \Omega$. Дисконтированный выигрыш игрока i в этой подыгре равен:

$$E_i^\omega(\eta^*) = \beta_i^\omega + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) E_i(\eta^*), \quad (2.4.9)$$

где $E_i(\eta^*) = (E_i^{\omega_1}(\eta^*), \dots, E_i^{\bar{\omega}}(\eta^*))^T$ и $E_i^\omega(\eta^*)$ — дисконтированный выигрыш игрока i в подыгре G_σ^ω , начинающейся из состояния ω , когда игроки реализуют

η^* . Уравнение (2.4.9) может быть записано в векторной форме:

$$E_i(\eta^*) = \beta_i + \delta\Pi(\eta^*)E_i(\eta^*),$$

или

$$E_i(\eta^*) = (\mathbb{I} - \delta\Pi(\eta^*))^{-1} \beta_i.$$

Учитывая, что β_i удовлетворяет (2.4.2), получаем

$$E_i(\eta^*) = (\mathbb{I} - \delta\Pi(\eta^*))^{-1} (\mathbb{I} - \delta\Pi(\eta^*)) \sigma_i = \sigma_i.$$

Это равенство доказывает, что σ -регуляризация игры G удовлетворяет принципу позиционной состоятельности. \square

Определение 2.7 и теорема 2.3 предлагают метод построения динамически устойчивых платежей в каждом состоянии стохастической игры. Выплаты компонент процедуры распределения дележа $\beta_i^{\omega_1}, \dots, \beta_i^{\bar{\omega}}$ в состояниях $\omega_1, \dots, \bar{\omega}$ предполагают, что игрок i получает ожидаемый дисконтированный выигрыш в игре G_σ (подыгре G_σ^ω), равной соответствующей компоненте кооперативного решения в игре G_c (подыгре G_c^ω). Более того, ожидаемая сумма будущих выплат в каждой подыгре соответствует одному и тому же принципу оптимальности на протяжении всей игры при реализации кооперативного решения.

§ 2.4.2 Стратегическая устойчивость кооперативного решения в стохастических играх бесконечной продолжительности

Принцип стратегической устойчивости гарантирует, что на протяжении игры индивидуальное отклонение от кооперативного решения в σ -регуляризованной игре не принесет игрокам выгоду. Другими словами, условие стратегической устойчивости гарантирует существование равновесия по Нэшу в σ -регуляризованной игре, при котором выигрыши игроков совпадают с соответствующими компонентами выбранного дележа σ (по которому произведена σ -регуляризация игры). Этот принцип устойчивой кооперации был сформулирован в работе [38]. В свою очередь, переформулируем его для стохастических игр, а также найдем условия, при которых равновесие по Нэшу является абсолютным равновесием (subgame perfect Nash equilibrium, см. [208] или [209]) в σ -регуляризованной

игре с вышеописанными выигрышами. Свойство абсолютного равновесия существенно для динамических игр, поскольку позволяет гарантировать существование равновесия по Нэшу в любой подыгре, при котором выигрыши игроков совпадают с кооперативными. Можно сравнить анализ, приведенный в этом разделе, со стандартным анализом, используемым в области повторяющихся игр. Условие стратегической устойчивости для стохастических (или в широком понимании, динамических) игр соответствует условию существования ситуации (абсолютного) равновесия по Нэшу в стратегиях наказания. Существенная разница заключается в том, что в нашей постановке игроки сначала проводят σ -регуляризацию стохастической игры для того, чтобы выбранное ими кооперативное решение было позиционно состоятельным.

Предположим, что игроки пришли к кооперативному соглашению, то есть нашли кооперативное решение, максимизирующее суммарный дисконтированный выигрыш игроков в игре. Если какой-либо игрок отклоняется от кооперативного решения, то остальные игроки переключаются с кооперативного поведения на стратегию наказания отклонившегося игрока. Строгое определение стратегии поведения, используемой игроками в ситуации равновесия по Нэшу приведено ниже (см. формулу (2.4.12)). Здесь предполагается, что стохастическая игра является игрой с совершенной информацией, то есть, все игроки знают, в каком состоянии они находятся в текущем шаге и историю этого шага.

Дадим определение равновесия по Нэшу в σ -регуляризованной стохастической игре. Обозначим через $E_i^{\omega, \sigma}$ дисконтированный выигрыш игрока i в σ -регуляризации подыгры G^ω , начинающейся из состояния $\omega \in \Omega$.

Определение 2.8. *Равновесием по Нэшу в σ -регуляризованной стохастической игре G_σ называется ситуация в стратегиях поведения $\varphi^* = (\varphi_1^*, \dots, \varphi_n^*)$ такая, что для любого игрока $i \in N$ и любого состояния $\omega \in \Omega$, справедливо неравенство*

$$E_i^{\omega, \sigma}(\varphi_i^*, \varphi_{N \setminus i}^*) \geq E_i^{\omega, \sigma}(\varphi_i, \varphi_{N \setminus i}^*) \quad (2.4.10)$$

при любой стратегии $\varphi_i \in \Phi_i$ игрока i .

Мы предполагаем, что стратегия поведения имеет следующую структуру. Если история шага k такова, что все игроки до этого шага использовали только

кооперативные стратегии, то на шаге k все игроки также разыгрывают коррелированные кооперативные стратегии. Если же до шага k наблюдается индивидуальное отклонение игрока $z \in N$, то тогда коалиция $N \setminus z$ начинает с шага k наказывать игрока z . Мы предполагаем, что при реализации наказания игрок $z \in N$ может гарантировать себе минимаксное значение в любой подыгре.⁹ Заметим, что если мы строим равновесие по Нэшу в игре, то нам необходимо рассмотреть только индивидуальные отклонения от кооперативного решения.¹⁰ Если от коалиции N отклоняется более чем один игрок, то игрок может реализовать любую стратегию из своего множества стратегий.

Теперь выпишем условие, при котором существует равновесие по Нэшу с кооперативными выигрышами. Введем обозначение:

$$F(\{i\}) \equiv (F^{\omega_1}(\{i\}), \dots, F^{\bar{\omega}}(\{i\}))^T, \\ F^{\omega}(\{i\}) = \max_{\hat{a}_i^{\omega} \in \Delta(A_i^{\omega})} \left\{ K_i^{\omega}(\hat{a}_i^{\omega}, a_{N \setminus i}^{\omega*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, (\hat{a}_i^{\omega}, a_{N \setminus i}^{\omega*})) V^{\omega'}(\{i\}) \right\}.$$

Неравенство

$$\sigma_i = (\mathbb{I} - \delta \Pi(\eta^*))^{-1} \beta_i \geq F(\{i\}), \quad (2.4.11)$$

сравнивает два выигрыша игрока i для каждой подыгры: 1) в левой части неравенства — выигрыш, когда игроки реализуют кооперативное решение, 2) в правой части — выигрыш игрока при отклонении на текущем шаге, суммированный с ожидаемым выигрышем в наказании. Если первый выигрыш не меньше, чем второй, то игрок не получит прибыль при отклонении. Если это неравенство выполнено для любого игрока и любого состояния, то принцип стратегической устойчивости кооперативного решения удовлетворен. Этот результат сформулирован в следующей теореме.

Теорема 2.4. *Если в σ -регуляризованной игре G_{σ} такой, что $\bar{\sigma} = \pi_0 \sigma$, неравенство (2.4.11) выполняется для любого игрока $i \in N$ и любого состояния*

⁹Строгое определение описанной стратегии поведения приведено в доказательстве теоремы 2.4.

¹⁰Если же мы хотим при этом доказать, что равновесие по Нэшу является абсолютным (subgame perfect), то нам нужно доказать, что неравенство (2.4.10) выполняется при любой истории текущего шага и на любом шаге. В связи с этим, нам надо определить стратегию игрока в равновесии по Нэшу, даже если более одного игрока отклоняется от кооперативной ситуации. Стратегия (2.4.20) определяет поведения игрока при любой возможной истории шага.

$\omega \in \Omega$, тогда ситуация в стратегиях поведения $\widehat{\varphi}$ вида (2.4.12) является равновесием по Нэшу с выигрышами игроков $(\bar{\sigma}_1, \dots, \bar{\sigma}_n)$.

Доказательство. Рассмотрим ситуацию в стратегиях поведения $\widehat{\varphi} = (\widehat{\varphi}_1, \dots, \widehat{\varphi}_n)$, где стратегия $\widehat{\varphi}_i$, $i \in N$, такова что

$$\widehat{\varphi}_i(h(k)) = \begin{cases} a_i^{\omega*}, & \text{если } \omega(k) = \omega, h(k) \subset h^*; \\ \hat{a}_i^\omega(z), & \text{если } \omega(k) = \omega, \text{ и } \exists l \in [1, k-1], \\ & z \in N, i \neq z: h(l) \subset h^*, \text{ и} \\ & (\omega(l), a(l)) \notin h^*, \text{ но} \\ & (\omega(l), (a_z^*(l), a_{N \setminus z}(l))) \in h^*; \\ \text{любая} & \text{в противном случае,} \end{cases} \quad (2.4.12)$$

где $a_i^{\omega*}$ — кооперативная стратегия игрока i , $\hat{a}_i^\omega(z) \in \Delta(A_i^\omega)$ — стратегия наказания, реализуемая игроком i такая, что совместно со стратегиями $\hat{a}_{i'}^\omega(z) \in \Delta(A_{i'}^\omega)$ игроков $i' \neq i$, $i' \in N \setminus z$, формирует (в чистых или смешанных стратегиях) стратегию коалиции $N \setminus z$ против игрока z .¹¹

Идея доказательства теоремы близка к доказательству народной теоремы для стохастических игр (см. [88]), но использует структуру стратегии (2.4.12). Заметим, что мы не определяем стратегии игроков, когда в истории наблюдаются отклонения более чем одного игрока, потому что мы не будем доказывать условия абсолютного равновесия, а ограничимся только условием, гарантирующим существование равновесия по Нэшу. Если более одного игрока отклоняется от кооперативного решения, игрок может выбрать любую стратегию из множества своих стратегий. Сейчас мы докажем, что $\widehat{\varphi} = (\widehat{\varphi}_1, \dots, \widehat{\varphi}_n)$, определенная в (2.4.12), является равновесием по Нэшу в стохастической игре G_σ . Посчитаем выигрыш игрока при условии, что все игроки используют стратегии (2.4.12) и реализуют кооперативный режим игры, то есть, реализуют ситуацию η^* . В этом случае дисконтированный выигрыш игрока i в подыгре G_σ^ω , $\omega \in \Omega$ равен

$$E_i^\omega(\widehat{\varphi}) = E_i^\omega(\eta^*).$$

¹¹Заметим, что стратегии игроков из коалиции $N \setminus z$ коррелированные.

Пусть $E_i(\widehat{\varphi})$ есть вектор $(E_i^{\omega_1}(\widehat{\varphi}), \dots, E_i^{\omega_n}(\widehat{\varphi}))^T$. Тогда для игрока $i \in N$ имеет место равенство:

$$E_i(\widehat{\varphi}) = (\mathbb{I} - \delta\Pi(\eta^*))^{-1}\beta_i. \quad (2.4.13)$$

Рассмотрим ситуацию $(\varphi_z, \widehat{\varphi}_{N \setminus z})$, когда некоторый игрок z отклоняется от стратегии $\widehat{\varphi}_z$. Пусть шаг k такой, что для любого шага $l \in [1, k-1]$ история $h(l) \subset h^*$, но $(\omega(k), a(k)) \notin h^*$ и $(\omega(k), (a_z^*(k), a_{N \setminus z}(k))) \in h^*$. Не умаляя общности, мы положим $\omega(k) = \omega$, тогда первое индивидуальное отклонение игрока z наступает на шаге k . Вычислим дисконтированный выигрыш игрока z в игре G_σ в ситуации $(\varphi_z, \widehat{\varphi}_{N \setminus z})$:

$$\overline{E}_z^\sigma(\varphi_z, \widehat{\varphi}_{N \setminus z}) = \pi_0 E_z^\sigma(\varphi_z, \widehat{\varphi}_{N \setminus z}),$$

где

$$E_z^\sigma(\varphi_z, \widehat{\varphi}_{N \setminus z}) = E_z^{\sigma, [1, k-1]}(\varphi_z, \widehat{\varphi}_{N \setminus z}) + \delta^{k-1} \Pi^{k-1}(\varphi_z, \widehat{\varphi}_{N \setminus z}) E_z^{\sigma, [k, \infty)}(\varphi_z, \widehat{\varphi}_{N \setminus z}). \quad (2.4.14)$$

Первое слагаемое в правой части (2.4.14) — ожидаемый выигрыш игрока z на первых $k-1$ шагах игры G_σ , второе слагаемое — ожидаемый выигрыш игрока z подыгре игры G_σ , начинающейся с шага k , где $E_z^{\sigma, [k, \infty)}(\varphi_z, \widehat{\varphi}_{N \setminus z})$ — вектор $(E_z^{\sigma, 1}(\varphi_z, \widehat{\varphi}_{N \setminus z}), \dots, E_z^{\sigma, \bar{\omega}}(\varphi_z, \widehat{\varphi}_{N \setminus z}))^T$, и $E_z^{\sigma, \omega}(\varphi_z, \widehat{\varphi}_{N \setminus z})$ — ожидаемый выигрыш игрока z в регуляризованной подыгре G_σ^ω , начинающейся из состояния ω . Так как до шага $k-1$ не наблюдается отклонений от кооперативной ситуации η^* , то имеют место равенства:

$$E_z^{\sigma, [1, k-1]}(\varphi_z, \widehat{\varphi}_{N \setminus z}) = E_z^{\sigma, [1, k-1]}(\eta^*),$$

$$\Pi^{k-1}(\varphi_z, \widehat{\varphi}_{N \setminus z}) = \Pi^{k-1}(\eta^*).$$

Теперь найдем дисконтированный выигрыш игрока z в подыгре G_σ^ω , начинающейся с шага k , когда состояние $\omega(k)$ есть ω . Справедлива формула:

$$E_z^{\sigma, \omega}(\varphi_z, \widehat{\varphi}_{N \setminus z}) = K_z^\omega(\widehat{a}_z^\omega, a_{N \setminus z}^{\omega^*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, (\widehat{a}_z^\omega, a_{N \setminus z}^{\omega^*})) V^{\omega'}(\{z\}), \quad (2.4.15)$$

где $\widehat{a}_z^\omega \in \Delta(A_z^\omega)$ и $a_z^\omega \neq a_z^{\omega^*}$. Игроки из коалиции $N \setminus z$ наказывают отклонившегося игрока z , предоставляя возможность игроку z получить свой минимальный выигрыш по определению стратегии $\widehat{\varphi}_i$. В равенстве (2.4.15) значение

характеристической функции $V^{\omega'}(\{z\})$ определяется по формуле (2.3.4). Так как ожидаемые выигрыши игрока z в ситуациях $\widehat{\varphi}$ и $(\varphi_z, \widehat{\varphi}_{N \setminus z})$ совпадают до шага $k-1$, то отклонение может увеличить выигрыш игрока z только в подыгре G_σ^ω , $\omega \in \Omega$. В частности, в ситуации $(\varphi_z, \widehat{\varphi}_{N \setminus z})$ игрок z может гарантировать себе следующий выигрыш, начиная с шага k :

$$F(\{z\}) = \max_{\widehat{a}_z^\omega \in \Delta(A_z^\omega)} \left\{ K_z^\omega(\widehat{a}_z^\omega, a_{N \setminus z}^{\omega*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, (\widehat{a}_z^\omega, a_{N \setminus z}^{\omega*})) V^{\omega'}(\{z\}) \right\}. \quad (2.4.16)$$

По определению ПРД дисконтированный выигрыш игрока z в регуляризованной подыгре G_σ^ω в ситуации $\widehat{\varphi}(\cdot)$ может быть найден по формуле

$$E_z^\sigma(\widehat{\varphi}) = (\mathbb{I} - \delta \Pi(\eta^*))^{-1} \beta_z = \sigma_z, \quad (2.4.17)$$

где $E_z^\sigma(\widehat{\varphi}) = (E_z^{\sigma, \omega_1}(\widehat{\varphi}), \dots, E_z^{\sigma, \bar{\omega}}(\widehat{\varphi}))^T$. Учитывая неравенство (2.4.11), из (2.4.16), (2.4.17) и рассуждений, приведенных выше, получаем

$$E_z^\sigma(\widehat{\varphi}) \geq E_z^\sigma(\varphi_z, \widehat{\varphi}_{N \setminus z}),$$

что выполнено, если неравенство

$$\sigma_z = (\mathbb{I} - \delta \Pi(\eta^*))^{-1} \beta_z \geq F(\{z\}), \quad (2.4.18)$$

справедливо.

Если неравенство (2.4.18) выполнено для любого $z \in N$, то ни один игрок не имеет выгоды от отклонения от кооперативного режима ни в какой подыгре σ -регуляризации игры G . Это доказывает, что ситуация $\widehat{\varphi}(\cdot) = (\widehat{\varphi}_1(\cdot), \dots, \widehat{\varphi}_n(\cdot))$ со стратегиями, определенными равенством равенством (2.4.12), является равновесием по Нэшу. При этом дисконтированный выигрыш игрока i в игре G_σ в ситуации $\widehat{\varphi}$ равен $\bar{\sigma}_i$, где $\bar{\sigma}_i = \pi_0 \sigma_i$, и $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\bar{\omega}})$ состоит из i -ых компонент дележей $\sigma^{\omega_1}, \dots, \sigma^{\bar{\omega}}$, посчитанных для кооперативных подыгр $G^{\omega_1}, \dots, G^{\bar{\omega}}$ соответственно. \square

Отметим, что стратегии игроков, используемые в режиме наказания, не являются индивидуально рациональными, а именно, игроку i , производящему наказание отклонившегося игрока z , приходится использовать стратегии, минимизирующие выигрыш игрока z в подыгре, что может быть невыгодно игроку i и

может служить поводом для отклонения от ситуации (2.4.12). Поэтому ситуация в стратегиях поведения (2.4.12) не является абсолютным равновесием по Нэшу. Для того, чтобы построить абсолютное равновесие по Нэшу в σ -регуляризации игры G с выигрышами игроков $(\bar{\sigma}_1, \dots, \bar{\sigma}_n)$, определим новую ситуацию в стратегиях поведения следующим образом. Если история шага является кооперативной и содержится в h^* , то все игроки используют кооперативные стратегии, в противном случае все игроки реализуют стратегии, являющиеся частью равновесия по Нэшу в текущей подыгре¹². Обозначим равновесие по Нэшу в игре G через $\eta^{ne} = (\eta_1^{ne}, \dots, \eta_n^{ne})$, где $\eta_i^{ne}(\omega) = \hat{a}_i^{\omega, ne} \in \Delta(A_i^\omega)$. Для удобства введем обозначение:

$$Q(\{i\}) \equiv (Q^{\omega_1}(\{i\}), \dots, Q^{\bar{\omega}}(\{i\}))^T,$$

$$Q^\omega(\{i\}) = \max_{\hat{a}_i^\omega \in \Delta(A_i^\omega)} \left\{ K_i^\omega(\hat{a}_i^\omega, a_{N \setminus i}^{\omega*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, (\hat{a}_i^\omega, a_{N \setminus i}^{\omega*})) E_i^{\omega'}(\eta^{ne}) \right\},$$

и

$$\sigma_i = (\mathbb{I} - \delta \Pi(\eta^*))^{-1} \beta_i \geq Q(\{i\}). \quad (2.4.19)$$

Следующая теорема гарантирует существование абсолютного равновесия по Нэшу. Выполнение неравенства (2.4.19) гарантирует выполнение условия стратегической устойчивости кооперативного решения, при этом равновесие по Нэшу является абсолютным.

Теорема 2.5. *Если в σ -регуляризованной игре G_σ , такой что $\bar{\sigma} = \pi_0 \sigma$, неравенство (2.4.19) выполнено для любого игрока $i \in N$, то существует ситуация абсолютного равновесия по Нэшу в стратегиях поведения $\tilde{\varphi}$ с выигрышами игроков, равными $(\bar{\sigma}_1, \dots, \bar{\sigma}_n)$.*

Доказательство. Доказательство теоремы 2.5 проводится подобно доказательству теоремы 2.4 с использованием структуры «новой» ситуации в стратегиях поведения, которую обозначим через $\tilde{\varphi} = (\tilde{\varphi}_1, \dots, \tilde{\varphi}_n)$, где стратегия $\tilde{\varphi}_i$ игрока $i \in N$ такова, что

$$\tilde{\varphi}_i(h(k)) = \begin{cases} a_i^{\omega*}, & \text{если } \omega(k) = \omega, h(k) \subset h^*; \\ \hat{a}_i^{\omega, ne}, & \text{если } h(k) \not\subset h^*, \end{cases} \quad (2.4.20)$$

¹²В случае, если существует несколько равновесий по Нэшу, будем считать, что игроками выбрано одно равновесие по Нэшу для реализации режима наказания. Заметим, что в этом случае игроки используют коррелированные стратегии [67].

где $\hat{a}_i^{\omega, ne} \in \Delta(A_i^\omega)$ — стратегия игрока i , используемая при наказании, которая может быть как чистой, так и смешанной. Заметим, что в случае, когда в истории шага наблюдается одновременное отклонение более одного игрока, все игроки переключаются на реализацию равновесия η^{ne} . \square

§ 2.4.3 Защита от иррационального поведения кооперативного решения в стохастических играх бесконечной продолжительности

Принципы позиционной состоятельности и стратегической поддержки предполагают, что игроки являются рациональными. Но в реальных ситуациях кооперация может быть разрушена по причинам, которые нельзя считать рациональными. Например, игрок может совершать иррациональные поступки для получения дополнительной прибыли, нарушая кооперативные договоренности. Такие действия игроков могут привести к нарушению условий кооперации и переходу игроков к игре, где каждый игрок играет индивидуально. В этом случае, будем считать, что кооперация прервалась при «иррациональном поведении» игроков¹³.

Д. Янг предложил условие¹⁴, при котором даже если иррациональное поведение наблюдается в игре, игрок уверен, что получит не меньший выигрыш, чем когда бы он играл всю игру индивидуально [223]. Поясним это высказывание. Рассмотрим два сценария: 1) игрок играет кооперативно, как и все другие игроки, до некоторого периода k , после этого кооперация прерывается; 2) игрок играет индивидуально на протяжении всей игры. Если выигрыш игрока в первом сценарии не меньше, чем выигрыш при втором сценарии, то принцип защиты от иррационального поведения удовлетворен. Сформулируем условие защиты от иррационального поведения.

Определение 2.9. Дележ $\bar{\sigma} = \pi_0 \sigma$ и соответствующая ему ПРД удовлетворяют принципу защиты от иррационального поведения, если

$$E_i^{\sigma, [1, k]} + \delta^k \Pi^k(\eta^*) V(\{i\}) \geq V(\{i\}), \text{ для всех } i \in N \text{ и } k = 1, 2, \dots, \quad (2.4.21)$$

¹³ Аналогичные рассуждения можно привести в случае повторяющейся игры.

¹⁴ Так называемое условие Янга или принцип защиты от иррационального поведения был сформулирован для линейно-квадратичных игр в работах [53, 133].

где $E_i^{\sigma[1,k]}$ — ожидаемый выигрыш игрока i на первых k шагах σ -регуляризации G_σ .

Здесь требуется сделать предположение о том, что игроки перед началом каждого шага знают о том, разрушилась кооперация или нет, то есть, задержки информации не предполагается. В левой части неравенства (2.4.21) первое слагаемое равно ожидаемому выигрышу игрока i на первых k шагах игры, если игроки реализуют вектор кооперативных стратегий η^* , и произведена σ -регуляризация игры G . Второе слагаемое — ожидаемый выигрыш игрока i в игре, начинающейся с шага $k + 1$, когда кооперация распалась. В правой части неравенства (2.4.21) — выигрыш игрока i в случае, когда он играет индивидуально всю игру.

Теорема 2.6. *Если неравенство*

$$(\mathbb{I} - \delta\Pi(\eta^*))(\sigma_i - V(\{i\})) \geq 0 \quad (2.4.22)$$

выполнено для любого игрока $i \in N$, тогда дележ $\bar{\sigma} = \pi_0\sigma$ и соответствующая ему ПРД $\{\beta_i\}_{i \in N}$ удовлетворяют принципу защиты от иррационального поведения.

Доказательство. Покажем, что условие (2.4.22) является достаточным для (2.4.21). Доказательство проведем методом математической индукции. Сначала запишем неравенство (2.4.21) для $k = 1$. Затем перепишем условие (2.4.22), используя определение σ_i и ПРД (2.4.5). Получаем:

$$V(\{i\}) \leq \beta_i + \delta\Pi(\eta^*)V(\{i\}). \quad (2.4.23)$$

Эти неравенства совпадают, что доказывает Теорему при $k = 1$.

Предположим, что из (2.4.22) следует (2.4.21) для $k = l$. Переписав (2.4.21) для $k = l$, получаем:

$$V(\{i\}) \leq \beta_i + \dots + \delta^{l-1}\Pi^{l-1}(\eta^*)\beta_i + \delta^l\Pi^l(\eta^*)V(\{i\}). \quad (2.4.24)$$

Прделаем те же преобразования для $k = l + 1$. Неравенство (2.4.21) для $k = l + 1$:

$$V(\{i\}) \leq \beta_i + \dots + \delta^l\Pi^l(\eta^*)\beta_i + \delta^{l+1}\Pi^{l+1}(\eta^*)V(\{i\}). \quad (2.4.25)$$

Далее для того, чтобы доказать, что если неравенство (2.4.22) справедливо, то и неравенство (2.4.21) имеет место для $k = l + 1$, рассмотрим преобразованную правую часть (2.4.25):

$$\beta_i + \delta \Pi(\eta^*) \{ \beta_i + \delta \Pi(\eta^*) \beta_i + \dots + \delta^{l-1} \Pi^{l-1}(\eta^*) \beta_i + \delta^l \Pi^l(\eta^*) V(\{i\}) \}.$$

Учитывая (2.4.24), выражение в фигурных скобках не меньше, чем $V(\{i\})$. Следовательно, правая часть неравенства (2.4.25) не меньше, чем $\beta_i + \delta \Pi(\eta^*) V(\{i\})$. Из (2.4.2) и (2.4.22) получаем неравенство (2.4.21) для $k = l + 1$, которое доказывает теорему. \square

Следствие 2.1. *Для выполнения принципа защиты от иррационального поведения достаточно, чтобы для любого $i \in N$ имело место неравенство*

$$K_i(\tilde{a}) - \beta_i \leq \delta (\sigma_i^{\min} - V^{\max}(\{i\})), \quad (2.4.26)$$

где $K_i(\tilde{a}) = \left(\max_{a_i^{\omega_1} \in A_i^{\omega_1}} K_i^{\omega_1}(a_i^{\omega_1}, a_{N \setminus i}^{\omega_1*}), \dots, \max_{a_i^{\bar{\omega}} \in A_i^{\bar{\omega}}} K_i^{\bar{\omega}}(a_i^{\bar{\omega}}, a_{N \setminus i}^{\bar{\omega}*}) \right)^T$, и $\max_{a_i^{\omega} \in A_i^{\omega}} K_i^{\omega}(a_i^{\omega}, a_{N \setminus i}^{\omega*})$ — это максимальный выигрыш игрока i , получаемый им при отклонении от ситуации $a^{\omega*}$, которая является частью ситуации η^* , удовлетворяющей условию (2.3.1), $\tilde{a}_i = \arg \max_{a_i^{\omega} \in A_i^{\omega}} K_i^{\omega}(a_i^{\omega}, a_{N \setminus i}^{\omega*})$ для каждого состояния $\omega \in \Omega$ и игрока $i \in N$, $\sigma_i^{\min} = \left(\min_{\omega \in \Omega} \sigma_i^{\omega}, \dots, \min_{\omega \in \Omega} \sigma_i^{\omega} \right)^T$, $V^{\max}(\{i\}) = \left(\max_{\omega \in \Omega} V^{\omega}(\{i\}), \dots, \max_{\omega \in \Omega} V^{\omega}(\{i\}) \right)^T$.

Доказательство. Пусть выполнено достаточное условие (2.4.26), которое можно представить следующим образом:

$$\beta_i + \delta \sigma_i^{\min} \geq K_i(\tilde{a}) + \delta V^{\max}(\{i\}). \quad (2.4.27)$$

Произведем оценки левой и правой частей неравенства (2.4.27). Так как матрица переходных вероятностей $\Pi(\eta^*)$ является стохастической матрицей, то получаем:

$$\beta_i + \delta \sigma_i^{\min} = \beta_i + \delta \Pi(\eta^*) \sigma_i^{\min} \leq \beta_i + \delta \Pi(\eta^*) \sigma_i. \quad (2.4.28)$$

Для правой части неравенства (2.4.27) верно следующее равенство:

$$K_i(\tilde{a}) + \delta V^{\max}(\{i\}) = K_i(\tilde{a}) + \delta \Pi(\tilde{\eta}) V^{\max}(\{i\}), \quad (2.4.29)$$

где $\Pi(\tilde{a})$ — стохастическая матрица, а $\tilde{\eta} = (\tilde{\eta}_i : i \in N)$ — ситуация в стационарных стратегиях такая, что

$$\tilde{\eta}_j = \begin{cases} \arg \max_{\eta_i \in H_i} \Pi(\eta_i, \eta_{N \setminus i}^*) V(\{i\}), & \text{если } j = i \\ \eta_j^*, & \text{если } j \neq i \end{cases}$$

Получаем неравенство:

$$K_i(\tilde{a}) + \delta \Pi(\tilde{\eta}) V^{\max}(\{i\}) = \max_{a_i \in A_i} K_i(a_i, a_{N \setminus i}^*) + \delta \max_{\eta_i \in \Xi_i} \left\{ \Pi(\eta_i, \eta_{N \setminus i}^*) V(\{i\}) \right\} \geq \max_{\eta_i \in H_i} \left\{ K_i(a_i, a_{N \setminus i}^*) + \delta \Pi(\eta_i, \eta_{N \setminus i}^*) V(\{i\}) \right\}. \quad (2.4.30)$$

Из неравенств (2.4.27)–(2.4.30) следует неравенство (2.4.22). Следовательно, по Теореме 2.6 принцип защиты от иррационального поведения выполнен. \square

§ 2.5 Существование кооперативного решения в стохастических играх бесконечной продолжительности, удовлетворяющего принципам устойчивой кооперации

В этом разделе опишем проблему существования устойчивого кооперативного решения. Следует отметить, что принцип оптимальности, который игроки решили реализовать в кооперативной игре, должен содержать непустое подмножество дележей. Принципы оптимальности такие как, вектор Шепли, ES-значение, n -ядро всегда существуют, то есть, их можно определить для любой подыгры стохастической игры, используя значения характеристической функции, удовлетворяющей формулам (2.3.3), (2.3.4) и (2.3.8).

Существование позиционно состоятельного кооперативного решения следует из Теоремы 2.3 и метода построения ПРД для кооперативного решения $\bar{\sigma}$. При заданном кооперативном решении $\bar{\sigma}$ регуляризация стохастической игры определяет новые функции выигрыша для того, чтобы кооперативное решение было бы позиционно состоятельным. Таким образом, дисконтированные выигрыши игроков в σ -регуляризации исходной игры равны соответствующим компонентам кооперативного решения $\bar{\sigma}$ и удовлетворяют условию позиционной состоятельности (динамической устойчивости).

Следовательно, если игроки провели модификацию изначально заданной стохастической игры, построив ее σ -регуляризацию, то нетрудно доказать, что позиционно состоятельное решение $\bar{\sigma}$ существует.

Для того, чтобы проверить, удовлетворяет ли кооперативное решение $\bar{\sigma}$ принципам стратегической поддержки и защиты от иррационального поведения, нам необходимо решить систему неравенств:

$$\begin{cases} \sigma_i = (\mathbb{I} - \delta\Pi(\eta^*))^{-1}\beta_i \geq F(\{i\}), & i \in N, \\ (\mathbb{I} - \delta\Pi(\eta^*)) (\sigma_i - V(\{i\})) \geq 0, & i \in N. \end{cases}$$

Эти неравенства можно считать неравенствами относительно дисконтирующего фактора δ . Они схожи с ограничениями, которые накладываются на δ , чтобы доказать, что набор стратегий является абсолютным равновесием по Нэшу в повторяющейся игре. Эта система нелинейная относительно параметра δ . В общем случае невозможно выписать ограничения на δ в явном виде.

Тем не менее, можно утверждать, что устойчивое кооперативное решение существует, например, когда кооперативная ситуация совпадает с равновесием по Нэшу в стохастической игре, и игроки симметричны. В этом случае, вектор Шепли удовлетворяет принципам устойчивой кооперации. В частности, далее мы найдем решение этой системы неравенств для одного класса стохастических игр двух лиц с двумя состояниями.

Пример 3. *Стохастическая игра конкуренции между асимметричными фирмами.* Рассмотрим дуополию Курно в динамике с асимметричными фирмами. Опишем ее с помощью стохастической игры типа «Дилемма заключенного». Пусть множество состояний будет $\Omega = \{\omega_1, \omega_2\}$, где в ω_j определена игра $\langle N, A_1^{\omega_j}, A_2^{\omega_j}, K_1^{\omega_j}, K_2^{\omega_j} \rangle$, $j = 1, 2$, и $A_i^{\omega_j} = \{C_j, D_j\}$ — множество стратегий игрока $i = 1, 2$. Стратегии C_j и D_j соответствуют действиям «кооперироваться» и «отклоняться», соответственно. Для состояния ω_1 выигрыши игроков определены матрицей:

$$\begin{array}{cc} & C_1 & D_1 \\ C_1 & (7, 7) & (1, 8) \\ D_1 & (8, 1) & (4, 5) \end{array}$$

и для состояния ω_2 :

$$\begin{array}{cc} & C_2 & D_2 \\ C_2 & (9, 9) & (1, 10) \\ D_2 & (16.5, 1) & (6, 5) \end{array}$$

Состояние ω_1 соответствует состоянию рынка с низким спросом, а состояние ω_2 — состоянию рынка с высоким спросом. Обе одношаговые игры имеют единственное равновесие по Нэшу, когда обе фирмы «отклоняются» с выигрышами (4, 5) и (6, 5) в состояниях ω_1 и ω_2 соответственно. Ситуация, максимизирующая сумму дисконтированных выигрышей игроков, — это набор стратегий игроков такой, что в любом состоянии оба игрока «кооперируются» и получают выигрыши (7, 7) и (9, 9) в состояниях ω_1 и ω_2 соответственно. При кооперативном поведении выигрыши игроков равны, но в равновесиях по Нэшу игроки получают неодинаковые выигрыши. В частности, в состоянии с низким спросом фирма 1 имеет меньший выигрыш, чем фирма 2, тогда как в состоянии с высоким спросом фирма 2 имеет меньший выигрыш, чем фирма 1. Это может быть обусловлено техническими характеристиками производства. Например, фирма 2 может обладать технологией производства, которая является более эффективной в производстве небольших объемов выпуска.

В состоянии ω_2 фирмы отличаются в выигрышах, когда одна фирма «кооперируется», а вторая «отклоняется». В частности, выигрыш фирмы 1, когда она отклоняется, больше, чем выигрыш фирмы 1 при отклонении. Таким образом, асимметричность игроков очевидным образом повлияет на дележ суммарного кооперативного дисконтированного выигрыша. Другой особенностью состояния ω_2 является то, что когда обе фирмы «кооперируются», их суммарный выигрыш ненамного больше, чем выигрыш в ситуации (D_2, C_2) (18 против 17.5). Поэтому, если вероятность перехода из ситуации (D_2, C_2) в состояние ω_1 больше, чем из ситуации (C_2, C_2) , то фирмы могут согласиться на разыгрывание ситуации (D_2, C_2) , чтобы избежать перехода из состояния с высоким спросом в состояние с низким спросом.

Пусть вероятности перехода из состояний ω_1 и ω_2 следующие:

$$\begin{pmatrix} (0.3, 0.7) & (0.9, 0.1) \\ (0.4, 0.6) & (0.3, 0.7) \end{pmatrix}, \quad \begin{pmatrix} (0.9, 0.1) & (0.4, 0.6) \\ (0.1, 0.9) & (0.3, 0.7) \end{pmatrix}$$

где элемент (k, l) матрицы состоит из вероятностей перехода из состояния ω_j в состояния ω_1, ω_2 при условии, что фирма 1 выбирает стратегию k , а фирма 2 — стратегию l . Можно заметить, что вероятность перехода в состояние ω_1 в ситуации (C_2, C_2) намного выше, чем вероятность перехода в это состояние в ситуации (D_2, C_2) (0.9 против 0.1). Пусть дисконт-фактор равен $\delta = 0.99$ и вектор начального распределения по состояниям равен $\pi_0 = (0.5, 0.5)$.

Кооперативная игра. Сначала построим кооперативную игру G_c на основе стохастической игры G . Для этого нам необходимо найти кооперативное решение и посчитать значения характеристической функции для игры и любой ее подыгры (начинающейся из состояний ω_1 и ω_2). Сначала определим кооперативное решение $\eta^* = (\eta_1^*, \eta_2^*)$, следуя (2.2.4) и (2.3.1). Получим единственную ситуацию, в которой $\eta_1^* = (C_1, D_2)$ — стратегия фирмы 1, $\eta_2^* = (C_1, C_2)$ — стратегия фирмы 2, дают максимальный суммарный дисконтированный выигрыш игроков $\bar{V}(\{1, 2\}) = \pi_0 V(\{1, 2\}) = 1704.61$. Согласно этой ситуации, в состоянии ω_1 набор кооперативных стратегий (оба игрока кооперируются) дает выигрыш 7 каждой фирме. В состоянии ω_2 , при кооперативном поведении игроков фирма 1 отклоняется, а фирма 2 кооперируется, при этом, фирма 2 получает выигрыш, меньший своего выигрыша при равновесии по Нэшу. Но это будет компенсировано фирмой 1 при дележе суммарного выигрыша. Таким образом, значения характеристической функции для максимальной коалиции есть

$$V(\{1, 2\}) = \begin{pmatrix} E_1^{\omega_1}(\eta^*) + E_2^{\omega_1}(\eta^*) \\ E_1^{\omega_2}(\eta^*) + E_2^{\omega_2}(\eta^*) \end{pmatrix} = \begin{pmatrix} 1702.43 \\ 1706.80 \end{pmatrix}.$$

По определению (2.3.8) значения характеристической функции для пустой коалиции равны нулю:

$$V(\emptyset) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Посчитаем значения характеристической функции $V(S) = (V^{\omega_1}(S), V^{\omega_2}(S))^T$

для коалиций $S = \{1\}$ и $S = \{2\}$, используя (2.3.3):

$$V(\{1\}) = \begin{pmatrix} 538.60 \\ 540.60 \end{pmatrix}, \quad V(\{2\}) = \begin{pmatrix} 500.00 \\ 500.00 \end{pmatrix}.$$

Это выигрыши фирм в равновесии по Нэшу, когда обе фирмы отклоняются в каждом состоянии, то есть, при реализации ситуаций (D_1, D_1) и (D_2, D_2) .

Используя (2.3.7), можно рассчитать значения $\bar{V}(S)$ для игры и всех коалиций:

$$\bar{V}(\emptyset) = 0.00, \quad \bar{V}(\{1\}) = 539.60, \quad \bar{V}(\{2\}) = 500.00, \quad \bar{V}(\{1, 2\}) = 1704.61.$$

Таким образом, мы определили кооперативную стохастическую подыгру $G_c^{\omega_j}$ как пару $\langle N, V^{\omega_j}(\cdot) \rangle$, $j = 1, 2$, и кооперативную стохастическую игру G_c как $\langle N, \bar{V}(\cdot) \rangle$.

Вектор Шепли. Предположим, что игроки выбрали в качестве решения кооперативной стохастической игры G_c вектор Шепли, т.е. они вычисляют вектор Шепли и в качестве кооперативного решения в любой подыгре $G_c^{\omega_j}$, $j = 1, 2$. Для игры двух лиц вектор Шепли вычисляется по формуле:

$$\sigma_i^{\omega_j} = V^{\omega_j}(\{i\}) + \frac{V^{\omega_j}(\{1, 2\}) - V^{\omega_j}(\{1\}) - V^{\omega_j}(\{2\})}{2},$$

где $i = 1, 2$ и $j \in \{1, 2\}$. Вектор Шепли в подыграх равен:

$$\sigma_1 = \begin{pmatrix} 870.516 \\ 873.698 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 831.916 \\ 833.098 \end{pmatrix}.$$

Учитывая начальное распределение π_0 на множестве состояний, вычислим вектор Шепли $\bar{\sigma}$ во всей игре G_c по определению 2.4:

$$\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2) = (872.107, 832.507).$$

Позиционная состоятельность. Теперь будем проверять, удовлетворяет ли вектор Шепли принципам устойчивой кооперации. Начнем с принципа позиционной состоятельности. Если фирмы получают пошаговые выплаты в соответствии со своими изначально заданными функциями выигрышей, то их дисконтированные выигрыши во всей игре будут равны $\bar{E}_1(\eta^*) = 1526.809$ и

$\bar{E}_2(\eta^*) = 177.805$, что отличается от компонент вектора Шепли, которые равны $\bar{\sigma}_1 = 872.107$ и $\bar{\sigma}_2 = 832.507$. Определим процедуру распределения дележа или трансфертные выплаты игрокам, чтобы они в итоге получили компоненты вектора Шепли, а также процедура распределения дележа удовлетворяла принципу позиционной состоятельности. Используя то, что $\bar{\sigma} = \pi_0\sigma$ и уравнение (2.4.2), получим ПРД:

$$\beta_1 = (\mathbb{I} - \delta\Pi(\eta^*))\sigma_1 = \begin{pmatrix} 6.500 \\ 9.052 \end{pmatrix}, \quad \beta_2 = (\mathbb{I} - \delta\Pi(\eta^*))\sigma_2 = \begin{pmatrix} 7.500 \\ 8.448 \end{pmatrix}.$$

Определим σ -регуляризацию стохастической игры G , используя ПРД и определение 2.7. Для этого переопределим функции выигрышей игроков начальной игры во всех состояниях в случае, когда игроки реализуют η^* , заменив выигрыши соответствующими компонентами ПРД. Следовательно, выигрыши игроков в состояниях ω_1 и ω_2 равны соответственно:

$$\begin{pmatrix} (6.500, 7.500) & (1, 8) \\ (8, 1) & (4, 5) \end{pmatrix}, \\ \begin{pmatrix} (9, 9) & (1, 10) \\ (9.052, 8.448) & (6, 5) \end{pmatrix}.$$

В регуляризованной игре в состоянии с низким спросом (состояние ω_1) обе фирмы в кооперации реализуют стратегии «кооперироваться» и получают выигрыши (6.5, 7.5). Напомним, что их выигрыши в исходной игре были равны (7, 7): фирма 1 передает 0.5 фирме 2 на каждом шаге в этом состоянии. В состоянии с высоким спросом (состояние ω_2) фирма 1 играет стратегию «отклоняться», тогда как фирма 2 «кооперируется», как предписывает η^* . В состоянии ω_2 выигрыши игроков теперь равны (9.052, 8.448). В начальной стохастической игре они были следующими: (16.5, 1). Таким образом, фирма 2 получает $16.5 - 9.052 = 7.448$ от фирмы 1 на каждом шаге в этом состоянии. Если регуляризация начальной игры проведена методом, описанным выше, то вектор Шепли и соответствующая ему ПРД удовлетворяют принципу позиционной состоятельности.

Стратегическая устойчивость. Теперь проверим условие стратегической устойчивости вектора Шепли, а именно, проверим, имеют ли фирмы выгоду от

отклонения от кооперативного решения в индивидуальном порядке. Рассмотрим сначала состояние ω_1 . Так как в этом состоянии ситуация, реализуемая при кооперации, не является равновесием по Нэшу, то игроки могут получить прибыль от отклонения. Проверим справедливость неравенства:

$$\sigma_i^{\omega_1} \geq F^{\omega_1}(\{i\}),$$

для $i = 1, 2$, где

$$F^{\omega_1}(\{i\}) = \max_{\substack{a_i^{\omega_1} \in A_i^{\omega_1} \\ a_i^{\omega_1} \neq a_i^{\omega_1^*}}} \left\{ K_i^{\omega_1}(a_i^{\omega_1}, a_{N \setminus i}^{\omega_1^*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega_1, (a_i^{\omega_1}, a_{N \setminus i}^{\omega_1^*})) V^{\omega'}(\{i\}) \right\}.$$

Неравенство (2.4.11) для фирмы 1 записывается таким образом:

$$870.516 \geq 8 + 0.99 \begin{pmatrix} 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 538.60 \\ 540.60 \end{pmatrix} = 542.402,$$

и для фирмы 2:

$$831.916 \geq 8 + 0.99 \begin{pmatrix} 0.9 & 0.1 \end{pmatrix} \begin{pmatrix} 500.00 \\ 500.00 \end{pmatrix} = 503.$$

В состоянии ω_2 ситуация (D_2, C_2) является равновесием по Нэшу, поэтому игроки не могут увеличить свой выигрыш, отклонившись. Следовательно, можно сделать вывод о том, что неравенство (2.4.11) справедливо и для состояния ω_2 . Условие стратегической устойчивости выполнено.

Условие защиты от иррационального поведения. Для того, чтобы проверить это условие, необходимо сравнить выигрыши игроков в двух случаях:

- 1) фирма играет индивидуально в течение всей игры,
- 2) фирма кооперируется с другой фирмой до некоторого шага, после этого играет индивидуально.

Заметим, что во втором случае, когда фирмы кооперируются, то они получают выигрыши в соответствии с ПРД, построенной на основе заранее выбранного дележа.

Если выигрыш игрока в случае 1) не больше, чем выигрыш в случае 2), тогда принцип защиты от иррационального поведения удовлетворен. Это доказано в теореме 2.6, так как

$$(\mathbb{I} - \delta\Pi(\eta^*))(\sigma_1 - V(\{1\})) = (\mathbb{I} - \delta\Pi(\eta^*))(\sigma_2 - V(\{2\})) = \begin{pmatrix} 2.500 \\ 3.448 \end{pmatrix} \geq 0.$$

Таким образом, проведена регуляризация исходной игры, чтобы гарантировать позиционную состоятельность вектора Шепли, показано, что выполняются принципы стратегической устойчивости и защиты от иррационального поведения в данной игре.

§ 2.6 Сильное трансферабельное равновесие в стохастических играх бесконечной продолжительности

Теорему 2.4 можно обобщить на случай, когда отклоняется несколько игроков, то есть доказать, что при выполнении условия, аналогичного неравенству (2.4.11), в σ -регуляризации G_σ стохастической игры G существует сильное трансферабельное равновесие с выигрышами $(\bar{\sigma}_1, \dots, \bar{\sigma}_n)$. В этом случае игроки могут реализовать специальным образом построенную ситуацию в стратегиях наказания, где в качестве наказания отклонившейся коалиции не отклонившиеся игроки будут реализовывать стратегии наказания, позволяющие отклонившейся коалиции получить минимаксный выигрыш в любой подыгре.

Дадим определение сильному трансферабельному равновесию¹⁵ и докажем теорему, аналогичную теореме 2.4.

Определение 2.10. [46] Назовем ситуацию $\tilde{\varphi} = (\tilde{\varphi}_1, \dots, \tilde{\varphi}_n)$ сильным трансферабельным равновесием в регуляризованной игре G_σ , если для любой коалиции $S \subseteq N$, $S \neq \emptyset$, неравенство

$$\sum_{i \in S} \bar{E}_i^\sigma(\tilde{\varphi}) \geq \sum_{i \in S} \bar{E}_i^\sigma(\varphi_S, \tilde{\varphi}_{N \setminus S}) \quad (2.6.1)$$

выполняется для любой стратегии поведения коалиции S : $\varphi_S = (\varphi_i : i \in S) \in \prod_{i \in S} \Phi_i$. Здесь $\bar{E}_i^\sigma(\cdot)$ — дисконтированный выигрыш игрока i в σ -регуляризации игры G .

¹⁵Определение сильного равновесия по Нэшу можно также найти в [66].

Докажем теорему, позволяющую получить условие на параметры игры, при котором в регуляризованной игре G_σ существует трансферабельной равновесие с выигрышами игроков, равными соответствующим компонентам дележа, для которого произведена регуляризации начальной стохастической игры.

Теорема 2.7. *Если в регуляризованной игре G_σ такой, что кооперативное решение удовлетворяет условию $\bar{\sigma} = \pi_0\sigma$, выполнено неравенство:*

$$\sum_{i \in S} \beta_i \geq (\mathbb{I} - \delta\Pi(\eta^*))\tilde{F}(S) \quad (2.6.2)$$

для любой коалиции $S \subseteq N$, $S \neq \emptyset$, где $\tilde{F}(S) = (\tilde{F}^{\omega_1}(S), \dots, \tilde{F}^{\omega_n}(S))^T$,

$$\tilde{F}^\omega(S) = \max_{\substack{a_S^\omega \in \prod_{i \in S} \Delta(A_i^\omega) \\ a_S^\omega \neq a_S^{\omega*}}} \left\{ \sum_{i \in S} K_i^\omega(a_S^\omega, a_{N \setminus S}^{\omega*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) V^{\omega'}(S) \right\},$$

тогда в регуляризованной игре G_σ существует сильное трансферабельное равновесие с выигрышами игроков $(\bar{\sigma}_1, \dots, \bar{\sigma}_n)$.

Доказательство. Доказательство теоремы повторяет доказательство Теоремы 2.4, используя вместо стратегии (2.4.12) следующую стратегию поведения φ'_i , $i \in N$:

$$\varphi'_i(h(k)) = \begin{cases} a_i^{\omega*}, & \text{если } \omega(k) = \omega, h(k) \subset h^*; \\ a_i^{\omega'}(S), & \text{если } \omega(k) = \omega \text{ и } \exists l \in [1, k-1], \\ & S \subset N, i \notin S: h(l) \subset h^* \text{ и} \\ & (\omega(l), a(l)) \notin h^*, \text{ но} \\ & (\omega(l), (a_S^*(l), a_{N \setminus S}(l))) \in h^*; \\ \text{любая} & \text{в противном случае,} \end{cases} \quad (2.6.3)$$

где $a_i^{\omega*}$ — стратегия игрока i в кооперативном режиме, тогда как стратегия $a_i^{\omega'}(S) \in \Delta(A_i^\omega)$ — стратегия игрока i в режиме наказания, которая совместно со стратегиями $a_{i'}^{\omega'}(S) \in \Delta(A_{i'}^\omega)$ игроков $i' \neq i$, $i' \in N \setminus S$, образует стратегию коалиции $N \setminus S$ против коалиции S , позволяя коалиции S получить минимаксное значение $V^\omega(S)$ в подыгре G^ω . \square

§ 2.7 Стохастическая игра с одним поглощающим состоянием

В этом параграфе рассмотрим стохастическую игру двух лиц с двумя состояниями. Множество игроков есть $N = \{1, 2\}$. Пусть состояние ω_1 определено биматричной игрой:

$$\omega_1 : \begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} (a, a+1) & (c, b) \\ (b, c) & (d+1, d) \end{array} \right) \end{array} \quad (2.7.1)$$

В этом состоянии игроки имеют две стратегии C (кооперироваться) и D (не кооперироваться). Константы удовлетворяют неравенствам:

$$b > a + 1, \quad a > d + 1, \quad d > c > 0.$$

Также предположим, что

$$2a + 1 > b + c. \quad (2.7.2)$$

Из неравенства (2.7.2) следует, что игроки получают больший суммарный выигрыш, кооперируясь, чем если они играют, не кооперируясь. Игра в состоянии ω_1 представляет игру «Дилемма заключенного» с асимметричными игроками: в ситуации (C, C) выигрыш игрока 1 меньше, чем игрока 2, но в ситуации (D, D) выигрыш игрока 2 меньше выигрыша игрока 1. Если реализуется ситуация (C, C) или (D, D) в состоянии ω_1 , то игровой процесс остается в этом же состоянии в вероятностью 1. Если реализуется ситуация (C, D) или (D, C) , то осуществляется переход в состояние ω_2 , которое является «поглощающим», то есть, это состояние будет реализовываться на всех последующих шагах игры с вероятностью 1. В состоянии ω_2 оба игрока имеют единственную стратегию D , и их выигрыши равны d :

$$\omega_2 : \begin{array}{c} D \\ D \end{array} \left(d, d \right) \quad (2.7.3)$$

Матрицы переходных вероятностей из состояний ω_1 и ω_2 есть

$$\begin{pmatrix} (1, 0) & (0, 1) \\ (0, 1) & (1, 0) \end{pmatrix}, \quad (0, 1).$$

Дисконтирующий фактор равен $\delta \in (0, 1)$ и вектор начального распределения на множестве состояний равен $\pi_0 = (1, 0)$, то есть, стохастическая игра начинается из состояния ω_1 .

Кооперативная игра. По описанной некооперативной игре построим кооперативную игру, вычислив значения характеристических функций для всех коалиций, игры и любой ее подыгры. После этого покажем, как определить регуляризацию игры, чтобы вектор Шепли, выбранный игроками в качестве решения игры, был бы позиционно состоятелен. Условие стратегической устойчивости даст нижнюю границу для дисконтирующего фактора.

Первый шаг — определить кооперативную игру G_c по некооперативной стохастической игре G . В частности, нам необходимо найти вектор кооперативных стратегий и посчитать значения характеристических функций.

Вычислим кооперативное решение $\eta^* = (\eta_1^*, \eta_2^*)$ по формулам (2.3.1) и (2.2.4). В кооперативной ситуации оба игрока выбирают стратегию C в состоянии ω_1 и D в состоянии ω_2 . Суммарный выигрыш игроков в ситуации η^* равен значению характеристической функции для коалиции N :

$$\bar{V}(\{1, 2\}) = \bar{E}_1(\eta^*) + \bar{E}_2(\eta^*) = 2a + 1 + \delta(2a + 1) + \dots = \frac{2a + 1}{1 - \delta}. \quad (2.7.4)$$

В частности, для подыгр значения $V^\omega(\{1, 2\})$ равны

$$V(\{1, 2\}) = \begin{pmatrix} V^{\omega_1}(\{1, 2\}) \\ V^{\omega_2}(\{1, 2\}) \end{pmatrix} = \begin{pmatrix} \frac{2a + 1}{1 - \delta} \\ \frac{2d}{1 - \delta} \end{pmatrix}. \quad (2.7.5)$$

Теперь мы можем посчитать значения характеристических функций для коалиций $\{1\}$ и $\{2\}$ для двух подыгр, используя (2.3.4):

$$\begin{aligned} V^{\omega_1}(\{1\}) &= \frac{d + 1}{1 - \delta}, \\ V^{\omega_1}(\{2\}) &= \frac{d}{1 - \delta}, \\ V^{\omega_2}(\{1\}) &= V^{\omega_2}(\{2\}) = \frac{d}{1 - \delta}. \end{aligned}$$

По определению (2.3.8) значения характеристических функций для пустых коалиций равны нулю:

$$V(\emptyset) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Следуя (2.3.7), вычислим значения $\bar{V}(\cdot)$ всех возможных коалиций, учитывая, что вектор начального распределения по состояниям равен $\pi_0 = (1, 0)$:

$$\bar{V}(\emptyset) = 0, \quad \bar{V}(\{1\}) = \frac{d+1}{1-\delta}, \quad \bar{V}(\{2\}) = \frac{d}{1-\delta}, \quad \bar{V}(\{1, 2\}) = \frac{2a+1}{1-\delta}.$$

Таким образом, мы определили кооперативные подыгры $G_c^{\omega_j}$ в виде $\langle N, V^{\omega_j}(\cdot) \rangle$, $j = 1, 2$, и кооперативную игру G_c в виде $\langle N, \bar{V}(\cdot) \rangle$.

Вектор Шепли. Предположим, что в качестве решения кооперативной стохастической игры G_c и любой ее подыгры $G_c^{\omega_j}$, $j = 1, 2$ выбран вектор Шепли, который равен:

$$\sigma_i^{\omega_j} = V^{\omega_j}(\{i\}) + \frac{V^{\omega_j}(\{1, 2\}) - V^{\omega_j}(\{1\}) - V^{\omega_j}(\{2\})}{2},$$

где $i = 1, 2$. Вектора Шепли в подыграх равны:

$$\sigma_1 = \begin{pmatrix} \sigma_1^{\omega_1} \\ \sigma_1^{\omega_2} \end{pmatrix} = \begin{pmatrix} \frac{a+1}{1-\delta} \\ \frac{d}{1-\delta} \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} \sigma_2^{\omega_1} \\ \sigma_2^{\omega_2} \end{pmatrix} = \begin{pmatrix} \frac{a}{1-\delta} \\ \frac{d}{1-\delta} \end{pmatrix}.$$

Учитывая вектор начального распределения π_0 , вычислим вектор Шепли $\bar{\sigma}$ в игре G_c согласно определению 2.4:

$$\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2) = \left(\frac{a+1}{1-\delta}, \frac{a}{1-\delta} \right).$$

Позиционная состоятельность вектора Шепли. Будем последовательно проверять, удовлетворяет ли вектор Шепли принципам устойчивой кооперации. Начнем с принципа позиционной состоятельности. Если игроки будут получать выплаты в соответствии с изначально заданными функциями выигрышей, то дисконтированные выигрыши игроков будут равны $\frac{a}{1-\delta}$ и $\frac{a+1}{1-\delta}$ для первого и второго игроков соответственно, в то время как компоненты вектора Шепли во всей игре равны $\frac{a+1}{1-\delta}$ и $\frac{a}{1-\delta}$. Для того, чтобы вектор Шепли был динамически устойчивым, необходимо провести регуляризацию игры, учитывая, что $\bar{\sigma} = \pi_0 \sigma$

и определить ПРД по формуле (2.4.2):

$$\beta_1 = (\mathbb{I} - \delta\Pi(\eta^*))\sigma_1 = \begin{pmatrix} 1 - \delta & 0 \\ 0 & 1 - \delta \end{pmatrix} \begin{pmatrix} \frac{a+1}{1-\delta} \\ \frac{d}{1-\delta} \end{pmatrix} = \begin{pmatrix} a+1 \\ d \end{pmatrix},$$

$$\beta_2 = (\mathbb{I} - \delta\Pi(\eta^*))\sigma_2 = \begin{pmatrix} 1 - \delta & 0 \\ 0 & 1 - \delta \end{pmatrix} \begin{pmatrix} \frac{a}{1-\delta} \\ \frac{d}{1-\delta} \end{pmatrix} = \begin{pmatrix} a \\ d \end{pmatrix}.$$

Определим σ -регуляризацию начальной стохастической игры G , используя ПРД и определение 2.7. Переопределим функции выигрыша игроков в состоянии ω_1 с помощью ПРД, когда игроки реализуют ситуацию η^* . Таким образом, выигрыши игроков в состоянии ω_1 будут такими:

$$\begin{pmatrix} (a+1, a) & (c, b) \\ (b, c) & (d+1, d) \end{pmatrix}.$$

В состоянии ω_1 , когда оба игрока реализуют ситуацию η^* (оба игрока играют C в состоянии ω_1), их выигрыши будут равны $(a+1, a)$. Так как их выигрыши в начальной игре в этом состоянии были равны $(a, a+1)$, то игрок 2 передает выигрыш, равный единице, игроку 1. Если провести регуляризацию с помощью метода, описанного выше, то вектор Шепли и соответствующая ему ПРД удовлетворяют принципу позиционной состоятельности (см. теорему 2.3).

Стратегическая устойчивость вектора Шепли. Теперь проверим выполняется ли принцип стратегической устойчивости вектора Шепли, проверив, выгодно ли игрокам отклоняться от кооперативного решения. Рассмотрим возможные отклонения игроков в состоянии ω_1 (в состоянии ω_2 игроки имеют единственную стратегию). В этом состоянии ситуация, которая реализуется при кооперации, не является равновесием по Нэшу, поэтому по крайней мере одному из игроков может быть выгодно отклонение от кооперативного поведения. Проверим, справедливо ли неравенство

$$\sigma_i^{\omega_1} \geq F^{\omega_1}(\{i\}), \quad (2.7.6)$$

для любого $i = 1, 2$, где

$$F^{\omega_1}(\{i\}) = \max_{a_i^{\omega_1} \in \Delta(A_i^{\omega_1})} \left\{ K_i^{\omega_1}(a_i^{\omega_1}, a_{N \setminus i}^{\omega_1^*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega_1, (a_i^{\omega_1}, a_{N \setminus i}^{\omega_1^*})) V^{\omega'}(\{i\}) \right\}.$$

Для игрока 1, неравенство (2.7.6) эквивалентно следующему:

$$\frac{a+1}{1-\delta} \geq b + \delta d + \delta^2 d + \dots = b + \frac{\delta d}{1-\delta},$$

для игрока 2:

$$\frac{a}{1-\delta} \geq b + \delta d + \delta^2 d + \dots = b + \frac{\delta d}{1-\delta}.$$

Эти два неравенства дают ограничения на дисконтирующий фактор δ , при которых условие стратегической устойчивости вектора Шепли выполняется:

$$\delta \geq \frac{b-a}{b-d}.$$

Защита от иррационального поведения. Для того, чтобы проверить, выполняется ли принцип защиты от иррационального поведения, необходимо сравнить выигрыши игрока в случаях:

- 1) когда игрок играет индивидуально на протяжении всей игры,
- 2) когда игрок кооперируется с другим игроком до некоторого шага, а потом играет индивидуально.

Если выигрыш в 2) не меньше, чем выигрыш в 1), тогда принцип защиты от иррационального поведения выполняется. Это подтверждается теоремой 2.6, потому как

$$(\mathbb{I} - \delta\Pi(\eta^*)) (\sigma_1 - V(\{1\})) = \begin{pmatrix} 1-\delta & 0 \\ 0 & 1-\delta \end{pmatrix} \begin{pmatrix} \frac{a+1}{1-\delta} - \frac{d+1}{1-\delta} \\ \frac{d}{1-\delta} - \frac{d}{1-\delta} \end{pmatrix} = \begin{pmatrix} a-d \\ 0 \end{pmatrix} \geq 0,$$

$$(\mathbb{I} - \delta\Pi(\eta^*)) (\sigma_2 - V(\{2\})) = \begin{pmatrix} 1-\delta & 0 \\ 0 & 1-\delta \end{pmatrix} \begin{pmatrix} \frac{a}{1-\delta} - \frac{d}{1-\delta} \\ \frac{d}{1-\delta} - \frac{d}{1-\delta} \end{pmatrix} = \begin{pmatrix} a-d \\ 0 \end{pmatrix} \geq 0.$$

Оба игрока имеют больший выигрыш в случае, если сначала они получают выплаты в соответствии с ПРД, а потом играют индивидуально, по сравнению с выигрышем в игре, когда они всю игру играют индивидуально.

Выводы. В качестве заключения, можно сформулировать условия, при которых вектор Шепли в описанной стохастической игре удовлетворяет трем принципам устойчивой кооперации (позиционная состоятельность, стратегическая устойчивость, защита от иррационального поведения):

1. Дисконтирующий фактор должен быть $\delta \geq \frac{b-a}{b-d}$.
2. Стохастическая игра должна быть σ -регуляризованной, то есть, выигрыши игроков в состоянии ω_1 должны быть следующими:

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} (a+1, a) & (c, b) \\ (b, c) & (d+1, d) \end{array} \right) \end{array}$$

а в состоянии ω_2 должны остаться неизменными.

§ 2.8 Сильная позиционная состоятельность c -ядра в стохастических играх бесконечной продолжительности

Теперь предположим, что решение кооперативной стохастической игры представляет собой множество, т.е. является подмножеством множества дележей, не вырожденным в точку. Пусть таким решением будет для определенности c -ядро. Сформулируем проблему сильной позиционной состоятельности c -ядра и предложим достаточные условия сильной позиционной состоятельности c -ядра для стохастических игр бесконечной продолжительности вида (2.1.1).

Предположим, что c -ядра стохастической игры G_c и любой подыгры G_c^ω , $\omega \in \Omega$, непусты. Кооперируясь, игроки договариваются о совместной реализации кооперативного решения η^* и рассчитывают получить компоненты дележа, принадлежащего c -ядру \overline{CO} . Попадая по ходу игры в состояния $\omega \in \Omega$, игрок $i \in N$ выбирает стратегию $a_i^{\omega*}$ в соответствии с кооперативной стратегией η_i^* и получает выигрыш $K_i^\omega(a^{\omega*})$. Если игроки произведут пересчет решения, т.е. найдут решение кооперативной подыгры G_c^ω , то текущим решением будет c -ядро CO^ω . Было бы разумным потребовать, чтобы выигрыш, полученный игроком в реализующемся состоянии ω , суммированный с ожидаемой суммой любых дележей из c -ядер $CO^{\omega'}$, $\omega' \in \Omega$, следующих за состоянием ω , являлась бы дележом c -ядра CO^ω . Если это свойство выполнено для любого состояния $\omega \in \Omega$, то c -ядро кооперативной стохастической игры G_c является сильно позиционно состоятельным.

Для нахождения сильно позиционно состоятельного c -ядра определим так называемое ожидаемое c -ядро в состоянии ω , т.е. определим множество ожидаемых дележей, принадлежащих c -ядрам, которые являются решениями следующим подыгр. Для каждого состояния $\omega \in \Omega$ определим *ожидаемое c -ядро*:

$$EC(\omega) = \left\{ \sigma(\omega) = \sum_{\omega' \in \Omega} p(\omega'|\omega, a^{\omega*}) \sigma^{\omega'} \mid \sigma^{\omega'} \in CO^{\omega'} \right\}. \quad (2.8.1)$$

Множеству $EC(\omega)$ принадлежат вектора $\sigma(\omega)$, которые являются математическими ожиданиями всевозможных наборов дележей из c -ядер подыгр, начинающихся из следующих за ω состояний, относительно вероятностного распределения $\{p(\omega'|\omega, a^{\omega*}), \omega' \in \Omega\}$.

Вспомним определение 2.5 процедуры распределения дележа. Первое условие (2.4.1) в определении можно назвать условием «достижимости процедуры распределения дележа», поскольку оно позволяет гарантировать, что в любом реализовавшемся состоянии сумма выплат игрокам равна сумме полученных ими выигрышей при реализации кооперативных стратегий. Второе условие гарантирует игрокам получение компонент заранее выбранного дележа из c -ядра кооперативной игры G_c в смысле математического ожидания, если выплаты игрокам на протяжении всей игры будут производиться в соответствии с процедурой распределения $\{\beta^\omega : \omega \in \Omega\}$.

Теперь зададим процедуру распределения дележа $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n)$, где $\bar{\sigma}_i = \pi_0 \sigma_i$, $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in CO^\omega$, таким образом, чтобы c -ядро было сильно позиционно состоятельно.

Определение 2.11. Назовем c -ядро \overline{CO} кооперативной стохастической игры G_c сильно позиционно состоятельным, если для любого дележа из c -ядра \overline{CO} существует процедура распределения $\{\beta^\omega : \omega \in \Omega\}$ такая, что для любого состояния $\omega \in \Omega$ имеют место включения

$$\beta^\omega \oplus \delta EC(\omega) \subset CO^\omega, \quad (2.8.2)$$

$$B^\omega \in CO^\omega, \quad \omega \in \Omega \quad (2.8.3)$$

где

$$\beta^\omega \oplus \delta EC(\omega) = \left\{ \beta^\omega + \delta \sigma(\omega) : \sigma(\omega) \in EC(\omega) \right\}.$$

При этом, процедуру распределения $\{\beta^\omega : \omega \in \Omega\}$ также назовем *сильно позиционно состоятельной*.

Условие (2.8.2) означает, что множество векторов, равных сумме процедуры распределения дележа в состоянии ω и дележа из ожидаемого c -ядра для этого состояния, содержится в c -ядре подыгры, начинающейся из состояния ω . Это условие накладывает ограничения на выплаты игрокам в реализуемых состояниях, и чаще всего не выполняется для произвольной игры, если выплаты игрокам производятся в соответствии с изначально заданными функциями выигрышей.

Наложим дополнительные ограничения на характеристические функции подыгр, начинающихся из состояний множества Ω , чтобы получить достаточные условия сильной позиционной состоятельности c -ядра. Обозначим через $EV^\omega(S)$ ожидаемое значение характеристической функции, вычисленное для коалиции $S \subseteq N$, для подыгр, следующих за подыгрой в ω :

$$EV^\omega(S) = \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) V^{\omega'}(S).$$

Обозначим через

$$\Delta V^\omega(S) = V^\omega(S) - \delta EV^\omega(S)$$

разность между значениями характеристической функции в состоянии ω и ожидаемым значением характеристической функции. Обозначим через ΔCO^ω — аналог c -ядра, построенного по функции $\Delta V^\omega(S)$. Сформулируем достаточное условие сильной позиционной состоятельности процедуры распределения дележа и c -ядра \overline{CO} .

Теорема 2.8. Пусть для любого состояния $\omega \in \Omega$ c -ядро CO^ω и множество ΔCO^ω непусты. Если для любого состояния $\omega \in \Omega$ процедура распределения $\{\beta^\omega : \omega \in \Omega\}$ дележа из c -ядра \overline{CO} удовлетворяет условиям:

$$\beta^\omega \in \Delta CO^\omega, \tag{2.8.4}$$

$$B^\omega \in CO^\omega, \quad \omega \in \Omega, \tag{2.8.5}$$

то этот дележ из c -ядра \overline{CO} и соответствующая процедура $\{\beta^\omega : \omega \in \Omega\}$ *сильно позиционно состоятельны*.

Доказательство. Покажем, что произвольный вектор $\beta^\omega \in \Delta CO^\omega$, удовлетворяющий условиям (2.8.4) и (2.8.5), является сильно позиционно состоятельной процедурой распределения некоторого дележа $\bar{\sigma} \in \overline{CO}$, то есть, что условия (2.8.2) и (2.8.3) из определения 2.11 выполнены. Условие (2.8.5) совпадает с (2.8.3), поэтому осталось показать, что включение (2.8.2) выполняется для любого состояния $\omega \in \Omega$. В состоянии ω возьмем произвольный вектор $\sigma(\omega) \in EC(\omega)$ и найдем сумму $\beta^\omega + \delta\sigma(\omega)$. Проверим, принадлежит ли последний вектор c -ядру \overline{CO} . Сначала вычислим сумму всех компонент этого вектора:

$$\begin{aligned} \sum_{i \in N} \beta_i^\omega + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) \sum_{i \in N} \sigma_i^{\omega'} &= \\ &= V^\omega(N) - \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) V^{\omega'}(N) + \\ &+ \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) \sum_{i \in N} \sigma_i^{\omega'} = V^\omega(N), \end{aligned}$$

что означает выполнение свойства коллективной рациональности.

Теперь пусть $S \subset N$:

$$\begin{aligned} \sum_{i \in S} \beta_i^\omega + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) \sum_{i \in S} \sigma_i^{\omega'} &\geq \\ &\geq V^\omega(S) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) V^{\omega'}(S) - \\ &- \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) V^{\omega'}(S) = V^\omega(S). \end{aligned}$$

В силу произвольности выбора состояния $\omega \in \Omega$, можно сделать вывод о том, что дележ из c -ядра кооперативной игры G_c и его ПРД $\{\beta^\omega : \omega \in \Omega\}$ сильно позиционно состоятельны. \square

В случае непустоты аналогов c -ядер ΔCO^ω для любого состояния ω теорема 2.8 дает метод построения сильно позиционно состоятельной процедуры распределения дележей из c -ядра. Стоит отметить, что в общем случае не все дележи из c -ядра могут быть реализованы с помощью определенной выше процедуры распределения $\{\beta^\omega : \omega \in \Omega\}$.

Глава 3

Динамические игры, разыгрываемые на деревьях событий

§ 3.1 Определение динамической игры, разыгрываемой на дереве событий

В этой главе мы рассмотрим класс динамических игр, в которых игровой процесс происходит на так называемом «дереве событий». Дерево событий представляет собой конечный древовидный или связный ациклический граф. Основное свойство дерева событий — это то, что вероятности перехода из одной вершины дерева в другую заданы изначально и не меняются в течение игры и они не зависят от стратегий игроков, выбранных на этом шаге. Подробное описание данного класса игр можно найти в работах [103, 226]. Обозначения, используемые в работе для этого класса игр, также изначально введены в вышеупомянутых работах.

Пусть $\mathcal{T} = \{0, 1, \dots, T\}$ — это множество периодов игры. Стохастический процесс задан в виде дерева событий, которое имеет начальную вершину n^0 в период времени $t = 0$ и множество вершин $\mathcal{N}^t = \{n_1^t, \dots, n_{N_t}^t\}$ в период $t = 1, \dots, T$. Каждая вершина $n_i^t \in \mathcal{N}^t$ представляет собой возможное состояние стохастического процесса в период времени t . Структура древовидного графа позволяет аккумулировать информацию о том, как стохастический процесс развивается во времени, т.е. состояние меняется. Обозначим через $a(n_i^t) \in \mathcal{N}^{t-1}$ единственного предшественника вершины $n_i^t \in \mathcal{N}^t$ в дереве событий для любого $t = 1, \dots, T$ и через $\mathcal{S}(n_i^t) \subset \mathcal{N}^{t+1}$ множество всех вершин-последователей,

которые следуют за вершиной $n_i^t \in \mathcal{N}^t$ для всех $t = 0, \dots, T - 1$. Путь от корня дерева событий n^0 к терминальной вершине n_i^T будем называть *сценарием*. Каждый сценарий имеет заданную вероятность реализации, и сумма вероятностей реализаций всех возможных сценариев в дереве равна единице. Считаем, что задана вероятность $\pi(n_i^t)$ прохождения стохастического процесса через вершину n_i^t . Вероятность $\pi(n_i^t)$ равна сумме вероятностей реализации всех сценариев, содержащих вершину n_i^t . В частности, вероятность начальной вершины $\pi(n^0) = 1$, а вероятность любой терминальной вершины $\pi(n_i^T)$ равна вероятности реализации единственного сценария, заканчивающегося в вершине n_i^T . Заданные вероятности $\pi(n_i^t), n_i^t \in \mathcal{N}^t, t = 0, \dots, T - 1$, удовлетворяют условию:

$$\pi(n_i^t) = \sum_{n_k^{t+1} \in \mathcal{S}(n_i^t)} \pi(n_k^{t+1}). \quad (3.1.1)$$

Обозначим через $M = \{1, \dots, m\}$ множество игроков. Для каждого игрока $j \in M$ определим множество стратегий для каждой вершины дерева событий. В каждой вершине $n_i^t \in \mathcal{N}^t, t = 0, 1, \dots, T - 1$, задано множество стратегий $U_j^{n_i^t} \subset \mathbb{R}^{\mu_j^{n_i^t}}$ игрока j , где $\mu_j^{n_i^t}$ — заданное положительное число. Пусть $u_j(n_i^t) \in U_j^{n_i^t}$ — стратегия игрока j в вершине n_i^t , тогда $u(n_i^t) = (u_1(n_i^t), \dots, u_m(n_i^t))$ — ситуация в вершине n_i^t . Обозначим через $U^{n_i^t} = U_1^{n_i^t} \times \dots \times U_j^{n_i^t} \times \dots \times U_m^{n_i^t}$ множество ситуаций в вершине n_i^t .

Также определено множество состояний $X \subset \mathbb{R}^p$, где p — заданное положительное число. Будем считать, что множество состояний X одинаково для всех вершин $n_i^t \in \mathcal{N}^t, t = 0, 1, \dots, T$. Для каждой вершины n_i^t определена функция перехода из одного состояния в другое: $f^{n_i^t}(\cdot, \cdot) : X \times U^{n_i^t} \mapsto X$. Уравнение динамики состояния определяется следующим образом:

$$x(n_i^t) = f^{a(n_i^t)}(x(a(n_i^t)), u(a(n_i^t))), \quad (3.1.2)$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, n_i^t \in \mathcal{N}^t, t = 1, \dots, T. \quad (3.1.3)$$

В каждой нетерминальной вершине $n_i^t, t = 0, \dots, T - 1$, выигрыш игрока j определяется состоянием и стратегиями всех игроков, т. е. это функция $\phi_j^{n_i^t}(x(n_i^t), u(n_i^t))$. В терминальной вершине n_i^T выигрыш игрока j определяется функцией $\Phi_j^{n_i^T}(x(n_i^T))$, которая напрямую не зависит от стратегий игроков.

Будем предполагать, что игрок $j \in M$ максимизирует свой суммарный ожидаемый выигрыш, дисконтированный с помощью дисконт фактора λ_j ($0 < \lambda_j < 1$). Определим многошаговую игру, используя уравнения динамики состояния и функции выигрышей игроков:

$$\begin{aligned}\mathbf{x} &= \{x(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T\}, \\ \mathbf{u} &= \{u(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T - 1\},\end{aligned}$$

при этом выигрыш игрока j обозначим через $J_j(\mathbf{x}, \mathbf{u})$ и будем вычислять по формуле

$$\begin{aligned}J_j(\mathbf{x}, \mathbf{u}) &= \sum_{t=0}^{T-1} \lambda_j^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \phi_j^{n_i^t}(x(n_i^t), u(n_i^t)) \\ &\quad + \lambda_j^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) \Phi_j^{n_i^T}(x(n_i^T)), \quad j \in M,\end{aligned}\tag{3.1.4}$$

при условиях

$$x(n_i^t) = f^{a(n_i^t)}(x(a(n_i^t)), u(a(n_i^t))),\tag{3.1.5}$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, t = 1, \dots, T,$$

$$x(n^0) = x^0.\tag{3.1.6}$$

Определение 3.1. *S-адаптивная стратегия игрока j , заданная в каждой вершине дерева событий, — это вектор допустимых стратегий игрока j , т. е. $\mathbf{u}_j = \{u_j(n_i^t) : u_j(n_i^t) \in U_j^{n_i^t}, n_i^t \in \mathcal{N}^t, t = 0, \dots, T - 1\}$.*

Обозначим через $\mathbf{u} = (\mathbf{u}_j : j \in M)$ набор *S*-адаптивных стратегий игроков из множества M и назовем его ситуацией. Теперь мы можем определить игру в нормальной форме следующим образом: $\langle M, \{\mathbf{u}_j\}_{j \in M}, \{W_j(\mathbf{u}, x^0)\}_{j \in M} \rangle$, где $W_j(\mathbf{u}, x^0)$ — функция выигрыша игрока j , $W_j(\mathbf{u}, x^0) = J_j(\mathbf{x}, \mathbf{u})$, \mathbf{x} — набор состояний системы, являющийся единственным решением уравнений (3.1.2) при начальном состоянии x^0 и условии, что реализуется ситуация \mathbf{u} .

Если игра разыгрывается некооперативно, т. е. каждый игрок максимизирует свою функцию выигрыша (3.1.4), то в качестве решения игры мы будем рассматривать равновесие по Нэшу в классе *S*-адаптивных стратегий:

Определение 3.2. *S-адаптивное равновесие по Нэшу* — это ситуация в *S-адаптивных стратегиях* $\mathbf{u}^N = (\mathbf{u}_j : j \in M)$ такая, что для любого игрока $j \in M$ и любой *S-адаптивной стратегии* \mathbf{u}_j справедливо неравенство:

$$W_j(\mathbf{u}^N, x^0) \geq W_j((\mathbf{u}_j, \mathbf{u}_{-j}^N), x^0),$$

где $(\mathbf{u}_j, \mathbf{u}_{-j}^N)$ — это ситуация в *S-адаптивных стратегиях*, в которой любой игрок $i \in M$, $i \neq j$, реализует равновесную по Нэшу *S-адаптивную стратегию*, а игрок j — стратегию \mathbf{u}_j .

Если игроки приняли решение о кооперации, тогда они будут максимизировать сумму дисконтированных ожидаемых выигрышей игроков во всей игре, т. е. можно сформулировать задачу:

$$\max_{\mathbf{u}_j: j \in M} \sum_{j \in M} W_j(\mathbf{u}, x^0).$$

Обозначим через $\mathbf{u}^*(x^0)$ вектор кооперативных стратегий, решающих поставленную задачу, т. е.

$$\mathbf{u}^*(x^0) = \arg \max_{\mathbf{u}=(\mathbf{u}_j: j \in M)} \sum_{j \in M} W_j(\mathbf{u}, x^0). \quad (3.1.7)$$

Кооперативную траекторию, которая определяется кооперативными стратегиями $\mathbf{u}^*(x^0)$ и уравнениями (3.1.2), обозначим через $\mathbf{x}^* = \{x^*(n_l^t) : n_l^t \in \mathcal{N}^t, t = 0, 1, \dots, T\}$.

В дальнейшем, нам также понадобится определение подыгры, начинающейся в вершине $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T - 1$ из состояния $x^*(n_l^t)$. Эта подыгра разыгрывается на поддереве событий $\Gamma(n_l^t)$ начального дерева событий. Корневой вершиной поддерева $\Gamma(n_l^t)$ является вершина n_l^t . Определим выигрыш игрока $j \in M$ в этой подыгре следующим образом:

$$W_j(\mathbf{u}(x^*(n_l^t)), x^*(n_l^t)) = \sum_{\theta=t}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \phi_j^{n_l^\theta}(x(n_l^\theta), u(x(n_l^\theta))) + \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \Phi_j^{n_l^T}(x(n_l^T)), \quad (3.1.8)$$

где $\mathcal{N}_\Gamma^\theta = \mathcal{N}^\theta \cap \Gamma(n_l^t)$, $\mathbf{u}(x^*(n_l^t)) = (\mathbf{u}_j(x^*(n_l^t)) : j \in M)$ — это ситуация в *S-адаптивных стратегиях*, $\mathbf{u}_j(x^*(n_l^t))$ — *S-адаптивная стратегия* игрока j в

подыгре, начинающейся из вершины n_i^t с начальным состоянием $x^*(n_i^t)$, $n_i^t \in \mathcal{N}^t$, $t = 1, \dots, T - 1$. Число $\pi(n_i^\theta | n_i^t)$ — условная вероятность того, что вершина n_i^θ реализуется при условии, что подыгра начинается из вершины n_i^t . Условная вероятность $\pi(n_i^\theta | n_i^t)$ может быть вычислена по формуле: $\pi(n_i^\theta | n_i^t) = \pi(n_i^\theta) / \pi(n_i^t)$, если $\pi(n_i^t) \neq 0$. В противном случае, подыгра, начинающаяся с вершины n_i^t , не может реализоваться.

В случае, когда игроки играют некооперативно в подыгре, начинающейся из вершины n_i^t из состояния $x^*(n_i^t)$, то в качестве решения мы также рассмотрим S -адаптивное равновесие по Нэшу согласно определению 3.2. Обозначим через $\mathbf{u}^N(x^*(n_i^t)) = (\mathbf{u}_j^N(x^*(n_i^t)) : j \in M)$ ситуацию равновесия по Нэшу в этой подыгре. Выигрыш игрока j в подыгре, начинающейся из вершины n_i^t из состояния $x^*(n_i^t)$, в ситуации S -адаптивного равновесия по Нэшу равен $W_j(\mathbf{u}^N(x^*(n_i^t)), x^*(n_i^t))$. Обозначим через $\mathbf{u}_j^N(x^*(n_i^t), [n_v^\tau, n_w^T])$ траекторию оптимальных стратегий $\mathbf{u}_j^N(x^*(n_i^t))$, которые реализуются на пути от вершины $n_v^\tau \in \mathcal{N}^\tau$, $\tau > t$ до терминальной вершины $n_w^T \in \mathcal{N}^T$.

Если игроки играют кооперативно в подыгре, начинающейся из вершины n_i^t и состояния $x^*(n_i^t)$, то они максимизируют сумму дисконтированных выигрышей, полученных игроками в этой подыгре:

$$\max_{\mathbf{u}_j(x^*(n_i^t)):j \in M} \sum_{j \in M} W_j(\mathbf{u}(x^*(n_i^t)), x^*(n_i^t)).$$

Кооперативные стратегии в подыгре, начинающейся из вершины n_i^t из состояния $x^*(n_i^t)$, определяются следующим образом:

$$\mathbf{u}^*(x^*(n_i^t)) = \arg \max_{\mathbf{u}_j(x^*(n_i^t)):j \in M} \sum_{j \in M} W_j(\mathbf{u}(x^*(n_i^t)), x^*(n_i^t)). \quad (3.1.9)$$

Таким образом, выигрыш игрока j в кооперативной подыгре, начинающейся из вершины n_i^t и начального состояния $x^*(n_i^t)$, $n_i^t \in \mathcal{N}^t$, $t = 1, \dots, T$, равен $W_j(\mathbf{u}^*(x^*(n_i^t)), x^*(n_i^t))$. Обозначим траекторию или «след» стратегий $\mathbf{u}_j^*(x^*(n_i^t))$ на пути от вершины $n_v^\tau, n_v^\tau \in \mathcal{N}^\tau$, $\tau > t$ до вершины $n_w^T \in \mathcal{N}^T$, через $\mathbf{u}_j^*(x^*(n_i^t), [n_v^\tau, n_w^T])$.

Замечание 3.1. Траектории $\mathbf{u}_j^N(x^*(n_i^t), [n_v^\tau, n_w^T])$ и $\mathbf{u}_j^*(x^*(n_i^t), [n_v^\tau, n_w^T])$ при $\tau > t$ в общем случае не совпадают. Причина этого заключается в том, что тра-

ектория $\mathbf{u}_j^N(x^*(n_l^t), [n_v^\tau, n_w^T])$ находится в предположении, что игроки кооперировались только в интервале времени $[0, t]$, тогда как $\mathbf{u}_j^N(x^*(n_v^\tau), [n_v^\tau, n_w^T])$ рассчитывается в предположении, что кооперация между игроками имеет место на интервале $[0, \tau]$, где $\tau > t$.

Замечание 3.2. Предположим, что кооперативное равновесие и равновесие по Нэшу во всей игре и любой ее подыгре единственны. Для единственности кооперативного равновесия мы требуем строгой вогнутости целевой функции, а также выпуклости и компактности множества стратегий каждого игрока. Для доказательства единственности S-адаптивного равновесия по Нэшу используем то, что многошаговая игра определена как игра в нормальной форме, поэтому условия единственности являются стандартными для «классических» игр с непрерывными функциями выигрышами при заданных ограничениях [200].

§ 3.2 Позиционная состоятельность s -ядра в динамических играх, разыгрываемых на деревьях событий

Предположим, что игроки решили играть кооперативно в течение всей игры, а также использовать в качестве дележа суммарного выигрыша s -ядро. Выигрыши игроков во всей игре будут равны $W_j(\mathbf{u}^*(x^0), x^0)$, $j \in M$. Характеристическая функция $V(G; x^*(n_l^t))$ в подыгре, начинающейся из любой вершины n_l^t и состояния $x^*(n_l^t)$ определим следующим образом:

$$V(G; x^*(n_l^t)) : 2^M \rightarrow \mathbb{R}, \quad (3.2.1)$$

где G — коалиция игроков, $G \subset M$, и $V(\emptyset; x^*(n_l^t)) = 0$. В частности, $V(G; x^0)$ — характеристическая функция игры в целом.

Обозначим через $Y(x^*(n_l^t))$ множество дележей:

$$Y(x^*(n_l^t)) = \{(y_1(x^*(n_l^t)), \dots, y_m(x^*(n_l^t))) \mid y_j(x^*(n_l^t)) \geq V(\{j\}; x^*(n_l^t)), \\ \forall j \in M, \text{ и } \sum_{j \in M} y_j(x^*(n_l^t)) = V(M; x^*(n_l^t))\}.$$

Как характеристическая функция, так и дележ, и множество дележей, зависят от начальной вершины и начального состояния этой подыгры. Мы опустим вершину в обозначении аргумента для простоты, оставив только начальное состояние подыгры.

Определение 3.3. Дележ $y(x^*(n_i^t)) = (y_1(x^*(n_i^t)), \dots, y_m(x^*(n_i^t)))$ называется неустойчивым относительно коалиции G , если $V(G; x^*(n_i^t)) > \sum_{j \in G} y_j(x^*(n_i^t))$. Дележ $y(x^*(n_i^t))$ неустойчив, если существует коалиция G такая, что дележ $y(x^*(n_i^t))$ неустойчив относительно G . В противном случае $y(x^*(n_i^t))$ устойчив.

Множество $C(x^0)$ устойчивых дележей — s -ядро игры, начинающейся в вершине n^0 из состояния x^0 , которое определяется следующим образом:

$$C(x^0) = \left\{ (y_1, \dots, y_m) \mid \sum_{j \in G} y_j \geq V(G; x^0) \text{ для } \forall G \subset M, \right. \\ \left. \text{и } \sum_{j \in M} y_j = V(M; x^0) \right\}. \quad (3.2.2)$$

Аналогичным образом определим s -ядро для каждой подыгры, начинающейся в вершине n_i^t из состояния $x^*(n_i^t)$:

$$C(x^*(n_i^t)) = \left\{ (y_1, \dots, y_m) \mid \sum_{j \in G} y_j \geq V(G; x^*(n_i^t)) \text{ } \forall G \subset M, \right. \\ \left. \text{и } \sum_{j \in M} y_j = V(M; x^*(n_i^t)) \right\}. \quad (3.2.3)$$

Предположим, что s -ядро игры $C(x^0)$ и s -ядро $C(x^*(n_i^t))$ подыгры, начинающейся с любой вершины n_i^t из состояния $x^*(n_i^t)$, непусты. В противном случае дальнейшие построения теряют смысл. Условия непустоты s -ядра в данной работе не исследуются, поскольку эта проблема достаточно хорошо изучена и изложена в литературе по нединамическим кооперативным играм [51, 176].

Для построения характеристической функции $V(G; x^*(n_i^t))$ необходимо сделать предположения о том, что принимать за «силу» коалиции, т. е. как рассчитывать значения функции для любой коалиции G , принимая во внимание

поведение игроков вне этой коалиции, т.е. игроков из коалиции $N \setminus G$. Существует несколько подходов к построению характеристической функции для кооперативной игры, если изначально задана некооперативная игра. Эти подходы подробно описаны в Главе 1. В рамках данной главы мы будем использовать γ -подход и строить γ -характеристическую функцию [81]. Согласно предположениям данного подхода, когда образуется коалиция G , остальные игроки из множества $N \setminus G$ не объединяются в одну коалицию, а играют равновесные по Нэшу стратегии в некооперативной игре, в которой коалиция G представляет собой одного игрока. Следовательно, значение γ -характеристической функции для коалиции G равняется суммарному выигрышу игроков коалиции G в ситуации равновесия по Нэшу, когда игроки коалиции G максимизируют их суммарный выигрыш, а все остальные игроки максимизируют свои индивидуальные выигрыши. c -ядро, рассчитанное по значениям γ -характеристической функции, называется γ -ядро [81]. Определим, какое c -ядро называется позиционно состоятельным, введя понятие процедуры распределения дележа для данного класса игр.

Определение 3.4. Набор выплат игрокам $(\{\beta_j(x^*(n_i^t))\}_{n_i^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M)$, определенных для каждой вершины n_i^t и всех игроков $j \in M$, называется процедурой распределения дележа $y(x^0) = (y_1(x^0), \dots, y_m(x^0))$, если для любого игрока $j \in M$ имеет место равенство:

$$y_j(x^0) = \sum_{\theta=0}^T \lambda_j^\theta \sum_{n_i^\theta \in \mathcal{N}^\theta} \pi(n_i^\theta) \beta_j(x^*(n_i^\theta)), \quad (3.2.4)$$

$$\sum_{j \in M} \beta_j(x^*(n_i^t)) = \sum_{j \in M} \phi_j^{n_i^t}(x^*(n_i^t), u^*(x^*(n_i^t))), \quad (3.2.5)$$

$$\sum_{j \in M} \beta_j(x^*(n_i^T)) = \sum_{j \in M} \Phi_j^{n_i^T}(x^*(n_i^T)), \quad (3.2.6)$$

где равенство (3.2.5) имеет место для всех $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T - 1$ и равенство (3.2.6) — для всех $n_i^T \in \mathcal{N}^T$.

Условие (3.2.4) означает, что ожидаемая сумма выплат игроку $j \in M$ равна соответствующей компоненте дележа из c -ядра, который игроки решили реализовать в игре. Равенства (3.2.5), (3.2.6) — условия «допустимости» процедуры

распределения дележа, т. е. сумма выплат игрокам в любой вершине дерева событий равна сумме выигрышей, которые игроки получают в этой вершине, реализуя кооперативные стратегии $u_j^*(x^*(n_l^t))$, $j \in M$, во всех вершинах $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T - 1$.

Определение 3.5. Дележ $y(x^0) \in C(x^0)$ и соответствующая ему процедура распределения дележа

$$\left(\left\{ \beta_j(x^*(n_l^t)) \right\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M \right)$$

называются *позиционно состоятельными* (динамически устойчивыми), если для любой вершины $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T$ с состоянием $x^*(n_l^t)$ существует дележ $y(x^*(n_l^t)) = (y_1(x^*(n_l^t)), \dots, y_m(x^*(n_l^t))) \in C(x^*(n_l^t))$, удовлетворяющий условию:

$$\sum_{\theta=0}^{t-1} \lambda_j^\theta \sum_{n_k^\theta \in \mathcal{N}^\theta} \pi(n_k^\theta) \beta_j(x^*(n_k^\theta)) + \lambda_j^t \sum_{n_k^t \in \mathcal{N}^t} \pi(n_k^t) y_j(x^*(n_l^t)) = y_j(x^0). \quad (3.2.7)$$

Определение 3.6. C -ядро $C(x^0)$ — позиционно состоятельное решение, если любой дележ $y(x^0)$ из C -ядра $C(x^0)$ является позиционно состоятельным.

Теорема 3.1. Если C -ядро $C(x^0)$ игры и C -ядра $C(x^*(n_l^t))$ всех ее подыгр, начинающихся в вершинах $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T$ из состояний $x^*(n_l^t)$, непусты, тогда C -ядро $C(x^0)$ позиционно состоятельно, если соответствующая процедура распределения дележа $y(x^0) \in C(x^0)$ для любого $t = 0, \dots, T - 1$ вычисляется по формуле:

$$\beta_j(x^*(n_l^t)) = y_j(x^*(n_l^t)) - \lambda_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \pi(n_k^{t+1} | n_l^t) y_j(x^*(n_k^{t+1})), \quad (3.2.8)$$

и для $t = T$ по формуле:

$$\beta_j(x^*(n_l^T)) = y_j(x^*(n_l^T)), \quad (3.2.9)$$

где $y(x^*(n_l^t)) = (y_1(x^*(n_l^t)), \dots, y_m(x^*(n_l^t))) \in C(x^*(n_l^t))$ для любой вершины $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T$ и $\pi(n_k^{t+1} | n_l^t)$ — условная вероятность того, что реализуется вершина n_k^{t+1} , если вершина n_l^t уже реализовалась.

Доказательство. Сначала покажем, что выплаты $\beta_j(x^*(n_l^t))$, определяемые формулами (3.2.8) и (3.2.9), являются компонентами процедуры распределения дележа $y(x^0)$ из s -ядра $C(x^0)$. Мы должны показать, что условия (3.2.4)–(3.2.6) выполнены. Доказательство равенства (3.2.4) следует из метода построения процедуры распределения дележа, а именно, из (3.2.8) и (3.2.9), а также учитывая, что $\pi(n_l^t)\pi(n_k^{t+1}|n_l^t) = \pi(n_k^{t+1})$ для любой вершины $n_k^{t+1} \in \mathcal{N}^{t+1}$, $t = 0, \dots, T-1$:

$$\begin{aligned} \sum_{\theta=0}^T \lambda_j^\theta \sum_{n_k^\theta \in \mathcal{N}^\theta} \pi(n_k^\theta) \beta_j(x^*(n_k^\theta)) &= \sum_{\theta=0}^{T-1} \lambda_j^\theta \sum_{n_k^\theta \in \mathcal{N}^\theta} \pi(n_k^\theta) \left\{ y_j(x^*(n_l^t)) \right. \\ &\quad \left. - \lambda_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \pi(n_k^{t+1}|n_l^t) y_j(x^*(n_k^{t+1})) \right\} + \lambda_j^T \sum_{n_k^T \in \mathcal{N}^T} \pi(n_k^T) y_j(x^*(n_l^T)) \\ &= \pi(n^0) y_j(x^0) - \lambda_j \sum_{n_k^1 \in \mathcal{S}(n^0)} \pi(n^0) \pi(n_k^1|n^0) y_j(x^*(n_k^1)) + \\ &\quad + \lambda_j \sum_{n_k^1 \in \mathcal{S}(n^0)} \pi(n_k^1) y_j(x^*(n_k^1)) - \dots + \lambda_j^T \sum_{n_k^T \in \mathcal{S}(n_l^{T-1})} \pi(n_k^T) y_j(x^*(n_k^T)) - \\ &\quad - \lambda_j^T \sum_{n_k^T \in \mathcal{S}(n_l^{T-1})} \pi(n_l^{T-1}) \pi(n_k^T|n_l^{T-1}) y_j(x^*(n_k^T)) = y_j(x^0), \end{aligned}$$

так как $\pi(n^0) = 1$.

Теперь покажем, что условия (3.2.5) и (3.2.6) выполняются, если в качестве процедуры распределения дележа $y(x^0)$ выбран набор выплат, удовлетворяющих уравнениям (3.2.8) и (3.2.9). Для терминальных вершин условие (3.2.6) очевидным образом выполнено. Подробнее рассмотрим условие (3.2.5) для произвольной нетерминальной вершины n_l^t :

$$\begin{aligned} \sum_{j \in M} \beta_j(x^*(n_l^t)) &= \sum_{j \in M} y_j(x^*(n_l^t)) - \sum_{j \in M} \lambda_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \pi(n_k^{t+1}|n_l^t) y_j(x^*(n_k^{t+1})) = \\ &= V(M; x^*(n_l^t)) - \left(\sum_{j \in M} W_j(\mathbf{u}^*(x^*), x^*(n_l^t)) - \sum_{j \in M} \phi_j^{n_l^t}(x^*(n_l^t), \mathbf{u}^*(x^*)) \right). \end{aligned}$$

Принимая во внимание, что $V(M; x^*(n_l^t)) = \sum_{j \in M} W_j(\mathbf{u}^*(x^*), x^*(n_l^t))$ и метод построения характеристической функции, мы получаем равенство (3.2.5).

Если c -ядро $C(x^0)$ игры и c -ядра $C(x^*(n_l^t))$ всех ее подыгр, начинающихся в вершинах $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$, непусты, мы всегда можем найти хотя бы один дележ $y(x^*(n_l^t)) \in C(x^*(n_l^t))$ и используя дележи $y(x^*(n_l^t))$ для всех вершин $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T$, мы можем найти процедуру распределения дележа $(\{\beta_j(x^*(n_l^t))\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M)$, используя формулы (3.2.8) и (3.2.9) для любого дележа из c -ядра $C(x^0)$. Доказательство того, что такая процедура распределения дележа позиционно состоятельна, т. е. того, что она удовлетворяет условию (3.2.7), может быть легко получено, если подставить $\beta_j(x^*(n_k^\theta))$ из (3.2.8) и (3.2.9) в равенства (3.2.7) и, учитывая $\pi(n^0) = 1$ и $\pi(n_l^t)\pi(n_k^{t+1}|n_l^t) = \pi(n_k^{t+1})$ для любой вершины $n_k^{t+1} \in \mathcal{N}^t$, $t = 0, \dots, T$:

$$\begin{aligned} & \sum_{\theta=0}^{t-1} \lambda_j^\theta \sum_{n_k^\theta \in \mathcal{N}^\theta} \pi(n_k^\theta) \beta_j(x^*(n_k^\theta)) + \lambda_j^t \sum_{n_k^t \in \mathcal{N}^t} \pi(n_k^t) y_j(x^*(n_l^t)) = \pi(n^0) y_j(x^0) \\ & - \lambda_j \sum_{n_k^1 \in \mathcal{S}(n^0)} \pi(n^0) \pi(n_k^1 | n^0) y_j(x^*(n_k^1)) + \lambda_j \sum_{n_k^1 \in \mathcal{S}(n^0)} \pi(n_k^1) y_j(x^*(n_k^1)) \\ & - \dots - \lambda_j^t \sum_{n_k^t \in \mathcal{N}^t} \pi(n_k^{t-1}) \pi(n_k^t | n_k^{t-1}) y_j(x^*(n_l^t)) + \lambda_j^t \sum_{n_k^t \in \mathcal{N}^t} \pi(n_k^t) y_j(x^*(n_l^t)) \\ & = y_j(x^0). \end{aligned}$$

Последнее равенство позволяет получить равенство (3.2.7), также учитывая, что $\pi(n^0) = 1$ и $\pi(n_l^t)\pi(n_k^{t+1}|n_l^t) = \pi(n_k^{t+1})$ для любой вершины $n_k^{t+1} \in \mathcal{N}^t$, $t = 0, \dots, T$. \square

Реализация ПРД $(\{\beta_j(x^*(n_l^t))\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M)$ требует определения трансферных платежей между игроками в каждой вершине $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$ в соответствии с правилом:

$$\omega_j(n_l^t, x^*(n_l^t)) = \beta_j(x^*(n_l^t)) - \phi_j^{n_l^t}(x^*(n_l^t), u^*(x^*(n_l^t))), \quad (3.2.10)$$

и для всех вершин $n_l^T \in \mathcal{N}^T$:

$$\omega_j(n_l^T, x^*(n_l^T)) = \beta_j(x^*(n_l^T)) - \Phi_j^{n_l^T}(x^*(n_l^T)), \quad (3.2.11)$$

где $\omega_j(n_l^t, x^*(n_l^t))$ — трансферный платеж, который делает игрок j в вершине n_l^t при реализации кооперативной траектории $x^*(n_l^t)$, такой, что

$$\sum_{j \in M} \omega_j(n_l^t, x^*(n_l^t)) = 0$$

для любой вершины n_i^t кооперативной траектории $x^*(n_i^t)$. Естественно, трансферная выплата $\omega_j(n_i^t, x^*(n_i^t))$ может быть как положительной, так и отрицательной, так и равняться нулю. Это зависит от знаков в правых частях выражений (3.2.10)-(3.2.11).

К изложенному в этом разделе имеется два комментария:

1. В работе не затрагивается проблема выбора некоторого дележа из c -ядра в случае, когда оно содержит более одного дележа. Проблема выбора дележа изучается в области классических кооперативных игр (например, см. [100]). Для демонстрации теоретических результатов в следующей секции мы рассмотрим пример, где в качестве дележа из c -ядра в каждой вершине выбирается n -ядро.¹ [206]. Известно, что n -ядро единственно, и если c -ядро непусто, то n -ядро принадлежит c -ядру (см. [176]).
2. Мы предполагаем, что игроки используют γ -подход для построения характеристической функции и, следовательно, вычисляют так называемое γ -ядро. Теоретические результаты этой главы остаются справедливыми при построении α - или β -характеристических функций, вычисляя α - и β -ядра соответственно.

§ 3.3 Позиционно состоятельное c -ядро в игре совместного управления загрязнением окружающей среды

Продемонстрируем процедуру построения позиционно состоятельного (динамически устойчивого) c -ядра на примере задачи сокращения выбросов, которая представляет стохастическую версию детерминированной модели с тремя участниками, представленной в статье [98]. Обозначим через $M = \{1, 2, 3\}$ множество игроков и через $\mathcal{T} = \{0, 1, 2\}$ — множество периодов. Пусть $u(n_i^t) = (u_1(n_i^t), u_2(n_i^t), u_3(n_i^t))$ — вектор выбросов некоторого загрязняющего вещества

¹Для определения n -ядра, требуется ввести понятие эксцесса коалиции G . Это разница между суммой выигрышей игроков этой коалиции согласно некоторому дележу и значением характеристической функции этой коалиции. Если мы расставим все эксцессы в порядке возрастания от меньшего в большему и найдем минимальный вектор на множестве дележей, то получим n -ядро [176].

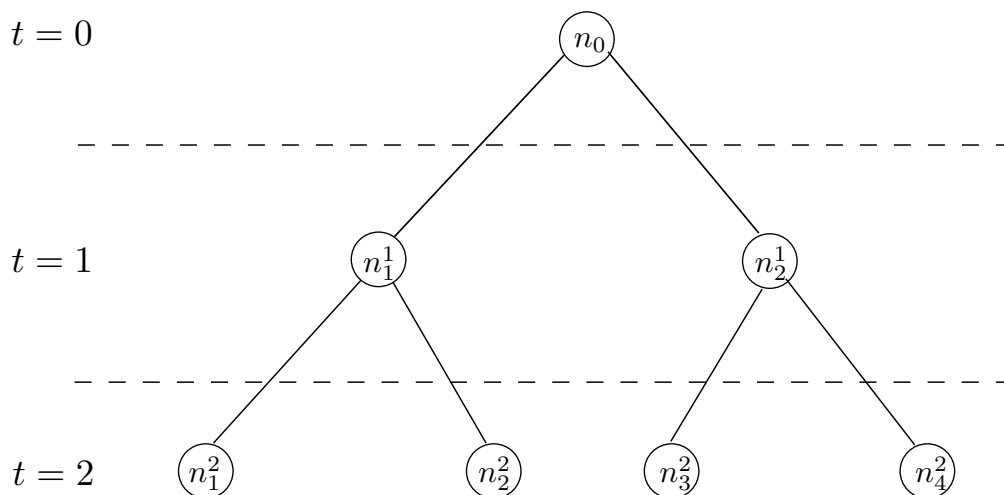


Рис. 3.1: Дерево событий.

в вершине n_i^t в период t , и $x(n_i^t)$ — «запас» загрязняющего вещества в этой вершине. Изменение этого запаса для нетерминальных вершин n_i^t описывается разностным уравнением:

$$x(n_i^t) = (1 - \delta(n_i^t))x(a(n_i^t)) + \sum_{j \in M} u_j(n_i^t), \quad (3.3.1)$$

для начальной вершины n^0 :

$$x(n^0) = (1 - \delta(n^0))x_{-0} + \sum_{j \in M} u_j(n^0), \quad (3.3.2)$$

где x_{-0} задано, $\delta(n_i^t)$ ($0 < \delta(n_i^t) < 1$) случайный уровень абсорбции загрязняющего вещества природой в вершине n_i^t . Мы предполагаем, что $\delta(n_i^t)$ может принимать два значения: $\delta(n_i^t) \in \{\underline{d}, \bar{d}\}$, где $\underline{d} < \bar{d}$. Дерево событий изображено на рис. 3.1. Пусть вершины n^0, n_1^1, n_2^1, n_3^2 соответствуют низкому уровню абсорбции загрязняющего вещества \underline{d} , а вершины n_2^1, n_2^2, n_4^2 соответствуют высокому уровню абсорбции загрязняющего вещества \bar{d} .

Издержки загрязнения — возрастающая выпуклая функция запаса загрязняющего вещества. Предположим, что она имеет вид: $D_j(x(n_i^t)) = \alpha_j x^2(n_i^t)$, $j \in M$, где α_j — положительная константа. Издержки от совершения выбросов — квадратичная функция $C_j(u_j(n_i^t)) = \frac{\gamma_j}{2} (u_j(n_i^t) - e)^2$, где e и γ_j — положительные константы.

Общие дисконтированные издержки, которые минимизирует игрок $j \in M$,

заданы функцией $J_j(\mathbf{x}, \mathbf{u})$, которая равна

$$\sum_{t=0}^{T-1} \lambda_j^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) (C_j(u_j(n_i^t)) + D_j(x(n_i^t))) + \lambda_j^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) D_j(x(n_i^T)),$$

где $\mathbf{x} = \{x(n_i^t)\}$ and $\mathbf{u} = \{\mathbf{u}(n_i^t)\}$, $n_i^t \in \mathcal{N}^t$, $\lambda_j \in (0, 1)$ — дисконт фактор игрока j . При этом дано уравнение (3.3.1) с начальным запасом $x_{-0} = 0$ до начала игры, а также имеются ограничения: $u_j(n_i^t) \in [0, e]$ для любого игрока $j \in M$ и любой вершины $n_i^t \in \mathcal{N}^T$.

Используем следующие параметры для численного примера:

$$\begin{aligned} \alpha_1 &= 1, \alpha_2 = 1.5, \alpha_3 = 1.8, \\ \gamma_1 &= 2, \gamma_2 = 1.5, \gamma_3 = 1.4, \\ \underline{\delta} &= 0.15, \bar{\delta} = 0.45, e = 5, \lambda_1 = \lambda_2 = \lambda_3 = 0.9, \\ \pi(n_1^1) &= 0.3, \pi(n_2^1) = 0.7, \pi(n_1^2) = 0.2, \\ \pi(n_2^2) &= 0.1, \pi(n_3^2) = 0.3, \pi(n_4^2) = 0.4. \end{aligned}$$

Используя (3.1.7) и (3.1.9), вычислим кооперативные стратегии для всех возможных подыгр и всей игры. Получаем следующую кооперативную траекторию²: $(x^*(n^0), x^*(n_1^1), x^*(n_2^1), x^*(n_1^2), x^*(n_2^2), x^*(n_3^2), x^*(n_4^2)) = (0.44621, 0.71208, 0.73186, 0.60527, 0.39165, 0.62208, 0.40252)$.

Значения γ -характеристической функции даны в табл. 3.1. Нетрудно заметить, что построенная функция для любой подыгры является субаддитивной, что не является выполнимым в общем случае при построении γ -характеристической функции.

Непустые s -ядра всей игры и всех подыгр, которые рассчитаны по формулам (3.2.2) и (3.2.3), представлены в табл. 3.2. Заметим, что неравенства в (3.2.2) и (3.2.3) нужно заменить на противоположные, потому что мы рассматриваем задачу минимизации затрат в отличие от задачи максимизации прибыли, решаемой ранее. Предполагая, что игроки договорились использовать n -ядро во всех подыграх для построения динамически устойчивой процедуры распределения дележа. Компоненты n -ядер для всех подыгр представлены в табл. 3.3.

²Вычисления этого раздела были произведены в программе Maple с использованием пакетов *Optimization* и *Simplex* [135].

Период	$t = 0$	$t = 1$		$t = 2$			
Вершина	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2
$V(\{1\})$	33.51298	23.60711	19.89234	4.43294	1.85601	4.11655	1.72354
$V(\{2\})$	50.39239	32.77819	31.00779	6.64941	2.78401	6.17483	2.58531
$V(\{3\})$	50.97086	34.33383	32.20934	7.97930	3.340813	7.40980	3.10237
$V(\{1, 2\})$	76.31863	43.69330	41.51454	2.45773	1.02901	2.50077	1.04703
$V(\{1, 3\})$	74.57715	42.49885	40.49344	2.24939	0.94178	2.30923	0.96684
$V(\{2, 3\})$	101.36325	67.11203	63.21713	14.62871	6.12482	13.58463	5.68768
$V(\{1, 2, 3\})$	112.16237	61.35616	59.92551	1.57532	0.65956	1.66402	0.69670

Таблица 3.1: Характеристические функции для всех подыгр и игры в целом.

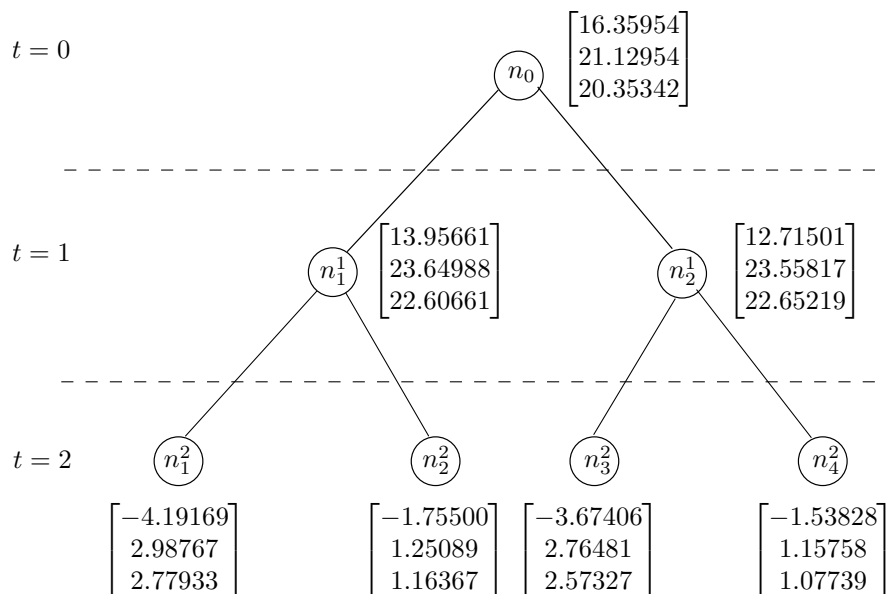


Рис. 3.2: Процедура распределения дележа.

На основе n -ядра мы определим позиционно состоятельный дележ, используя уравнения (3.2.8) и (3.2.9) из теоремы 3.1. В результате, мы получаем процедуру распределения дележа $\beta_j(x^*(n_l^t))$, $j \in M$, которая представлена на рис. 3.2. Возле каждой вершины располагается соответствующий вектор ПРД $\beta(\cdot)$.

Для пояснения результата рассмотрим выигрыш первого игрока. В соответствии с n -ядром первый игрок получает выигрыш 25.92624 во всей игре (см. Табл. 3.3). Процедура распределения дележа предписывает первому игроку выплаты в каждой вершине таким образом, чтобы их ожидаемая сумма равнялась

n^0
$\{y = (y_1, y_2, y_3) : y_3 = 112.162 - y_1 - y_2,$ $61.192 \leq y_1 + y_2 \leq 76.319, 10.799 \leq y_1 \leq$ $\leq 33.513, 37.585 \leq y_2 \leq 50.392\}$
n_1^1
$\{y = (y_1, y_2, y_3) : y_3 = 61.356 - y_1 - y_2,$ $27.022 \leq y_1 + y_2 \leq 43.693, -5.756 \leq y_1 \leq$ $\leq 23.607, 18.857 \leq y_2 \leq 32.778\}$
n_2^1
$\{y = (y_1, y_2, y_3) : y_3 = 59.926 - y_1 - y_2,$ $27.716 \leq y_1 + y_2 \leq 41.515, -3.291 \leq y_1 \leq$ $\leq 19.892, 19.432 \leq y_2 \leq 31.008\}$
n_1^2
$\{y = (y_1, y_2, y_3) : y_3 = 1.575 - y_1 - y_2,$ $-6.404 \leq y_1 + y_2 \leq 2.458, -13.05 \leq y_1 \leq$ $\leq 4.433, -0.674 \leq y_2 \leq 6.649\}$
n_2^2
$\{y = (y_1, y_2, y_3) : y_3 = 0.660 - y_1 - y_2,$ $-2.681 \leq y_1 + y_2 \leq 1.029, -5.465 \leq y_1 \leq$ $\leq 1.856, -0.282 \leq y_2 \leq 2.784\}$
n_3^2
$\{y = (y_1, y_2, y_3) : y_3 = 1.664 - y_1 - y_2,$ $-5.746 \leq y_1 + y_2 \leq 2.501, -11.92 \leq y_1 \leq$ $\leq 4.117, -0.645 \leq y_2 \leq 6.175\}$
n_4^2
$\{y = (y_1, y_2, y_3) : y_3 = 0.697 - y_1 - y_2,$ $-2.406 \leq y_1 + y_2 \leq 1.047, -4.991 \leq y_1 \leq$ $\leq 1.724, -0.270 \leq y_2 \leq 2.585\}$

Таблица 3.2: C -ядра.

$t = 0$			$t = 1$		
n^0			n_1^1	n_2^1	
25.926			10.915		10.507
43.989			25.818		25.220
42.247			24.623		24.199
$t = 2$					
n_1^2	n_2^2	n_3^2	n_4^2		
-4.192	-1.76	-3.67			-1.54
2.988	1.251	2.765			1.158
2.779	1.164	2.573			1.077

Таблица 3.3: N -ядра для игры и всех подыгр.

25.92624, а именно:

$$\begin{aligned}
25.92624 = & 16.35954 + 0.3 \cdot 0.9 \cdot 13.95661 + 0.7 \cdot 0.9 \cdot 12.71501 \\
& - 0.2 \cdot (0.9)^2 \cdot 4.19169 - 0.1 \cdot (0.9)^2 \cdot 1.75500 \\
& - 0.3 \cdot (0.9)^2 \cdot 3.67406 - 0.4 \cdot (0.9)^2 \cdot 1.53828.
\end{aligned}$$

Подобным же образом, в каждой подыгре сумма ожидаемых выплат β_j должна равняться соответствующей компоненте n -ядра. Например, в вершине n_1^1 игрок 1 получает 10.91510 в соответствии с n -ядром этой подыгры. Несложно проверить, что это число равняется ожидаемой сумме компонент процедуры распределения дележа, которые он получит в подыгре, начинающейся из вершины n_1^1 , а именно:

$$10.91510 = 13.95661 - (2/3 \cdot 0.9 \cdot 4.19169) - (1/3 \cdot 0.9 \cdot 1.75500),$$

где $2/3$ и $1/3$ вероятность реализации вершин n_1^2 и n_2^2 из вершины n_1^1 соответственно. Отметим, что процедура распределения дележа β_j является позиционно состоятельной и позволяет реализовать дележ (в нашем случае это n -ядро) в динамике. В качестве первоначального дележа из s -ядра может быть выбран любой другой дележ, процедура распределения может быть построена аналогичным образом. Тем самым мы показали позиционную состоятельность s -ядра при условии реализации дележей с помощью ПРД, найденной по формулам (3.2.8) и (3.2.9).

§ 3.4 Абсолютное ε -равновесие в динамических играх, разыгрываемых на деревьях событий

В этом параграфе рассмотрим игру, разыгрываемую на дереве событий, описанную в параграфе § 3.1. Поскольку эта игра имеет конечную продолжительность, то в общем случае нельзя гарантировать существование абсолютного равновесия по Нэшу с кооперативными выигрышами. Поэтому проверим условия существования приближенного или ε -равновесия (см. [128, 141]) для рассматриваемого класса игр.

Мы можем рассматривать игры на деревьях событий как игры в развернутой форме с полной информацией. Это означает, что каждый игрок знает не только текущую вершину $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T$ дерева событий и стратегии, который он реализовал на пути от начальной вершины n^0 до вершины $a(n_l^t)$, а также и стратегии всех остальных игроков во всех предыдущих периодах. Пусть этот путь будет следующим: $(n^0, n_{i_1}^1, \dots, n_{i_{t-1}}^{t-1}, n_l^t)$, обозначим его через $P(n_l^t)$. Набор реализованных ситуаций на пути $P(a(n_l^t))$ ситуаций назовем *историей вершины* n_l^t и обозначим её через

$$H(n_l^t) = ((n^0, u(n^0)), (n_{i_1}^1, u(n_{i_1}^1)), \dots, (n_{i_{t-1}}^{t-1}, u(n_{i_{t-1}}^{t-1}))).$$

Определение 3.7. *Стратегия поведения $\sigma_j = \{\sigma_j(n_l^t) : n_l^t \in \mathcal{N}^t, t = 0, \dots, T-1\}$ игрока $j \in M$ в игре G , разыгрываемой на дереве событий, — отображение, которое ставит в соответствие каждой вершине n_l^t и истории $H(n_l^t)$ стратегию $u_j(n_l^t) \in U_j^{n_l^t}$, т. е.*

$$\sigma_j(n_l^t) : H(n_l^t) \longrightarrow U_j^{n_l^t}.$$

Другими словами, стратегия поведения предписывает игроку, какую стратегию выбрать в каждой вершине n_l^t дерева событий в зависимости от истории этой вершины. Обозначим через Σ_j множество стратегий поведения игрока j в игре G , через $\sigma = (\sigma_1, \dots, \sigma_m)$ ситуацию в стратегиях поведения и через $\Sigma = \Sigma_1 \times \dots \times \Sigma_m$ множество ситуаций в стратегиях поведения. Для каждой ситуации в стратегиях поведения можно посчитать ожидаемый выигрыш любого игрока в любой подыгре, включая игру в целом. Для того, чтобы избежать

введения новых обозначений, мы обозначим выигрыш игрока j в подыгре, начинающейся из вершины n_i^t из состояния $x(n_i^t)$, как функцию ситуации в стратегиях поведения через

$$W_j(\sigma, x(n_i^t)) = W_j(\mathbf{u}(n_i^t), x(n_i^t)),$$

где $\mathbf{u}(n_i^t)$ — набор стратегий в подыгре, начинающейся из вершины n_i^t , который единственным образом определяется ситуацией σ .

Определение 3.8. Ситуация в стратегиях поведения $\hat{\sigma}$ является ε -равновесием, если для любого игрока $j \in M$ и любой стратегии $\sigma_j \in \Sigma_j$, справедливо неравенство:

$$W_j(\hat{\sigma}, x^0) \geq W_j((\hat{\sigma}_{-j}, \sigma_j), x^0) - \varepsilon. \quad (3.4.1)$$

Определение 3.9. Ситуация $\hat{\sigma}$ является абсолютным ε -равновесием, если для любого игрока $j \in M$, любой вершины $n_i^t \in \mathcal{N}^t$, любой стратегии $\sigma_j \in \Sigma_j$ и любой истории $H(n_i^t)$, справедливо неравенство:

$$W_j(\hat{\sigma}, x(n_i^t)|H(n_i^t)) \geq W_j((\hat{\sigma}_{-j}, \sigma_j), x(n_i^t)|H(n_i^t)) - \varepsilon,$$

где $W_j(\hat{\sigma}, x(n_i^t)|H(n_i^t))$ — выигрыш игрока j в подыгре, начинающейся в вершине n_i^t из состояния $x(n_i^t)$ с историей $H(n_i^t)$, когда игроки реализуют ситуацию в стратегиях поведения $\hat{\sigma}$.

Для того, чтобы стратегически поддержать кооперацию в динамической игре с конечной продолжительностью, разыгрываемой на дереве событий, мы построим приближенное равновесие в стратегиях поведения. Идея построения приближенного равновесия близка к тому, что сделано для повторяющихся игр с конечной продолжительностью. Отличие модели данной главы от выше упомянутой теории заключается в том, что мы дополнительно рассматриваем вектор состояния, изменения которого заданы уравнениями (3.1.5) и (3.1.6). Прежде чем строго определить стратегию поведения, сделаем два предположения:

1. Игроки решили реализовать кооперативную ситуацию \mathbf{u}^* , определяемую условием (3.1.7).

2. Если игрок j отклоняется от кооперативной стратегии $u_j^*(a(n_l^t))$ в вершине $a(n_l^t)$ и играет стратегию $u_j(a(n_l^t)) \neq u_j^*(a(n_l^t))$, тогда кооперация рушится, и все игроки начинают играть равновесные по Нэшу стратегии в подыгре, начинающейся из вершины n_l^t из состояния $x(n_l^t) = f^{a(n_l^t)}(x^*(a(n_l^t)), (u_{-j}^*(a(n_l^t)), u_j(a(n_l^t))))$.

Если кооперация распадается, выигрыш игрока $p \in M$ в подыгре, начинающейся из вершины n_l^t из состояния $x(n_l^t)$, равен $W_p(\mathbf{u}^N(x(n_l^t)), x(n_l^t))$, это есть выигрыш игрока p , который он может получить, если все игроки играют равновесные по Нэшу стратегии. Обозначим через $\hat{\sigma} = (\hat{\sigma}_p : p \in M)$ ситуацию в стратегиях поведения, которая предписывает игроку $p \neq j$ играть кооперативную стратегию $u_p^*(n_l^t)$ в вершине n_l^t , если история этой вершины не содержит отклонений от кооперативной траектории, и играть $u_p^N(n_l^t)$ в противном случае. Обозначим через $\hat{\mathbf{u}}(x(n_l^t)) = \{\hat{u}(n_l^\theta) : n_l^\theta \in \Gamma(n_l^t)\}$ набор стратегий, соответствующих ситуации $\hat{\sigma}$, в соответствии с которой «режим наказания» этой стратегии используется в подыгре, начинающейся из вершины n_l^t из состояния $x(n_l^t)$. Если в истории вершины n_l^θ наблюдались индивидуальные отклонения игрока j , мы обозначим набор стратегий, соответствующих ситуации $\hat{\sigma}$ через $\hat{\mathbf{u}}^j(n_l^\theta)$.

Стратегия наказания любого игрока состоит из двух типов поведения или двух режимов:

Номинальный режим. Если история вершины n_l^t совпадает с историей

$$H^*(n_l^t) = ((n^0, u^*(n^0)), (n_{i_1}^1, u^*(n_{i_1}^1)), \dots, (n_{i_{t-1}}^{t-1}, u^*(n_{i_{t-1}}^{t-1}))), \quad (3.4.2)$$

когда все игроки играют свои кооперативные стратегии на пути $P(n_{i_{t-1}}^{t-1})$ от вершины n^0 до вершины $n_{i_{t-1}}^{t-1}$, тогда игрок $p \in M$ играет стратегию $u_p^*(n_l^t)$ в вершине n_l^t .

Режим наказания. Если история вершины n_l^t такова, что существует вершина n на пути $P(n_{i_{t-1}}^{t-1})$ такая, что $u(n) \neq u^*(n)$, тогда стратегия игрока p — равновесная по Нэшу стратегия, рассчитанная для подыгры, начинающейся из вершины, следующей за n и соответствующего состояния. Здесь история вершины n_l^t такая, что существует вершина n и по крайней мере

один отклонившийся игрок $j \in M$, $j \neq p$, т.е. история $H(n)$ вершины n — часть истории $H^*(n_l^t)$, а пара $(n, u(n))$ не является частью истории $H^*(n_l^t)$. Но если заменить стратегию $u_j(n)$ игрока j в вершине n кооперативной стратегией $u_j^*(n)$, тогда пара $(n, (u_{-j}(n), u_j^*(n)))$ станет $(n, u^*(n))$ и будет частью истории $H^*(n_l^t)$.

Строго говоря, стратегия наказания игрока $p \in M$ определяется следующим образом:

$$\hat{\sigma}_p(H(n_l^t)) = \begin{cases} u_p^*(n_l^t), & \text{если } H(n_l^t) = H^*(n_l^t), \\ \hat{u}_p(n_l^t), & \text{если } \exists \text{ вершина } n \text{ на пути } P(n_l^t) \\ & \text{такая, что } u(n) \neq u^*(n), \end{cases} \quad (3.4.3)$$

где $\hat{u}_p(n_l^t)$ — стратегия игрока p в вершине n_l^t . Стратегия $\hat{u}_p(n_l^t)$ является реализацией стратегии наказания в подыгре, начинающейся из единственной вершины, принадлежащей множеству $\mathcal{S}(n) \cap P(n_l^t)$. Стратегия $\hat{u}_p(n_l^t)$ совпадает с $u_p^N(n_l^t)$, рассчитанной как часть равновесия по Нэшу в подыгре, начинающейся из вершины $n_1 = \mathcal{S}(n) \cap P(n_l^t)$ из состояния $x(n_l^t)$.

Для того, чтобы избежать усложнения обозначений, опустим аргумент состояния в обозначении стратегии наказания, но заметим, что они очевидным образом зависят от него. Пусть вершина n_1 следует сразу после вершины n , в которой игрок j отклоняется. Набор стратегий $(u_{-j}^*(n), u_j(n))$ реализуется, и состояние в вершине n_1 может быть вычислено с помощью уравнения $x(n_1) = f^n(x^*(n), (u_{-j}^*(n), u_j(n)))$. Стратегия $\hat{u}_p(n_1)$ — часть набора $\hat{u}(n_1) = (u_p^N(n_1) : p \in M)$, где $u_p^N(n_1)$ — стратегия игрока p в вершине n_1 . Теперь в подыгре, начинающейся из вершины $n_l^t \in \mathcal{N}^t$ и состояния $x(n_l^t)$, набор стратегий, наказывающих индивидуальное отклонение игрока j , определяется следующим образом:

$$\hat{\mathbf{u}}^j(n_l^t) = (\hat{u}_p^j(n_l^t) : p \in M),$$

где $\hat{\mathbf{u}}_p^j(n_l^t) = \{\hat{u}_p^j(n_l^\theta) : n_l^\theta \in \Gamma(n_l^t)\}$. Этот набор стратегий порождает траекторию состояний, соответствующих равновесию по Нэшу в этой подыгре:

$$\hat{\mathbf{x}}^j(n_l^t) = \{\hat{x}^j(n_l^\theta) : n_l^\theta \in \Gamma(n_l^t)\}.$$

Для того чтобы определить стратегии наказания, найдем m ситуаций для каждой подыгры, в каждой из которых наказывается за отклонение один из игроков.

Теорема 3.2. *Для любого $\varepsilon \geq \tilde{\varepsilon}$ в динамической игре, разыгрываемой на дереве событий, существует абсолютное ε -равновесие в стратегиях наказания с выигрышами игроков $(W_1(\mathbf{u}^*, x^0), \dots, W_m(\mathbf{u}^*, x^0))$, и*

$$\tilde{\varepsilon} = \max_{j \in M} \max_{\substack{n_i^t \in \mathcal{N}^t \\ t=0, \dots, T-1}} \varepsilon_j(n_i^t), \quad (3.4.4)$$

где

$$\begin{aligned} \varepsilon_j(n_i^t) = & \max_{u_j(n_i^t) \in U_j^{n_i^t}} \left\{ \phi_j^{n_i^t}(x^*(n_i^t), (u_{-j}^*(n_i^t), u_j(n_i^t))) - \phi_j^{n_i^t}(x^*(n_i^t), u^*(n_i^t)) + \right. \\ & + \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_i^\theta | n_i^t) \left(\phi_j^{n_i^\theta}(\hat{x}^j(n_i^\theta), \hat{u}^j(n_i^\theta)) - \phi_j^{n_i^\theta}(x^*(n_i^\theta), u^*(n_i^\theta)) \right) + \\ & \left. + \lambda_j^{T-t} \sum_{n_i^T \in \mathcal{N}_\Gamma^T} \pi(n_i^T | n_i^t) \left(\Phi_j^{n_i^T}(\hat{x}(n_i^T)) - \Phi_j^{n_i^T}(x^*(n_i^T)) \right) \right\}, \quad (3.4.5) \end{aligned}$$

и $\hat{u}^j(n_i^\theta)$ — набор стратегий в вершине n_i^θ , соответствующий стратегии поведения $\hat{\sigma}$, определяемой формулой (3.4.3), и когда «режим наказания» стратегии начинается в подыгре, начинающейся из вершины, принадлежащей множеству $\mathcal{S}(n_i^t)$ из состояния $f^{n_i^t}(x(n_i^t), (u_{-j}^*(n_i^t), u_j(n_i^t)))$. Таким образом, разности во второй и третьей строке равенства (3.4.5) также зависят от стратегии $u_j(n_i^t)$. Состояние $\hat{x}^j(n_i^\theta)$, $n_i^\theta \in \Gamma(n_i^\theta)$, — получается единственным образом из уравнения состояния при наборе стратегий $\hat{\mathbf{u}}^j(n_i^t)$.

Доказательство. Рассмотрим стратегию наказания $\hat{\sigma} = (\hat{\sigma}_p : p \in M)$, определенную в (3.4.3), и подыгру, начинающуюся из произвольной вершины $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$. Рассмотрим все различные истории вершины n_i^t и вычислим выигрыш игрока j при индивидуальном отклонении в вершине n_i^t . Его выигрыш в этой подыгре, если все игроки придерживаются кооперативных стратегий,

будет равен

$$\begin{aligned}
W_j(\mathbf{u}^*(n_l^t), x^*(n_l^t)) &= \phi_j^{n_l^t}(x^*(n_l^t), u^*(n_l^t)) + \\
&+ \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \phi_j^{n_l^\theta}(x^*(n_l^\theta), u^*(n_l^\theta)) + \\
&+ \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \Phi_j^{n_l^T}(x^*(n_l^T)),
\end{aligned} \tag{3.4.6}$$

где $\mathcal{N}_\Gamma^\theta = \mathcal{N}^\theta \cap \Gamma(n_l^t)$, $\mathbf{u}^*(n_l^t) = (\mathbf{u}_j^*(n_l^t) : j \in M)$ — S -адаптивная кооперативная ситуация.

Пусть сначала история вершины n_l^t будет $H^*(n_l^t)$. Предположим, что игрок j отклоняется в вершине n_l^t от кооперативной траектории. В этом случае, он может выиграть следующий выигрыш в подыгре, начинающейся в вершине n_l^t , при условии, что игроки реализуют ситуацию $\hat{\sigma} = (\hat{\sigma}_p(\cdot) : p \in M)$, определяемую равенством (3.4.3):

$$\begin{aligned}
&\max_{u_j(n_l^t) \in U_j^{n_l^t}} \left\{ \phi_j^{n_l^t}(x^*(n_l^t), (u_{-j}^*(n_l^t), u_j(n_l^t))) + \right. \\
&+ \left. \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \phi_j^{n_l^\theta}(\hat{x}^j(n_l^\theta), \hat{u}^j(n_l^\theta)) + \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \Phi_j^{n_l^T}(\hat{x}^j(n_l^T)) \right\},
\end{aligned} \tag{3.4.7}$$

где равновесная по Нэшу стратегия (наказание для игрока j) начинает реализовываться игроками в вершинах множества $\mathcal{S}(n_l^t)$. Мы можем посчитать прибыль игрока j от отклонения в вершине n_l^t как разницу между (3.4.7) и (3.4.6), а именно:

$$\begin{aligned}
\varepsilon_j(n_l^t) &= \max_{u_j(n_l^t) \in U_j^{n_l^t}} \left\{ \phi_j^{n_l^t}(x^*(n_l^t), (u_{-j}^*(n_l^t), u_j(n_l^t))) - \phi_j^{n_l^t}(x^*(n_l^t), u^*(n_l^t)) + \right. \\
&+ \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \left(\phi_j^{n_l^\theta}(\hat{x}^j(n_l^\theta), \hat{u}^j(n_l^\theta)) - \phi_j^{n_l^\theta}(x^*(n_l^\theta), u^*(n_l^\theta)) \right) + \\
&+ \left. \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \left(\Phi_j^{n_l^T}(\hat{x}(n_l^T)) - \Phi_j^{n_l^T}(x^*(n_l^T)) \right) \right\},
\end{aligned} \tag{3.4.8}$$

Теперь предположим, что история вершины n_l^t не совпадает с историей $H^*(n_l^t)$. Это означает, что все игроки уже сменили «номинальный режим» на «режим

наказания». Игрок j не может получить прибыль от отклонения в вершине n_i^t , потому что игроки реализуют равновесные по Нэшу стратегии независимо от того, какой игрок (или группа игроков) отклонился в предыдущих вершинах.

Вычисляя максимальный выигрыш от отклонения для любой подыгры и любого игрока по формуле (3.4.8), мы получим значение $\tilde{\varepsilon}$ из формулировки теоремы, которое есть

$$\tilde{\varepsilon} = \max_{j \in M} \max_{\substack{n_i^t \in \mathcal{N}^t \\ t=0, \dots, T-1}} \varepsilon_j(n_i^t).$$

Для любого $\varepsilon \geq \tilde{\varepsilon}$ ситуация в стратегиях поведения, определяемая формулой (3.4.3), является ситуацией абсолютного ε -равновесия по построению. \square

Замечание 3.3. Величина $\tilde{\varepsilon}$ — максимальный выигрыш, который какой-либо игрок может достичь в игре, отклоняясь от кооперации. Число $\tilde{\varepsilon}$ зависит от функций выигрышей игроков, а также от множества других параметров игры, например, от количества игроков и продолжительности игры. Поэтому, непростой является задача нахождения верхней границы $\tilde{\varepsilon}$. Тем не менее, может быть предложен некоторый способ нахождения оценки этой величины в относительном измерении. Один способ — сравнить $\tilde{\varepsilon}$ с суммарным выигрышем игроков при кооперации, рассчитав так называемую «цену кооперации» или *CoC* (*cost of cooperation*):

$$\text{CoC} = \frac{\tilde{\varepsilon}}{\sum_{j \in M} W_j(\mathbf{u}^*, x^0)}.$$

§ 3.5 Построение абсолютного ε -равновесия в игре совместного управления загрязнением окружающей среды

Для иллюстрации теоретических результатов предыдущего параграфа рассмотрим модель совместного загрязнения странами некоторой территории [98]. Пусть $M = \{1, 2, 3\}$ — множество игроков, $\mathcal{T} = \{0, 1, 2, 3\}$ — множество периодов игры, $u(n_i^t) = (u_1(n_i^t), u_2(n_i^t), u_3(n_i^t))$ — вектор выбросов некоторого загрязняющего вещества в вершине n_i^t , $x(n_i^t)$ — «запас» загрязняющего вещества в вершине

n_i^t в период t . Динамика состояния x задается уравнением:

$$x(n_i^t) = (1 - \delta(a(n_i^t)))x(a(n_i^t)) + \sum_{j \in M} u_j(a(n_i^t)), \quad (3.5.1)$$

$$x(n^0) = x^0, \quad (3.5.2)$$

с заданным начальным запасом x^0 в вершине n^0 , и $\delta(n_i^t)$ ($0 < \delta(n_i^t) < 1$) случайный уровень абсорбции природой загрязняющего вещества в вершине n_i^t . Предполагаем, что $\delta(n_i^t)$ может принимать значения из множества $\{\underline{\delta}, \bar{\delta}\}$, где $\underline{\delta} < \bar{\delta}$. Дерево событий представлено на рис. 3.3. Пусть вершины n_2^1, n_2^2, n_4^2 соответствуют низкому уровню абсорбции $\underline{\delta}$, и вершины n^0, n_1^1, n_1^2, n_3^2 — высокому уровню абсорбции $\bar{\delta}$.

В каждой вершине $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, 3$ задана функция издержек игрока $j \in M$, являющаяся суммой двух функций: 1) $D_j(x(n_i^t)) = \alpha_j x^2(n_i^t)$, где $\alpha_j > 0$; 2) $C_j(u_j(n_i^t)) = \frac{\gamma_j}{2} (u_j(n_i^t) - e)^2$, где e и γ_j — положительные константы.

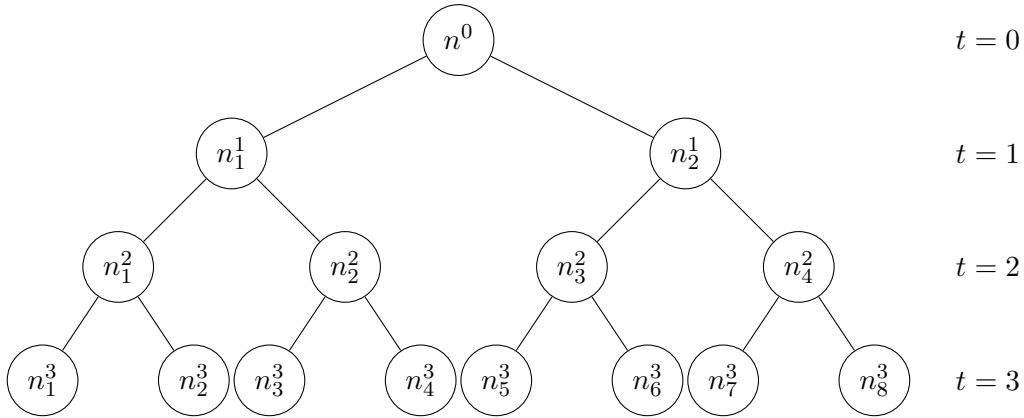


Рис. 3.3: Дерево событий для игры с $T = 3$.

Дисконтированные суммарные издержки $J_j(\mathbf{x}, \mathbf{u})$ игрока $j \in M$, которые он минимизирует, заданы формулой:

$$\sum_{t=0}^2 \lambda_j^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) (C_j(u_j(n_i^t)) + D_j(x(n_i^t))) + \lambda_j^3 \sum_{n_i^3 \in \mathcal{N}^3} \pi(n_i^3) D_j(x(n_i^3)),$$

где $\mathbf{x} = \{x(n_i^t)\}$ и $\mathbf{u} = \{u_j(n_i^t)\}$, $n_i^t \in \mathcal{N}^t$, $\lambda_j \in (0, 1)$ дисконт фактор игрока j , при заданных уравнении (3.5.1) и начальном «запасе» $x^0 = 0$, а также при ограничениях: $u_j(n_i^t) \in [0, e]$ для любого игрока $j \in M$ и любой вершины $n_i^t \in \mathcal{N}^t$, $t = 0, 1, 2$.

Используем следующие параметры игры:

$$\begin{aligned}\alpha_1 &= 0.1, \alpha_2 = 0.2, \alpha_3 = 0.3, \\ \gamma_1 &= 0.9, \gamma_2 = 0.8, \gamma_3 = 0.7, \\ \underline{\delta} &= 0.45, \bar{\delta} = 0.8, e = 30, \lambda_1 = \lambda_2 = \lambda_3 = 0.9, \\ \pi(n_1^1) &= 0.6, \pi(n_2^1) = 0.4, \\ \pi(n_1^2) &= 0.3, \pi(n_2^2) = 0.3, \pi(n_3^2) = 0.3, \pi(n_4^2) = 0.1, \\ \pi(n_1^3) &= 0.1, \pi(n_2^3) = 0.2, \pi(n_3^3) = 0.1, \pi(n_4^3) = 0.2, \\ \pi(n_5^3) &= 0.05, \pi(n_6^3) = 0.25, \pi(n_7^3) = 0.05, \pi(n_8^3) = 0.05.\end{aligned}$$

Воспользуемся формулами (3.1.7) и (3.1.9) и найдем кооперативные стратегии для игры и любой ее подыгры. Кооперативная траектория состояния будет следующей:

$$\begin{array}{cccccccc}x^*(n^0) & x^*(n_1^1) & x^*(n_2^1) & x^*(n_1^2) & x^*(n_2^2) & x^*(n_3^2) & x^*(n_4^2) & x^*(n_1^3) \\ 0 & 55.177 & 55.177 & 63.4188 & 63.4188 & 79.0484 & 79.0484 & 72.8623 \\ x^*(n_2^3) & x^*(n_3^3) & x^*(n_4^3) & x^*(n_5^3) & x^*(n_6^3) & x^*(n_7^3) & x^*(n_8^3) & \\ 72.8623 & 88.6125 & 88.6125 & 75.0804 & 75.0804 & 94.7123 & 94.7123 & \end{array}$$

По формуле (3.1.8) рассчитаем издержки игроков во всех подыграх, они представлены в табл. 3.4.

Таблица 3.4: Издержки игроков при кооперативном равновесии.

Период	$t = 0$	$t = 1$		$t = 2$			
Вершина	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2
$W_1(\mathbf{u}(\cdot), x^*(\cdot))$	249.371	213.06	242.807	122.401	161.771	149.748	201.349
$W_2(\mathbf{u}(\cdot), x^*(\cdot))$	381.031	344.644	394.864	214.701	279.02	267.534	351.835
$W_3(\mathbf{u}(\cdot), x^*(\cdot))$	517.495	479.554	550.625	308.23	398.087	386.625	504.397
	$t = 3$						
	n_1^3	n_2^3	n_3^3	n_4^3	n_5^3	n_6^3	n_7^3
	53.0891	53.0891	78.5218	78.5218	56.3706	56.3706	89.7041
	106.178	106.178	157.044	157.044	112.741	112.741	179.408
	159.267	159.267	235.565	235.565	169.112	169.112	269.112

Для каждого игрока $j \in M$ и вершины $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T - 1$ нам требуется решить оптимизационную задачу, определенную в (3.4.8), где \max требуется

заменить на \min . Найдя эти издержки, мы можем найти разности между выигрышами при кооперации и при отклонении для каждой вершины и каждого игрока (см. табл. 3.5). На основе этих разностей мы определяем $\varepsilon_j(n_i^t)$ для всех $j \in M$ и $n_i^t \in \mathcal{N}^t$, $t = 0, 1, 2$ (см. табл. 3.6).

Таблица 3.5: Максимальный выигрыш от отклонения для всех подыгр.

Период	$t = 0$	$t = 1$		$t = 2$			
Вершина	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2
Игрок 1	52.0734	45.6117	53.2126	23.4217	34.6419	24.8694	39.5753
Игрок 2	-24.5520	5.18347	11.0904	16.4602	24.3455	17.4776	27.8126
Игрок 3	-111.9650	-38.4121	-33.8928	10.2658	15.1836	10.9003	17.346

Таблица 3.6: Значения $\varepsilon_j(\cdot)$ для всех игроков и подыгр.

Период	$t = 0$	$t = 1$		$t = 2$			
Вершина	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2
ε_1	52.0734	45.6117	53.2126	23.4217	34.6419	24.8694	39.5753
ε_2	0	5.18347	11.0904	16.4602	24.3455	17.4776	27.8126
ε_3	0	0	0	10.2658	15.1836	10.9003	17.346

Из табл. 3.6 нетрудно заметить, что в вершине n^0 только игрок 1 имеет выгоду от отклонения от кооперативного поведения. В момент времени $t = 1$ игроки 1 и 2 могут выиграть при отклонении, а в период 2 все игроки имеют положительный выигрыш, отклоняясь от кооперации, что еще раз подчеркивает невозможность реализовать наказание отклоняющегося игрока на последнем шаге. Самый большой выигрыш от отклонения имеет игрок 1 в вершине n_2^1 . Следовательно, получаем значение $\tilde{\varepsilon}$, равное 53.2126. «Цена кооперации» в этой игре равна $53.2126 / (249.371 + 381.031 + 517.495)$ или приблизительно 4.63%.

Игра относится к классу экологических игр с отрицательным воздействием. В общем случае, для таких играх показано, что сложно достичь кооперации в такой постановке [109], поэтому результаты, полученные выше в этом примере, не являются неожиданными. Чтобы получить некоторое представление о том, что происходит в этой игре при ее бóльшей продолжительности, рассмотрим игру с 11 периодами времени, т. е. $\mathcal{T} = \{0, 1, \dots, 10\}$. Остальные параметры игры остаются такими же, как и раньше.

Дерево событий для этой игры изображено на рис. 3.4. Это так называемое бинарное дерево, т. е. из любой вершины в периодах $t = 0, \dots, 9$ всегда можно перейти в две последующие вершины. Условная вероятность реализации левой вершины равна $\frac{1}{3}$ и правой — $\frac{2}{3}$. Например, $\pi(n_1^1) = \frac{1}{3}$ и $\pi(n_2^1) = \frac{2}{3}$ в период 1, и вероятности для $t = 2$ равны $\pi(n_1^2) = \frac{1}{9}$, $\pi(n_2^2) = \frac{2}{9}$, $\pi(n_3^2) = \frac{2}{9}$, $\pi(n_4^2) = \frac{4}{9}$. Начальная вершина n^0 и все «левые» вершины имеют низкий уровень абсорбции загрязняющего вещества $\underline{\delta}$ и все «правые» вершины имеют высокий уровень абсорбции загрязняющего вещества $\bar{\delta}$.

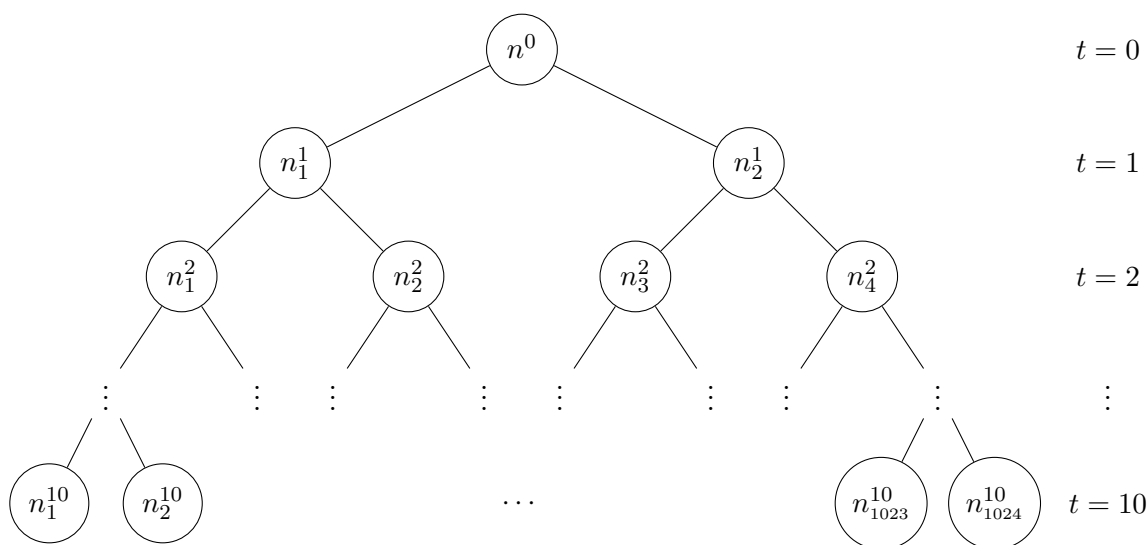


Рис. 3.4: Дерево событий в примере с $T = 10$.

Как и в предыдущем примере, найдем кооперативные выигрыши и максимальные выигрыши игроков при отклонении от кооперативного поведения для всех подыгр и всех игроков. В табл. 3.7 и 3.8 приведены только максимальные выигрыши игроков при отклонении для каждого периода времени и значения $\varepsilon_j(n_i^t)$, поскольку всего вершин равно 2047. По этим таблицам нетрудно заметить, что первый период времени, когда какой-либо игрок имеет положительный выигрыш при отклонении — период 0 (игрок 1). Интересно, что игроку 2 имеет смысл отклониться от кооперативной стратегии в периоды 8 и 9, а игроку 3 — только в последний период 9, на котором игроки в последний раз выбирают стратегии, поскольку в период 10 все игроки получают только терминальный выигрыши. Величина $\tilde{\varepsilon}$ равняется 132.87, и цена кооперации равна 4.37%. При кооперативном поведении суммарное загрязнение меньше, чем при

равновесии по Нэшу. Этот результат обнадеживает не только с экономической точки зрения, но и с точки зрения охраны окружающей среды.

Таблица 3.7: Максимальный выигрыш от отклонения от кооперации в период t для игры с 11 периодами.

Период t	0	1	2	3	4	5	6	7	8	9
Игрок 1	117.05	131.23	132.87	128.121	120.169	110.289	98.8774	85.8045	69.719	43.7459
Игрок 2	-162.213	-138.843	-118.765	-99.4618	-79.2663	-57.3184	-33.1637	-6.82434	19.3776	30.7436
Игрок 3	-497.143	-458.461	-414.163	-364.781	-309.905	-249.008	-181.732	-108.818	-33.8152	19.174

Таблица 3.8: Максимальное значение ε_j для периода t для игры с 11 периодами.

Период t	0	1	2	3	4	5	6	7	8	9
$\max_{n_i^t \in \mathcal{N}^t} \varepsilon_1(n_i^t)$	117.05	131.23	132.87	128.121	120.169	110.289	98.8774	85.8045	69.719	45.1047
$\max_{n_i^t \in \mathcal{N}^t} \varepsilon_2(n_i^t)$	0	0	0	0	0	0	0	0	19.3776	30.7436
$\max_{n_i^t \in \mathcal{N}^t} \varepsilon_3(n_i^t)$	0	0	0	0	0	0	0	0	0	19.174

На Рисунке 3.5 представлен график зависимости $\tilde{\varepsilon}$ и дисконт фактора $\lambda_j = \lambda$, $j = 1, 2, 3$ для игры с 11 периодами времени. Чем больше дисконт фактор, тем больше $\tilde{\varepsilon}$.

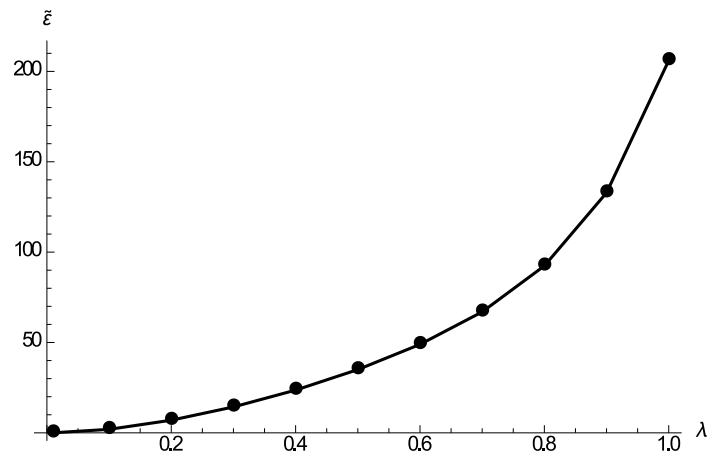


Рис. 3.5: Значения $\tilde{\varepsilon}$ как функции от λ .

Таблица 3.9 дает информацию о том, в какой первый период времени каждый игрок имеет выгоду от отклонения от кооперативной стратегии. Игрок 3 не будет отклоняться вплоть до последнего периода, если $\lambda \geq 0.5$. Игрок 1 имеет положительный выигрыш от отклонения, начиная с периода 1 для любых значений λ . При небольших значениях дисконт фактора игроки имеют меньшие выигрыши от отклонения по сравнению с выигрышами при более высоких

дисконт факторах. Тем не менее, при низких дисконт факторах игроки могут начать отклоняться от кооперации раньше, чем в играх с большими дисконт факторами.

Таблица 3.9: Первый период времени, когда игрок имеет положительный выигрыш при отклонении от кооперативной стратегии.

λ	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
Игрок 1	0	0	0	0	0	0	0	0	0	0	0
Игрок 2	0	0	0	0	1	1	5	7	7	8	8
Игрок 3	0	0	1	8	8	9	9	9	9	9	9

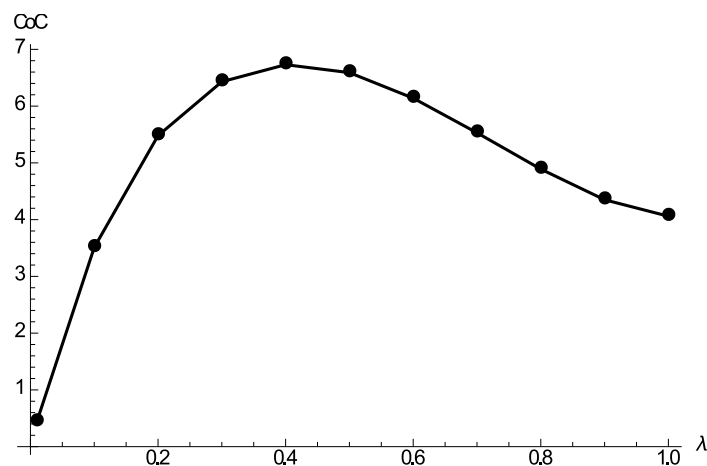


Рис. 3.6: Цена кооперации (в процентах) как функция дисконт фактора λ .

На Рисунке 3.6 представлен график цены кооперации как функции дисконт фактора λ . Можно заметить, что значения цены кооперации не превосходят 7% для любого λ .

§ 3.6 Условие стратегической устойчивости кооперативного решения в динамических играх, разыгрываемых на деревьях событий

Будем предполагать, что игроки согласились играть кооперативно на протяжении всей игры, т.е. они согласились образовать максимальную коалицию и максимизировать суммарный выигрыш коалиции. Пусть также выплаты игрокам в каждой вершине дерева событий производятся в соответствии с ПРД, чтобы по-

лучить позиционно состоятельный дележ. Таким образом, заданная изначально функция выигрыша игрока $j \in M$ в вершине n_i^t , а именно, $\phi_j^{n_i^t}(x^*(n_i^t), u^*(n_i^t))$, будет заменена компонентой процедуры распределения дележа $\beta_j(n_i^t)$ для всех $j \in M$ и $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$. Аналогично, $\Phi_j^{n_i^T}(x^*(n_i^T))$ заменяется $\beta_j(n_i^T)$, $j \in M$ для всех $n_i^T \in \mathcal{N}^T$. Теперь функция выигрыша игрока $j \in M$ в подыгре, начинающейся с вершины $n_i^t \in \mathcal{N}^t$ из состояния $x^*(n_i^t)$, определена таким образом:

$$\hat{W}_j(\mathbf{u}(x^*(n_i^t)), x^*(n_i^t)) = \sum_{\theta=t}^T \lambda_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_i^\theta | n_i^t) \beta_j(n_i^\theta). \quad (3.6.1)$$

Когда произошло перераспределение выплат игрокам, можно поставить вопрос о стратегической поддержке кооперативного решения, как и в Главе 2. В случае игры с конечной продолжительностью, покажем, что существует абсолютное ε -равновесие по Нэшу с выигрышами, равными выигрышам в некооперативной игре, в которой выплаты производятся в соответствии с процедурой распределения дележа.

Рассмотрим динамическую игру, разыгрываемую на дереве событий, как игру в развернутой форме и предположим, что игроки в каждой вершине обладают следующей информацией при выборе своей стратегии. Каждый игрок знает вершину дерева $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T$, в которой он находится, знает историю этой вершины, т. е. свои стратегии и стратегии всех игроков, которые были выбраны на пути от начальной вершины n^0 до вершины $a(n_i^t)$. Напомним, что $P(n_i^t) = (n^0, n_{i_1}^1, \dots, n_{i_{t-1}}^{t-1}, n_i^t)$ — это единственный путь от начальной вершины n^0 к вершине n_i^t .

Будем рассматривать игру в стратегиях поведения $\sigma_j = \left\{ \sigma_j^{n_i^t} \right\}_{n_i^t \in \mathcal{N}^t, t=0, \dots, T-1}$, где $\sigma_j^{n_i^t}: H(n_i^t) \rightarrow U_j^{n_i^t}$ и $H(n_i^t)$ есть история вершины n_i^t . Для заданной ситуации в стратегиях поведения $\sigma = (\sigma_j : j \in M)$ можно вычислить выигрыш любого игрока во всей игре и любой ее подыгре. Выигрыш игрока j в подыгре, начинающейся в вершине n_i^t из состояния $x(n_i^t)$ равен

$$\hat{W}_j(\sigma, x(n_i^t)) = \hat{W}_j(\mathbf{u}(n_i^t), x(n_i^t)),$$

где $\mathbf{u}(n_i^t)$ — набор управлений в подыгре, n_i^t , начинающейся в вершине n_i^t из

состояния $x(n_i^t)$, соответствующий ситуации σ .

Теперь подробно опишем стратегию наказания. Для стратегической поддержки кооперации в конечной динамической игре, разыгрываемой на дереве событий, построим приближенное равновесие (ε -равновесие) в стратегиях поведения. Сделаем следующие предположения о поведении игроков:

1. Игроки хотят реализовать кооперативную ситуацию \mathbf{u}^* , определяемую формулой (3.1.7).
2. Игроки предполагают в игре в целом получить выигрыши в соответствии с дележом $y(x^0)$, при условии реализации ситуации \mathbf{u}^* .
3. Выплаты игрокам в каждой вершине реализованной траектории производятся в соответствии с процедурой распределения дележа, определяемой рекуррентно формулами (3.2.8) и (3.2.9) на основе дележа $y(x^0)$.
4. Если игрок j отклоняется от кооперации в вершине $a(n_i^t)$, т. е. выбирает стратегию $u_j(a(n_i^t))$, отличную от $u_j^*(a(n_i^t))$, максимальная коалиция распадается и все игроки используют равновесные по Нэшу стратегии, начиная со следующей вершины и до конца игры. Таким образом, игроки разыгрывают S -адаптивное равновесие по Нэшу в подыгре, начинающейся из вершины n_i^t в состоянии $x(n_i^t) = f^{a(n_i^t)}(x^*(a(n_i^t)), (u_{-j}^*(a(n_i^t)), u_j(a(n_i^t))))$.
5. Если более одного игрока отклоняется в вершине $a(n_i^t)$, то максимальная коалиция распадается, и в вершине n_i^t игроки переключаются на равновесные по Нэшу стратегии, так что оставшуюся часть игры они играют некооперативно.

Как и в предыдущем параграфе, строим ситуацию в стратегиях поведения, включающих два режима:

Номинальный режим. Если история вершины n_i^t совпадает с кооперативной историей $H^*(n_i^t)$ вершины n_i^t , тогда игрок $p \in M$ выбирает стратегию $u_p^*(n_i^t)$ в вершине n_i^t .

Режим наказания. Если история вершины n_i^t такова, что существует вершина n на пути $P(n_{i_{t-1}}^{t-1})$ и при этом $u(n) \neq u^*(n)$, тогда стратегия игрока p

— равновесная по Нэшу стратегия в подыгре, начинающейся из вершины, сразу следующей за n из соответствующего состояния. В этом случае история вершины n_i^t такова, что существует n и по крайней мере один игрок $j \in M, j \neq p$, что история $H(n)$ вершины n — это часть истории $H^*(n_i^t)$, а пара $(n, u(n))$ не является частью истории $H^*(n_i^t)$.

Игрок $p \in M$ будет использовать следующую стратегию наказания:

$$\hat{\sigma}_p(H(n_i^t)) = \begin{cases} u_p^*(n_i^t), & \text{если } H(n_i^t) = H^*(n_i^t), \\ \hat{u}_p(n_i^t), & \text{если } \exists \text{ вершина } n \text{ на пути } P(n_i^t), \\ & \text{такая что } u(n) \neq u^*(n), \end{cases} \quad (3.6.2)$$

где $\hat{u}_p(n_i^t)$ — стратегия игрока p в вершине n_i^t . Стратегия $\hat{u}_p(n_i^t)$, реализующая наказание в подыгре, начинающейся из единственной вершины, принадлежащей множеству $\mathcal{S}(n) \cap P(n_i^t)$. Стратегия $\hat{u}_p(n_i^t)$ совпадает с равновесной по Нэшу стратегией $u_p^N(n_i^t)$ в подыгре, начинающейся из вершины $n_1 = \mathcal{S}(n) \cap P(n_i^t)$ в состоянии $x(n_1)$.

Для того, чтобы избежать дополнительных обозначений, мы опять опустим аргумент состояния в обозначении стратегии наказания, которая естественным образом зависит от состояния, в котором происходит отклонение. Пусть вершина n_1 следует после вершины n , в которой игрок j отклоняется. Реализуется набор стратегий $(u_{-j}^*(n), u_j(n))$, а состояние в вершине n_1 находится с помощью уравнения динамики $x(n_1) = f^n(x^*(n), (u_{-j}^*(n), u_j(n)))$. Стратегия $\hat{u}_p(n_1)$ — часть ситуации $\hat{u}(n_1) = (u_p^N(n_1) : p \in M)$, где $u_p^N(n_1)$ — стратегия игрока p в вершине n_1 .

Теперь в подыгре, начинающейся в вершине $n_i^t \in \mathcal{N}^t$ и состоянии $x(n_i^t)$, набор стратегий, наказывающих индивидуальное отклонение игрока j , будет таким:

$$\hat{\mathbf{u}}^j(n_i^t) = (\hat{\mathbf{u}}_p^j(n_i^t) : p \in M),$$

где $\hat{\mathbf{u}}_p^j(n_i^t) = \{\hat{u}_p^j(n_i^\theta) : n_i^\theta \in \Gamma(n_i^t)\}$. Это набор стратегий определяет траекторию состояний, соответствующих равновесию по Нэшу в этой подыгре:

$$\hat{\mathbf{x}}^j(n_i^t) = \{\hat{x}^j(n_i^\theta) : n_i^\theta \in \Gamma(n_i^t)\}.$$

Для построения стратегий наказания мы определяем m ситуаций (каждая из которых наказывает определенного игрока) в каждой подыгре. Сформулируем основной результат этого параграфа.

Теорема 3.3. *Рассмотрим динамическую игру, разыгрываемую на дереве событий, в которой выигрыши игрокам осуществляются в соответствии с процедурой распределения дележа, определяемой рекуррентно формулами (3.2.8) и (3.2.9) на основе дележа $y(n^0)$. В этой игре для любого $\varepsilon \geq \hat{\varepsilon}$ существует абсолютное ε -равновесие в стратегиях наказания с выигрышами игроков $y_1(n^0), \dots, y_m(n^0)$, и*

$$\hat{\varepsilon} = \max_{j \in M} \max_{\substack{n_i^t \in \mathcal{N}^t \\ t=0, \dots, T-1}} \varepsilon_j(n_i^t), \quad (3.6.3)$$

где

$$\begin{aligned} \varepsilon_j(n_i^t) = & \max_{u_j(n_i^t) \in U_j^{n_i^t}} \left\{ \phi_j^{n_i^t}(x^*(n_i^t), (u_{-j}^*(n_i^t), u_j(n_i^t))) - \beta_j(n_i^t) + \right. \\ & + \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_i^\theta | n_i^t) \left(\phi_j^{n_i^\theta}(\hat{x}^j(n_i^\theta), \hat{u}^j(n_i^\theta)) - \beta_j(n_i^\theta) \right) + \\ & \left. + \lambda_j^{T-t} \sum_{n_i^T \in \mathcal{N}_\Gamma^T} \pi(n_i^T | n_i^t) \left(\Phi_j^{n_i^T}(\hat{x}(n_i^T)) - \beta_j(n_i^T) \right) \right\}, \quad (3.6.4) \end{aligned}$$

где $\hat{u}^j(n_i^\theta)$ — ситуация в вершине n_i^θ , определяемая ситуацией в стратегиях поведения $\hat{\sigma}$ по формуле (3.6.2) и когда режим наказания этой стратегии «включается» в подыгре, начинающейся из вершины множества $\mathcal{S}(n_i^t)$ и в состоянии $f^{n_i^t}(x(n_i^t), (u_{-j}^*(n_i^t), u_j(n_i^t)))$. Таким образом, разности во второй и третьей строках также зависят от стратегии $u_j(n_i^t)$. Состояние $\hat{x}^j(n_i^\theta)$, $n_i^\theta \in \Gamma(n_i^\theta)$, находится единственным образом из уравнений динамики состояния по стратегиям $\hat{\mathbf{u}}^j(n_i^t)$.

Доказательство. Рассмотрим стратегию поведения $\hat{\sigma} = (\hat{\sigma}_p : p \in M)$, определенную равенством (3.6.2), и подыгру, начинающуюся из вершины $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T - 1$. Рассмотрим всевозможные истории вершины n_i^t и посчитаем выигрыш игрока j , отклоняющегося в вершине n_i^t . В случае, если он не отклоняется в этой вершине, а продолжает реализовывать кооперативную стратегию

вместо с остальными игроками, его выигрыш равен

$$\begin{aligned} \hat{W}_j(\mathbf{u}^*(n_l^t), x^*(n_l^t)) = & \beta_j(n_l^t) + \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \beta_j(n_l^\theta) + \\ & + \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \beta_j(n_l^T), \end{aligned} \quad (3.6.5)$$

где $\mathcal{N}_\Gamma^\theta = \mathcal{N}^\theta \cap \Gamma(n_l^t)$, $\mathbf{u}^*(n_l^t) = (\mathbf{u}_j^*(n_l^t) : j \in M)$ — S -адаптивная ситуация, максимизирующая суммарный выигрыш игроков. При подсчете выигрыша используется ПРД, компоненты которой игрок получает вместо выигрышей, определенных изначально заданными функциями выигрышей ϕ_j и Φ_j .

Рассмотрим случай, когда история вершины n_l^t совпадает с кооперативной историей $H^*(n_l^t)$. Предположим, что игрок j отклоняется в вершине n_l^t от кооперативной траектории. В этом случае он может гарантировать следующий выигрыш в подыгре, начинающейся из вершины n_l^t , при условии, что все игроки реализуют стратегии поведения $\hat{\sigma} = (\hat{\sigma}_p(\cdot) : p \in M)$, определенные в (3.6.2):

$$\begin{aligned} \max_{u_j(n_l^t) \in U_j^{n_l^t}} \left\{ \phi_j^{n_l^t}(x^*(n_l^t), (u_{-j}^*(n_l^t), u_j(n_l^t))) + \right. \\ + \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \phi_j^{n_l^\theta}(\hat{x}^j(n_l^\theta), \hat{u}^j(n_l^\theta)) + \\ \left. + \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \Phi_j^{n_l^T}(\hat{x}^j(n_l^T)) \right\}, \end{aligned} \quad (3.6.6)$$

где режим наказания (т. е. равновесные по Нэшу стратегии) начинает применяться в вершинах из множества $\mathcal{S}(n_l^t)$. Если происходит отклонение от кооперативной траектории, то выигрыши игроков начинают вычисляться в соответствии с изначально заданными функциями выигрышей $\phi_j^{n_l^t}(\cdot)$ и $\Phi_j^{n_l^T}(\cdot)$.

Теперь найдем выигрыш игрока j от отклонения в вершине n_l^t как разность

между его выигрышами без отклонения и с отклонением в вершине n_i^t :

$$\begin{aligned} \varepsilon_j(n_i^t) = \max_{u_j(n_i^t) \in U_j^{n_i^t}} & \left\{ \phi_j^{n_i^t}(x^*(n_i^t), (u_{-j}^*(n_i^t), u_j(n_i^t))) - \beta_j(n_i^t) + \right. \\ & + \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_i^\theta | n_i^t) \left(\phi_j^{n_i^\theta}(\hat{x}^j(n_i^\theta), \hat{u}^j(n_i^\theta)) - \beta_j(n_i^\theta) \right) + \\ & \left. + \lambda_j^{T-t} \sum_{n_i^T \in \mathcal{N}_\Gamma^T} \pi(n_i^T | n_i^t) \left(\Phi_j^{n_i^T}(\hat{x}(n_i^T)) - \beta_j(n_i^T) \right) \right\}. \end{aligned} \quad (3.6.7)$$

Далее предположим, что история вершины n_i^t не совпадает с $H^*(n_i^t)$. Это означает, что все игроки уже переключились с номинального режима на режим наказания. Игрок j не получит выигрыша от отклонения в вершине n_i^t , потому как все игроки используют равновесные по Нэшу стратегии независимо от того, какой игрок (или группа игроков) отклонился в предыдущих вершинах.

Вычислим максимальный выигрыш от отклонения среди всех подыгр и всех игроков, который вычисляется по формуле (3.6.7). Мы получим значение $\hat{\varepsilon}$ из утверждения теоремы:

$$\hat{\varepsilon} = \max_{j \in M} \max_{\substack{n_i^t \in \mathcal{N}^t \\ t=0, \dots, T-1}} \varepsilon_j(n_i^t).$$

Для любого $\varepsilon \geq \hat{\varepsilon}$ ситуация в стратегиях поведения (3.6.2) является ситуацией абсолютного ε -равновесия по Нэшу по построению. \square

§ 3.7 Абсолютное ε -равновесия с использованием процедуры распределения дележа в игре управления загрязнением окружающей среды

Для иллюстрации результатов предыдущего параграфа рассмотрим игру 3 лиц, представленную в параграфах § 3.3 и § 3.5, но с некоторыми изменениями. Изначальная теоретико-игровая модель игры взята из работы [98]. Пусть множество игроков будет $M = \{1, 2, 3\}$, множество периодов $\mathcal{T} = \{0, 1, \dots, 5\}$. Также, как и в предыдущих примерах $u(n_i^t) = (u_1(n_i^t), u_2(n_i^t), u_3(n_i^t))$ — вектор выбросов некоторого загрязняющего вещества в вершине n_i^t , $x(n_i^t)$ — «запас» загрязняющего вещества в вершине n_i^t в период t . Динамика состояния x задается урав-

нением:

$$x(n_i^t) = (1 - \delta(a(n_i^t)))x(a(n_i^t)) + \sum_{j \in M} u_j(a(n_i^t)), \quad (3.7.1)$$

с заданным начальным «запасом» загрязняющего вещества x^0 в вершине n^0 , и $\delta(n_i^t)$ ($0 < \delta(n_i^t) < 1$) случайный уровень абсорбции природой загрязняющего вещества в вершине n_i^t . Предположим, что $\delta(n_i^t) \in \{\underline{\delta}, \bar{\delta}\}$, где $\underline{\delta} < \bar{\delta}$. Предположим также, что дерево событий — бинарное дерево, т. е. в каждый период времени $t = 0, \dots, 4$ любая вершина имеет две вершины-последователя (см. Рис. 3.7). Условная вероятность реализации левой вершины-последователя равна $\frac{1}{4}$ для любой вершины и $\frac{3}{4}$ для правой вершины-последователя. Например, $\pi(n_1^1) = \frac{1}{4}$ и $\pi(n_2^1) = \frac{3}{4}$ в период 1, а вероятности равны $\pi(n_1^2) = \frac{1}{16}$, $\pi(n_2^2) = \frac{3}{16}$, $\pi(n_3^2) = \frac{3}{16}$, $\pi(n_4^2) = \frac{9}{16}$ для $t = 2$. Корневая вершина n^0 и все левосторонние вершины имеют низкий уровень естественной абсорбции $\underline{\delta}$ загрязняющего вещества, и все правосторонние вершины имеют высокий уровень абсорбции $\bar{\delta}$.

В каждой вершине $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, 4$ функция издержек игрока $j \in M$, являющаяся суммой двух функций: 1) $D_j(x(n_i^t)) = \alpha_j x^2(n_i^t)$, где $\alpha_j > 0$; 2) $C_j(u_j(n_i^t)) = \frac{\gamma_j}{2} (u_j(n_i^t) - e)^2$, где e и γ_j — положительные константы.

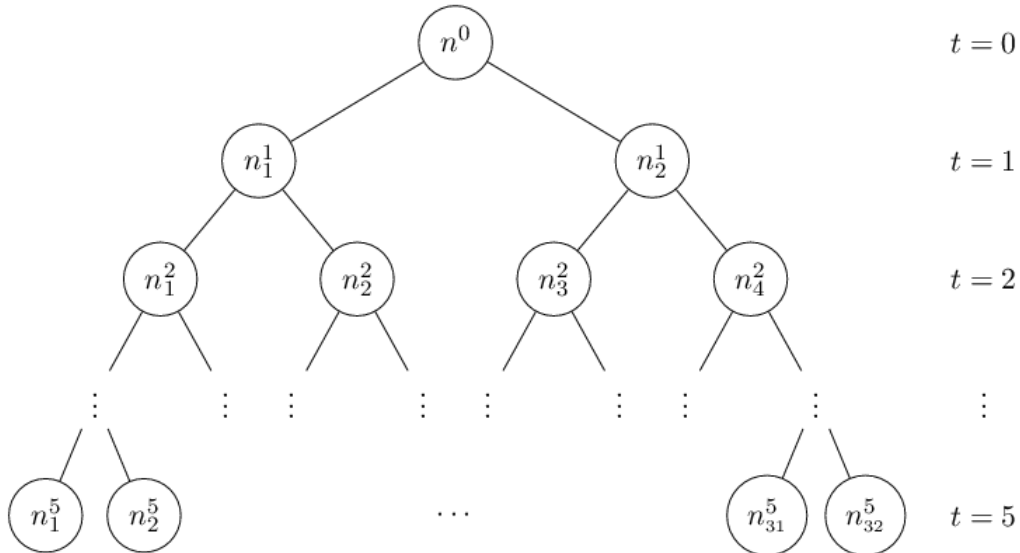


Рис. 3.7: Дерево событий для $T = 5$.

Суммарные дисконтированные издержки $J_j(\mathbf{x}, \mathbf{u})$ игрока $j \in M$, которые он

стремится минимизировать, заданы формулой

$$\sum_{t=0}^4 \lambda_j^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) (C_j(u_j(n_i^t)) + D_j(x(n_i^t))) + \lambda_j^5 \sum_{n_i^5 \in \mathcal{N}^5} \pi(n_i^5) D_j(x(n_i^5)),$$

где $\mathbf{x} = \{x(n_i^t)\}$ и $\mathbf{u} = \{u(n_i^t)\}$, $n_i^T \in \mathcal{N}^t$, $\lambda_j \in (0, 1)$ — дисконтирующий фактор игрока j , при условии, что выполнено (3.7.1), и начальный «запас» $x_0 = 0$. Также имеются ограничения на стратегии игроков: $u_j(n_i^t) \in [0, e]$ для любого игрока $j \in M$ и любой вершины $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, 4$.

Используем следующие параметры для численного моделирования:

$$\alpha_1 = 0.1, \alpha_2 = 0.2, \alpha_3 = 0.3,$$

$$\gamma_1 = 0.9, \gamma_2 = 0.8, \gamma_3 = 0.7,$$

$$\underline{\delta} = 0.45, \bar{\delta} = 0.8, e = 30, \lambda_1 = \lambda_2 = \lambda_3 = 0.9.$$

В таблицах 3.10 и 3.11 представлены результаты вычислений максимальных выигрышей от отклонения в соответствии с формулой (3.6.4) и максимальных выигрышей от отклонения в случае, когда игроки не используют ПРД для определения выплат игрокам, а игроки получают выплаты в соответствии с их изначально заданными функциями выигрышей. Если быть точными, в таблице представлены максимальные выигрыши от отклонений за период для каждого $t = 0, 1, 2, 3, 4$.

Таблица 3.10: Максимальный выигрыш от отклонения в период t для игры с 6 периодами, когда игроки используют ПРД.

Период t	0	1	2	3	4
Игрок 1	-128.351	-96.4928	-59.492	-23.5305	-0.794881
Игрок 2	-118.967	-73.6691	-28.0224	12.5117	26.5563
Игрок 3	-105.177	-33.8832	25.4111	67.5529	61.8055

Суммарные издержки трех игроков во всей игре равны 2148.7. Таблицы 3.10 и 3.11 показывают преимущество использования ПРД в этой игре. Когда игроки используют ПРД, максимальный выигрыш от отклонения или $\hat{\epsilon}$ равен 67.5529 в сравнении с 108.902 в игре, где не используется ПРД для перераспределения выигрышей. Нетрудно заметить, что первый период, когда какой-либо игрок

Таблица 3.11: Максимальный выигрыш от отклонения в период t для игры с 6 периодами, когда игроки не используют ПРД.

Период t	0	1	2	3	4
Игрок 1	104.714	108.902	96.9909	75.0803	40.6216
Игрок 2	-103.318	-60.0593	-18.0474	18.2751	28.5478
Игрок 3	-353.892	-251.621	-141.231	-41.2657	17.8045

имеет выгоду от отклонения — период $t = 2$, тогда как в игре без использования ПРД, такой период — это $t = 0$. Игроки 1 и 2 имеют меньшие выигрыши от отклонения в игре с ПРД, чем в игре без ПРД. Но игрок 3 имеет больше оснований для отклонения от кооперативной траектории (имеет большие выигрыши от отклонения), если в игре используется ПРД для определения выплат игрокам.

§ 3.8 Динамические игры, разыгрываемые на бинарных деревьях событий

§ 3.8.1 Динамические игры с линейной динамикой состояния

В этом параграфе сделаем некоторые предположения о дереве событий и функциях выигрышей игроков. После этого рассмотрим некооперативный и кооперативный варианты игры, а также произведем сравнение выигрышей игроков в этих случаях с помощью так называемой «цены анархии» [117].

Пусть уравнение динамики состояния имеет вид:

$$x(n_i^t) = A(a(n_i^t))x(a(n_i^t)) + \sum_{j \in M} B_j(a(n_i^t))u_j(a(n_i^t)), \quad (3.8.1)$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, t = 1, \dots, T. \quad (3.8.2)$$

где $x(n_i^t) \in X$, $u_j(a(n_i^t)) \in U_j^{a(n_i^t)}$ для всех вершин $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T$, матрицы $A(a(n_i^t))$ и $B_j(a(n_i^t))$ имеют размерности $q \times q$ и $q \times r$ соответственно. Компоненты вектора состояния $x(n_i^t)$ вершины n_i^t являются линейными комбинациями компонент состояния предшествующей вершины $a(n_i^t)$ и стратегий игроков, которые были ими выбраны в предыдущей вершине. Коэффициенты линейных комбинаций — это строки матриц $A(a(n_i^t))$ и $B_j(a(n_i^t))$, $j \in M$.

В каждой вершине n_i^t , $t = 0, \dots, T-1$, выигрыш игрока j — функция состояния и стратегий всех игроков, выбранных в этой вершине:

$$\begin{aligned} \phi_j^{n_i^t}(x(n_i^t), u(n_i^t)) &= \frac{1}{2}u'_j(n_i^t)C_{1j}(n_i^t)u_j(n_i^t) + \sum_{k \neq j} u'_k(n_i^t)C_{2j}(n_i^t)u_j(n_i^t) + \\ &+ c'_j(n_i^t)u_j(n_i^t) + d'_j(n_i^t)x(n_i^t), \end{aligned}$$

где $C_{1j}(n_i^t)$, $C_{2j}(n_i^t)$ — матрицы размерностей $r \times r$, $c_j(n_i^t) \in \mathbb{R}^r$ и $d_j(n_i^t) \in \mathbb{R}^q$. Предполагаем, что матрицы C_{1j} , $j \in M$, — симметричные и отрицательно определенные.

В терминальной вершине n_i^T , выигрыш игрока j определяется функцией

$$\Phi_j^{n_i^T}(x(n_i^T)) = d'_j(n_i^T)x(n_i^T).$$

Пусть игрок $j \in M$ максимизирует свой ожидаемый выигрыш, дисконтированный³ с помощью фактора ϱ_j ($0 < \varrho_j < 1$). Уравнения состояния и функции выигрышей игроков определяют многошаговую игру, где

$$\begin{aligned} \mathbf{x} &= \{x(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T\}, \\ \mathbf{u} &= \{u(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T-1\}, \end{aligned}$$

и $J_j(\mathbf{x}, \mathbf{u})$ — выигрыш игрока $j \in M$:

$$J_j(\mathbf{x}, \mathbf{u}) = \sum_{t=0}^{T-1} \varrho_j^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \phi_j^{n_i^t}(x(n_i^t), u(n_i^t)) + \varrho_j^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) \Phi_j^{n_i^T}(x(n_i^T)), \quad (3.8.3)$$

при условии

$$x(n_i^t) = A(a(n_i^t))x(a(n_i^t)) + \sum_{j \in M} B_j(a(n_i^t))u_j(a(n_i^t)), \quad (3.8.4)$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, t = 1, \dots, T, \quad (3.8.5)$$

$$x(n^0) = x^0. \quad (3.8.6)$$

Целевая функция (3.8.3) — строго вогнутая функция стратегий игроков.

Ранее мы определили класс S -адаптивных стратегий, в которых и будем решать игру. Напомним, какой информацией обладает игрок, выбирая стратегию

³В этом параграфе будем обозначать дисконт-фактор игрока j через ϱ_j , а не через λ_j , как было в предыдущих параграфах Главы 3, поскольку в рамках этого параграфа буква λ_j будет использоваться для обозначения множителя Лагранжа.

в любой вершине. Если стратегии (или управления) зависят только от времени и не зависят от текущего состояния системы, то мы называем их программными (в зарубежной литературе в этом случае называть информационную структуру игры “open loop information structure”). В случае, когда стратегии зависят не только от времени, но и от состояния системы, мы называем их позиционными (в зарубежной литературе в этом случае называть информационную структуру игры “closed loop information structure”). В теории динамических игр, разыгрываемых на деревьях событий, используются S -адаптивные стратегии, которые являются программными стратегиями, адаптированными к стохастическому процессу, определяющему динамику игры на дереве событий. Индекс S отображает стохастическую (stochastic) составляющую в стратегии. Если рассмотреть произвольную вершину дерева событий, то мы знаем историю этой вершины и текущее состояние системы.

Так же как и в предыдущих параграфах, рассмотрим два варианта игры: кооперативный и некооперативный. В некооперативном варианте игры в качестве равновесия в игре будем рассматривать S -адаптивное равновесие по Нэшу. При кооперативном варианте, игроки максимизируют сумму дисконтированных выигрышей во всей игре

$$\max_{(\mathbf{u}_j: j \in M)} \sum_{j \in M} W_j(\mathbf{u}, x^0).$$

Результирующий вектор кооперативных стратегий также обозначим через \mathbf{u}^* . Это набор управлений, рассчитанный для игры, начинающейся с вершины n^0 из состояния x^0 :

$$\mathbf{u}^* = \arg \max_{(\mathbf{u}_j: j \in M)} \sum_{j \in M} W_j(\mathbf{u}, x^0). \quad (3.8.7)$$

Кооперативная траектория состояний, определяемая единственным образом по стратегиям \mathbf{u}^* , есть $\mathbf{x}^* = \{x^*(n_l^t) : n_l^t \in \mathcal{N}^t, t = 0, 1, \dots, T\}$.

Замечание 3.4. В дополнение к замечанию 3.2 предположим, что для заданного набора параметров игры с линейной функцией динамики состояния (а именно, для матриц $A(a(n))$, $B_j(a(n))$, $C_{1j}(n)$, $C_{2j}(n)$ и векторов $c_j(n)$, $d_j(n)$ для всех вершин $n \in \mathcal{N}$ и любого игрока $j \in M$) наборы стратегий в любой вершине n дерева событий, являющиеся компонентами S -адаптивного равно-

весия по Нэшу и решения задачи максимизации суммарного дисконтированного выигрыша, принадлежат множеству U^n . Также получаемые из уравнения (3.8.1) единственным образом состояния $x(n)$ в каждой вершине n принадлежат множеству X .

§ 3.8.2 Динамические игры, разыгрываемые на бинарных деревьях событий, с симметричными игроками

Для того, чтобы получить аналитический вид оптимальных стратегий в явном виде, рассмотрим случай бинарного дерева, для вершин которого справедливо следующее правило: любая нетерминальная вершина имеет два ребра, т. о. из каждой вершины дерева событий возможна реализация двух событий в следующий период времени, причем условные вероятности реализации этих сценариев не зависят от вершины. В параграфе § 3.8 делается предположение о линейности по состоянию системы, а именно, правая часть уравнения динамики состояния и функции выигрышей игроков — линейные функции состояния. Также предположим, что множество игроков однородно, т.е. все игроки симметричны и имеют одинаковые функции выигрышей и множества стратегий. Эти условия помогут вычислить в явном виде цену анархии как меру различия в поведении игроков (кооперативное или некооперативное), а также проанализировать эту величину в зависимости от параметров модели.

Итак, каждая вершина в любой период времени $t = 1, \dots, T - 1$ имеет два ребра. Условная вероятность реализации левого ребра равна p и правого ребра — $1 - p$. Например, в период времени $t = 1$ существует две вершины n_1^1 и n_2^1 с вероятностями реализации $\pi(n_1^1) = p$ и $\pi(n_2^1) = 1 - p$ соответственно. Вершина n_1^1 имеет два ребра, т.е. две вершины n_1^2 и n_2^2 следуют за ней с вероятностями реализации $\pi(n_1^2) = p^2$ и $\pi(n_2^2) = p(1 - p)$. Бинарное дерево игры с T периодами содержит $2^{T+1} - 1$ вершин, включая корень дерева.

Уравнения динамики состояния будут следующими:

$$x(n_i^t) = \begin{cases} A_0 x(a(n_i^t)) + \sum_{j \in M} B_0 u_j(a(n_i^t)), & \text{если } a(n_i^t) = n^0, \\ A_1 x(a(n_i^t)) + \sum_{j \in M} B_1 u_j(a(n_i^t)), & \text{если } a(n_i^t) = n_k^{t-1} \text{ и} \\ & k = 2z + 1, z = 0, 1, \dots, \\ A_2 x(a(n_i^t)) + \sum_{j \in M} B_2 u_j(a(n_i^t)), & \text{если } a(n_i^t) = n_k^{t-1} \text{ и} \\ & k = 2z, z = 1, 2, \dots, \end{cases} \quad (3.8.8)$$

при этом $u_j(a(n_i^t)) \in U_j^{a(n_i^t)}$ для любой вершины $n_i^t \in \mathcal{N}^t, t = 1, \dots, T$.

Так как по предположению игроки симметричны, то функция выигрыша игрока $j \in M$ в любой нетерминальной вершине n_i^t имеет вид:

$$\phi_j^{n_i^t}(x(n_i^t), u(n_i^t)) = \frac{1}{2} u'_j(n_i^t) C_1 u_j(n_i^t) + \sum_{k \neq j} u'_k(n_i^t) C_2 u_j(n_i^t) + c' u_j(n_i^t) + d' x(n_i^t), \quad (3.8.9)$$

где C_1, C_2, c и d не зависят от вершины $n_i^t \in \mathcal{N}^t$ и от номера игрока.

В терминальной вершине n_i^T выигрыш игрока j определяется функцией

$$\Phi_j^{n_i^T}(x(n_i^T)) = d' x(n_i^T). \quad (3.8.10)$$

Симметричные игроки имеют одинаковый дисконт фактор выигрышей, равный $\varrho_j = \varrho, j \in M$.

§ 3.8.3 S -адаптивное равновесие по Нэшу в динамических играх, разыгрываемых на деревьях событий

Найдем S -адаптивное равновесие по Нэшу в игре, разыгрываемой на бинарном дереве событий с симметричными игроками, описанной в параграфе § 3.8.2. Игроки стремятся максимизировать функции выигрышей (3.8.3), где ϕ_j и Φ_j определены формулами (3.8.9) и (3.8.10) при условии, что выполнены условия (3.8.5), (3.8.6) и (3.8.8).

Сформулируем необходимые и достаточные условия для нахождения S -адаптивного равновесия по Нэшу с использованием принципа максимума Понтрягина [5]. Для игрока j определим функцию Гамильтона для корня дерева n^0 и

любой другой вершины $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T - 1$:

$$\begin{aligned}
 H_j^{n^0} &= \frac{1}{2} u'_j(n^0) C_1 u_j(n^0) + \sum_{k \neq j} u'_k(n^0) C_2 u_j(n^0) + c' u_j(n^0) + d' x^0 + \\
 &+ \varrho [p \lambda'_j(n_1^1) + (1 - p) \lambda'_j(n_2^1)] \left[A_0 x^0 + \sum_{j \in M} B_0 u_j(n^0) \right], \\
 H_j^{n_l^t} &= \frac{1}{2} u'_j(n_l^t) C_1 u_j(n_l^t) + \sum_{k \neq j} u'_k(n_l^t) C_2 u_j(n_l^t) + c' u_j(n_l^t) + d' x(n_l^t) + \\
 &+ \varrho [p \lambda'_j(n_1^{t+1}) + (1 - p) \lambda'_j(n_2^{t+1})] \left[A_l x(n_l^t) + \sum_{j \in M} B_l u_j(n_l^t) \right], \quad l = 1, 2, \\
 &j \in M,
 \end{aligned}$$

где $n_l^t \in \{n_1^t, n_2^t\}$, индекс $l = 1$ соответствует любой нечетной вершине уровня t , а индекс $l = 2$ соответствует любой четной вершине уровня t , а также $\lambda_j(n_l^t)$ — множители Лагранжа из пространства \mathbb{R}^q .

Теорема 3.4. *Если матрица $C_1 + (m - 1)C_2$ обратима, то стратегии*

$$u_j^{tN}(n^0) = - \left[c' + \varrho d' \sum_{\tau=0}^{T-1} (\varrho A)^\tau B_0 \right] (C_1 + (m - 1)C_2)^{-1}, \quad (3.8.11)$$

$$\begin{aligned}
 u_j^{tN}(n_l^t) &= - \left[c' + \varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_l \right] (C_1 + (m - 1)C_2)^{-1}, \quad (3.8.12) \\
 &t = 1, \dots, T - 1, \quad l = 1, 2, \quad j \in M,
 \end{aligned}$$

образуют S -адаптивное равновесие по Нэшу в динамической игре, разыгрываемой на бинарном дереве с симметричными игроками, где $A = pA_1 + (1 - p)A_2$. Соответствующие множители Лагранжа имеют вид:

$$\lambda_j^{tN}(n_l^t) = d' \left[I + \varrho \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau A_l \right], \quad t = 1, \dots, T - 1, \quad (3.8.13)$$

$$\lambda_j^{tN}(n_l^T) = d', \quad (3.8.14)$$

где I — единичная матрица размерности q . При этом выигрыш игрока $j \in M$

в равновесии по Нэшу равен

$$\begin{aligned}
J_j(\mathbf{x}^N, \mathbf{u}^N) = & d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 + u_j^N(n^0) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n^0) + \\
& + \left[c' + m \varrho d' \sum_{t=0}^{T-1} (\varrho A)^t B_0 \right] u_j^N(n^0) + \\
& + \sum_{t=1}^{T-1} \varrho^t \left(p u_j^N(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_1^t) + \right. \\
& \quad \left. + (1-p) u_j^N(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_2^t) \right) + \\
& + \sum_{t=1}^{T-1} \varrho^t \left(p \left[c' + m \varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_1 \right] u_j^N(n_1^t) + \right. \\
& \quad \left. + (1-p) \left[c' + m \varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_2 \right] u_j^N(n_2^t) \right). \quad (3.8.15)
\end{aligned}$$

Доказательство. Следующие условия должны быть выполнены для существования S -адаптивного равновесия по Нэшу для любого $l = 1, 2$:

$$\begin{aligned}
0 &= \frac{\partial H_j^{n^0}}{\partial u_j(n^0)} = u_j'(n^0) C_1 + \sum_{k \neq j} u_k'(n^0) C_2 + c' + \\
& \quad + \varrho [p \lambda_j'(n_1^1) + (1-p) \lambda_j'(n_2^1)] B_0, \\
0 &= \frac{\partial H_j^{n_i^t}}{\partial u_j(n_i^t)} = u_j'(n_i^t) C_1 + \sum_{k \neq j} u_k'(n_i^t) C_2 + c' + \\
& \quad + \varrho [p \lambda_j'(n_1^{t+1}) + (1-p) \lambda_j'(n_2^{t+1})] B_l, \quad n_i^t \in \mathcal{N}^t, \quad t = 1, \dots, T-1, \\
\lambda_j'(n_i^t) &= \frac{\partial H_j^{n_i^t}}{\partial x(n_i^t)} = d' + \varrho [p \lambda_j'(n_1^{t+1}) + (1-p) \lambda_j'(n_2^{t+1})] A_l, \\
& \quad n_i^t \in \mathcal{N}^t, \quad t = 1, \dots, T-1, \\
\lambda_j'(n_i^T) &= \frac{\partial \Phi_j^{n_i^T}}{\partial x(n_i^T)} = d', \quad n_i^T \in \mathcal{N}^T
\end{aligned}$$

с начальным условием на состояние системы: $x(n^0) = x^0$.

Из двух последних уравнений и структуры бинарного дерева получаем рекуррентным образом выражения для множителей Лагранжа в виде (3.8.13) и (3.8.14). Далее подставим эти выражения из (3.8.13) и (3.8.14) в первые два уравнения и решая уравнения $\partial H_j^{n^0} / \partial u_j(n^0) = 0$ и $\partial H_j^{n_i^t} / \partial u_j(n_i^t) = 0$, мы по-

лучим выражения для оптимальных стратегий в S -адаптивном равновесии по Нэшу, определенном в (3.8.11) и (3.8.12). Можно заметить, что стратегия игрока зависит от периода времени и типа вершины, т.е. она зависит от того, является эта вершина с четным или нечетным индексом.

Выигрыш игрока $j \in M$ в S -адаптивном равновесии по Нэшу вычисляется по формуле (3.8.3). Подставляя в правую часть (3.8.3) выражения для S -адаптивных равновесных по Нэшу стратегий и соответствующую траекторию состояний, получаем:

$$\begin{aligned}
J_j(\mathbf{x}^N, \mathbf{u}^N) &= \sum_{t=0}^{T-1} \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \left[\frac{1}{2} u_j'^N(n_i^t) C_1 u_j^N(n_i^t) + \sum_{k \neq j} u_k'^N(n_i^t) C_2 u_j^N(n_i^t) + \right. \\
&\quad \left. + c' u_j^N(n_i^t) + d' x^N(n_i^t) \right] + \varrho^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) d' x^N(n_i^T) = \\
&= \sum_{t=0}^{T-1} \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \left[u_j'^N(n_i^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_i^t) + c' u_j^N(n_i^t) \right. \\
&\quad \left. + d' x^N(n_i^t) \right] + \varrho^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) d' x^N(n_i^T) \\
&= u_j'^N(n^0) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n^0) + c' u_j^N(n^0) + \\
&\quad + \sum_{t=1}^{T-1} \varrho^t \left[p u_j'^N(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_1^t) + \right. \\
&\quad \left. + (1-p) u_j'^N(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_2^t) \right] + \\
&\quad + c' \sum_{t=1}^{T-1} \varrho^t [p u_j^N(n_1^t) + (1-p) u_j^N(n_2^t)] + \sum_{t=0}^T \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) d' x^N(n_i^t) \\
&= u_j'^N(n^0) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n^0) + c' u_j^N(n^0) + \\
&\quad + \sum_{t=1}^{T-1} \varrho^t \left[p u_j'^N(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_1^t) + \right. \\
&\quad \left. + (1-p) u_j'^N(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_2^t) \right] +
\end{aligned}$$

$$\begin{aligned}
& +c' \sum_{t=1}^{T-1} \varrho^t [pu_j^N(n_1^t) + (1-p)u_j^N(n_2^t)] + \\
& +d' \left[x^0 + \varrho \sum_{t=0}^{T-1} (\varrho A)^t [A_0 x^0 + mB_0 u_j^N(n^0)] + \right. \\
& \quad \left. + m \sum_{t=2}^T \varrho^t \sum_{\tau=0}^{t-2} A^\tau [pB_1 u_j^N(n_1^{t-\tau-1}) + (1-p)B_2 u_j^N(n_2^{t-\tau-1})] \right] = \\
= & d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 + u_j^N(n^0) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^N(n^0) + \\
& + \left[c' + m\varrho d' \sum_{t=0}^{T-1} (\varrho A)^t B_0 \right] u_j^N(n^0) + \\
& + \sum_{t=1}^{T-1} \varrho^t \left(pu_j^N(n_1^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^N(n_1^t) + \right. \\
& \quad \left. + (1-p)u_j^N(n_2^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^N(n_2^t) \right) + \\
& + \sum_{t=1}^{T-1} \varrho^t \left(p \left[c' + m\varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_1 \right] u_j^N(n_1^t) + \right. \\
& \quad \left. + (1-p) \left[c' + m\varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_2 \right] u_j^N(n_2^t) \right).
\end{aligned}$$

Теорема доказана. □

Замечание 3.5. Отметим три факта об S -адаптивном равновесии по Нэшу:

1. Множители Лагранжа напрямую не зависят от количества игроков.
2. Если A_1 , A_2 и d отрицательно определены, тогда на траекториях, состоящих только из четных или нечетных вершин, мы получаем:

$$\lambda_j^N(n_l^1) \geq \dots \geq \lambda_j^N(n_l^T), l = 1, 2.$$

При этом множитель Лагранжа игрока $j \in M$ является невозрастающей функцией периода времени.

3. Если A_1 , A_2 , B_l , c , d неотрицательны, и матрицы C_1 , C_2 отрицательно определены, тогда имеет место похожее свойство для стратегий:

$$u_j^N(n_l^1) \geq \dots \geq u_j^N(n_l^{T-1}), l = 1, 2.$$

Более того, если $B_0 = B_l$, тогда $u_j^N(n^0) \geq u_j^N(n_l^1) \geq \dots \geq u_j^N(n_l^{T-1})$.

Заметим, что предположения, сделанные в пунктах 2 и 3 замечания 3.5, являются стандартными для динамических игр, рассматриваемых в экономике и менеджменте, в том числе, в задачах охраны окружающей среды, одна из которых будет рассмотрена ниже.

§ 3.8.4 Кооперативное решение динамической игры, разыгрываемой на бинарном дереве событий

Предположим, что игроки максимизируют суммарный дисконтированный выигрыш:

$$J_M(\mathbf{x}, \mathbf{u}) = \sum_{t=0}^{T-1} \varrho^t \sum_{n_l^t \in \mathcal{N}^t} \pi(n_l^t) \phi_M^{n_l^t}(x(n_l^t), u(n_l^t)) + \varrho^T \sum_{n_l^T \in \mathcal{N}^T} \pi(n_l^T) \Phi_M^{n_l^T}(x(n_l^T)) \quad (3.8.16)$$

при условиях (3.8.5), (3.8.6), (3.8.8), где для любой вершины $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T-1$:

$$\phi_M^{n_l^t}(x(n_l^t), u(n_l^t)) = \sum_{j \in M} \left(\frac{1}{2} u'_j(n_l^t) C_1 u_j(n_l^t) + \sum_{k \neq j} u'_k(n_l^t) C_2 u_j(n_l^t) + c' u_j(n_l^t) \right) + md'x(n_l^t),$$

и для любой вершины $n_l^T \in \mathcal{N}^T$:

$$\Phi_M^{n_l^T}(x(n_l^T)) = md'x(n_l^T).$$

Для максимальной коалиции M определим функцию Гамильтона для корня дерева n^0 и для любой промежуточной вершины $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T-1$:

$$H_M^{n^0} = \sum_{j \in M} \left(\frac{1}{2} u'_j(n^0) C_1 u_j(n^0) + \sum_{k \neq j} u'_k(n^0) C_2 u_j(n^0) + c' u_j(n^0) \right) + md'x^0 + \\ + \varrho [p \lambda'_M(n_1^1) + (1-p) \lambda'_M(n_2^1)] \left[A_0 x^0 + \sum_{j \in M} B_0 u_j(n^0) \right],$$

$$H_M^{n_l^t} = \sum_{j \in M} \left(\frac{1}{2} u'_j(n_l^t) C_1 u_j(n_l^t) + \sum_{k \neq j} u'_k(n_l^t) C_2 u_j(n_l^t) + c' u_j(n_l^t) \right) + m d' x(n_l^t) + \varrho [p \lambda'_M(n_1^{t+1}) + (1-p) \lambda'_M(n_2^{t+1})] \left[A_l x(n_l^t) + \sum_{j \in M} B_l u_j(n_l^t) \right], l = 1, 2.$$

Теорема 3.5. Если матрица $C_1 + (m-1)(C_2 + C'_2)$ обратима, стратегии

$$u_j^*(n^0) = - \left[c' + \varrho m d' \sum_{\tau=0}^{T-1} (\varrho A)^\tau B_0 \right] (C_1 + (m-1)(C_2 + C'_2))^{-1}, \quad (3.8.17)$$

$$u_j^*(n_l^t) = - \left[c' + \varrho m d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_l \right] (C_1 + (m-1)(C_2 + C'_2))^{-1}, \quad (3.8.18)$$

где $t = 1, \dots, T-1$, $l = 1, 2$, образуют кооперативное решение в динамической игре, разыгрываемой на бинарном дереве событий с симметричными игроками. Соответствующие множители Лагранжа равны

$$\lambda_M^*(n_l^t) = m d' \left[I + \varrho \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau A_l \right], \quad t = 1, \dots, T-1, \quad (3.8.19)$$

$$\lambda_M^*(n_l^T) = m d'. \quad (3.8.20)$$

При этом выигрыш игрока $j \in M$ при кооперативном решении равен

$$J_j(\mathbf{x}^*, \mathbf{u}^*) = d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 - u_j^*(n^0) \left[\frac{1}{2} C_1 + (m-1) C'_2 \right] u_j^*(n^0) - \sum_{t=1}^{T-1} \varrho^t \left(p u_j^*(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C'_2 \right] u_j^*(n_1^t) + (1-p) u_j^*(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C'_2 \right] u_j^*(n_2^t) \right). \quad (3.8.21)$$

Доказательство. Следующие условия должны выполняться, чтобы существо-

вали кооперативные стратегии, т. е. стратегии из (3.8.7) для любого $l = 1, 2$:

$$0 = \frac{\partial H_M^{n^0}}{\partial u_j(n^0)} = u'_j(n^0)C_1 + \sum_{k \neq j} u'_k(n^0)(C_2 + C'_2) + c' + \\ + \varrho [p\lambda'_M(n_1^1) + (1-p)\lambda'_M(n_2^1)] B_0, \\ 0 = \frac{\partial H_M^{n_l^t}}{\partial u_j(n_l^t)} = u'_j(n_l^t)C_1 + \sum_{k \neq j} u'_k(n_l^t)(C_2 + C'_2) + c' + \\ + \varrho [p\lambda'_M(n_1^{t+1}) + (1-p)\lambda'_M(n_2^{t+1})] B_l, \\ n_l^t \in \mathcal{N}^t, \quad t = 1, \dots, T-1,$$

$$\lambda'_M(n_l^t) = \frac{\partial H_M^{n_l^t}}{\partial x(n_l^t)} = md' + \varrho [p\lambda'_M(n_1^{t+1}) + (1-p)\lambda'_M(n_2^{t+1})] A_l, \\ n_l^t \in \mathcal{N}^t, \quad t = 1, \dots, T-1, \\ \lambda'_M(n_l^T) = \frac{\partial \Phi_M^{n_l^T}}{\partial x(n_l^T)} = md', \quad n_l^T \in \mathcal{N}^T$$

с начальным условием на состояние системы $x(n^0) = x^0$.

Из последних двух уравнений и структуры бинарного дерева, можно рекуррентно получить выражения для множителей Лагранжа форм (3.8.19) и (3.8.20). Подставляя выражения для множителей Лагранжа из (3.8.19) и (3.8.20) и решая систему первых двух уравнений $\partial H_M^{n^0}/\partial u_j(n^0) = 0$ и $\partial H_M^{n_l^t}/\partial u_j(n_l^t) = 0$ для всех вершин $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T-1$, получаем оптимальные стратегии игроков вида (3.8.17) и (3.8.18). Можно заметить, что кооперативная стратегия игрока зависит от периода времени и от типа вершины (четная или нечетная), т. е. от того, в четной или нечетной вершине в текущий момент времени находится игровой процесс.

Суммарный выигрыш любого игрока $j \in M$ в кооперативной ситуации определим по формуле (3.8.3), подставляя в него кооперативные стратегии и соответствующую траекторию состояний. Он будет равен

$$J_j(\mathbf{x}^*, \mathbf{u}^*) = \sum_{t=0}^{T-1} \varrho^t \sum_{n_l^t \in \mathcal{N}^t} \pi(n_l^t) \left[\frac{1}{2} u_j^*(n_l^t) C_1 u_j^*(n_l^t) + \sum_{k \neq j} u_k^*(n_l^t) C_2 u_j^*(n_l^t) + \right. \\ \left. + c' u_j^*(n_l^t) + d' x^*(n_l^t) \right] + \varrho^T \sum_{n_l^T \in \mathcal{N}^T} \pi(n_l^T) d' x^*(n_l^T) =$$

$$\begin{aligned}
&= \sum_{t=0}^{T-1} \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \left[u_j^*(n_i^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_i^t) + \right. \\
&\quad \left. + c' u_j^*(n_i^t) + d' x^*(n_i^t) \right] + \varrho^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) d' x^*(n_i^T) = \\
&= u_j^*(n^0) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n^0) + c' u_j^*(n^0) + \\
&\quad + \sum_{t=1}^{T-1} \varrho^t \left[p u_j^*(n_1^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_1^t) \right. \\
&\quad \quad \left. + (1-p) u_j^*(n_2^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_2^t) \right] + \\
&\quad + c' \sum_{t=1}^{T-1} \varrho^t [p u_j^*(n_1^t) + (1-p) u_j^*(n_2^t)] + \sum_{t=0}^T \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) d' x^*(n_i^t) \\
&= u_j^*(n^0) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n^0) + c' u_j^*(n^0) + \\
&\quad + \sum_{t=1}^{T-1} \varrho^t \left[p u_j^*(n_1^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_1^t) + \right. \\
&\quad \quad \left. + (1-p) u_j^*(n_2^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_2^t) \right] + \\
&\quad + c' \sum_{t=1}^{T-1} \varrho^t [p u_j^*(n_1^t) + (1-p) u_j^*(n_2^t)] + \\
&\quad + d' \left[x^0 + \varrho \sum_{t=0}^{T-1} (\varrho A)^t [A_0 x^0 + m B_0 u_j^*(n^0)] + \right. \\
&\quad \quad \left. + m \sum_{t=2}^T \varrho^t \sum_{\tau=0}^{t-2} A^\tau [p B_1 u_j^*(n_1^{t-\tau-1}) + (1-p) B_2 u_j^*(n_2^{t-\tau-1})] \right] \\
&= d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 + u_j^*(n^0) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n^0) \\
&\quad + \left[c' + m \varrho d' \sum_{t=0}^{T-1} (\varrho A)^t B_0 \right] u_j^*(n^0) + \\
&\quad + \sum_{t=1}^{T-1} \varrho^t \left(p u_j^*(n_1^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_1^t) + \right. \\
&\quad \quad \left. + (1-p) u_j^*(n_2^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_2^t) \right) +
\end{aligned}$$

$$\begin{aligned}
& + \sum_{t=1}^{T-1} \varrho^t \left(p \left[c' + m\varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_1 \right] u_j^*(n_1^t) + \right. \\
& \quad \left. + (1-p) \left[c' + m\varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_2 \right] u_j^*(n_2^t) \right) = \\
= & d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 + u_j^*(n^0) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^*(n^0) - \\
& - u_j^*(n^0) [C_1 + (m-1)(C_2 + C'_2)] u_j^*(n^0) + \\
& + \sum_{t=1}^{T-1} \varrho^t \left(p u_j^*(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^*(n_1^t) + \right. \\
& \quad \left. + (1-p) u_j^*(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^*(n_2^t) \right) - \\
& - \sum_{t=1}^{T-1} \varrho^t \left(p u_j^*(n_1^t) [C_1 + (m-1)(C_2 + C'_2)] u_j^*(n_1^t) + \right. \\
& \quad \left. + (1-p) u_j^*(n_2^t) [C_1 + (m-1)(C_2 + C'_2)] u_j^*(n_2^t) \right) = \\
= & d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 - u_j^*(n^0) \left[\frac{1}{2} C_1 + (m-1) C'_2 \right] u_j^*(n^0) - \\
& - \sum_{t=1}^{T-1} \varrho^t \left(p u_j^*(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C'_2 \right] u_j^*(n_1^t) + \right. \\
& \quad \left. + (1-p) u_j^*(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C'_2 \right] u_j^*(n_2^t) \right)
\end{aligned}$$

Теорема доказана. □

Замечание 3.6. *Справедливы три утверждения о кооперативном равновесии:*

1. *Если A_1 , A_2 и d неотрицательны, тогда на траекториях, состоящих либо только из четных вершин, либо только из нечетных вершин, имеет место неравенство:*

$$\lambda_j^*(n_l^1) \geq \dots \geq \lambda_j^*(n_l^T), l = 1, 2.$$

При этом, множитель Лагранжа игрока $j \in M$ — невозрастающая функция времени.

2. Если A_1, A_2, B_l, c, d неотрицательны, и C_1, C_2 отрицательно определены, тогда аналогичное свойство справедливо и для кооперативных стратегий:

$$u_j^*(n_l^1) \geq \dots \geq u_j^*(n_l^{T-1}), l = 1, 2.$$

Более того, если $B_0 = B_l$, тогда $u_j^*(n^0) \geq u_j^*(n_l^1) \geq \dots \geq u_j^*(n_l^{T-1})$.

3. Для любой вершины $n \in \mathcal{N} \setminus \mathcal{N}^T$ и любого игрока $j \in M$ кооперативные стратегии $u_j^*(n)$ и стратегии, определяющие S -адаптивное равновесие по Нэшу, $u_j^N(n)$ удовлетворяют линейному матричному уравнению:

$$u_j^*(n) = [mu_j'^N(n) (C_1 + (m-1)C_2) + (m-1)c'] \cdot (C_1 + (m-1)(C_2 + C_2'))^{-1}. \quad (3.8.22)$$

Выразив из (3.8.22) S -адаптивную равновесную стратегию игрока через кооперативную и подставив это выражение в (3.8.3), получим выигрыш любого игрока $j \in M$ при S -адаптивном равновесии по Нэшу в альтернативной форме:

$$\begin{aligned} J_j(\mathbf{x}^N, \mathbf{u}^N) &= d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 + u_j'^N(n^0) \left[\frac{1}{2} C_1 + (m-1)C_2 \right] u_j^N(n^0) - \\ &\quad - u_j'^*(n^0) [C_1 + (m-1)(C_2 + C_2')] u_j^N(n^0) + \\ &\quad + \sum_{t=1}^{T-1} \varrho^t \left(pu_j'^N(n_1^t) \left[\frac{1}{2} C_1 + (m-1)C_2 \right] u_j^N(n_1^t) + \right. \\ &\quad \left. + (1-p)u_j'^N(n_2^t) \left[\frac{1}{2} C_1 + (m-1)C_2 \right] u_j^N(n_2^t) \right) - \\ &\quad - \sum_{t=1}^{T-1} \varrho^t \left(pu_j'^*(n_1^t) [C_1 + (m-1)(C_2 + C_2')] u_j^N(n_1^t) + \right. \\ &\quad \left. + (1-p)u_j'^*(n_2^t) [C_1 + (m-1)(C_2 + C_2')] u_j^N(n_2^t) \right). \end{aligned}$$

При этом выигрыш игрока при S -адаптивном равновесии по Нэшу является функцией как S -адаптивных равновесных по Нэшу, так и кооперативных стратегий.

§ 3.8.5 Цена анархии в динамической игре, разыгрываемой на бинарном дереве событий

В этом разделе сравним выигрыши игроков при S -адаптивном равновесии по Нэшу и кооперативном равновесии. В качестве меры разности между выигрышами игроков в этих двух ситуациях будем использовать цену анархии [117].

Определение 3.10. В динамической игре, разыгрываемой на дереве событий, с функцией выигрыша j -го игрока $W_j(\mathbf{u}, x^0) = J_j(\mathbf{x}, \mathbf{u})$, определяемой по формуле (3.8.3) при условиях (3.8.4), (3.8.5), (3.8.6), цена анархии в задаче максимизации вычисляется по формуле:

$$PoA = \frac{\max_{\mathbf{u}} \sum_{j \in M} W_j(\mathbf{u}, x^0)}{\sum_{j \in M} W_j(\mathbf{u}^N, x^0)}.$$

В случае максимизации выигрышей, бинарного дерева и симметричных игроков, цена анархии вычисляется по формуле:

$$PoA = \frac{J_j(\mathbf{x}^*, \mathbf{u}^*)}{J_j(\mathbf{x}^N, \mathbf{u}^N)}, \quad (3.8.23)$$

где $J_j(\mathbf{x}^*, \mathbf{u}^*)$ — выигрыш любого игрока $j \in M$ в кооперативном равновесии, определяемый по формуле (3.8.21), и $J_j(\mathbf{x}^N, \mathbf{u}^N)$ — выигрыш любого игрока $j \in M$ в S -адаптивном равновесии по Нэшу, вычисляемый по формуле (3.8.15).

Очевидно, что цена анархии не меньше единицы, если выигрыши $J_j(\mathbf{x}^*, \mathbf{u}^*)$ и $J_j(\mathbf{x}^N, \mathbf{u}^N)$ неотрицательны, и ее можно вычислить в явном виде.

Замечание 3.7. Предложим другую формулу для вычисления цены анархии. Зная $u_j^N(n_l^t)$ и $u_j^*(n_l^t) \in \mathbb{R}^r$, $t = 0, \dots, T-1$ и $l = 1, 2$, сформируем вектора кооперативных стратегий в виде:

$$\mathbf{u}_j^*(l) = (u_j^*(n^0), u_j^*(n_l^1), \dots, u_j^*(n_l^{T-1})) \in \mathbb{R}^{rT}, \quad l = 1, 2,$$

и вектора, формирующие S -адаптивное равновесие по Нэшу, в виде:

$$\mathbf{u}_j^N(l) = (u_j^N(n^0), u_j^N(n_l^1), \dots, u_j^N(n_l^{T-1})) \in \mathbb{R}^{rT}, \quad l = 1, 2.$$

Обозначив через $C = \frac{1}{2}C_1 + (m-1)C'_2$, сформируем блочно-диагональную матрицу \mathbf{C} размерности $rT \times rT$:

$$\mathbf{C} = \begin{pmatrix} C & 0 & \dots & 0 \\ 0 & C & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C \end{pmatrix}.$$

Таким образом, получаем формулу для вычисления цены анархии:

$$PoA = \frac{p\mathbf{u}'_j^*(1)\mathbf{C}\mathbf{u}_j^*(1) + (1-p)\mathbf{u}'_j^*(2)\mathbf{C}\mathbf{u}_j^*(2)}{p\Psi(1) + (1-p)\Psi(2)},$$

где

$$\Psi(k) = \mathbf{u}'_j^*(k)\mathbf{C}\mathbf{u}_j^*(k) - (\mathbf{u}_j^*(k) - \mathbf{u}_j^N(k))' \mathbf{C} (\mathbf{u}_j^*(k) - \mathbf{u}_j^N(k)), \quad k = 1, 2.$$

§ 3.9 Цена анархии в одной игре охраны окружающей среды

§ 3.9.1 Модель

В этом разделе проиллюстрируем результаты предыдущего параграфа на примере игры совместного загрязнения окружающей среды m странами, которые являются игроками. Эти страны совместно загрязняют одну географическую территорию.

В последние три десятилетия значительная часть литературы по математическому моделированию посвящена вопросам, связанным с управлением глобальным загрязнением окружающей среды. Предполагается, что выбросы участников (фирмы, регионы, страны) влияют на общие с другими участниками территории и на издержки всех игроков, участников конфликта. В такой постановке задачи было показано, что, когда игроки выбирают уровни выбросов некооперативно, каждый игрок загрязняет окружающую среду больше, чем если бы они играли в кооперации. Основная причина этого заключается в том, что при кооперации (совместной оптимизации) каждый игрок поглощает издержки всех игроков, в то время как если игрок ведет себя индивидуально, он учитывает только свои издержки.

Рассмотрим простую модель, которая хорошо вписывается в общий контекст имеющейся литературы, например, см. [109] и [125] для изучения приложений динамических игр в области управления загрязнением. Представленная в этом разделе модель предполагает наличие стандартных для этой области моделирования предположений: 1) загрязнения необходимы для производства товаров или услуг, они являются сопутствующими при получении доходов; 2) загрязнения со временем аккумулируются и «запас» загрязняющего вещества вредит окружающей среде и приносит издержки всем игрокам; 3) природа «очищается сама», т. е. часть загрязняющего вещества уничтожается сама собой, но уровень, с которым океаны или леса самоочищаются со временем, неизвестен; 4) каждый игрок стремится максимизировать дисконтированную сумму выигрышей, получаемую за некоторый конечный промежуток, и в каждый момент времени выигрыш игрока равен разности прибыли и издержек.

Предположим, что игра разыгрывается на бинарном дереве с T периодами, начиная с периода $t = 0$ и с условной вероятностью p реализации нечетной вершины и $1 - p$ реализации четной вершины.

Пусть $u_j(n_i^t) \in \mathbb{R}_+$ — выбросы игрока j в вершине $n_i^t \in \mathcal{N}^t$, и $x(n_i^t) \in \mathbb{R}_+$ — «запас» загрязняющего вещества с начальным состоянием $x(n^0) = 0$.⁴ Уравнение состояния для любой вершины $n_i^t \in \mathcal{N}^t$, $t = 1, \dots, T$ имеет форму:

$$x(n_i^t) = \begin{cases} \sum_{j \in M} u_j(a(n_i^t)) & \text{если } a(n_i^t) = n^0, \\ \delta_1 x(a(n_i^t)) + \sum_{j \in M} u_j(a(n_i^t)) & \text{если } a(n_i^t) \text{ — нечетная вершина,} \\ \delta_2 x(a(n_i^t)) + \sum_{j \in M} u_j(a(n_i^t)) & \text{если } a(n_i^t) \text{ — четная вершина,} \end{cases} \quad (3.9.1)$$

где δ_1 и δ_2 — положительные константы из единичного интервала, представляющие естественные уровни абсорбции для нечетных и четных вершин соответственно.

Если предположить, что зависимость между уровнем производства и уровнем загрязнения игрока j монотонно возрастающая, то выигрыш игрока j является функцией выбросов $f_j(u)$ всех игроков. «Запас» загрязняющего вещества наносит вред окружающей среде, который примерно равен $\kappa x(n_i^t)$, где $\kappa > 0$.

⁴Предположение о положительном начальном состоянии не внесет значительных изменений при нахождении решения.

Мы рассматриваем линейные издержки, но могут быть рассмотрены и другие функциональные зависимости. Предположим, что имеется взаимосвязь между загрязнениями игроков в функции выигрыша. Представленная модель соответствует модели олигополии Курно.

Рассмотрим индустрию из m олигополистических фирм, производящих однородный продукт и конкурирующих по Курно, имеющих линейные обратные функции спроса:

$$\mathcal{P}(n_l^t) = \left(\alpha - \gamma \sum_{i \in M} u_i(n_l^t) \right). \quad (3.9.2)$$

Доход фирмы в вершине n_l^t , $t = 0, 1, \dots, T - 1$, определяется функцией

$$f_j^{n_l^t}(u(n_l^t)) = u_j(n_l^t) \left(\alpha - \gamma \sum_{i \in M} u_i(n_l^t) \right), \quad (3.9.3)$$

и выигрыш (разность дохода и издержек) в этой вершине будет равен

$$\phi_j^{n_l^t}(x(n_l^t), u_j(n_l^t)) = u_j(n_l^t) \left(\alpha - \gamma \sum_{i \in M} u_i(n_l^t) \right) - \kappa x(n_l^t), \quad (3.9.4)$$

где α , γ и κ — положительные константы. В терминальной вершине n_l^T выигрыш игрока j определяется функцией

$$\Phi_j^{n_l^T}(x(n_l^T)) = -\kappa x(n_l^T). \quad (3.9.5)$$

Замечание 3.8. В литературе часто делается предположение о том, что функция дохода зависит только от выбросов, т.е. эта функция имеет форму $f_j(u_j)$. Но мы рассмотрим более общий случай, когда функция дохода зависит от стратегий других игроков.

§ 3.9.2 Основные результаты

Из Теорем 3.4 и 3.5 мы получаем стратегии, определяющие S -адаптивное равновесие по Нэшу и кооперативное равновесие:

$$u_j^N(n_l^t) = \frac{1}{(m+1)\gamma} (\alpha - \rho\kappa\Delta(t)), \quad (3.9.6)$$

$$u_j^*(n_l^t) = \frac{1}{2m\gamma} (\alpha - m\rho\kappa\Delta(t)), \quad (3.9.7)$$

для любого игрока $j \in M$, $n_i^t \in \mathcal{N}^t$, $t = 0, 1, \dots, T-1$, где

$$\Delta(t) = \frac{1 - [\varrho(p\delta_1 + (1-p)\delta_2)]^{T-t}}{1 - \varrho(p\delta_1 + (1-p)\delta_2)}.$$

Разность между стратегией, составляющей S -адаптивное равновесие по Нэшу, и кооперативной стратегией равна

$$u_j^N(n_i^t) - u_j^*(n_i^t) = \frac{m-1}{2m(m+1)\gamma} (\alpha + m\varrho\kappa\Delta(t)), \quad (3.9.8)$$

которая положительна в каждой вершине бинарного дерева.

Из Теорем 3.4 и 3.5 получим выражения для выигрышей любого игрока $j \in M$ в S -адаптивном равновесии по Нэшу и кооперативном решении:

$$\begin{aligned} J_j(\mathbf{x}^N, \mathbf{u}^N) &= \sum_{t=0}^{T-1} \varrho^t [m^2\gamma u_j^N(n_i^t) - \alpha(m-1)] u_j^N(n_i^t) = \\ &= m\gamma \sum_{t=0}^{T-1} \varrho^t [2u_j^*(n_i^t) - u_j^N(n_i^t)] u_j^N(n_i^t), \end{aligned} \quad (3.9.9)$$

$$J_j(\mathbf{x}^*, \mathbf{u}^*) = m\gamma \sum_{t=0}^{T-1} \varrho^t [u_j^*(n_i^t)]^2. \quad (3.9.10)$$

Заметив, что

$$2u_j^*(n_i^t) - u_j^N(n_i^t) = \frac{1}{m(m+1)\gamma} (\alpha - m^2\varrho\kappa\Delta(t)),$$

мы можем записать разность между выигрышами в S -адаптивном равновесии по Нэшу и кооперативном решении в виде:

$$J_j(\mathbf{x}^N, \mathbf{u}^N) - J_j(\mathbf{x}^*, \mathbf{u}^*) = -m\gamma \sum_{t=0}^{T-1} \varrho^t [u_j^N(n_i^t) - u_j^*(n_i^t)]^2, \quad (3.9.11)$$

которая всегда отрицательна.

Замечание 3.9. Так как $\Delta(t)$ убывает по t , то достаточно предположить, что $\alpha > m\varrho\kappa\Delta(0)$ для того, чтобы оптимальные стратегии для обеих равновесий были бы неотрицательны. Более того, чтобы функция $J_j(\mathbf{x}^N, \mathbf{u}^N)$ была положительна, достаточным условием является неравенство $\alpha > m^2\varrho\kappa\Delta(0)$. Будем придерживаться этого предположения и далее.

Теорема 3.6. Если $\alpha > m^2 \rho \kappa \Delta(0)$, в игре с выигрышами, определяемыми формулой (3.9.4) для нетерминальных вершин и формулой (3.9.5) для терминальных вершин и уравнениями состояний (3.9.1), тогда цена анархии PoA определяется формулой

$$PoA = 1 + \frac{(m-1)^2 \sum_{t=0}^{T-1} \varrho^t (\alpha + m \rho \kappa \Delta(t))^2}{4m \sum_{t=0}^{T-1} \varrho^t (\alpha - m^2 \rho \kappa \Delta(t)) (\alpha - \rho \kappa \Delta(t))} \quad (3.9.12)$$

и имеет следующие границы:

$$\begin{aligned} \underline{PoA} &= 1 + \frac{(m-1)^2 (\alpha + m \rho \kappa)^2}{4m (\alpha - m^2 \rho \kappa) (\alpha - \rho \kappa)} \leq PoA \leq \\ &\leq 1 + \frac{(m-1)^2 (\alpha + m \rho \kappa \Delta(0))^2}{4m (\alpha - m^2 \rho \kappa \Delta(0)) (\alpha - \rho \kappa \Delta(0))} = \overline{PoA}. \end{aligned} \quad (3.9.13)$$

Доказательство. Вычислим PoA , используя формулу (4.1.21):

$$\begin{aligned} PoA &= \frac{J_j(\mathbf{x}^*, \mathbf{u}^*)}{J_j(\mathbf{x}^N, \mathbf{u}^N)} = 1 + \frac{J_j(\mathbf{x}^*, \mathbf{u}^*) - J_j(\mathbf{x}^N, \mathbf{u}^N)}{J_j(\mathbf{x}^N, \mathbf{u}^N)} = \\ &= 1 + \frac{m\gamma \sum_{t=0}^{T-1} \varrho^t [u_j^N(n_t^t) - u_j^*(n_t^t)]^2}{m\gamma \sum_{t=0}^{T-1} \varrho^t [2u_j^*(n_t^t) - u_j^N(n_t^t)] u_j^N(n_t^t)} = \\ &= 1 + \frac{(m-1)^2 \sum_{t=0}^{T-1} \varrho^t (\alpha + m \rho \kappa \Delta(t))^2}{4m \sum_{t=0}^{T-1} \varrho^t (\alpha - m^2 \rho \kappa \Delta(t)) (\alpha - \rho \kappa \Delta(t))}. \end{aligned}$$

Здесь $u_j^N(n_t^t) - u_j^*(n_t^t)$ — убывающая функция аргумента t , а $(2u_j^*(n_t^t) - u_j^N(n_t^t))$ и $u_j^N(n_t^t)$ — возрастающие функции аргумента t . Вычислим верхнюю границу цены анархии:

$$\begin{aligned} PoA &\leq 1 + \frac{m\gamma \sum_{t=0}^{T-1} \varrho^t [u_j^N(n^0) - u_j^*(n^0)]^2}{m\gamma \sum_{t=0}^{T-1} \varrho^t [2u_j^*(n^0) - u_j^N(n^0)] u_j^N(n^0)} \\ &= 1 + \frac{m\gamma \cdot \frac{1-\varrho^T}{1-\varrho} \cdot \frac{(m-1)^2}{4m^2(m+1)^2\gamma^2} \cdot (\alpha + m \rho \kappa \Delta(0))^2}{m\gamma \frac{1-\varrho^T}{1-\varrho} \cdot \frac{(\alpha - m^2 \rho \kappa \Delta(0)) (\alpha - \rho \kappa \Delta(0))}{m(m+1)^2\gamma^2}} \\ &= 1 + \frac{(m-1)^2 (\alpha + m \rho \kappa \Delta(0))^2}{4m (\alpha - m^2 \rho \kappa \Delta(0)) (\alpha - \rho \kappa \Delta(0))}. \end{aligned}$$

Вычислим нижнюю границу цены анархии:

$$\begin{aligned}
PoA &\geq 1 + \frac{m\gamma \sum_{t=0}^{T-1} \varrho^t [u_j^N(n^{T-1}) - u_j^*(n^{T-1})]^2}{m\gamma \sum_{t=0}^{T-1} \varrho^t [2u_j^*(n^{T-1}) - u_j^N(n^{T-1})] u_j^N(n^{T-1})} \\
&= 1 + \frac{m\gamma \cdot \frac{1-\varrho^T}{1-\varrho} \cdot \frac{(m-1)^2}{4m^2(m+1)^2\gamma^2} \cdot (\alpha + m\rho\kappa\Delta(T-1))^2}{m\gamma \frac{1-\varrho^T}{1-\varrho} \cdot \frac{(\alpha - m^2\rho\kappa\Delta(T-1))(\alpha - \rho\kappa\Delta(T-1))}{m(m+1)^2\gamma^2}} \\
&= 1 + \frac{(m-1)^2(\alpha + m\rho\kappa\Delta(T-1))^2}{4m(\alpha - m^2\rho\kappa\Delta(T-1))(\alpha - \rho\kappa\Delta(T-1))}.
\end{aligned}$$

Учитывая, что $\Delta(T-1) = 1$, получим нижнюю границу цены анархии:

$$PoA \geq 1 + \frac{(m-1)^2(\alpha + m\rho\kappa)^2}{4m(\alpha - m^2\rho\kappa)(\alpha - \rho\kappa)}.$$

□

Цена анархии и ее границы зависят от параметров модели, но не зависят от γ . Напомним, что этот параметр появляется в обратной функции спроса (3.9.2), где $\mathcal{P}(n_j^t)$ интерпретируется как стоимость за единицу выбросов (или что тоже самое, производство) в вершине n_j^t . Чем больше значение γ , тем меньше цена $\mathcal{P}(n_j^t)$, которую игроки могут получить. Из (3.9.9) и (3.9.10) несложно заметить, что когда мы вычисляем отношение выигрышей для вычисления цены анархии, параметр γ исчезает. С экономической точки зрения, это связано с предположением, сделанным здесь, также как и с литературой в этой области в целом о том, что продукт является гомогенным, т.е. влияние дополнительных выбросов (производство), произведенных любым игроком, имеет тоже самое влияние на цену. Если продукт был бы несовершенен взаимозаменяемым, то, очевидно, результат был бы иным.

Следующее утверждение показывает, как другие параметры влияют на цену анархии и ее границы. Параметры динамики загрязнения (δ_1, δ_2, p) не представляют интереса для анализа, поскольку они задают стохастический процесс и не имеют прямого отношения к игрокам.

Утверждение 3.1. *Анализ параметров. Для $\alpha > m^2\rho\kappa\Delta(0)$, справедливы утверждения:*

1. $P_{\circ A}$ и $\overline{P_{\circ A}}$ — возрастающие функции по T , при этом

$$\lim_{T \rightarrow \infty} P_{\circ A} = \lim_{T \rightarrow \infty} \overline{P_{\circ A}} = 1 + \frac{(m-1)^2 (\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) + m\varrho\kappa)^2}{4m (\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - m^2\varrho\kappa) (\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - \varrho\kappa)}.$$

2. $P_{\circ A}$, $\underline{P_{\circ A}}$ и $\overline{P_{\circ A}}$ — возрастающие функции количества игроков, т.е. параметра m .

3. $P_{\circ A}$, $\underline{P_{\circ A}}$ и $\overline{P_{\circ A}}$ — убывающие функции по α .

4. $P_{\circ A}$, $\underline{P_{\circ A}}$ и $\overline{P_{\circ A}}$ — возрастающие функции по κ .

5. $P_{\circ A}$, $\underline{P_{\circ A}}$ и $\overline{P_{\circ A}}$ — возрастающие функции по ϱ .

Доказательство. 1. Сначала рассмотрим верхнюю границу из неравенства (3.9.13) как функцию параметра $\Delta(0)$. Частная производная этой границы по этому параметру равна

$$\frac{\alpha\kappa\varrho(m-1)^2(m+1)^2\Delta(0)(\alpha - m\varrho\kappa\Delta(0))(\alpha + m\varrho\kappa\Delta(0))}{4(\alpha - \varrho\kappa\Delta(0))^2(\alpha - m^2\varrho\kappa\Delta(0))}.$$

Она положительна, потому что $\alpha > m^2\varrho\kappa\Delta(0)$. Таким образом, верхняя граница — возрастающая функция параметра $\Delta(0)$. Более того, $\Delta(0)$ — возрастающая функция параметра T . Получаем, что верхняя граница — возрастающая функция параметра T . Нижняя граница цены анархии не зависит от T .

Рассмотрим предел цены анархии и ее верхнюю границу в случае, когда T стремится к бесконечности. Верхний предел равен

$$\lim_{T \rightarrow \infty} \left[1 + \frac{(m-1)^2(\alpha + m\varrho\kappa\Delta(0))^2}{4m(\alpha - m^2\varrho\kappa\Delta(0))(\alpha - \varrho\kappa\Delta(0))} \right] = 1 + \frac{(m-1)^2(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) + m\varrho\kappa)}{4m(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - m^2\varrho\kappa)(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - \varrho\kappa)},$$

учитывая, что $\lim_{T \rightarrow \infty} \Delta(0) = 1/[1 - \varrho(p\delta_1 + (1-p)\delta_2)]$ и $\varrho(p\delta_1 + (1-p)\delta_2) \in (0, 1)$.

Теперь рассмотрим цену анархии. Так как $\lim_{T \rightarrow \infty} \Delta(t) = 1/(1 - \varrho(p\delta_1 + (1-p)\delta_2))$, мы найдем предел суммы в числителе:

$$\begin{aligned} \lim_{T \rightarrow \infty} \sum_{t=0}^{T-1} \varrho^t (\alpha + m\varrho\kappa\Delta(t))^2 &= \sum_{t=0}^{\infty} \varrho^t \left(\alpha + \frac{m\varrho\kappa}{(1 - \varrho(p\delta_1 + (1-p)\delta_2))} \right)^2 \\ &= \frac{(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) + m\varrho\kappa)^2}{(1 - \varrho)(1 - \varrho(p\delta_1 + (1-p)\delta_2))^2}. \end{aligned}$$

Предел суммы в знаменателе равен

$$\begin{aligned} \sum_{t=0}^{T-1} \varrho^t (\alpha - m^2\varrho\kappa\Delta(t))(\alpha - \varrho\kappa\Delta(t)) &= \\ &= \frac{(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - m^2\varrho\kappa)(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - \varrho\kappa) \sum_{t=0}^{\infty} \varrho^t}{(1 - \varrho(p\delta_1 + (1-p)\delta_2))^2} \\ &= \frac{(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - m^2\varrho\kappa)(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - \varrho\kappa)}{(1 - \varrho)(1 - \varrho(p\delta_1 + (1-p)\delta_2))^2}. \end{aligned}$$

Учитывая найденные значения пределов, мы получаем

$$\begin{aligned} \lim_{T \rightarrow \infty} P_{oA} &= 1 + \\ &+ \frac{(m-1)^2 (\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) + m\varrho\kappa)^2}{4m (\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - m^2\varrho\kappa) (\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - \varrho\kappa)}. \end{aligned}$$

2. Рассмотрим выражение цены анархии P_{oA} , второе слагаемое в формуле следующее:

$$\frac{(m-1)^2 \sum_{t=0}^{T-1} \varrho^t (\alpha + m\varrho\kappa\Delta(t))^2}{4m \sum_{t=0}^{T-1} \varrho^t (\alpha - m^2\varrho\kappa\Delta(t))(\alpha - \varrho\kappa\Delta(t))}. \quad (3.9.14)$$

Множитель $(m-1)^2/4m$ возрастает с ростом m для $m > 1$, множитель $\sum_{t=0}^{T-1} \varrho^t (\alpha + m\varrho\kappa\Delta(t))^2$ возрастает с ростом m , и множитель $\sum_{t=0}^{T-1} \varrho^t (\alpha - m^2\varrho\kappa\Delta(t))(\alpha - \varrho\kappa\Delta(t))$ убывает с ростом m . Следовательно, цена анархии P_{oA} возрастает с ростом m . Аналогичный анализ можно провести для $\underline{P_{oA}}$ и $\overline{P_{oA}}$, и это доказывает, что $\underline{P_{oA}}$ и $\overline{P_{oA}}$ возрастают с ростом m .

3. Сначала рассмотрим цену анархии P_{oA} . Второе слагаемое (3.9.14) можно

записать следующим образом:

$$\frac{(m-1)^2 \sum_{t=0}^{T-1} \varrho^t (\alpha - m^2 \varrho \kappa \Delta(t)) (\alpha - \varrho \kappa \Delta(t)) + (m+1)^2 \alpha \varrho \kappa \sum_{t=0}^{T-1} \varrho^t \Delta(t)}{4m \sum_{t=0}^{T-1} \varrho^t (\alpha - m^2 \varrho \kappa \Delta(t)) (\alpha - \varrho \kappa \Delta(t))} = \frac{(m-1)^2}{4m} \left(1 + \frac{(m+1)^2 \alpha \varrho \kappa \sum_{t=0}^{T-1} \varrho^t \Delta(t)}{\sum_{t=0}^{T-1} \varrho^t (\alpha - m^2 \varrho \kappa \Delta(t)) (\alpha - \varrho \kappa \Delta(t))} \right).$$

Последняя дробь убывает с ростом α , следовательно, цена анархии P_{oA} строго убывает по α . Аналогичный результат справедлив и для $\underline{P_{oA}}$, и $\overline{P_{oA}}$. Нижняя и верхняя границы цены анархии также убывающие функции по α .

4. Рассмотрим цену анархии P_{oA} . В выражении P_{oA} второе слагаемое можно записать в виде (3.9.14). Числитель этой дроби возрастает по κ , а знаменатель убывает по κ . Следовательно, цена анархии P_{oA} возрастает по параметру κ . Аналогичные рассуждения можно привести для нижней и верхней границ $\underline{P_{oA}}$, $\overline{P_{oA}}$. Обе границы возрастают по κ .
5. Числитель второго слагаемого в выражении $\underline{P_{oA}}$ из (3.9.13) — строго возрастающая функция по ϱ , тогда как знаменатель — строго убывающая функция по ϱ , потому что частная производная по ϱ отрицательна для $\alpha > m^2 \varrho \kappa \Delta(0)$. Следовательно, $\underline{P_{oA}}$ — строго возрастающая функция по ϱ . Аналогичный результат имеет место и для $\overline{P_{oA}}$. Числитель второго слагаемого в выражении для $\overline{P_{oA}}$ (3.9.13) строго возрастает по $\varrho \Delta(0)$, а знаменатель строго убывает по $\varrho \Delta(0)$. Следовательно, $\overline{P_{oA}}$ — строго возрастающая функция по $\varrho \Delta(0)$ и ϱ .

Так как $(\alpha - m^2 \varrho \kappa \Delta(t)) (\alpha - \varrho \kappa \Delta(t)) = (\alpha - m \varrho \kappa \Delta(t))^2 - (m-1)^2 \alpha \varrho \kappa \Delta(t)$ и $(\alpha + m \varrho \kappa \Delta(t))^2 = (\alpha - m \varrho \kappa \Delta(t))^2 + 4m \alpha \varrho \kappa \Delta(t)$, то выражение для P_{oA}

может быть записано следующим образом:

$$PoA = \frac{\frac{(m+1)^2}{4m} \cdot \sum_{t=0}^{T-1} \varrho^t (\alpha - m\varrho\kappa\Delta(t))^2}{\sum_{t=0}^{T-1} \varrho^t [(\alpha - m\varrho\kappa\Delta(t))^2 - (m-1)^2\alpha\varrho\kappa\Delta(t)]}. \quad (3.9.15)$$

Для $\alpha > m^2\varrho\kappa\Delta(0)$, каждое слагаемое в знаменателе в выражении PoA убывает по $\varrho\Delta(t)$ (и, следовательно, по ϱ) быстрее, чем слагаемое в числителе с тем же номером. Следовательно, отношение, а именно, PoA строго возрастает по ϱ .

□

§ 3.9.3 Интерпретация результатов

Теперь приведем интерпретацию результатов.

Влияние продолжительности игры T : чем больше периодов длится игра, тем больше суммарные потери, которые игроки понесут, если они не будут кооперироваться. Так как для любой заданной продолжительности игры кооперация приводит к потерям, не большим, чем при отсутствии кооперации, то увеличивая продолжительность игры, мы тем самым увеличиваем длину интервала, которому принадлежит цена анархии. Интересно отметить, что результат пункта 1 показывает, что цена анархии ограничена сверху в случае, когда игра имеет бесконечную продолжительность. Также мы можем отметить, что нижняя граница \underline{PoA} равна цене анархии в одношаговой детерминированной игре. Нетрудно заметить, что для $T = 1$ и $t = 0$ мы получаем

$$\Delta(0) = \frac{1 - \varrho(p\delta_1 + (1-p)\delta_2)}{1 - \varrho(p\delta_1 + (1-p)\delta_2)} = 1,$$

и \underline{PoA} тогда равна PoA . В одношаговой игре дерево игры не имеет значения, потому как она происходит в начальной вершине и заканчивается. Это позволяет сделать вывод о том, что оценка влияния изменения параметров на нижнюю границу цены анархии эквивалентна анализу на чувствительность к изменению параметров цены анархии в статической игре.

Влияние количества игроков m : Результаты показывают, что чем больше количество игроков, тем больше стоимость отсутствия кооперации. Эта подобно

тому, что происходит в сетевых играх, где “эгоистическое” поведение приводит к большим перегрузкам и, следовательно, к увеличению общих затрат. В рассматриваемой постановке, когда игра играется некооперативно, каждый игрок игнорирует влияние своих выбросы на других игроков, и загрязняет больше. Следовательно, чем больше количество игроков, тем больше общее загрязнение (которое качественно аналогично увеличению перегрузок в играх маршрутизации), и тем больше издержки игрока. При кооперации негативное воздействие на окружающую среду учитывается игроками, если выбросы выбираются игроками из условия совместной оптимизации. Это приводит к снижению общего загрязнения окружающей среды и снижению издержек по сравнению с некооперативным сценарием игры.

Влияние α : Введем дополнительное обозначение:

$$\mathcal{P}(n_i^t) = \left(\alpha - \gamma \sum_{i \in M} u_i(n_i^t) \right).$$

Если считать это ценой, то параметр α является мерой того, сколько максимально готов заплатить покупатель за единицу выбросов (или продукции). Результаты показывают, что увеличение значения α уменьшает значение цены анархии, то есть, кооперация дает меньшую выгоду в относительном выражении, когда потребители готовы платить больше. Для того, чтобы объяснить этот результат, напомним, что доходы в вершине n_i^t определяются функцией

$$f_j^{n_i^t}(u(n_i^t)) = u_j(n_i^t) \left(\alpha - \gamma \sum_{i \in M} u_i(n_i^t) \right).$$

Подставляя некооперативные и кооперативные оптимальные стратегии (выбросы) в эту функцию, получаем:

$$\begin{aligned} f_j^{n_i^t}(u^N(n_i^t)) &= \frac{(\alpha - \rho\kappa\Delta(t)) (\alpha + m\rho\kappa\Delta(t))}{(m+1)^2\gamma}, \\ f_j^{n_i^t}(u^*(n_i^t)) &= \frac{(\alpha - m\rho\kappa\Delta(t)) (\alpha + m\rho\kappa\Delta(t))}{4m\gamma}. \end{aligned}$$

Тогда маргинальные доходы равны

$$\frac{\partial f_j^{n_i^t}(u^N(n_i^t))}{\partial u_j^N(n_i^t)} = \rho\kappa\Delta(t),$$

$$\frac{\partial f_j^{n_i^t}(u^*(n_i^t))}{\partial u_j^*(n_i^t)} = \frac{2m\alpha - (\alpha - m\rho\kappa\Delta(t))(m+1)}{2m}.$$

Это показывает, что маргинальные доходы не зависят от α при некооперативном оптимальном поведении игроков, и они возрастают по этому параметру при оптимальном кооперативном поведении игроков. Это означает, что существует положительный стимул к увеличению выбросов при кооперативном сценарии игры, что, в свою очередь, приводит со временем к более высоким издержкам.

Влияние κ : Кооперация между игроками означает, что игрок, определяя оптимальную стратегию (уровень выбросов), учитывает издержки всех игроков, а не только свои издержки, как в случае некооперативного поведения. Это объединение издержек игроков в свою очередь приводит к снижению выбросов в случае кооперативного поведения игроков по сравнению с некооперативным поведением. Следовательно, чем больше предельные издержки κ , тем больше разность между оптимальными выбросами при кооперации и в её отсутствии. Это приводит к большим издержкам при отсутствии кооперации между игроками в задачах охраны окружающей среды.

Влияние ρ : Чем больше дисконт-фактор ρ , тем больше цена анархии. Увеличение ρ означает, что мы придаем меньший вес будущим доходам и кооперативное решение может дать меньшую выгоду по сравнению с некооперативным.

§ 3.10 Позиционно состоятельный вектор Шепли в игре, разыгрываемой на дереве событий со случайным временем окончания

§ 3.10.1 Описание игры

В этом параграфе рассматривается модель игры, разыгрываемой на дереве событий, в которой момент окончания игры является случайным, в отличие от предыдущих параграфов этой главы. Для такой игры рассмотрен кооператив-

ный вариант и предложен метод построения позиционно состоятельного вектора Шепли. Хотелось бы отметить, что проблема динамической устойчивости (позиционной состоятельности — для игр, развивающихся на графах) кооперативных решений дифференциальных игр со случайной продолжительностью впервые сформулирована в работе [47]. Впоследствии эта работа продолжена в статье [49]. Принципы устойчивой кооперации для дифференциальных игр со случайной продолжительностью были сформулированы в [55], а уравнения Гамильтона-Якоби- Беллмана для задач данного типа были получены в [132].

Стохастический процесс определяется заданием двух вероятностей для каждой вершины $n_i^t \in \mathcal{N}^t$, $t \in \mathcal{T}$. Эти вероятности — $\tau(n_i^t)$ и $\pi(n_i^t)$. В любой вершине $n_i^t \in \mathcal{N}^t$ игра может окончиться с вероятностью $\tau(n_i^t) \in [0, 1]$.⁵ В частности, $\tau(n^0) = 0$. Если стохастический процесс не заканчивается в вершине n_i^t , то он переходит в одну из вершин, следующих за n_i^t и принадлежащих множеству $\mathcal{S}(n_i^t)$. Обозначим через $\pi(n_i^t)$ вероятность прохождения через вершину n_i^t , которая равна сумме вероятностей всех сценариев, содержащих эту вершину. В частности, $\pi(n^0) = 1$ и $\pi(n_i^T)$ есть вероятность единственного сценария, который заканчивается в вершине n_i^T . Заметим, что $\tau(n_i^T) = \pi(n_i^T)$ для любой терминальной вершины $n_i^T \in \mathcal{N}^T$, то есть, если реализовалась некоторая вершина из множества \mathcal{N}^T , то в ней стохастический процесс и закончится. Заданные вероятности удовлетворяют условию:

$$\pi(n_i^t) = \tau(n_i^t) + \sum_{n \in \mathcal{S}(n_i^t)} \pi(n), \quad (3.10.1)$$

которое означает, что вероятность прохождения через вершину n_i^t равна сумме вероятности, что стохастический процесс закончится в этой вершине, и сумме вероятностей прохождения через вершины, следующие за вершиной n_i^t .

Также, как и ранее, $M = \{1, \dots, m\}$ — множество игроков, $u_j(n_i^t)$ — стратегия игрока j в вершине n_i^t из множества $U_j^{n_i^t} \subset \mathbb{R}^{m_j}$, $u(n_i^t) = (u_1(n_i^t), \dots, u_m(n_i^t))$ ситуация в вершине n_i^t . Для каждой вершины $n_i^t \in \mathcal{N}^t$, $t = 0, 1, \dots, T$, множество $U_j^{n_i^t}$ — множество стратегий игрока j в вершине n_i^t . Также $U^{n_i^t} = U_1^{n_i^t} \times$

⁵Определяя вероятности $\tau(n_i^t)$ для каждой вершины $n_i^t \in \mathcal{N}^t$ для всех $t = 0, \dots, T$ мы определяем вероятностное распределение момента окончания стохастического процесса на множестве вершин дерева событий. Для вероятностного распределения имеет место равенство: $\sum_{n \in \mathcal{N}} \tau(n) = 1$.

$\dots \times U_j^{n_i^t} \times \dots \times U_m^{n_i^t}$ — множество ситуаций в вершине n_i^t , $X \subset \mathbb{R}^p$ — множество состояний, одинаковое для всех вершин n_i^t .

Функция перехода $f^{n_i^t}(\cdot, \cdot) : X \times U^{n_i^t} \mapsto X$ для вершины n_i^t определяет динамику состояний

$$x(n_i^t) = f^{a(n_i^t)}(x(a(n_i^t)), u(a(n_i^t))), \quad (3.10.2)$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, \quad t = 1, \dots, T, \quad (3.10.3)$$

где $x(n_i^t) \in X$ для любой вершине n_i^t , $t = 0, 1, \dots, T$.

В вершине n_i^t функция выигрыша игрока j определяется функцией $\phi_j^{n_i^t}(x(n_i^t), u(n_i^t))$, если стохастический процесс не заканчивается в этой вершине; в противном случае, она определяется функцией $\Phi_j^{n_i^t}(x(n_i^t))$. В частности, для $t = T$ выигрыш игрока j есть $\Phi_j^{n_i^T}(x(n_i^T))$. Информация о том, что вершина n_i^t становится терминальной, узнается игроками перед тем, как игроки выбирают стратегии в этой вершине, но не в вершине $a(n_i^t)$. Заметим, что в любой нетерминальной вершине, выигрыш игрока зависит от состояния и стратегий игроков.

Напомним, что ρ_j ($0 < \rho_j < 1$) — дисконт-фактор игрока $j \in M$ и предположим, что игроки максимизируют суммарный ожидаемый дисконтированный выигрыш. Уравнения состояния и функции выигрышей игроков определяют многошаговую игру, где

$$\mathbf{x} = \{x(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T\},$$

$$\mathbf{u} = \{u(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T - 1\},$$

и $J_j(\mathbf{x}, \mathbf{u})$ — функция выигрыша игрока j :

$$\begin{aligned} J_j(\mathbf{x}, \mathbf{u}) &= \sum_{t=0}^{T-1} \rho_j^t \sum_{n_i^t \in \mathcal{N}^t} (\pi(n_i^t) - \tau(n_i^t)) \phi_j^{n_i^t}(x(n_i^t), u(n_i^t)) + \\ &+ \sum_{t=1}^T \rho_j^t \sum_{n_i^t \in \mathcal{N}^t} \tau(n_i^t) \Phi_j^{n_i^t}(x(n_i^t)), \quad j \in M, \end{aligned} \quad (3.10.4)$$

при условии, что

$$x(n_i^t) = f^{a(n_i^t)}(x(a(n_i^t)), u(a(n_i^t))), \quad (3.10.5)$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, t = 1, \dots, T,$$

$$x(n^0) = x^0. \quad (3.10.6)$$

Пусть $\mathbf{u} = (\mathbf{u}_j : j \in M)$ — ситуация в S -адаптивных стратегиях, которые удовлетворяют определению 3.1.

Мы можем определить игру в нормальной форме с функцией выигрыша игрока j , равной $W_j(\mathbf{u}, x^0) = J_j(\mathbf{x}, \mathbf{u})$, $j \in M$, где \mathbf{x} — траектория состояний, единственным образом получаемая из \mathbf{u} из уравнений (3.10.5), начинающаяся из состояния x^0 .

Если игра разыгрывается некооперативно, то в качестве принципа оптимальности выберем S -адаптивное равновесие по Нэшу, определение которого дано выше (см. определение 3.2).

Заметим, что при S -адаптивной информационной структуре стратегии и уравнения состояния определены на множестве вершин дерева событий.

Если игроки кооперируются, то они максимизируют сумму дисконтированных выигрышей в игре:

$$\max_{\mathbf{u}_j: j \in M} \sum_{j \in M} W_j(\mathbf{u}, x^0).$$

В результате решения этой задачи максимизации мы получаем вектор стратегий игроков \mathbf{u} , определяющий их поведение в игре, начинающейся из вершины n^0 и состояния x^0 , то есть,

$$\mathbf{u}^* = \arg \max_{\mathbf{u}_j: j \in M} \sum_{j \in M} W_j(\mathbf{u}, x^0). \quad (3.10.7)$$

Также как и ранее, $\mathbf{x}^* = \{x^*(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, 1, \dots, T\}$ — кооперативная траектория состояний, соответствующая набору стратегий \mathbf{u}^* .

В дальнейшем нам понадобится определение подыгры, начинающейся из вершины $n_i^t \in \mathcal{N}^t$ и состояния $x^*(n_i^t)$, $t = 1, \dots, T - 1$. Определим её также, как и

ранее. Выигрыш игрока $j \in M$ в этой подыгре равен

$$\begin{aligned} W_j(\mathbf{u}(n_i^t), x^*(n_i^t)) &= \sum_{\theta=t}^{T-1} \rho_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} (\pi(n_i^\theta | n_i^t) - \tau(n_i^\theta | n_i^t)) \phi_j^{n_i^\theta}(x(n_i^\theta), u(n_i^\theta)) + \\ &+ \sum_{\theta=t+1}^T \rho_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} \tau(n_i^\theta | n_i^t) \Phi_j^{n_i^\theta}(x(n_i^\theta)), \end{aligned} \quad (3.10.8)$$

где $\mathcal{N}_\Gamma^\theta = \mathcal{N}^\theta \cap \Gamma(n_i^t)$, $\mathbf{u}(n_i^t) = (\mathbf{u}_j(n_i^t) : j \in M)$ — ситуация в S -адаптивных стратегиях и $\mathbf{u}_j(n_i^t) = \{u_j(n_i^\theta) : n_i^\theta \in \Gamma(n_i^t)\}$ — S -адаптивная стратегия игрока j в подыгре, начинающейся в вершине n_i^t из состояния $x^*(n_i^t)$. Множитель $\pi(n_k^\theta | n_i^t)$ — условная вероятность прохождения через вершину n_k^θ , если подыгра начинается в вершине n_i^t . Множитель $\tau(n_k^\theta | n_i^t)$ — условная вероятность⁶ того, что игра закончится в вершине n_k^θ , когда подыгра начинается из вершины n_i^t .

В некооперативной подыгре, начинающейся из вершины n_i^t и состояния $x^*(n_i^t)$, набор стратегий $\mathbf{u}^N(x^*(n_i^t)) = (\mathbf{u}_j^N(x^*(n_i^t)) : j \in M)$ — S -адаптивное равновесие по Нэшу. Выигрыш игрока j в S -адаптивном равновесии по Нэшу равен $W_j(\mathbf{u}^N(x^*(n_i^t)), x^*(n_i^t))$.

В кооперативной подыгре, начинающейся из вершины n_i^t из состояния $x^*(n_i^t)$, ситуация в S -адаптивных стратегиях, максимизирующая суммарный дисконтированный выигрыш игроков в игре, есть

$$\mathbf{u}^*(n_i^t) = \arg \max_{\mathbf{u}_j(n_i^t) : j \in M} \sum_{j \in M} W_j(\mathbf{u}(n_i^t), x^*(n_i^t)). \quad (3.10.9)$$

Выигрыш игрока j в этой кооперативной подыгре равен $W_j(\mathbf{u}^*(n_i^t), x^*(n_i^t))$, $t = 1, \dots, T$.

Если говорить о единственности S -адаптивного равновесия по Нэшу и кооперативного равновесия, то считаем справедливым замечание 3.2 и выполнение условий, приведенных в этом замечании.

⁶Условные вероятности $\pi(n_k^\theta | n_i^t)$ и $\tau(n_k^\theta | n_i^t)$ вычисляются по формулам: $\pi(n_k^\theta | n_i^t) = \pi(n_k^\theta) / (\pi(n_i^t) - \tau(n_i^t))$ и $\tau(n_k^\theta | n_i^t) = \tau(n_k^\theta) / (\pi(n_i^t) - \tau(n_i^t))$ соответственно, если $\pi(n_i^t) \neq 0$ и $\pi(n_i^t) \neq \tau(n_i^t)$; в противном случае, подыгра, начинающаяся в вершине n_i^t не может произойти.

§ 3.10.2 Позиционно состоятельный вектор Шепли в динамических играх на деревьях событий

Предположим, что игроки согласились играть в кооперации и использовать вектор Шепли⁷ в качестве решения кооперативной игры при дележе совместного выигрыша $\sum_{j \in M} W_j(\mathbf{u}^*, x^0)$. Для всех коалиций $G \subset M$ определим значения характеристической функции⁸ $V(G; x^*(n_l^t))$ в подыгре, начинающейся в вершине n_l^t из состояния $x^*(n_l^t)$, для всех вершин $n_l^t \in \mathcal{N}^t, t = 0, 1, \dots, T - 1$:

$$V(G; x^*(n_l^t)) : 2^M \rightarrow \mathbb{R} \quad (3.10.10)$$

и $v(\emptyset; x^*(n_l^t)) = 0$.

Как и ранее в этой главе, мы будем строить γ -характеристическую функцию, предполагая, что когда коалиция G формируется, оставшиеся игроки не играют против этой коалиции, а играют индивидуальные оптимальные стратегии наилучшего ответа.⁹ Следовательно, значение γ -характеристической функции для коалиции G — выигрыш коалиции G в S -адаптивном равновесии по Нэшу в некооперативной игре между G (где все игроки максимизируют суммарный выигрыш, играя как один игрок), а все остальные игроки играют индивидуально, максимизируя свои собственные выигрыши.

Множество дележей $Y(x^*(n_l^t))$ в подыгре, начинающейся в вершине n_l^t из состояния $x^*(n_l^t)$, есть

$$Y(x^*(n_l^t)) = \left\{ (y_1(x^*(n_l^t)), \dots, y_m(x^*(n_l^t))) \mid y_j(x^*(n_l^t)) \geq V(\{j\}; x^*(n_l^t)) \right. \\ \left. \text{для всех } j \in M, \text{ и } \sum_{j \in M} y_j(x^*(n_l^t)) = V(M; x^*(n_l^t)) \right\}.$$

⁷Здесь в качестве решения кооперативной игры выбран вектор Шепли, но результаты этого раздела справедливы и в случае любого другого одноточечного принципа оптимальности. Если решение кооперативной игры — множество дележей, то необходимо сделать предположение о его непустоте для любой подыгры и воспользоваться результатами § 3.2.

⁸Мы определяем характеристическую функцию $v(G; x^*(n_l^t))$ для всех вершин $n_l^t \in \mathcal{N}^t$ на основании первоначально заданной игры, предполагая, что игроки используют кооперативные стратегии \mathbf{u}^* , двигаясь вдоль кооперативной траектории \mathbf{x}^* . Значения характеристической функции $V(G; x^*(n_l^t))$ также зависят от начальной для этой подыгры вершины n_l^t и кооперативных стратегий, но мы опустим их в обозначениях для простоты изложения.

⁹В этой главе используется γ -характеристическая функция, но можно использовать и другие методы построения характеристических функций, например, α -, β - и δ -характеристические функции. Определения этих функций можно найти в § 1.3 Главы 1.

Вектор Шепли $Sh(x^*(n_l^t)) = (Sh_1(x^*(n_l^t)), \dots, Sh_m(x^*(n_l^t)))$ — дележ в подыгре, начинающейся из вершины n_l^t и состояния $x^*(n_l^t)$:

$$Sh_j(x^*(n_l^t)) = \sum_{G \subseteq M: G \ni j} \frac{(m-g)!(g-1)!}{m!} (V(G; x^*(n_l^t)) - V(G \setminus j; x^*(n_l^t))), \quad (3.10.11)$$

где $g = |G|$.

Определение 3.11. Назовем $(\{\beta_j(x^*(n_l^t))\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M)$ процедурой распределения (ПРД) вектора Шепли $Sh(x^0) = (Sh_1(x^0), \dots, Sh_m(x^0))$, где

$$\beta_j(x^*(n_l^t)) = \begin{cases} \beta_j^c(x^0), & \text{если } n_l^t = n^0, \\ \beta_j^c(x^*(n_l^t)), & \text{если процесс не закончится в } n_l^t, t = 1, \dots, T-1, \\ \beta_j^r(x^*(n_l^t)), & \text{если процесс закончится в } n_l^t, t = 1, \dots, T-1, \\ \beta_j^r(x^*(n_l^T)), & \text{если } n_l^t = n_l^T \end{cases} \quad (3.10.12)$$

для всех $j \in M$, если следующие условия выполнены:¹⁰

$$Sh_j(x^0) = \sum_{t=0}^{T-1} \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} (\pi(n_l^t) - \tau(n_l^t)) \beta_j^c(x^*(n_l^t)) + \sum_{t=1}^T \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \tau(n_l^t) \beta_j^r(x^*(n_l^t)), \quad j \in M, \quad (3.10.13)$$

$$\sum_{j \in M} \beta_j^c(x^*(n_l^t)) = \sum_{j \in M} \phi_j^{n_l^t}(x^*(n_l^t), u^*(n_l^t)), \quad (3.10.14)$$

если процесс не закончится в $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$,

$$\sum_{j \in M} \beta_j^r(x^*(n_l^t)) = \sum_{j \in M} \Phi_j^{n_l^t}(x^*(n_l^t)), \quad (3.10.15)$$

если процесс закончится в $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T$.

В начальной вершине n^0 , в которой стохастический процесс не может закончиться, и в вершинах $n_l^T \in \mathcal{N}^T$, в которых процесс закончится с вероятностью единица, компоненты ПРД определены единственным образом, потому что в этих вершинах известно, продолжится игра или закончится наверняка. Для

¹⁰Индексы c и r в обозначении β_j в формуле (3.10.12) соответствуют словам «продолжить» (“continue”) и «закончиться» (“terminate”).

«промежуточных» вершин $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T - 1$, определены два возможных значения ПРД для каждого игрока j , что соответствует двум возможным сценариям течения игры: продолжения или окончания, а именно, (1) $\beta_j^c(x^*(n_l^t))$, если игра не заканчивается в вершине n_l^t , и (2) $\beta_j^\tau(x^*(n_l^t))$, если заканчивается. Условие (3.10.13) означает, что компонента вектора Шепли, соответствующая игроку $j \in M$, равна его дисконтированному выигрышу во всей игре. Условия (3.10.14), (3.10.15) можно назвать условиями «достижимости» процедуры распределения дележа, то есть, сумма выплат игрокам в любой вершине равна сумме выигрышей игроков, которые игроки получают в этой вершине при реализации ситуации $\mathbf{u}^*(n_l^t)$.

Определение 3.12. Вектор Шепли $Sh(x^0)$ и соответствующая ему процедура распределения дележа $(\{\beta_j(x^*(n_l^t))\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M)$, определяемая по формуле (3.10.12), называются позиционно состоятельными во всей игре, если для любой вершины $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T$, выполнено условие:

$$Sh_j(x^0) = \sum_{\theta=0}^{t-1} \rho_j^\theta \sum_{n_k^\theta \in \mathcal{N}^\theta} \{(\pi(n_k^\theta) - \tau(n_k^\theta)) \beta_j^c(x^*(n_k^\theta)) + \tau(n_k^\theta) \beta_j^\tau(x^*(n_k^\theta))\} + \rho_j^t \sum_{n_k^t \in \mathcal{N}^t} \{(\pi(n_k^t) - \tau(n_k^t)) Sh_j(x^*(n_k^t)) + \tau(n_k^t) Sh_j^\tau(x^*(n_k^t))\}, \quad (3.10.16)$$

где $Sh_j(x^*(n_k^t))$ — j -ая компонента вектора Шепли, вычисленного для подыгры, начинающейся в вершине n_k^t из состояния $x^*(n_k^t)$, в случае, если стохастический процесс не заканчивается в вершине n_k^t , и $Sh_j^\tau(x^*(n_k^t))$ — j -ая компонента вектора Шепли, вычисленного для подыгры, начинающейся в вершине n_k^t из состояния $x^*(n_k^t)$, в случае, если стохастический процесс заканчивается в этой вершине.¹¹

Если выигрыши в вершинах распределяются между игроками с помощью процедуры распределения дележа, динамическая устойчивость или позиционная состоятельность вектора Шепли $Sh(x^0)$ означает, что можно определить систему пошаговых выплат или трансферных платежей, которая гарантиру-

¹¹Если игра заканчивается в вершине n_k^t , игроки не выбирают стратегий в этой вершине, а компоненты вектора Шепли $Sh_j^\tau(x^*(n_k^t))$ равны соответствующим значениям их функций выигрыша $\Phi_j^{n_k^t}(x^*(n_k^t))$.

ет, что в каждой вершине дисконтированная сумма оставшихся выплат будет равна компонентам вектора Шепли в этой подыгре.

Теорема 3.7. *Вектор Шепли $Sh(x^0)$ позиционно состоятелен, если соответствующая процедура распределения дележа, определенная формулой (3.10.12), вычисляется по формулам для всех $t = 0, \dots, T - 1$:*

$$\begin{aligned} \beta_j^c(x^*(n_l^t)) = Sh_j(x^*(n_l^t)) - \rho_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \{ (\pi(n_k^{t+1}|n_l^t) - \tau(n_k^{t+1}|n_l^t)) Sh_j(x^*(n_k^{t+1})) + \\ + \tau(n_k^{t+1}|n_l^t) \Phi_j^{n_k^{t+1}}(x^*(n_k^{t+1})) \}, \end{aligned} \quad (3.10.17)$$

и для всех $t = 1, \dots, T$:

$$\beta_j^r(x^*(n_l^t)) = \Phi_j^{n_l^t}(x^*(n_l^t)). \quad (3.10.18)$$

Доказательство. Сначала докажем, что $\beta_j(x^*(n_l^t))$, определенные формулами (3.10.17) и (3.10.18), — процедура распределения дележа вектора Шепли $Sh(x^0)$, то есть, покажем, что условия (3.10.13)–(3.10.15) выполнены. Подставляя выражения для $\beta_j^c(x^*(n_l^t))$ и $\beta_j^r(x^*(n_l^t))$ из (3.10.17) и (3.10.18) в правую часть уравнения (3.10.13), и учитывая, что

$$\pi(n_k^{t+1}|n_l^t) - \tau(n_k^{t+1}|n_l^t) = \frac{\pi(n_k^{t+1}) - \tau(n_k^{t+1})}{\pi(n_l^t) - \tau(n_l^t)}, \quad (3.10.19)$$

получаем

$$\begin{aligned} \sum_{t=0}^{T-1} \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} (\pi(n_l^t) - \tau(n_l^t)) \left[Sh_j(x^*(n_l^t)) - \rho_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \{ (\pi(n_k^{t+1}|n_l^t) - \tau(n_k^{t+1}|n_l^t)) \cdot \right. \\ \left. \cdot Sh_j(x^*(n_k^{t+1})) + \tau(n_k^{t+1}|n_l^t) \Phi_j^{n_k^{t+1}}(x^*(n_k^{t+1})) \} \right] + \sum_{t=1}^T \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \tau(n_l^t) \Phi_j^{n_l^t}(x^*(n_l^t)) = \\ = (\pi(n^0) - \tau(n^0)) Sh_j(x^0) - \rho_j (\pi(n^0) - \tau(n^0)) \sum_{n_k^1 \in \mathcal{S}(n^0)} \{ (\pi(n_k^1|n^0) - \tau(n_k^1|n^0)) \cdot \end{aligned}$$

$$\begin{aligned}
& \cdot Sh_j(x^*(n_k^1)) + \tau(n_k^1|n^0)\Phi_j^{n_k^1}(x^*(n_k^1)) \} + \rho_j \sum_{n_k^1 \in \mathcal{N}^1} \tau(n_k^1)\Phi_j^{n_k^1}(x^*(n_k^1)) + \dots \\
& + \rho_j^{T-1} \sum_{n_l^{T-1} \in \mathcal{N}^{T-1}} (\pi(n_l^{T-1}) - \tau(n_l^{T-1})) \left[Sh_j(x^*(n_l^{T-1})) - \right. \\
& - \rho_j \sum_{n_k^T \in \mathcal{S}(n_l^{T-1})} \left\{ (\pi(n_k^T|n_l^{T-1}) - \tau(n_k^T|n_l^{T-1})) Sh_j(x^*(n_k^T)) + \right. \\
& \left. \left. + \tau(n_k^T|n_l^{T-1})\Phi_j^{n_k^T}(x^*(n_k^T)) \right\} \right] + \rho_j^T \sum_{n_l^T \in \mathcal{N}^T} \tau(n_l^T)\Phi_j^{n_l^T}(x^*(n_l^T)) = Sh_j(x^0).
\end{aligned}$$

Также используем условие: $Sh_j(x^*(n_k^T)) = \Phi_j^{n_k^T}(x^*(n_k^T))$ для любой вершины $n_k^T \in \mathcal{N}^T$.

Теперь покажем, что условия (3.10.14) и (3.10.15) выполнены, если мы используем процедуру распределения дележа, определенную формулами (3.10.17) и (3.10.18). Для терминальных вершин условие (3.10.15) выполнено, поскольку вектор Шепли является дележом и удовлетворяет условию эффективности. Проверим, выполнено ли условие (3.10.14) для начальной вершины и любой промежуточной вершины $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$:

$$\begin{aligned}
& \sum_{j \in M} \beta_j^c(x^*(n_l^t)) = \\
& = \sum_{j \in M} Sh_j(x^*(n_l^t)) - \sum_{j \in M} \rho_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \left\{ (\pi(n_k^{t+1}|n_l^t) - \tau(n_k^{t+1}|n_l^t)) Sh_j(x^*(n_k^{t+1})) \right. \\
& \left. + \tau(n_k^{t+1}|n_l^t)\Phi_j^{n_k^{t+1}}(x^*(n_k^{t+1})) \right\} = \\
& = V(M; x^*(n_l^t)) - \left\{ \sum_{j \in M} W_j(\mathbf{u}^*(n_l^t), x^*(n_l^t)) - \sum_{j \in M} \phi_j^{n_l^t}(x^*(n_l^t), u^*(n_l^t)) \right\},
\end{aligned}$$

и используя, что

$$V(M; x^*(n_l^t)) = \sum_{j \in M} W_j(\mathbf{u}^*(n_l^t), x^*(n_l^t)),$$

а также метод построения характеристической функции, получаем (3.10.14).

Наконец, докажем, что вектор Шепли и соответствующая ему процедура распределения дележа позиционно состоятельны, то есть, они удовлетворяют

условию (3.10.16). Это можно доказать, если подставить $\beta_j^c(x^*(n_k^\theta))$, определенную формулой (3.10.17), и $\beta_j^r(x^*(n_k^\theta))$, определенную формулой (3.10.18), в условие (3.10.16) и учесть, что $\pi(n^0) = 1$, $\tau(n^0) = 0$ и условие (3.10.19) для любой вершины $n_k^{t+1} \in \mathcal{N}^t, t = 0, \dots, T - 1$. Получаем

$$\begin{aligned} & \sum_{\theta=0}^{t-1} \rho_j^\theta \sum_{n_l^\theta \in \mathcal{N}^\theta} \left\{ (\pi(n_l^\theta) - \tau(n_l^\theta)) Sh_j(x^*(n_l^\theta)) - \right. \\ & - \rho_j \sum_{n_k^{\theta+1} \in \mathcal{S}(n_l^\theta)} [(\pi(n_k^{\theta+1}) - \tau(n_k^{\theta+1})) Sh_j(x^*(n_k^{\theta+1})) + \\ & + \tau(n_k^{\theta+1}) \Phi_j^{n_k^{\theta+1}}(x^*(n_k^{\theta+1}))] \left. \right\} + \rho_j^t \sum_{n_k^t \in \mathcal{N}^t} \left\{ (\pi(n_k^t) - \tau(n_k^t)) Sh_j(x^*(n_k^t)) + \right. \\ & \left. + \tau(n_k^t) \Phi_j^{n_k^t}(x^*(n_k^t)) \right\} = Sh_j(x^0). \end{aligned}$$

Это доказывает позиционную состоятельность вектора Шепли, когда соответствующая ему ПРД определяется формулами (3.10.17) и (3.10.18). \square

Реализация процедуры распределения дележа ($\{\beta_j(x^*(n_l^t))\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M$) требует определения трансферных платежей, осуществляемых игроками между собой по следующему правилу:

- если игра не заканчивается в вершине $n_l^t \in \mathcal{N}^t, t = 0, \dots, T - 1$, тогда

$$\omega_j(n_l^t, x^*(n_l^t)) = \beta_j^c(x^*(n_l^t)) - \phi_j^{n_l^t}(x^*(n_l^t), u^*(n_l^t)), \quad (3.10.20)$$

- если игра заканчивается в вершине $n_l^t \in \mathcal{N}^t, t = 1, \dots, T$, тогда

$$\omega_j(n_l^t, x^*(n_l^t)) = 0, \quad (3.10.21)$$

где $\omega_j(n_l^t, x^*(n_l^t))$ — трансферный платеж, который делает игрок j в вершине n_l^t на кооперативной траектории $x^*(n_l^t)$ такой, что $\sum_{j \in M} \omega_j(n_l^t, x^*(n_l^t)) = 0$ для любой вершины n_l^t при состоянии $x^*(n_l^t)$. Платеж $\omega_j(n_l^t, x^*(n_l^t))$ из (3.10.20) может быть любого знака, в зависимости от знака разности в правой части (3.10.20).

§ 3.10.3 Необходимые условия существования S -адаптивного равновесия по Нэшу

Для того, чтобы найти значения γ -характеристической функции, необходимо найти S -адаптивное равновесие по Нэшу в игре, разыгрываемой на дереве событий с неизвестным временем окончания. В этом разделе мы сформулируем необходимые условия S -адаптивного равновесия по Нэшу, используя принцип максимума Понтрягина.

Сначала запишем выигрыш игрока j вида (3.10.4) в эквивалентной форме:

$$\begin{aligned} J_j(\mathbf{x}, \mathbf{u}) &= \phi_j^{n^0}(x^0, u(n^0)) + \\ &+ \sum_{t=1}^{T-1} \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \left\{ (\pi(n_l^t) - \tau(n_l^t)) \phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) + \tau(n_l^t) \Phi_j^{n_l^t}(x(n_l^t)) \right\} + \\ &+ \rho_j^T \sum_{n_l^T \in \mathcal{N}^T} \tau(n_l^T) \Phi_j^{n_l^T}(x(n_l^T)), \quad j \in M, \end{aligned}$$

где $\tau(n_l^T) = \pi(n_l^T)$ для любой вершины $n_l^T \in \mathcal{N}^T$.

Для каждого игрока $j \in M$ запишем лагранжиан:

$$\begin{aligned} \mathcal{L} &= \phi_j^{n_0}(x^0, u(x^0)) + \tag{3.10.22} \\ &+ \sum_{t=1}^{T-1} \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \left\{ (\pi(n_l^t) - \tau(n_l^t)) \phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) + \tau(n_l^t) \Phi_j^{n_l^t}(x(n_l^t)) + \right. \\ &+ \left. \pi(n_l^t) \lambda_j(n_l^t) (f^{a(n_l^t)}(x(a(n_l^t)), u(a(n_l^t))) - x(n_l^t)) \right\} + \\ &+ \rho_j^T \sum_{n_l^T \in \mathcal{N}^T} \tau(n_l^T) \left\{ \Phi_j^{n_l^T}(x(n_l^T)) + \right. \\ &+ \left. \lambda_j(n_l^T) (f^{a(n_l^T)}(x(a(n_l^T)), u(a(n_l^T))) - x(n_l^T)) \right\} = \\ &= \phi_j^{n_0}(x^0, u(x^0)) + \\ &+ \sum_{t=1}^{T-1} \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \pi(n_l^t) \left\{ \frac{\pi(n_l^t) - \tau(n_l^t)}{\pi(n_l^t)} \phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) + \frac{\tau(n_l^t)}{\pi(n_l^t)} \Phi_j^{n_l^t}(x(n_l^t)) + \right. \end{aligned}$$

$$\begin{aligned}
& + \lambda_j(n_i^t) \left(f^{a(n_i^t)}(x(a(n_i^t)), u(a(n_i^t))) - x(n_i^t) \right) \Big\} + \\
& + \rho_j^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) \left\{ \Phi_j^{n_i^T}(x(n_i^T)) + \right. \\
& \left. + \lambda_j(n_i^T) \left(f^{a(n_i^T)}(x(a(n_i^T)), u(a(n_i^T))) - x(n_i^T) \right) \right\},
\end{aligned}$$

где $\pi(n_i^T) = \tau(n_i^T)$ для любой вершины $n_i^T \in \mathcal{N}^T$. Для любого игрока $j \in M$ множители Лагранжа обозначены через $\lambda_j(n_i^t)$ и заданы на множестве вершин периодов $t = 1, \dots, T$. Множитель Лагранжа $\lambda_j(n_i^t)$ имеет такую же размерность, что и $x(n_i^t)$. Определим пре-гамильтониан для любого игрока $j \in M$ и любой вершины $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T - 1$ следующим образом

$$\begin{aligned}
H_j^{n_i^t}(x(n_i^t), u(n_i^t), \lambda_j(\mathcal{S}(n_i^t))) &= \frac{\pi(n_i^t) - \tau(n_i^t)}{\pi(n_i^t)} \phi_j^{n_i^t}(x(n_i^t), u(n_i^t)) + \quad (3.10.23) \\
&+ \frac{\tau(n_i^t)}{\pi(n_i^t)} \Phi_j^{n_i^t}(x(n_i^t)) + \rho_j \sum_{\nu \in \mathcal{S}(n_i^t)} \frac{\pi(\nu)}{\pi(n_i^t)} \lambda_j(\nu) f^{n_i^t}(x(n_i^t), u(n_i^t)),
\end{aligned}$$

где $\lambda_j(\mathcal{S}(n_i^t))$ набор $\lambda_j(\nu)$, где $\nu \in \mathcal{S}(n_i^t)$.

Теорема 3.8. Пусть \mathbf{u}^N — S-адаптивное равновесие по Нэшу в игре, разыгрываемой на дереве событий с начальным состоянием x^0 , генерирующее траекторию $\mathbf{x}^N = \{x^N(n_i^t) : n_i^t \in \mathcal{N}^t, t = 1, \dots, T\}$. Тогда для любого игрока $j \in M$ множители Лагранжа $\lambda_j(n_i^t)$, $n_i^t \in \mathcal{N}^t$, $t = 1, \dots, T$ таковы, что выполняются следующие условия для любых $j \in M$, $u_j(n_i^t) = u_j^N(n_i^t)$, $x(n_i^t) = x^N(n_i^t)$:

$$0 = \frac{\partial H_j^{n_i^t}}{\partial u_j(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, \quad t = 0, \dots, T - 1, \quad (3.10.24)$$

$$\lambda_j(n_i^t) = \frac{\partial H_j^{n_i^t}}{\partial x(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, \quad t = 1, \dots, T - 1 \quad (3.10.25)$$

$$\lambda_j(n_i^T) = \frac{\partial \Phi_j^{n_i^T}(x(n_i^T))}{\partial x(n_i^T)}, \quad n_i^T \in \mathcal{N}^T. \quad (3.10.26)$$

Доказательство. Рассмотрим выражение (3.10.22) лагранжиана и перепишем

его в виде:

$$\begin{aligned} \mathcal{L} = & \sum_{t=0}^{T-1} \rho_j^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \left\{ H_j^{n_i^t}(x(n_i^t), u(n_i^t), \lambda_j(\mathcal{S}(n_i^t))) - \lambda_j(n_i^t)x(n_i^t) \right\} + \\ & + \rho_j^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) \left\{ \Phi_j^{n_i^T}(x(n_i^T)) - \lambda_j(n_i^T)x(n_i^T) \right\}. \end{aligned}$$

Учитывая выражение (3.10.23), уравнения (3.10.24)-(3.10.26) получаются приравниванием к нулю частных производных Лагранжиана по $x(n_i^t)$ и $u(n_i^t)$. \square

§ 3.10.4 Теоретико-игровая модель охраны окружающей среды со случайным временем окончания

Продemonстрируем теоретические результаты на примере из одной модели охраны окружающей среды, а именно, построим позиционно состоятельный вектор Шепли в стохастической версии детерминированной игры трех лиц, предложенной в [98]. Описание игры с четырьмя периодами $t = 0, \dots, 3$ можно найти в разделе § 3.5. Уравнения динамики состояния системы имеют тот же самый вид. Уровень абсорбции загрязняющего вещества также может принимать только два значения: $\delta(n_i^t) \in \{\underline{\delta}, \bar{\delta}\}$, где $\underline{\delta} < \bar{\delta}$. Дерево событий представлено на рис. 3.8. Вершины $n_2^1, n_2^2, n_4^2, n_2^3, n_4^3, n_6^3, n_8^3$ соответствуют низкому уровню абсорбции с δ , равной $\underline{\delta}$, и вершины $n_1^1, n_3^2, n_1^3, n_3^3, n_5^3, n_7^3$ соответствуют высокому уровню абсорбции с δ , равной $\bar{\delta}$.

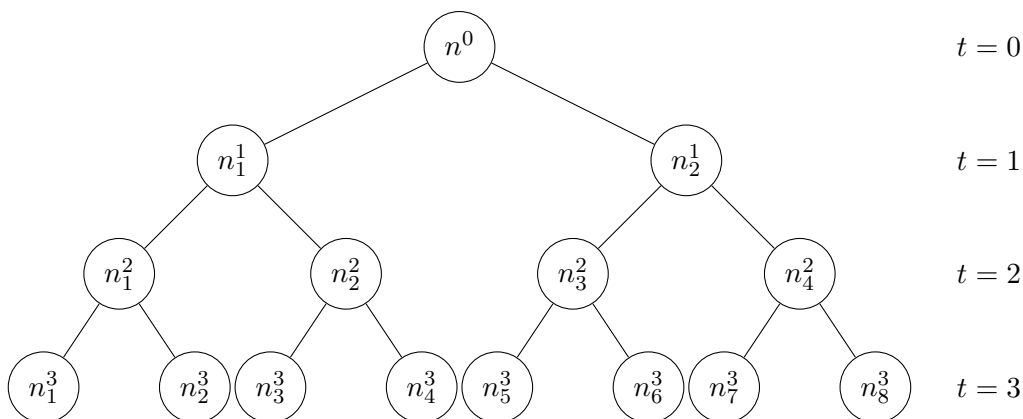


Рис. 3.8: Дерево событий.

Функция издержек от загрязнения — выпуклая возрастающая функция «запаса загрязняющего вещества»: $D_j(x(n_i^t)) = \alpha_j x^2(n_i^t), j \in M$, где α_j — поло-

жительный параметр. Издержки от объема выбросов $u_j(n_l^t)$ — квадратичная функция $C_j(u_j(n_l^t)) = \frac{\gamma_j}{2} (u_j(n_l^t) - e)^2$, где e и γ_j — положительные константы. Функции издержек заданы следующим образом: $\phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) = C_j(u_j(n_l^t)) + D_j(x(n_l^t))$ для любой вершины $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$, и $\Phi_j^{n_l^t}(x(n_l^t)) = D_j(x(n_l^t))$ для любой вершины $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T$.

Суммарные дисконтированные издержки $J_j(\mathbf{x}, \mathbf{u})$, которые игрок j стремится минимизировать, равны

$$\sum_{t=0}^2 \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} (\pi(n_l^t) - \tau(n_l^t)) (C_j(u_j(n_l^t)) + D_j(x(n_l^t))) + \sum_{t=1}^3 \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \tau(n_l^t) D_j(x(n_l^t)),$$

где $\mathbf{x} = \{x(n_l^t)\}$ и $\mathbf{u} = \{u(n_l^t)\}$, $n_l^t \in \mathcal{N}^t$ при условии, что динамика состояния задана уравнением (3.3.1), и начальный «запас загрязняющего вещества» равен $x^0 = 0$. Также имеют место ограничения: $u_j(n_l^t) \in [0, e]$ для игрока $j \in M$ и любой вершины $n_l^t \in \mathcal{N}^t$, $t = 0, 1, 2$.

Используем следующие параметры для иллюстрации результатов:

$$\alpha_1 = 0.1, \alpha_2 = 0.2, \alpha_3 = 0.3, \gamma_1 = 0.9, \gamma_2 = 0.8, \gamma_3 = 0.7,$$

$$\underline{\delta} = 0.45, \bar{\delta} = 0.8, e = 25, \lambda_1 = \lambda_2 = \lambda_3 = 0.9,$$

$$\pi(n_1^1) = 0.6, \pi(n_2^1) = 0.4, \pi(n_1^2) = 0.3, \pi(n_2^2) = \pi(n_3^2) = 0.25, \pi(n_4^2) = 0.1,$$

$$\pi(n_1^3) = 0.05, \pi(n_2^3) = 0.15, \pi(n_3^3) = 0.05, \pi(n_4^3) = 0.15,$$

$$\pi(n_5^3) = 0.05, \pi(n_6^3) = 0.15, \pi(n_7^3) = 0.025, \pi(n_8^3) = 0.025,$$

$$\tau(n_1^1) = \tau(n_2^1) = 0.05, \tau(n_1^2) = 0.1, \tau(n_2^2) = \tau(n_3^2) = \tau(n_4^2) = 0.05,$$

$$\tau(n_1^3) = 0.05, \tau(n_2^3) = 0.15, \tau(n_3^3) = 0.05, \tau(n_4^3) = 0.15,$$

$$\tau(n_5^3) = 0.05, \tau(n_6^3) = 0.15, \tau(n_7^3) = \tau(n_8^3) = 0.025.$$

По формулам (3.10.7) и (3.10.9), найдем кооперативные стратегии для игры

и всех ее подыгр. В результате получим кооперативную траекторию¹²:

$x^*(n^0)$	0.000
$x^*(n_1^1), x^*(n_2^1)$	47.106
$x^*(n_1^2), x^*(n_2^2)$	54.843
$x^*(n_3^2), x^*(n_4^2)$	68.085
$x^*(n_1^3), x^*(n_2^3)$	61.002
$x^*(n_3^3), x^*(n_4^3)$	74.622
$x^*(n_5^3), x^*(n_6^3)$	62.881
$x^*(n_7^3), x^*(n_8^3)$	79.790

Значения γ -характеристических функций представлены в табл. 3.12. Нетрудно заметить, что γ -характеристическая функция супераддитивна для любой подыгры.

Период	$t = 0$		$t = 1$		$t = 2$			
Вершина	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2	
$V(\{1\})$	125.635	117.503	135.606	82.087	107.905	101.619	135.337	
$V(\{2\})$	260.689	241.361	278.095	166.723	219.625	205.948	275.036	
$V(\{3\})$	409.197	374.297	430.416	255.002	336.795	314.146	420.966	
$V(\{1, 2\})$	375.811	351.253	405.044	245.272	322.236	303.808	404.320	
$V(\{1, 3\})$	508.782	473.494	545.609	329.056	432.681	407.230	542.561	
$V(\{2, 3\})$	631.725	588.827	678.576	409.931	538.771	507.561	675.823	
$V(\{1, 2, 3\})$	666.538	642.349	742.643	463.653	604.231	579.041	762.632	
	$t = 3$							
Вершина	n_1^3	n_2^3	n_3^3	n_4^3	n_5^3	n_6^3	n_7^3	n_8^3
$V(\{1\})$	37.212	37.212	55.684	55.684	39.540	39.540	63.664	63.664
$V(\{2\})$	74.424	74.424	111.369	111.369	79.080	79.080	127.329	127.329
$V(\{3\})$	111.636	111.636	167.053	167.053	118.620	118.620	190.993	190.993
$V(\{1, 2\})$	111.636	111.636	167.053	167.053	118.620	118.620	190.993	190.993
$V(\{1, 3\})$	148.847	148.847	222.737	222.737	158.160	158.160	254.658	254.658
$V(\{2, 3\})$	186.059	186.059	278.422	278.422	197.700	197.700	318.322	318.322
$V(\{1, 2, 3\})$	223.271	223.271	334.106	334.106	237.240	237.240	381.987	381.987

Таблица 3.12: Характеристические функции для игры и ее подыгр.

Вектора Шепли, рассчитанные по формуле (3.10.11) для игры и подыгр, представлены в табл. 3.13. На основе вектора Шепли построим позиционно состоятельную процедуру распределения дележа, используя уравнения (3.10.17) и

¹²Вычисления этого раздела были проведены в программе Mathematica 10.0.

(3.10.18) из теоремы 3.7. Процедура распределения вектора Шепли $\beta_j^c(x^*(n_i^t))$ и $\beta_j^\tau(x^*(n_i^t))$, $j \in M$ представлена на рис. 3.9. Отметим, что в начальной вершине n^0 определена только β^c , а в вершинах из множества \mathcal{N}^T — только β^τ , тогда как в промежуточных вершинах и β^c , и β^τ , найденные по формулам (3.10.17) и (3.10.18).

Период	$t = 0$		$t = 1$			$t = 2$		
Вершина	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2	
Sh_1	89.267	91.856	106.915	70.704	90.871	89.524	115.862	
Sh_2	218.265	211.452	244.643	153.459	199.776	191.853	252.342	
Sh_3	359.005	339.040	391.085	239.490	313.584	297.664	394.428	

$t = 3$								
Вершина	n_1^3	n_2^3	n_3^3	n_4^3	n_5^3	n_6^3	n_7^3	n_8^3
Sh_1	37.212	37.212	55.684	55.684	39.540	39.540	63.664	63.664
Sh_2	74.424	74.424	111.369	111.369	79.080	79.080	127.329	127.329
Sh_3	111.636	111.636	167.053	167.053	118.620	118.620	190.993	190.993

Таблица 3.13: Вектора Шепли игры и ее подыгр.

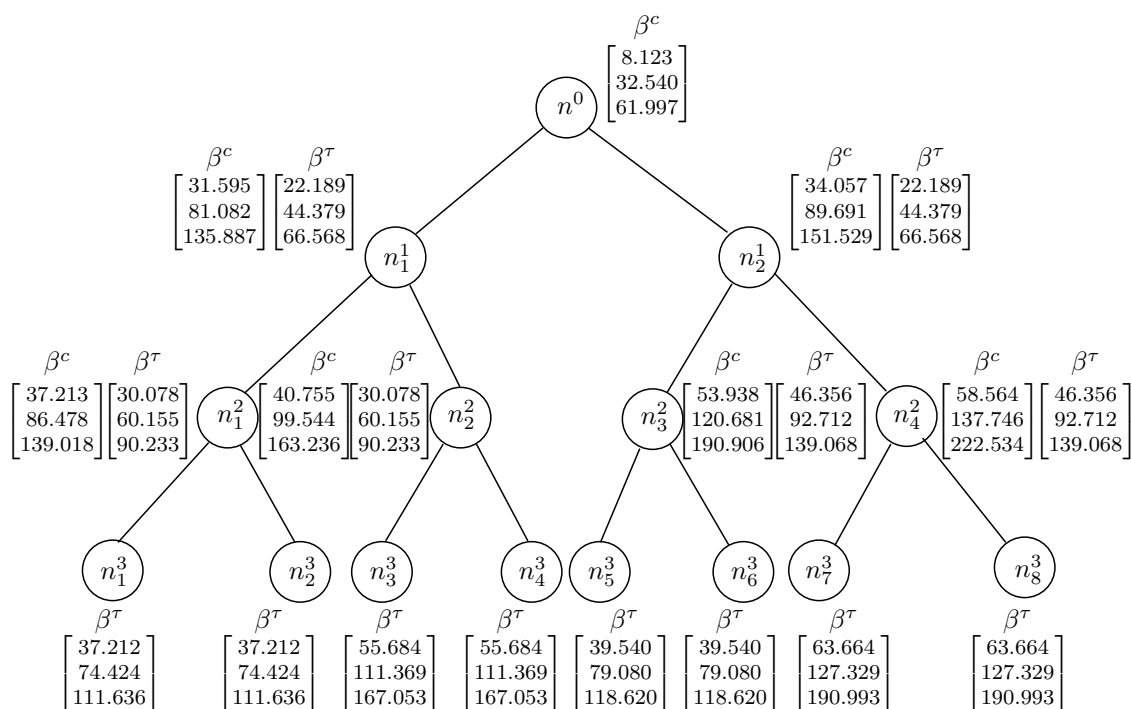


Рис. 3.9: Процедура распределения дележа.

Покажем, что первая компонента вектора Шепли равна ожидаемой сумме компонент процедуры распределения дележа. Игрок 1 должен получить 89.267

в качестве издержек в игре (см. табл. 3.13). Компоненты процедуры распределения дележа представлены на рис. 3.9. Первая компонента вектора Шепли может быть представлена следующим образом:

$$\begin{aligned}
 89.267 = & 8.123 + 0.9 \cdot \{ (0.6 - 0.05) \cdot 31.595 + 0.05 \cdot 22.189 + \\
 & + (0.4 - 0.05) \cdot 34.057 + 0.05 \cdot 22.189 \} \\
 + & 0.9^2 \cdot \{ (0.3 - 0.1) \cdot 37.213 + 0.1 \cdot 30.078 + (0.25 - 0.05) \cdot 40.755 + \\
 & + 0.05 \cdot 30.078 + (0.25 - 0.05) \cdot 53.938 \\
 & + 0.05 \cdot 46.356 + (0.1 - 0.05) \cdot 58.564 + 0.05 \cdot 46.356 \} \\
 + & 0.9^3 \cdot \{ 0.05 \cdot 37.212 + 0.15 \cdot 37.212 + 0.05 \cdot 55.684 + 0.15 \cdot 55.684 \\
 & + 0.05 \cdot 39.540 + 0.15 \cdot 39.540 + 0.025 \cdot 63.664 + 0.025 \cdot 63.664 \}.
 \end{aligned}$$

Аналогичные вычисления можно провести и для игроков 2 и 3. Если выплаты игрокам будут производиться в соответствии с ПРД, то вектор Шепли будет позиционно состоятелен.

Глава 4

Приложения стохастических игр

§ 4.1 Модели передачи данных в беспроводных сетях

В этом параграфе представлен метод моделирования работы сложной телекоммуникационной системы путем разбиения ее на небольшие подсистемы простых конфигураций и решения задачи оптимизации работы таких небольших подсистем. Описываемые сети имеют простую структуру, и передача данных может быть смоделирована стохастической игрой, где в качестве игроков выступают устройства передачи данных (например, роутеры). Состоянием в стохастической игре будет вектор, определяющий наличие пакета для передачи в каждой вершине сети.

Теоретико-игровое моделирование динамической передачи данных предполагает, что устройства передачи данных являются игроками, производящими некоторые действия. Конечно, данное представление технических устройств является условным и делается для того, чтобы определить правила передачи данных для максимизации пропускной способности сети.

Будем предполагать, что во всех играх игроки рациональны, то есть, они стремятся максимизировать свою суммарную ожидаемую прибыль.

§ 4.1.1 Игра «Дилемма пересылки»

Рассмотрим первую схему передачи данных, широко используемую при моделировании беспроводных сетей, — «Дилемму пересылки» или Forwarder's dilemma. Устройства передачи данных (хабы или роутеры) обозначены на рис. 4.1 вершинами 1 и 2. При теоретико-игровом моделировании предполагается, что в

вершинах 1 и 2 находятся игроки 1 и 2 соответственно. В этих вершинах в начале каждого периода, часто называемого в литературе как *time slot*, появляются пакеты данных единичной емкости (принимается для простоты) с вероятностями $a_1 \in [0, 1]$ и $a_2 \in [0, 1]$ соответственно. Цель игрока 1 — переслать пакет в вершину d_1 , что, как видно из схемы передачи данных, он может сделать только через игрока 2. Аналогично, цель игрока 2 — переслать пакет в вершину d_2 , что он может сделать только через игрока 1. Для успешной передачи данных игроки должны пересылать пакеты в пункты назначения друг через друга.

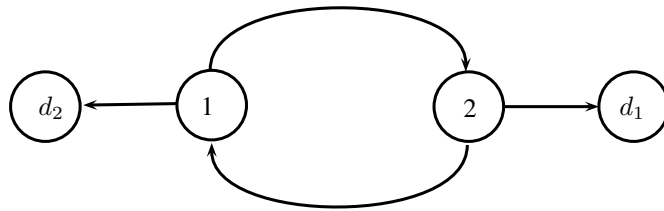


Рис. 4.1: Схема передачи данных для игры Forwarder's dilemma.

Введем систему поощрения за передачу пакета и издержки по пересылке пакетов, применяемую к игрокам:

- $c \in (0, 1)$ — издержки, которые несет игрок при пересылке одного пакета данных;
- 1 — вознаграждение игроку за успешную пересылку пакета.

Заметим, что при пересылке «чужого» пакета игрок, его пересылающий, несет издержки в размере c , вознаграждение за успешную пересылку пакета в размере 1 получает игрок, у которого этот пакет изначально появился. В случае отказа игроком переслать пакет другого игрока, этот пакет остается у игрока, у которого этот пакет изначально появился, и на следующем шаге новый пакет данных у него не появляется.

Состояние стохастической игры определяет наличие или отсутствие пакета данных у игроков 1 и 2. Множество состояний моделируемой игры есть $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$, где первый элемент вектора обозначает число пакетов у игрока 1, а второй элемент — число пакетов у игрока 2.

Рассмотрим последовательно каждое состояние игры, определим стратегии, выигрыши игроков в этих состояниях, а также вероятности перехода в состояния игры.

- Состояние $(0, 0)$.

Стратегия игрока 1 и 2 — W («ждать»). Выигрыши игроков — $(0, 0)$. Вероятности перехода из состояния $(0, 0)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ обозначим через $p_{1,1}$, $p_{1,2}$, $p_{1,3}$ и $p_{1,4}$ соответственно. Эти вероятности формируют вектор $p_1 = (p_{1,1}, p_{1,2}, p_{1,3}, p_{1,4})$, который равен $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

- Состояние $(0, 1)$.

Стратегия игрока 2 — T («посылать»). Игрок 1 имеет две стратегии: F («пересылать») и D («не пересылать»). Выигрыши игроков представлены в матрице:

$$\begin{array}{c} T \\ F \begin{pmatrix} -c, 1 \end{pmatrix} \\ D \begin{pmatrix} 0, -c \end{pmatrix} \end{array} \quad (4.1.1)$$

Вероятности перехода из состояния $(0, 1)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ при условии, что в игре реализовалась ситуация (F, T) , обозначим через $p_{2,1}(F, T)$, $p_{2,2}(F, T)$, $p_{2,3}(F, T)$ и $p_{2,4}(F, T)$ соответственно. Эти вероятности задают вектор переходных вероятностей $p_2(F, T) = (p_{2,1}(F, T), p_{2,2}(F, T), p_{2,3}(F, T), p_{2,4}(F, T))$, который равен $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

Вероятности перехода из состояния $(0, 1)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ при условии, что в игре реализовалась ситуация (D, T) , обозначим через $p_{2,1}(D, T)$, $p_{2,2}(D, T)$, $p_{2,3}(D, T)$ и $p_{2,4}(D, T)$ соответственно. Эти вероятности задают вектор переходных вероятностей $p_2(D, T) = (p_{2,1}(D, T), p_{2,2}(D, T), p_{2,3}(D, T), p_{2,4}(D, T))$, который равен $(0, 1 - a_1, 0, a_1)$.

- Состояние $(1, 0)$.

Стратегия игрока 1 — T («посылать»). Игрок 2 имеет две стратегии: F («пересылать») и D («не пересылать»). Выигрыши игроков представлены в матрице:

$$T \begin{pmatrix} F & D \\ (1, -c) & (-c, 0) \end{pmatrix} \quad (4.1.2)$$

Вектор вероятностей перехода из состояния $(1, 0)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ при условии, что в игре реализовалась ситуация (T, F) , обозначим через $p_3(T, F) = (p_{3,1}(T, F), p_{3,2}(T, F), p_{3,3}(T, F), p_{3,4}(T, F))$, он равен $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

Вектор вероятностей перехода из состояния $(1, 0)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ при условии, что в игре реализовалась ситуация (T, D) , обозначим через $p_3(T, D) = (p_{3,1}(T, D), p_{3,2}(T, D), p_{3,3}(T, D), p_{3,4}(T, D))$, он равен $(0, 0, 1 - a_2, a_2)$.

- Состояние $(1, 1)$.

В этом состоянии у каждого игрока в начале периода появился пакет данных для передачи. Стратегии любого игрока 1 или 2: F («пересылать») и D («не пересылать»). Таким образом, если каждый игрок пересылает пакет другому игроку, то пакеты считаются успешно доставленными и каждый из игроков получает выигрыш, равный единице за вычетом издержек по пересылке пакета другому игроку. Выигрыши игроков запишем в матрице:

$$\begin{matrix} & F & D \\ F & (1 - c, 1 - c) & (-c, 1) \\ D & (1, -c) & (0, 0) \end{matrix} \quad (4.1.3)$$

В этом состоянии есть четыре ситуации в чистых стратегиях. Вероятности перехода из состояния $(1, 1)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ представлены в табл. 4.1.

Будем предполагать, что игроки используют стационарные стратегии, которые зависят только от текущего состояния и не зависят от периода времени

Таблица 4.1: Вероятности перехода из состояния $(1, 1)$ в игре «Дилемма пересылки».

Ситуация	$p_4(\cdot) = (p_{4,1}(\cdot), p_{4,2}(\cdot), p_{4,3}(\cdot), p_{4,4}(\cdot))$
(F, F)	$((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$
(F, D)	$(0, 0, 1 - a_2, a_2)$
(D, F)	$(0, 1 - a_1, 0, a_1)$
(D, D)	$(0, 0, 0, 1)$

и его истории¹. Обозначим через $\eta_i \in H_i$ стратегию игрока $i = 1, 2$ в игре Γ . Стратегия η_i есть функция состояния, которая отображает его в множество вероятностных распределений на множестве чистых стратегий игрока i в этом состоянии. Пара стационарных стратегий (η_1, η_2) формируют ситуацию в игре Γ , которую обозначим через η .

В качестве выигрыша игрока i будем рассматривать ожидаемую дисконтированную сумму пошаговых выигрышей игроков:

$$E_i(\eta) = \pi_0(\mathbb{I} - \delta\Pi(\eta))^{-1}K_i(\eta), \quad (4.1.4)$$

где π_0 — вектор начального распределения по состояниям, \mathbb{I} — единичная матрица, размерность которой равна количеству состояний в игре, δ — дисконтирующий фактор, $\Pi(\eta)$ — матрица переходных вероятностей, строками которой являются вектора p_1, \dots, p_4 , определенные выше, $K_i(\eta)$ — вектор выигрышей в каждом состоянии игры при условии, что игроки реализуют ситуацию η .

Определение 4.1. Ситуация $\eta^N = (\eta_1^N, \eta_2^N)$ называется равновесием по Нэшу в стохастической игре Γ , если справедливы неравенства:

$$\begin{aligned} E_1(\eta_1^N, \eta_2^N) &\geq E_1(\eta_1, \eta_2^N) \text{ для } \forall \eta_1 \in H_1, \\ E_2(\eta_1^N, \eta_2^N) &\geq E_2(\eta_1^N, \eta_2) \text{ для } \forall \eta_2 \in H_2. \end{aligned}$$

Определение 4.2. Ситуация $\eta^* = (\eta_1^*, \eta_2^*)$ называется кооперативным решением в стохастической игре Γ , если справедливо неравенство:

$$\sum_{i=1}^2 E_i(\eta^*) \geq \sum_{i=1}^2 E_i(\eta) \text{ для } \forall \eta \in H_1 \times H_2.$$

¹Стационарные стратегии являются подклассом стратегий поведения, которые зависят от истории текущего периода. Использование стационарных стратегий оправдано для моделирования многошаговой передачи данных в беспроводных сетях, поскольку в качестве периода времени обычно берется 1 мс, и хранение истории текущего периода требует большого объема памяти.

Будем использовать следующую теорему существования равновесия по Нэшу в стохастической игре с конечным множеством стратегий каждого игрока.

Теорема 4.1. [91, 217] *Любая стохастическая игра многих лиц с ненулевой суммой с дисконт-фактором $\delta \in (0, 1)$ с конечными множествами состояний и стратегий игроков в этих состояниях имеет ситуацию равновесия по Нэшу в стационарных стратегиях.*

При вычислении выигрышей игроков по формуле (4.1.4) мы оцениваем ожидаемое число доставленных игроком пакетов данных при условии дисконтирования. Вычисление выигрышей игроков по формуле (4.1.4) позволит оценить эффективность работы протокола передачи данных (правила передачи данных в сети) при заданной конфигурации. Ситуация в игре определяет протокол передачи данных, поэтому сравнение эффективностей протоколов можно производить, сравнивая выигрыши игроков в ситуациях, определяющих эти протоколы.

Можно использовать следующий способ сравнения кооперативного и некооперативного поведения игроков. То есть, мы будем находить ситуации, оптимальные при кооперации и ее отсутствии и делать вывод об эффективности кооперации, при возможности вычисляя значение некоторого коэффициента. В качестве меры выгоды кооперативного поведения игроков по сравнению с некооперативным в игре Γ выберем отношение разности суммы выигрышей игроков при кооперации и при ее отсутствии к сумме выигрышей при отсутствии кооперации, или так называемую *стоимость отказа от кооперации*:

$$\Delta(\Gamma) = \frac{\sum_{i=1}^2 E_i(\eta^*) - \sum_{i=1}^2 E_i(\eta^N)}{\sum_{i=1}^2 E_i(\eta^N)}, \quad (4.1.5)$$

где η^* — ситуация в стационарных стратегиях, максимизирующая сумму $E_1(\eta) + E_2(\eta)$ на множестве $\eta \in H_1 \times H_2$; η^N — равновесие по Нэшу в стационарных стратегиях². Коэффициент не может быть вычислен, если $E_1(\eta^N) + E_2(\eta^N) = 0$.

Замечание 4.1. *Стоимость отказа от кооперации является преобразованной ценой анархии [117], которая равна отношению суммарного выигрыша иг-*

²Если существует несколько равновесий, то условно будем выбирать равновесие по Нэшу с наименьшим суммарным выигрышем

роков в кооперативном равновесии к их суммарному выигрышу в наилучшей ситуации равновесия по Нэшу. Мы посчитали использование стоимости отказа от кооперации вида (4.1.5), которая меньше цены анархии на единицу, более удобным в контексте рассматриваемой задачи.

Замечание 4.2. В этом параграфе мы не строим кооперативную игру, а именно, не определяем значения характеристической функции, позволяющей находить дележи или распределения совместно полученных выигрышей. Подробное описание построения кооперативной стохастической игры, а именно, характеристической функции и дележей, определяющих перераспределение выигрышей между игроками на основе тех или иных «правил», а также проверку их динамической устойчивости, можно найти в статье [163]. Для модели передачи данных в одной беспроводной сети в статье [28] была построена кооперативная игра, найден вектор Шепли для перераспределения выигрышей игроков, а также построена процедура распределения дележа, позволяющая достичь динамическую устойчивость вектора Шепли.

В силу сложности вычисления обратной матрицы в общем виде в формуле (4.1.4) в игре «Дилемма пересылки», найдем равновесие по Нэшу и кооперативное решение игры при заданных значениях параметров игры: $a_1 = 0.8$, $a_2 = 0.2$, $c = 0.1$, $\delta = 0.99$.

Запишем выигрыши игроков, вычисленные по формуле (4.1.4), в матрицу (4.1.6), строки (столбцы) соответствуют чистым стратегиям первого (второго) игрока. Стратегия игрока состоит из четырех элементов — стратегий игрока в всех состояниях в следующем порядке $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$.

$$\begin{array}{cccc}
 & WFFF & WTFD & WTDF & WTDD \\
 WFFF & (77.67, 12.33) & (64.23, 13.67) & (9.49, 18.39) & (-9.97, 20.30) \\
 WTFD & (78.58, 3.19) & (3.03, 0.12) & (43.20, 5.47) & (-0.34, 0.27) \\
 WTDF & (71.77, 10.79) & (62.16, 12.63) & (10.54, 17.84) & (-9.94, 20.20) \\
 WTTD & (79.70, -9.94) & (1.77, -0.22) & (77.80, -9.76) & (-0.21, 0.29)
 \end{array} \quad (4.1.6)$$

В игре «Дилемма пересылки» существует равновесие по Нэшу в чистых стационарных стратегиях $(WDTD, WTDD)$, которое предписывает обоим игрокам «ждать» в состоянии $(0, 0)$; «не пересылать» — игроку 1 и «посылать» — игроку 2 в состоянии $(0, 1)$; «посылать» — игроку 1 и «не пересылать» — игроку 2 в состоянии $(1, 0)$; «не пересылать» — обоим игрокам в состоянии $(1, 1)$. Таким образом, игроки не пересылают пакеты друг друга ни в каком состоянии. Выигрыши игроков позволяют сделать вывод о том, сколько пакетов игроков будет доставлено. Выигрыши игроков 1 и 2 в игре при реализации ситуации $(WDTD, WTDD)$ равны -0.21 и 0.29 соответственно, сумма выигрышей равна 0.08 . Если учесть, что выигрыш игрока при пересылке пакета данных равен единице, то можно сказать, что реализация оптимального в смысле равновесия по Нэшу поведения игроков приведет к «простою» системы.

Кооперативным решением в игре «Дилемма пересылки» является ситуация $(WFFF, WFFF)$, которая предписывает обоим игрокам «ждать» в состоянии $(0, 0)$; «пересылать» — игроку 1 и «посылать» — игроку 2 в состоянии $(0, 1)$; «посылать» — игроку 1 и «пересылать» — игроку 2 в состоянии $(1, 0)$; «пересылать» — обоим игрокам в состоянии $(1, 1)$. В этой ситуации игроки пересылают пакеты друг друга во всех состояниях. Выигрыши игроков 1 и 2 во всей игре при реализации ситуации $(WFFF, WFFF)$ равны 77.67 и 12.33 соответственно, сумма выигрышей равна 90 . Если учесть, что выигрыш игрока при пересылке пакета данных равен единице, то очевидно, что система не «простаивает», то есть, всегда, когда появляются пакеты, они доставляются в конечные вершины.

Стоимость отказа от кооперации в игре «Дилемма пересылки» вычислим по формуле (4.1.5), она равна 1124 . Это число показывает во сколько раз прибыль игроков от перехода к кооперативному поведению превосходит их выигрыш при некооперативном оптимальном поведении. Очевидно, что существует необходимость координации действий передающих устройств с целью увеличения пропускной способности системы конфигурации, представленной на рис. 4.1.

§ 4.1.2 Игра «Совместная пересылка пакета»

Рассмотрим вторую схему передачи данных, используемую при моделировании беспроводных сетей. Устройства передачи данных (игроки) обозначены на рис. 4.2 вершинами 1 и 2. В вершине s в начале каждого периода времени появляется пакет данных единичной емкости с вероятностью $a_1 \in [0, 1]$. Цель обоих игроков — переслать пакет в вершину d . Как видно из схемы передачи данных, пакет дойдет до вершины d , только если оба игрока его пересылают. В случае успешной доставки пакета оба игрока получают выигрыш 1 за вычетом издержек по пересылке пакета. Если игрок 1 будет пересылать пакет, а игрок 2 не будет этого делать, то только игрок 1 несет издержки по пересылке пакета, а недоставленный пакет возвращается в вершину s . Если же игрок 1 не пересылает пакет, то ни один из игроков не несет издержки, при этом пакет возвращается в вершину s .

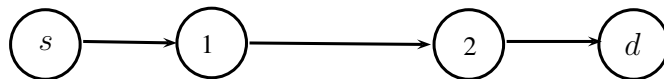


Рис. 4.2: Схема передачи данных для игры «Совместная пересылка пакета».

Состояние стохастической игры определяет наличие или отсутствие пакета данных в вершине s в начале каждого периода. Множество состояний игры есть $\{0, 1\}$.

Рассмотрим каждое состояние игры, определим стратегии, функции выигрышей игроков в этих состояниях, а также вероятности перехода.

- Состояние 0.

Стратегия игрока 1 и 2 — W («ждать»). Выигрыши игроков — $(0, 0)$. Вероятности перехода из состояния 0 в состояния 0, 1 обозначим через $p'_{1,1}$, $p'_{1,2}$ соответственно. Эти вероятности формируют вектор $p'_1 = (p'_{1,1}, p'_{1,2})$, он равен $(1 - a_1, a_1)$.

- Состояние 1.

В этом состоянии в вершине s в начале периода появился пакет данных для передачи. Стратегии любого игрока 1 или 2: F («пересылать») и D

(«не пересылать»). Таким образом, если каждый игрок пересылает пакет, то пакет считается успешно доставленным, и каждый из игроков получает выигрыш, равный единице за вычетом издержек по пересылке пакета другого игрока. Таким образом, выигрыши игроков будут следующими:

$$\begin{array}{cc} & F & D \\ \begin{array}{c} F \\ D \end{array} & \left(\begin{array}{cc} (1-c, 1-c) & (-c, 0) \\ (0, 0) & (0, 0) \end{array} \right) & \end{array} \quad (4.1.7)$$

В этом состоянии есть четыре ситуации в чистых стратегиях, вероятности перехода из состояния 1 в состояния 0 и 1 представлены в табл. 4.2.

Таблица 4.2: Вероятности перехода из состояния 1 в игре «Совместная пересылка пакета».

Ситуация	$p'_2(\cdot) = (p'_{2,1}(\cdot), p'_{2,2}(\cdot))$
(F, F)	$(1 - a_1, a_1)$
(F, D)	$(0, 1)$
(D, F)	$(0, 1)$
(D, D)	$(0, 1)$

У каждого игрока в этой стохастической игре есть две чистые стационарные стратегии. Пусть $\eta_{i,j}$ — j -ая стратегия игрока $i = 1, 2$, при этом, $j = 1, 2$. Тогда стратегия $\eta_{i,1}$ такова, что $\eta_{i,1}(1) = W$ и $\eta_{i,1}(2) = F$ для любого i . Стратегия $\eta_{i,2}$ такова, что $\eta_{i,2}(1) = W$ и $\eta_{i,2}(2) = D$ для любого i .

Пусть $\pi_0 = (\pi_{0,1}; 1 - \pi_{0,1})$. Вычислим выигрыши игроков в ситуациях, которые образуют эти стационарные стратегии, используя формулу (4.1.4), и запишем их в матрице:

$$\begin{array}{cc} & \eta_{2,1} & \eta_{2,2} \\ \begin{array}{c} \eta_{1,1} \\ \eta_{1,2} \end{array} & \left(\left(\frac{(1-c)((1-\delta)(1-\pi_{0,1})+a_1\delta)}{1-\delta}, \frac{(1-c)((1-\delta)(1-\pi_{0,1})+a_1\delta)}{1-\delta} \right), \left(-\frac{c((1-\delta)(1-\pi_{0,1})+a_1\delta)}{(1-\delta)(1-\delta(1-a_1))}, 0 \right) \right) & \\ & (0, 0) & (0, 0) \end{array} \quad (4.1.8)$$

Теорема 4.2. В игре «Совместная пересылка пакета» ситуации $(\eta_{1,1}, \eta_{2,1})$ и $(\eta_{1,2}, \eta_{2,2})$ являются равновесиями по Нэшу с выигрышами игроков $E_i(\eta_{1,1}, \eta_{2,1}) =$

$(1 - c)((1 - \delta)(1 - \pi_{0,1}) + a_1\delta)/(1 - \delta)$, $i = 1, 2$ и $E_i(\eta_{1,1}, \eta_{2,1}) = 0$, $i = 1, 2$ соответственно для любого вектора начального распределения π_0 . При этом ситуация $(\eta_{1,1}, \eta_{2,1})$ является кооперативным решением для любого π_0 .

Доказательство. Заметим, что выигрыши обоих игроков в ситуации $(\eta_{1,1}, \eta_{2,1})$ положительны, а выигрыш игрока 1 в ситуации $(\eta_{1,1}, \eta_{2,2})$ отрицателен. В остальных ситуациях выигрыши игроков равны нулю. Сначала покажем, что ситуация $(\eta_{1,1}, \eta_{2,1})$ — ситуация равновесия по Нэшу, и сумма выигрышей игроков в этой ситуации максимальна. Выигрыши игроков равны и положительны, при индивидуальном отклонении от этой ситуации игроки могут разве что уменьшить свои выигрыши: игрок 1 при индивидуальном отклонении на стратегию $\eta_{1,2}$ получит выигрыш ноль, и второй игрок при индивидуальном отклонении на стратегию $\eta_{2,2}$ также получит выигрыш ноль. В этой же ситуации сумма выигрышей игроков больше суммы их выигрышей в любой другой ситуации. Таким образом, ситуация $(\eta_{1,1}, \eta_{2,1})$ — равновесие по Нэшу и кооперативное решение.

Очевидно, что ситуация $(\eta_{1,2}, \eta_{2,2})$ также является равновесием по Нэшу с выигрышами игроков, равными нулю. Первый игрок при отклонении от стратегии $\eta_{1,2}$ на стратегию $\eta_{1,1}$ получит отрицательный выигрыш $-c((1 - \delta)(1 - \pi_{0,1}) + a_1\delta)/\{(1 - \delta)(1 - \delta(1 - a_1))\}$, а второй игрок при индивидуальном отклонении на стратегию $\eta_{2,1}$ получит такой же выигрыш ноль. Следовательно, $(\eta_{1,2}, \eta_{2,2})$ — равновесие по Нэшу. \square

Теорема 4.2 позволяет оценить необходимость кооперативного поведения игроков в игре «Совместная пересылка пакета». Как как в наихудшем равновесии по Нэшу сумма выигрышей игроков равна нулю, то стоимость отказа от кооперации вычислить невозможно. Конечно, при возможности воздействия на игроков, рекомендацией будет реализация равновесия $(\eta_{1,1}, \eta_{2,1})$ в игре, которая является кооперативным решением (социально-оптимальным поведением). Игра «Совместная пересылка пакета» является представителем такого класса игр, в котором равновесие по Нэшу совпадает с кооперативным равновесием, что на практике встречается нечасто.

Так как сумма выигрышей игроков в наихудшей ситуации равновесия по Нэшу равна нулю, то стоимость отказа от кооперации вида (4.1.5) не определена.

§ 4.1.3 Игра «Множественный доступ»

Рассмотрим третью схему передачи данных в беспроводных сетях. Устройства передачи данных (игроки) обозначены на рис. 4.3 вершинами 1 и 2. В каждой из этих вершинах в начале любого периода времени может появиться пакет данных единичной емкости с вероятностью $a_1 \in [0, 1]$ и $a_2 \in [0, 1]$ соответственно. Цель игрока $i = 1, 2$ — переслать пакет в вершину r_i . При этом, как видно из схемы передачи данных, пакет должен пройти общую для обоих игроков вершину единичной емкости. В случае, если оба игрока одновременно пересылают пакет, то эти пакеты возвращаются в начальные вершины. Для успешной доставки пакета необходимо, чтобы один из игроков пересылал пакет, а другой выбирал бы стратегию «ждать». При успешной доставке своего пакета данных, игрок получает выигрыш 1 за вычетом издержек (как и ранее, издержки равны $c \in (0, 1)$) по пересылке пакета.

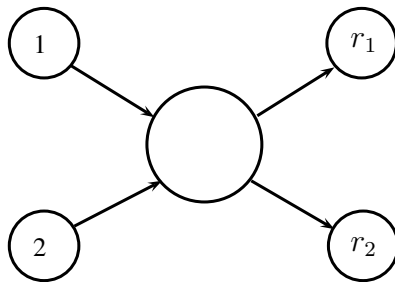


Рис. 4.3: Схема передачи данных для игры «Множественный доступ».

Состояние стохастической игры определяет наличие или отсутствие пакета данных у игроков 1 и 2. Множество состояний игры — $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$, где первый элемент вектора обозначает число пакетов у игрока 1, а второй элемент — число пакетов у игрока 2.

Рассмотрим все состояния игры, определим стратегии, выигрыши игроков в этих состояниях, а также вероятности перехода в другие состояния игры.

- Состояние $(0, 0)$.

Стратегия игрока 1 и 2 — W («ждать»). Выигрыши игроков — $(0, 0)$. Вероятности перехода из состояния $(0, 0)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ обозначим через $p''_{1,1}$, $p''_{1,2}$, $p''_{1,3}$ и $p''_{1,4}$ соответственно. Эти вероятности обра-

зуют вектор $p_1'' = (p_{1,1}'', p_{1,2}'', p_{1,3}'', p_{1,4}'')$, который равен $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

- Состояние $(0, 1)$.

Стратегия игрока 1 — W («ждать»). Стратегия игрока 2 — T («посылать»). В этом состоянии игрок 2 пересылает пакет, который успешно доставляется в конечную вершину r_2 . Выигрыши игроков — $(0, 1 - c)$.

Вероятности перехода из состояния $(0, 1)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ обозначим через $p_{2,1}'', p_{2,2}'', p_{2,3}'', p_{2,4}''$ соответственно. Эти вероятности задают вектор переходных вероятностей $p_2'' = (p_{2,1}'', p_{2,2}'', p_{2,3}'', p_{2,4}'')$, который равен $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

- Состояние $(1, 0)$.

Стратегия игрока 1 — T («посылать»). Стратегия игрока 2 — W («ждать»). В этом состоянии игрок 1 пересылает пакет, который успешно доставляется в конечную вершину r_1 . Выигрыши игроков — $(1 - c, 0)$.

Вероятности перехода из состояния $(1, 0)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ образуют вектор переходных вероятностей $p_3'' = (p_{3,1}'', p_{3,2}'', p_{3,3}'', p_{3,4}'')$, который равен $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

- Состояние $(1, 1)$.

В этом состоянии у каждого игрока в начале периода появился пакет данных для передачи. Стратегии игрока 1 или 2: T («посылать») и W («ждать»). Выигрыши игроков будут следующими:

$$\begin{array}{cc} & \begin{array}{cc} W & T \end{array} \\ \begin{array}{c} W \\ T \end{array} & \begin{pmatrix} (0, 0) & (0, 1 - c) \\ (1 - c, 0) & (-c, -c) \end{pmatrix} \end{array} \quad (4.1.9)$$

Здесь мы пренебрегаем издержками игрока за «простой» пакета данных, т. е. издержки, которые несет игрок, если пакет не был отправлен в текущий период времени.

В этом состоянии есть четыре ситуации в чистых стратегиях, запишем вероятности перехода из состояния $(1, 1)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ в табл. 4.3.

Таблица 4.3: Вероятности перехода из состояния $(1, 1)$ в игре «Множественный доступ».

Ситуация	$p_4''(\cdot) = (p_{4,1}'', p_{4,2}'', p_{4,3}'', p_{4,4}'')$
(W, W)	$(0, 0, 0, 1)$
(W, T)	$(0, 0, 1 - a_2, a_2)$
(T, W)	$(0, 1 - a_1, 0, a_1)$
(T, T)	$((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$

У игрока $i = 1, 2$ в этой стохастической игре есть две чистые стационарные стратегии: $\eta_{i,1}$ и $\eta_{i,2}$, представленные в табл. 4.4.

Таблица 4.4: Чистые стационарные стратегии игроков в игре «Множественный доступ».

Состояние	Игрок 1		Игрок 2	
	$\eta_{1,1}$	$\eta_{1,2}$	$\eta_{2,1}$	$\eta_{2,2}$
$(0, 0)$	W	W	W	W
$(0, 1)$	W	W	T	T
$(1, 0)$	T	T	W	W
$(1, 1)$	W	T	W	T

Пусть $\pi_0 = (0.25, 0.25, 0.25, 0.25)$. Вычислим выигрыши игроков в ситуациях, которые образуют стационарные стратегии, описанные в табл. 4.4, используя формулу (4.1.4), и запишем их в матрице:

$$\begin{array}{cc} & \eta_{2,1} & \eta_{2,2} \\ \begin{array}{c} \eta_{1,1} \\ \eta_{1,2} \end{array} & \left(\begin{array}{cc} (E_1(\eta_{1,1}, \eta_{2,1}), E_2(\eta_{1,1}, \eta_{2,1})) & (E_1(\eta_{1,1}, \eta_{2,2}), E_2(\eta_{1,1}, \eta_{2,2})) \\ (E_1(\eta_{1,2}, \eta_{2,1}), E_2(\eta_{1,2}, \eta_{2,1})) & (E_1(\eta_{1,2}, \eta_{2,2}), E_2(\eta_{1,2}, \eta_{2,2})) \end{array} \right), & (4.1.10) \end{array}$$

где

$$E_1(\eta_{1,1}, \eta_{2,1}) = \frac{(1 - c)(0.5(1 - \delta) + a_1\delta(1.5 - a_2))}{2(1 - \delta(1 - a_1a_2))},$$

$$E_2(\eta_{1,1}, \eta_{2,1}) = \frac{(1 - c)(0.5(1 - \delta) + a_2\delta(1.5 - a_1))}{2(1 - \delta(1 - a_1a_2))},$$

$$\begin{aligned}
E_1(\eta_{1,1}, \eta_{2,2}) &= \frac{0.5(1-c)(0.5 + (1-a_1)(a_2 - 0.5)\delta^2 + (1.5a_1 - a_1a_2 - a_2)\delta)}{(1-\delta)(1 - (1-a_1)a_2\delta)}, \\
E_2(\eta_{1,1}, \eta_{2,2}) &= \frac{(1-c)(0.5(1-\delta) + a_2\delta)}{1-\delta}, \\
E_1(\eta_{1,2}, \eta_{2,1}) &= \frac{(1-c)(0.5(1-\delta) + a_1\delta)}{1-\delta}, \\
E_2(\eta_{1,2}, \eta_{2,1}) &= \frac{0.5(1-c)(0.5 + (1-a_1)(a_2 - 0.5)\delta^2 + (1.5a_1 - a_1a_2 - a_2)\delta)}{(1-\delta)(1 - (1-a_1)a_2\delta)}, \\
E_1(\eta_{1,2}, \eta_{2,2}) &= \frac{(1-\delta)(0.25 - 0.5c) + a_1\delta(1-c-a_2)}{1-\delta}, \\
E_2(\eta_{1,2}, \eta_{2,2}) &= \frac{(1-\delta)(0.25 - 0.5c) + a_2\delta(1-c-a_1)}{1-\delta}.
\end{aligned}$$

Теорема 4.3. Пусть $\pi_0 = (0.25, 0.25, 0.25, 0.25)$. В игре «Множественный доступ» ситуации $(\eta_{1,1}, \eta_{2,2})$ и $(\eta_{1,2}, \eta_{2,1})$ являются равновесиями по Нэшу с выигрышами игроков

$$\begin{aligned}
E_1(\eta_{1,1}, \eta_{2,2}) &= \frac{0.5(1-c)(0.5 + (1-a_1)(a_2 - 0.5)\delta^2 + (1.5a_1 - a_1a_2 - a_2)\delta)}{(1-\delta)(1 - (1-a_1)a_2\delta)}, \\
E_2(\eta_{1,1}, \eta_{2,2}) &= \frac{(1-c)(a_2\delta + 0.5(1-\delta))}{1-\delta}
\end{aligned}$$

и

$$\begin{aligned}
E_1(\eta_{1,2}, \eta_{2,1}) &= \frac{(1-c)(a_1\delta + 0.5(1-\delta))}{1-\delta}, \\
E_2(\eta_{1,2}, \eta_{2,1}) &= \frac{0.5(1-c)(0.5 + (1-a_1)(a_2 - 0.5)\delta^2 + (1.5a_1 - a_1a_2 - a_2)\delta)}{(1-\delta)(1 - (1-a_1)a_2\delta)}.
\end{aligned}$$

Кооперативным решением является ситуация $(\eta_{1,1}, \eta_{2,2})$, если $a_2 > a_1$, и ситуация $(\eta_{1,2}, \eta_{2,1})$, если $a_2 < a_1$. В случае $a_1 = a_2$ обе ситуации $(\eta_{1,1}, \eta_{2,2})$, $(\eta_{1,2}, \eta_{2,1})$ являются кооперативными решениями.

Доказательство. $E_1(\eta_{1,1}, \eta_{2,1}) < E_1(\eta_{1,2}, \eta_{2,1})$, так как разность $E_1(\eta_{1,1}, \eta_{2,1}) - E_1(\eta_{1,2}, \eta_{2,1})$ равна

$$\frac{(1-c) \{-0.5(1-\delta)^2 - a_1\delta(1-\delta)(0.5 + 2a_2) - 2a_1^2a_2\delta^2\}}{2(1-\delta(1-a_1a_2))(1-\delta)} < 0.$$

Аналогично, $E_2(\eta_{1,1}, \eta_{2,1}) < E_2(\eta_{1,1}, \eta_{2,2})$, так как разность $E_2(\eta_{1,1}, \eta_{2,1}) - E_2(\eta_{1,1}, \eta_{2,2})$ равна

$$\frac{(1-c) \{-0.5(1-\delta)^2 - a_2\delta(1-\delta)(0.5 + 2a_1) - 2a_2^2a_1\delta^2\}}{2(1-\delta(1-a_1a_2))(1-\delta)} < 0.$$

Разность $E_1(\eta_{1,1}, \eta_{2,2}) - E_1(\eta_{1,2}, \eta_{2,2})$ положительна тогда и только тогда, когда

$$c > -\frac{\delta(1-a_1)(1-a_2)}{1-\delta+a_1\delta},$$

что всегда верно, поскольку выражение в правой части неравенства отрицательно, и $c \in (0, 1)$.

Аналогично, разность $E_2(\eta_{1,2}, \eta_{2,1}) - E_2(\eta_{1,2}, \eta_{2,2})$ положительна тогда и только тогда, когда

$$c > -\frac{\delta(1-a_2)(1-a_1)}{1-\delta+a_2\delta},$$

что всегда верно, поскольку выражение в правой части неравенства отрицательно, и $c \in (0, 1)$.

Следовательно, ситуации $(\eta_{1,2}, \eta_{2,1})$ и $(\eta_{1,1}, \eta_{2,2})$ являются равновесиями по Нэшу. Выигрыши игроков могут быть вычислены по формуле (4.1.4), результат представлен в формулировке теоремы.

Теперь проверим, какая ситуация является кооперативным решением. Очевидно, что сумма выигрышей игроков во всей игре в ситуациях $(\eta_{1,1}, \eta_{2,2})$ и $(\eta_{1,2}, \eta_{2,1})$ больше, чем сумма выигрышей в ситуациях $(\eta_{1,1}, \eta_{2,1})$ и $(\eta_{1,2}, \eta_{2,2})$. Найдем, при каких условиях кооперативным решением будет ситуация $(\eta_{1,1}, \eta_{2,2})$, и при каких условиях — $(\eta_{1,2}, \eta_{2,1})$. Для этого вычислим и преобразуем разность сумм выигрышей игроков в этих ситуациях:

$$\begin{aligned} E_1(\eta_{1,1}, \eta_{2,2}) + E_2(\eta_{1,1}, \eta_{2,2}) - (E_1(\eta_{1,2}, \eta_{2,1}) + E_2(\eta_{1,2}, \eta_{2,1})) = \\ = \frac{(a_2 - a_1)(1 - a_1)(1 - a_2)(\delta a_1 a_2 + 0.25(1 - \delta))\delta^2}{(1 - \delta)(1 - (1 - a_1)a_2\delta)(1 - (1 - a_2)a_1\delta)}. \end{aligned} \quad (4.1.11)$$

Эта разность положительна тогда и только тогда, когда $a_2 > a_1$. В этом случае ситуация $(\eta_{1,1}, \eta_{2,2})$ является кооперативным решением. Разность (4.1.11) отрицательна тогда и только тогда, когда $a_2 < a_1$, и в этом случае ситуация $(\eta_{1,2}, \eta_{2,1})$ является кооперативным решением. В случае равенства нулю разности (4.1.11) ($a_1 = a_2$) обе ситуации $(\eta_{1,1}, \eta_{2,2})$, $(\eta_{1,2}, \eta_{2,1})$ являются кооперативными решениями. \square

Из теоремы 4.3 следует, что кооперативное решение является и равновесием по Нэшу. Таким образом, игроки, максимизируя суммарный выигрыш, дей-

ствуют индивидуально-рационально, то есть, реализуют равновесные по Нэшу стратегии.

Но поскольку в игре «Множественный доступ» кооперативное решение является равновесием по Нэшу с большим суммарным выигрышем, то стоимость отказа от кооперации вида (4.1.5) равна 0 только в случае, когда $a_1 = a_2$. Это говорит об отсутствии необходимости координации действий передающих устройств, находящихся в вершинах 1 и 2, когда $a_1 = a_2$. Если же $a_1 > a_2$ или $a_1 < a_2$, то стоимость отказа от кооперации больше нуля. В этих случаях координация действий передающих устройств принесет увеличение прибыли.

Замечание 4.3. *Для двух из трех выше рассмотренных игр удалось получить теоретические результаты для игры общего вида, тогда как для одной игры найдено решение задачи для некоторого набора параметров. В двух из трех игр координация действий передающих устройств позволит увеличить выигрыши игроков, тем самым увеличив производительность сети. В одной игре кооперативное поведение игроков является индивидуально рациональным (при равных вероятностях появления пакетов у игроков), поэтому необходимости в осуществлении управления такой сетью нет. Можно сделать вывод о том, что прежде чем начать управление поведением участников сети, необходимо построить математическую модель, чтобы понять, нужно ли это управление. Но может получиться так, что даже если координация действий участников сети может привести к увеличению суммарного выигрыша, то это увеличение выигрыша может быть мало по сравнению с затратами на координацию действий игроков. Для того, чтобы понять, есть ли экономический смысл в управлении сетью, мы предлагаем вычислять стоимость отказа от кооперации, которая позволит произвести стоимостную оценку координации действий участников сети.*

§ 4.1.4 Игра «Множественный доступ» в случае неполной информации

В отличие от предыдущего параграфа, будем предполагать, что игроки не имеют информации о наличии пакета у другого игрока в текущий момент време-

ни, что означает, что игрок 1 в любой момент времени не различает, в каком из состояний множества $\{(0, 0), (0, 1)\}$ он находится, аналогично и для состояний из множества $\{(1, 0), (1, 1)\}$. Игрок 2 не различает состояния множества $\{(0, 0), (1, 0)\}$, а также состояния множества $\{(0, 1), (1, 1)\}$.

Пусть игрокам известен начальный вектор распределения по состояниям $\pi = (\pi_{(0,0)}, \pi_{(0,1)}, \pi_{(1,0)}, \pi_{(1,1)})$, где π_ω — вероятность того, что состояние ω реализуется в первый момент времени, $\sum_{\omega \in \Omega} \pi_\omega = 1$.

Замечание 4.4. *Исходя из постановки задачи, в качестве начального вектора распределения на множестве состояний π можно рассмотреть вектор $\pi = ((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$, который предполагает, что до хода случая в вершинах сети пакетов данных не было.*

Опишем процесс передачи данных в такой сети с помощью стохастической игры, для определения которой необходимо задать стратегии игроков и функции выигрышей в каждом состоянии и определить класс стратегий, в которых будет найдено решение игры.

Будем рассматривать класс стационарных стратегий, которые не зависят от времени, а зависят только от состояния системы (в дальнейшем, игры). В случае, когда у игрока $i = 1, 2$ нет пакета для пересылки, его единственная стратегия — ожидание, обозначим ее через w_i . Поскольку игроки не обладают полной информацией о наличии пакета у другого игрока, то стратегии игрока 1 в состояниях множества $\{(1, 0), (1, 1)\}$ одинаковые: w_1 (ожидание) и s_1 (отправка пакета). Тогда смешанной стратегией игрока 1 в состояниях множества $\{(1, 0), (1, 1)\}$ будет вектор $(1 - \xi_1, \xi_1)$, и $\xi_1 \in [0, 1]$ — вероятность выбора стратегии s_1 , т. е. вероятность отправки пакета игроком 1 в состояниях $(1, 0)$, $(1, 1)$. Таким образом, стационарная стратегия игрока 1 в игре определяется вероятностью $\xi_1 \in [0, 1]$. Аналогично, определим стратегию второго игрока. Стратегии игрока 2 в состояниях множества $\{(0, 1), (1, 1)\}$: w_2 (ожидание) и s_2 (отправка пакета). Тогда смешанной стратегией игрока 2 в состояниях множества $\{(0, 1), (1, 1)\}$ будет вектор $(1 - \xi_2, \xi_2)$, и $\xi_2 \in [0, 1]$ — вероятность отправки пакета игроком 2 в состояниях $(0, 1)$, $(1, 1)$. Таким образом, стационарная стратегия игрока 2 в игре определяется вероятностью $\xi_2 \in [0, 1]$.

Теперь последовательно рассмотрим все состояния игры, определим выигрыши игроков в этих состояниях как функции стратегий ξ_1 и ξ_2 игроков 1 и 2 соответственно и зададим вероятности перехода за один шаг из текущего состояния во все возможные состояния игры, которые в общем случае являются функциями стратегий игроков:

- Состояние $(0, 0)$.

Единственная стратегия игрока $i = 1, 2$ — стратегия w_i . Выигрыши игроков в этом состоянии: $K_1^{(0,0)}(\xi_1, \xi_2) = K_2^{(0,0)}(\xi_1, \xi_2) = 0$. Вероятности перехода из состояния $(0, 0)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ образуют вектор $p_{(0,0)}$, который равен

$$p_{(0,0)}(\xi_1, \xi_2) = ((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2).$$

- Состояние $(0, 1)$.

Единственная стратегия игрока 1 — стратегия w_1 . У игрока 2 — две чистые стратегии s_2 и w_2 . В этом состоянии игрок 2 выбирает стратегию s_2 с вероятностью $\xi_2 \in [0, 1]$. Выигрыши игроков равны $K_1^{(0,1)}(\xi_1, \xi_2) = 0$, $K_2^{(0,1)}(\xi_1, \xi_2) = \xi_2(1 - c) + (1 - \xi_2)(-d)$. Вероятности перехода из состояния $(0, 1)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ образуют вектор $p_{(0,1)}$, который равен

$$p_{(0,1)}(\xi_1, \xi_2) = (\xi_2(1 - a_1)(1 - a_2), \xi_2(1 - a_1)a_2 + (1 - \xi_2)(1 - a_1), \\ \xi_2a_1(1 - a_2), \xi_2a_1a_2 + (1 - \xi_2)a_1).$$

- Состояние $(1, 0)$.

У игрока 1 в этом состоянии две чистые стратегии s_1 и w_1 . Игрок 1 выбирает стратегию s_1 с вероятностью $\xi_1 \in [0, 1]$. Единственная стратегия игрока 2 — стратегия w_2 . Выигрыши игроков равны $K_1^{(1,0)}(\xi_1, \xi_2) = \xi_1(1 - c) + (1 - \xi_1)(-d)$, $K_2^{(1,0)}(\xi_1, \xi_2) = 0$. Вероятности перехода из состояния $(1, 0)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ образуют вектор $p_{(1,0)}$, который равен

$$p_{(1,0)}(\xi_1, \xi_2) = (\xi_1(1 - a_1)(1 - a_2), \xi_1a_1(1 - a_2) + (1 - \xi_1)(1 - a_2), \\ \xi_1(1 - a_1)a_2, \xi_1a_1a_2 + (1 - \xi_1)a_2).$$

- Состояние $(1, 1)$.

В этом состоянии у каждого игрока в начале периода есть пакет данных для передачи. Стратегии игрока 1: s_1 и w_1 , игрока 2: s_2 и w_2 . Игрок $i = 1, 2$ выбирает стратегию s_i с вероятностью $\xi_i \in [0, 1]$. Выигрыши игроков равны $K_1^{(1,1)}(\xi_1, \xi_2) = \xi_1\xi_2(-c-d) + \xi_1(1-\xi_2)(1-c) + (1-\xi_1)(-d)$, $K_2^{(1,1)}(\xi_1, \xi_2) = \xi_1\xi_2(-c-d) + \xi_2(1-\xi_1)(1-c) + (1-\xi_2)(-d)$.

Вероятности перехода из состояния $(1, 1)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ образуют вектор $p_{(1,1)}$, который равен

$$p_{(1,1)}(\xi_1, \xi_2) = (0, \xi_1(1-\xi_2)(1-a_1), (1-\xi_1)\xi_2(1-a_2), \\ (1-\xi_1)(1-\xi_2) + \xi_1\xi_2 + (1-\xi_1)\xi_2a_2 + \xi_1(1-\xi_2)a_1).$$

Стохастическая игра Γ передачи данных с сети с конфигурацией, изображенной на рис. 4.3, определяется множеством состояний Ω , множеством игроков $N = \{1, 2\}$, множествами стратегий каждого игрока в каждом состоянии из множества Ω , функциями выигрышей игроков K_i^ω , $i \in N$, в всех состояниях $\omega \in \Omega$, а также функциями перехода p_ω , которые представляют собой вероятностные распределения на множестве состояний Ω и заданы для каждого $\omega \in \Omega$.

Как уже говорилось выше, мы будем находить решения игры Γ в стационарных стратегиях. Стационарная стратегия η_i игрока i в игре Γ определяется вероятностью $\xi_i \in [0, 1]$, поэтому в дальнейшем, чтобы избежать громоздких обозначений, стационарную стратегию игрока i в игре Γ будем обозначать через ξ_i .

Вычислим математическое ожидание выигрыша игрока $i = 1, 2$ в игре Γ в предположении, что вектор начального распределения по состояниям π известен игрокам, по формуле [9, 11]:

$$E_i(\xi_1, \xi_2) = \pi(\mathbb{I} - \delta\Pi(\xi_1, \xi_2))^{-1}K_i(\xi_1, \xi_2),$$

где \mathbb{I} — единичная матрица размерности 4×4 , $\delta \in (0, 1)$ — дисконтирующий фактор, $\Pi(\xi_1, \xi_2)$ — матрица размерности 4×4 , строки которой — переходные

вероятности, т.е.,

$$\Pi(\xi_1, \xi_2) = \begin{pmatrix} p_{(0,0)}(\xi_1, \xi_2) \\ p_{(0,1)}(\xi_1, \xi_2) \\ p_{(1,0)}(\xi_1, \xi_2) \\ p_{(1,1)}(\xi_1, \xi_2) \end{pmatrix},$$

и вектор $K_i(\xi_1, \xi_2)$ выигрыша игрока i во всех состояниях:

$$K_i(\xi_1, \xi_2) = \begin{pmatrix} K_i^{(0,0)}(\xi_1, \xi_2) \\ K_i^{(0,1)}(\xi_1, \xi_2) \\ K_i^{(1,0)}(\xi_1, \xi_2) \\ K_i^{(1,1)}(\xi_1, \xi_2) \end{pmatrix}.$$

Решения игры. Цена анархии

Будем использовать некооперативный и кооперативный подходы при решении вышеописанной игры. Под некооперативным решением игры будем понимать равновесие по Нэшу. При нахождении кооперативного решения игры мы делаем предположение, что игроки объединяются в коалицию и максимизируют суммарный ожидаемый выигрыш коалиции в игре Γ . При этом игроки как и при некооперативном подходе, не обладают полной информацией о наличии пакета у другого игрока. Определим решение игры при кооперативном подходе следующим образом:

Определение 4.3. *Ситуация $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2)$ называется кооперативным решением в стохастической игре Γ , если справедливо неравенство:*

$$\sum_{i=1}^2 E_i(\tilde{\xi}) \geq \sum_{i=1}^2 E_i(\xi) \text{ для } \forall \xi \in [0, 1] \times [0, 1].$$

Кооперативное решение определяет «социальный оптимум» сети заданной конфигурации, когда ставится задача максимизации суммарного выигрыша системы или максимизации эффективности работы сети при ее централизованном управлении.

В качестве меры выгоды кооперативного поведения игроков по сравнению с некооперативным в игре Γ будем использовать *цену анархии*, которая равна отношению суммарного выигрыша игроков при кооперативном решении к их

суммарному выигрышу в наихудшей ситуации равновесия по Нэшу (равновесии по Нэшу с наименьшим суммарным выигрышем игроков):

$$PoA(\Gamma) = \frac{\sum_{i=1}^2 E_i(\tilde{\xi})}{\min_{\xi^* \in NE(\Gamma)} \sum_{i=1}^2 E_i(\xi^*)}, \quad (4.1.12)$$

где $NE(\Gamma)$ — множество равновесий по Нэшу в игре Γ . Цена анархии (4.1.12) может быть не определена, если сумма в знаменателе равна нулю.

Теорема 4.4. *В стохастической игре Γ передачи данных в сети с конфигурацией, изображенной на рис. 4.3, всегда существует равновесие по Нэшу и кооперативное решение.*

Доказательство. Поскольку множество игроков конечно, множества стратегий всех игроков во всех состояниях игры конечны, и игра решается в классе стационарных стратегий, то существование кооперативного равновесия следует из теоремы Шепли [210], а существование равновесия по Нэшу из теорем, доказанных в работах Финка и Такахаши [91, 217]. \square

В связи с тем, что в модели имеется шесть параметров $(a_1, a_2, c, d, \pi_0, \delta)$, один из которых — вектор, то записать стратегии, образующие равновесие по Нэшу или кооперативное решение в общем виде, не представляется возможным. Рассмотрим пример нахождения равновесия по Нэшу, кооперативного решения и цены анархии для одной сети.

Пусть вероятность появления пакета данных у игрока 1 больше, чем у игрока 2, а именно, $a_1 = 0.8$, $a_2 = 0.2$. Пусть издержки любого игрока по пересылке пакета равны $c = 0.1$. За простой пакета в один период времени игрок несет издержки в размере $d = 0.05$. Дисконтирующий фактор выигрышей обоих игроков $\delta = 0.99$. Предположим, что игра начинается из состояния, когда у обоих игроков имеется пакет для передачи $\pi = (0, 0, 0, 1)$.

Все вычисления проводились в программе Wolfram Mathematica 11.1 [126].

Функция выигрыша игрока 1 имеет вид:

$$181.152 / \left(-36.2304 + (-36.9142 - 10.734\xi_2)\xi_2 + (\xi_1)^2(1.42088 + \xi_2(-2.42088 + \xi_2)) + \xi_1(-3.1894 + \xi_2(53.5078 + 19.1515\xi_2)) \right) + \left(\xi_2(42.8875 + 27.0034\xi_2) + \xi_1(-3470.67 + (2899.78 - 2686.02\xi_2)\xi_2) + (\xi_1)^2(546.75 + \xi_2(1097.15 + 1578.1\xi_2)) \right) / \left(-36.2304 + (-36.9142 - 10.734\xi_2)\xi_2 + (\xi_1)^2(1.42088 + \xi_2(-2.42088 + \xi_2)) + \xi_1(-3.1894 + \xi_2(53.5078 + 19.1515\xi_2)) \right).$$

Функция выигрыша игрока 2 имеет вид:

$$181.152 / \left(-36.2304 + (-36.9142 - 10.734\xi_2)\xi_2 + (\xi_1)^2(1.42088 + \xi_2(-2.42088 + \xi_2)) + \xi_1(-3.1894 + \xi_2(53.5078 + 19.1515\xi_2)) \right) + \left((-3400.79 - 780.872\xi_2)\xi_2 + (\xi_1)^2(-7.10438 + (-435.212 - 222.744\xi_2)\xi_2) + \xi_1(15.947 + \xi_2(3308.04 + 1557.71\xi_2)) \right) / \left(-36.2304 + (-36.9142 - 10.734\xi_2)\xi_2 + (\xi_1)^2(1.42088 + \xi_2(-2.42088 + \xi_2)) + \xi_1(-3.1894 + \xi_2(53.5078 + 19.1515\xi_2)) \right).$$

Выигрыши игроков как функции вероятностей ξ_1 , ξ_2 изображены на рис. 4.4. Суммарный выигрыш игроков в игре Γ как функция стратегий ξ_1 и ξ_2 представлен на рис. 4.5.

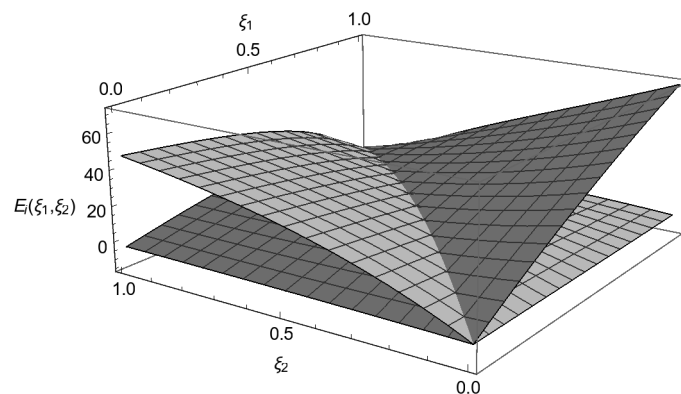


Рис. 4.4: Выигрыши игроков (игрок 1 — темно-серый, игрок 2 — светло-серый).

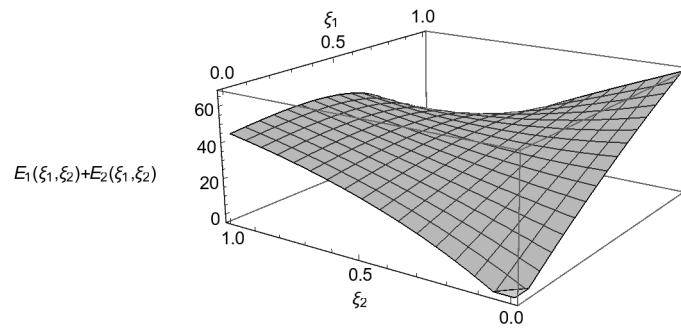


Рис. 4.5: Суммарный выигрыш игроков.

Равновесие по Нэшу определяется вероятностями $\xi_1^* = 0.955$, $\xi_2^* = 0.647$, что означает, что игрок 1, имея пакет для передачи, посылает его с вероятностью $\xi_1^* = 0.955$ и ждет с вероятностью $1 - \xi_1^* = 0.045$. При этом, игрок 2, имея пакет для передачи, посылает его с вероятностью $\xi_2^* = 0.647$ и ждет с вероятностью $1 - \xi_2^* = 0.353$. Выигрыши игроков в равновесии по Нэшу равны $E_1(\xi_1^*, \xi_2^*) = 23.538$, $E_2(\xi_1^*, \xi_2^*) = 0.496$.

Кооперативным решением в игре является ситуация $(\tilde{\xi}_1, \tilde{\xi}_2) = (1, 0)$, при которой выигрыши игроков равны $E_1(\tilde{\xi}_1, \tilde{\xi}_2) = 72.18$, $E_2(\tilde{\xi}_1, \tilde{\xi}_2) = -5$ с суммарным выигрышем игроков 67.18.

Воспользуемся формулой (4.1.12) и вычислим цену анархии в этой игре:

$$PoA(\Gamma) = \frac{67.18}{23.538 + 0.496} = 2.79521,$$

что говорит о необходимости координации стратегий игроков в данной сети.

§ 4.1.5 Игра «Посылай и пересылай»

В этом параграфе рассмотрим модель передачи данных с бóльшим количеством параметров и продемонстрируем построение кооперативной игры «Посылай и пересылай» (“Forward and transmit”) с последующей проверкой выполнения принципов устойчивой кооперации. Вершины 1 и 2 независимо генерируют пакеты данных единичной емкости в начале каждого промежутка времени

с вероятностями a_1 и a_2 соответственно, при условии, что их очереди пусты, то есть, не содержат пакета данных в конце предыдущего промежутка. Граф передачи данных изображен на рис. 4.6.

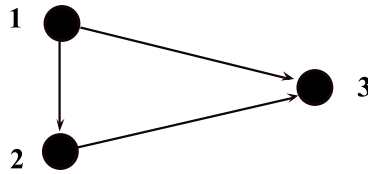


Рис. 4.6: Схема передачи данных для игры «Посылай или пересылай».

Сделаем некоторые предположения о работе системы:

1. Вершины 1 и 2 (игроки 1 и 2 соответственно) стремятся отправить пакеты данных единичной емкости в общий пункт — вершину 3.
2. Максимальная емкость любой вершины равна единице. Конечная вершина может принимать только один пакет в один промежуток времени. Предполагается, что многопакетная передача и одновременный прием и передача данных любой вершиной в любой промежуток времени невозможна.
3. Если игроки одновременно посылают пакеты в конечную вершину, то конечная вершина отвергает оба пакета, и они возвращаются в соответствующие начальные вершины, поэтому в следующий промежуток времени в вершинах 1 и 2 новые пакеты появиться не могут.
4. Все отправленные пакеты имеют одинаковый размер и требуется один промежуток времени для передачи пакета из одной вершины в другую, если эти вершины соединены дугой в графе (см. рис. 4.6).
5. Игрок 1 имеет возможность послать пакет напрямую в вершину 3 или переслать пакет игроку 2 для того, чтобы он послал этот пакет в конечную вершину 3 в следующий промежуток времени.
6. Если игрок 1 (вершина 1) посылает пакет в вершину 2, которая уже имеет пакет в своей очереди, то игрок 2 отвергает пакет игрока 1. В противном случае, игрок 2 решает принимать или отвергать пакет, посланный игроком 1.

Введем систему выигрышей и издержек:

- $f \geq 0$ — выигрыш игрока 1 или 2 за каждую успешную доставку пакета в конечную вершину 3.
- Игрок 1 получает выигрыш $c \geq 0$ от игрока 2 за доставку пакета игроку 2, который впоследствии получит выигрыш f только в случае успешной доставки этого пакета в конечную вершину в следующий промежуток времени.
- Задержка отправки пакета длиной в один промежуток времени приносит издержки в размере $d \geq 0$ игроку, в чьей очереди находится данный пакет (независимо от того, в какой вершине он изначально появился).
- D_{ij} — издержки по пересылке одного пакета из вершины i в вершину j , соединенных дугой.

Предположим, что выигрыши игроков дисконтируются с одинаковым фактором, равным $0 < \delta < 1$. Как и в предыдущих случаях, смоделируем проблему передачи данных с помощью стохастической игры. Состоянием в игре будет пара (Q_1, Q_2) , где $Q_i \in \{0, 1\}$ — число пакетов в вершине i , $i = 1, 2$. Таким образом, в игре возможны четыре состояния: $Q = \{(0, 0); (0, 1); (1, 0); (1, 1)\}$.

Последовательно рассмотрим каждое состояние и определим для него множества стратегий игроков, функции выигрышей игроков и вероятности перехода из этого состояния в любое состояние из множества Q при условии реализации любой возможной ситуации в игре:

1. Состояние $(0, 0)$. Игроки 1 и 2 имеют единственную стратегию — W («ждать»). Выигрыши игроков в этом состоянии равны $(0, 0)$. Вероятности перехода образуют вектор $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.
2. Состояние $(0, 1)$. Игрок 1 имеет единственную стратегию — W («ждать»), игрок 2 имеет единственную стратегию — $T3$ («посылать в вершину 3»). Выигрыши игроков в этом состоянии равны $(0, f - D_{23})$. Вероятности перехода образуют вектор $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

3. Состояние $(1, 0)$: Игрок 1 имеет две стратегии: 1) $T3$ («послать в вершину 3»), 2) $T2$ («переслать в вершину 2»). Игрок 2 также имеет две стратегии: 1) Ac («принять пакет из вершины 1»), 2) Rej («отвергнуть пакет из вершины 1»).

Выигрыши игроков представлены в матрице:

$$\begin{array}{cc} & Ac & Rej \\ \begin{array}{c} T3 \\ T2 \end{array} & \begin{pmatrix} (f - D_{13}, 0) & (f - D_{13}, 0) \\ (c - D_{12}, -c) & (-d - D_{12}, 0) \end{pmatrix} \end{array}$$

В этом состоянии есть четыре ситуации в чистых стратегиях. Вероятности перехода из состояния $(1, 0)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ представлены в табл. 4.5.

Таблица 4.5: Вероятности перехода из состояния $(1, 0)$ в игре «Посылай и пересылай».

Ситуация	$p_3(\cdot) = (p_{3,1}(\cdot), p_{3,2}(\cdot), p_{3,3}(\cdot), p_{3,4}(\cdot))$
$(T3, Ac)$	$((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$
$(T3, Rej)$	$((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$
$(T2, Ac)$	$(0, 1 - a_1, 0, a_1)$
$(T2, Rej)$	$(0, 0, 1 - a_2, a_2)$

4. Состояние $(1, 1)$. Игрок 1 имеет две стратегии: 1) $T3$ («переслать в вершину 3»), 2) W («ждать»). Игрок 2 имеет две стратегии: 1) $T3$ («переслать в вершину 3»), 2) W («ждать»). Выигрыши игроков представлены в матрице:

$$\begin{array}{cc} & T3 & W \\ \begin{array}{c} T3 \\ W \end{array} & \begin{pmatrix} (-d - D_{13}, -d - D_{23}) & (f - D_{13}, -d) \\ (-d, f - D_{23}) & (-d, -d) \end{pmatrix} \end{array}$$

В состоянии $(1, 1)$ есть четыре ситуации в чистых стратегиях. Вероятности перехода из состояния $(1, 1)$ в состояния $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ записаны в табл. 4.6.

Таблица 4.6: Вероятности перехода из состояния $(1, 1)$ в игре «Посылай или пересылай».

Ситуация	$p_4(\cdot) = (p_{4,1}(\cdot), p_{4,2}(\cdot), p_{4,3}(\cdot), p_{4,4}(\cdot))$
$(T3, T3)$	$(0, 0, 0, 1)$
$(T3, W)$	$(0, 1 - a_1, 0, a_1)$
$(W, T3)$	$(0, 0, 1 - a_2, a_2)$
(W, W)	$(0, 0, 0, 1)$

Обозначим чистую стратегию игрока i в стохастической игре через $\eta_i = (\eta_{ij} : j \in Q)$, где η_{ij} — стратегия игрока i в состоянии j . Ситуация в игре — $\eta = (\eta_1, \eta_2)$. Обозначим множество чистых стационарных стратегий игрока i через Ξ_i , $i = 1, 2$. Множества стратегий есть

$$\Xi_1 = \{(W, W, T3, W), (W, W, T3, T3), (W, W, T2, W), (W, W, T2, T3)\},$$

$$\Xi_2 = \{(W, T3, Ac, T3), (W, T3, Ac, W), (W, T3, Rej, T3), (W, T3, Rej, W)\}.$$

Например, чистая стратегия $\eta_1^1 = (W, W, T3, W)$ игрока 1 предписывает ему выбирать стратегию $T3$ в состоянии $(1, 0)$ и стратегию W в состояниях $(0, 0)$, $(0, 1)$ и $(1, 1)$. Каждый игрок имеет четыре чистые стационарные стратегии в стохастической игре. Следовательно, в стохастической игре будет 16 ситуаций в чистых стационарных стратегиях.

Для каждой ситуации в стационарных стратегиях вероятности перехода могут быть записаны в виде матрицы $\Pi(\eta)$. Для примера, рассмотрим ситуацию $\eta = (\eta_1, \eta_2) = ((W, W, T3, T3), (W, T3, Ac, T3))$. Для этой ситуации матрица вероятностей перехода равна

$$\Pi(\eta) = \begin{pmatrix} (1 - a_1)(1 - a_2) & (1 - a_1)a_2 & a_1(1 - a_2) & a_1a_2 \\ (1 - a_1)(1 - a_2) & (1 - a_1)a_2 & a_1(1 - a_2) & a_1a_2 \\ (1 - a_1)(1 - a_2) & (1 - a_1)a_2 & a_1(1 - a_2) & a_1a_2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Для каждой ситуации $\eta \in \Xi = \prod_{i=1}^2 \Xi_i$ мы можем посчитать ожидаемые выигрыши игроков в подыграх, которые обозначены через $E_i(\eta) = (E_i^{(0,0)}(\eta), E_i^{(0,1)}(\eta), E_i^{(1,0)}(\eta), E_i^{(1,1)}(\eta))'$, где

$$E_i(\eta) = (\mathbb{I}_4 - \delta\Pi(\eta))^{-1}K_i(\eta), \quad (4.1.13)$$

при этом $K_i(\eta)$, $\Pi(\eta)$ определены выше.

Ожидаемый выигрыш игрока i во всей игре, включая ход случая для выбора начального состояния игры, равен

$$\bar{E}_i(\eta) = \pi^0 E_i(\eta), \quad (4.1.14)$$

где $\pi^0 = (\pi_{(0,0)}^0, \pi_{(0,1)}^0, \pi_{(1,0)}^0, \pi_{(1,1)}^0)$ — вектор распределения начального состояния, где π_k^0 — вероятность того, что на первом шаге стохастической игры будет реализовано состояние $k \in Q$. Вектор π^0 задан.

Мы будем рассматривать игру в кооперативной постановке, что означает, что стратегии игроков координируются одним центром (стратегии коррелированы [67]) для того, чтобы улучшить работу сети. Координация действий передающих устройств полезна для увеличения скорости передачи данных и пропускной способности сети. Для решения кооперативной версии стохастической игры предположим, что игрок имеет информацию не только о состоянии своей очереди, но и о состоянии очереди другого игрока.

Алгоритм решения кооперативной стохастической игры

В этом разделе мы опишем последовательно этапы решения стохастической игры «Посылай или пересылай» для сети, представленной на рис. 4.6.

1. Для каждого состояния $k \in Q$ и каждой ситуации в чистых стационарных стратегиях $\eta = (\eta_1, \eta_2)$, $\eta_i \in \Xi_i$, $i = 1, 2$, посчитаем ожидаемый выигрыш игрока $E_i^k(\eta)$ в подыгре G^k по формуле (4.1.13) и ожидаемые выигрыши игроков $\bar{E}_i(\eta)$ во всей игре по формуле (4.1.14).
2. Найдем кооперативную ситуацию $\bar{\eta}$ из условия (2.3.1).
3. Вычислим значения характеристических функций $v^k(S)$ для любого состояния $k \in Q$ и любой коалиции $S \subset N$, используя уравнения (2.3.3), (2.3.4), (2.3.8). Затем вычислим значения $\bar{v}(S)$ для всех коалиций $S \subset N$ по формуле (2.3.7).
4. Найдем вектора Шепли $\alpha^k = (\alpha_1^k, \dots, \alpha_n^k)$ для любой подыгры G^k , начинающейся из состояния $k \in Q$ по формуле [211]:

$$\alpha_i^k = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (v^k(S \cup \{i\}) - v^k(S)). \quad (4.1.15)$$

И найдем вектор Шепли для всей игры $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_2)$ по формуле $\bar{\alpha}_i = \pi^0 \alpha_i$.

5. Вычислим компоненты ПРД β_i^k , $i = 1, 2$, $k \in Q$ по формуле (2.4.2).
6. Для того, чтобы сделать вектор Шепли позиционно состоятельным, определим α -регуляризацию G_α , переопределив функции выигрышей по формуле (2.4.8).
7. Проверим, существует ли равновесие по Нэшу в игре G_α в стратегиях поведения с выигрышами $(\bar{\alpha}_1, \bar{\alpha}_2)$. Для этого используем неравенство (2.4.11).

Рассмотрим численный пример игры передачи данных в беспроводной сети «Посылай и пересылай». Определим параметры для моделирования. Пусть вероятность появления пакета в вершине 1 больше, чем в вершине 2: $a_1 = 0.4$, $a_2 = 0.1$. Дисконтирующий фактор равен $\delta = 0.99$, что можно интерпретировать следующим образом: вероятность окончания игры на каждом шаге равна 0.01. Вознаграждения и издержки игроков следующие: $f = 1$, $d = 0.1$, $c = 0.3$, $D_{12} = 0.1$, $D_{13} = 0.6$, $D_{23} = 0.2$. Можно заметить, что стоимость пересылки пакета из вершины 1 в вершину 3 в три раза больше стоимости пересылки пакета из вершины 2 в вершину 3. Таким образом, кооперация вершин 1 и 2 может быть взаимовыгодна. Пусть игра может начаться из любого состояния с равной вероятностью, т. е. $\pi^0 = (0.25, 0.25, 0.25, 0.25)$.

Таблицы 4.7 и 4.8 содержат значения ожидаемых выигрышей игроков

$$E_i(\eta) = \left(E_i^{(0,0)}(\eta), E_i^{(0,1)}(\eta), E_i^{(1,0)}(\eta), E_i^{(1,1)}(\eta) \right)'$$

для любой ситуации η в чистых стационарных стратегиях для всех игроков $i = 1, 2$, и сумму этих выигрышей (последний столбец), равную $\bar{E}_1 + \bar{E}_2$, с учетом вектора начального распределения на множестве состояний π^0 .

Кооперативная ситуация, максимизирующая суммарный выигрыш игроков, есть

$$\bar{\eta} = \eta^{11} = (0, 0, 1, 1),$$

в которой стратегия первого игрока $\eta_1^{11} = (0, 0)$ предписывает ему «посылать пакет в вершину 2» в состоянии $(1, 0)$, если в вершине 1 есть пакет, а в вершине 2 пакета нет. Стратегия второго игрока $\eta_2^{11} = (1, 1)$ предписывает ему «принять

Таблица 4.7: Выигрыши игроков в игре «Посылай и пересылай» в ситуациях η^1, \dots, η^8 .

η	$E_1(\eta)$	$E_2(\eta)$	$\bar{E}_1 + \bar{E}_2$
$\eta^1 = (1, 1, 1, 1)$	-53.0129	-22.9935	-81.7048
	-53.0129	-22.1935	
	-52.6129	-22.9935	
	-70.1000	-30.0000	
$\eta^2 = (1, 1, 1, 0)$	15.8400	6.76818	23.1355
	15.8400	7.56818	
	16.2400	6.76818	
	16.2400	7.27732	
$\eta^3 = (1, 0, 1, 1)$	14.7353	7.9200	23.1855
	14.7353	8.7200	
	15.1353	7.9200	
	14.8563	8.7200	
$\eta^4 = (1, 0, 1, 0)$	-5.10968	-7.02581	-13.8016
	-5.10968	-6.22581	
	-4.70968	-7.02581	
	-10.0000	-10.0000	
$\eta^5 = (1, 1, 0, 1)$	-53.0129	-22.9935	-81.7048
	-53.0129	-22.1935	
	-52.6129	-22.9935	
	-70.0000	-30.0000	
$\eta^6 = (1, 1, 0, 0)$	15.8400	6.76818	23.1355
	15.8400	7.56818	
	16.2400	6.76818	
	16.2400	7.27732	
$\eta^7 = (1, 0, 0, 1)$	14.7353	7.9200	23.1855
	14.7353	8.7200	
	15.1353	7.9200	
	14.8563	8.7200	
$\eta^8 = (1, 0, 0, 0)$	-5.10968	-7.02581	-13.8016
	-5.10968	-6.22581	
	-4.70968	-7.02581	
	-10.0000	-10.0000	

пакет» от игрока 1 в состоянии $(1, 0)$. Если игра находится в состоянии $(1, 1)$, игрок 1 «ждет», а игрок 2 «посылает пакет» в вершину 3.

Таблица 4.8: Выигрыши игроков в игре «Посылай и пересылай» в ситуациях η^9, \dots, η^{16} .

η	$E_1(\eta)$	$E_2(\eta)$	$\bar{E}_1 + \bar{E}_2$
$\eta^9 = (0, 1, 1, 1)$	-64.7464	-27.3398	-94.3241
	-64.7464	-26.5398	
	-65.9794	-27.9446	
	-70.0000	-30.0000	
$\eta^{10} = (0, 1, 1, 0)$	11.5347	13.4228	25.4926
	11.5347	14.2228	
	11.8060	13.6218	
	12.0060	13.8218	
$\eta^{11} = (0, 0, 1, 1)$	4.90248	21.4759	26.9472
	4.90248	22.2759	
	5.04298	21.8337	
	4.87602	22.4792	
$\eta^{12} = (0, 0, 1, 0)$	-8.93504	-8.73596	-18.1458
	-8.93504	-7.93596	
	-9.06741	-8.97396	
	-10.0000	-10.0000	
$\eta^{13} = (0, 1, 0, 1)$	-64.2491	-26.728	-93.4537
	-64.2491	-25.928	
	-65.4128	-27.248	
	-70.0000	-30.000	
$\eta^{14} = (0, 1, 0, 0)$	-8.4855	5.60847	-2.43475
	-8.4855	6.40847	
	-8.8128	5.57380	
	-7.6828	6.13680	
$\eta^{15} = (0, 0, 0, 1)$	-18.532	7.9200	-10.426
	-18.532	8.7200	
	-19.010	7.9200	
	-18.910	8.7200	
$\eta^{16} = (0, 0, 0, 0)$	-10.559	-8.8313	-19.49526
	-10.559	-8.0313	
	-10.917	-9.0826	
	-10.000	-10.000	

Максимальный суммарный выигрыш игроков в игре равен

$$\max_{\eta \in \Xi} \sum_{i \in N} \bar{E}_i(\eta) = \sum_{i \in N} \bar{E}_i(\bar{\eta}) = 26.9472.$$

Вычислим значения характеристических функций для подыгр по формулам (2.3.3), (2.3.4), (2.3.8):

$$\begin{aligned}v(\{1\}) &= (-5.10968, -5.10968, -4.70968, -10.0)', \\v(\{2\}) &= (-8.735960, -7.93596, -8.97396, -10.0)', \\v(\{1, 2\}) &= (26.3784, 27.1784, 26.8766, 27.3553)',\end{aligned}$$

Характеристическая функция для всей игры находится по формуле (2.3.7):

$$\begin{aligned}\bar{v}(\{1\}) &= -6.23226, \\ \bar{v}(\{2\}) &= -8.91147, \\ \bar{v}(\{1, 2\}) &= 26.9472.\end{aligned}$$

Далее вычислим вектора Шепли для всех подыгр и игры, используя формулу (4.1.15):

- для подыгр:

$$\begin{aligned}\alpha_1 &= (15.0023, 15.0023, 15.5705, 13.6776)', \\ \alpha_2 &= (11.376, 12.176, 11.3062, 13.6776)',\end{aligned}$$

- для игры:

$$\begin{aligned}\bar{\alpha}_1 &= 14.8132, \\ \bar{\alpha}_2 &= 12.134.\end{aligned}$$

Компоненты процедуры распределения дележа β_1 для игрока 1 и β_2 для игрока 2 находятся по формуле (2.4.2) на основе векторов Шепли:

- $\beta_1 = (0, 0, 1.24274, -1.54974)'$,
- $\beta_2 = (0, 0.8, -1.34274, 2.24974)'$,

где β_i^k — выплата игроку i в состоянии k . Напомним, что выигрыши игроков в состояниях, определенные изначально заданными функциями выигрыша, следующие:

- $K_1 = (0, 0, 0.2, -0.1)'$,

$$\bullet K_2 = (0, 0.8, -0.3, 0.8)',$$

Следует отметить, что в состояниях $(0, 0)$ и $(0, 1)$ компоненты ПРД совпадают с выигрышами, заданными функциями K_1 и K_2 . В состояниях $(1, 0)$ и $(1, 1)$ происходит перераспределение выигрышей игроков. В состоянии $(1, 0)$ игроки получают суммарный выигрыш -0.1 и в соответствии с ПРД игрок 1 получает 1.24274 вместо 0.2 , а игрок 2 получает -1.34274 вместо -0.3 . Таким образом, игрок 2 платит 1.04274 игроку 1 для того, чтобы кооперативное решение (вектор Шепли) стало позиционно состоятельным. В состоянии $(1, 1)$ игроки получают суммарный выигрыш 0.7 и в соответствии с ПРД игрок 1 получает -1.54974 вместо -0.1 , а игрок 2 получает 2.24974 вместо 0.8 . Таким образом, игрок 1 платит 1.44974 игроку 2 для того, чтобы кооперативное решение стало динамически устойчивым.

Следовательно, вектор Шепли $\bar{\alpha} = (14.8132, 12.134)'$ динамически устойчив, если выплаты игрокам в состояниях производятся в соответствии с ПРД $\beta_1 = (0, 0, 1.24274, -1.54974)'$, и $\beta_2 = (0, 0.8, -1.34274, 2.24974)'$.

Теперь исследуем проблему стратегической поддержки кооперативного решения. Сначала вычислим значения $F^k(\{i\})$ для $i = 1, 2$ и $k \in Q$, определенные в Теореме 2.4:

$$1. F(\{1\}) = (-5.10968, -5.10968, -4.70968, -5.28632)',$$

$$2. F(\{2\}) = (-8.73596, -7.93596, -8.97396, -8.1858)'$$

Теперь проверим, выполняются ли неравенства (2.4.11) для любого игрока и любого состояния. Для игрока 1 неравенство (2.4.11) принимает форму:

$$\begin{pmatrix} 0 \\ 0 \\ 1.24274 \\ -1.54974 \end{pmatrix} \geq \begin{pmatrix} -0.186662 \\ -0.186662 \\ 0.418853 \\ -0.566649 \end{pmatrix}$$

Неравенство не является верным. В состоянии $(1, 1)$ игрок 1 имеет возможность увеличить свой выигрыш, отклонившись, поскольку его выигрыш в соответствии с ПРД в этом состоянии равен -1.54974 , что меньше его выигрыша

–0.566649 в случае отклонения. Это означает, что кооперация не может быть стратегически поддержана равновесием по Нэшу вида (2.4.12).

Для игрока 2 неравенство (2.4.11) принимает вид:

$$\begin{pmatrix} 0 \\ 0.8 \\ -1.34274 \\ 2.24974 \end{pmatrix} \geq \begin{pmatrix} -0.0718427 \\ 0.728157 \\ -1.01842 \\ 0.620393 \end{pmatrix}$$

Опять же, неравенство не выполнено. В состоянии (1, 0) игрок 2 может увеличить свой выигрыш, отклонившись, поскольку его выигрыш в этом состоянии в соответствии с ПРД равен -1.34274 , что меньше его выигрыша -1.01842 в случае отклонения. Можно утверждать, что ситуация (2.4.12) не является равновесием по Нэшу, поэтому кооперативное решение не может быть поддержано ситуацией (2.4.12).

§ 4.1.6 Стохастическая игра передачи данных при наличии буферов конечной емкости

Рассмотрим схему передачи данных типа АЛОНА [62, 131, 205], представленную на рис. 4.7. Игроки 1 и 2 (вершины 1 и 2) стремятся переслать как можно больше пакетов в вершины r_1 и r_2 соответственно на бесконечном промежутке времени. Все пакеты имеют единичный размер. Согласно схеме передачи данных для того, чтобы попасть конечную вершину пакет должен пройти через вершину единичной емкости, общую для обоих игроков. Игрок $i = 1, 2$ имеет буфер емкости k_i , что означает, что он может хранить от 0 до k_i пакетов единичного размера в каждый момент времени. В начале каждого периода игрок 1 и 2 может получить пакет данных единичного размера с вероятностями $\nu_1 \in (0, 1)$ и $\nu_2 \in (0, 1)$ соответственно, если на начало этого периода у него в буфере содержится меньше, чем k_i пакетов.

В каждый момент времени игрок $i = 1, 2$ может послать один пакет к конечную вершину r_i . В случае, если оба игрока одновременно посылают пакеты, они возвращаются в начальные вершины. Для успешной доставки пакета в конечную вершину необходимо, чтобы только один игрок посылал пакет. Если пакет

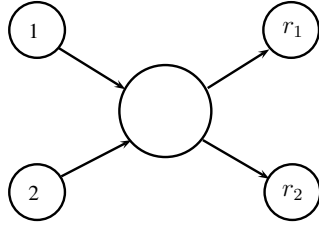


Рис. 4.7: Схема передачи данных.

успешно доставлен в конечную вершину, то игрок, которому принадлежит пакет, получает выигрыш 1 за вычетом издержек по пересылке, которые равны $c \in (0, 1)$. Также игрок несет издержки в размере $d \in [0, 1)$ за задержку одного пакета в начальной вершине за каждую единицу времени, $d \ll 1$.

Под состоянием системы в момент времени t будем понимать пару $(\omega_1(t), \omega_2(t))$, где $\omega_i(t) \in \{0, 1, \dots, k_i\}$ — число пакетов в буфере игрока i в момент времени t , $i = 1, 2$. Множество состояний системы в любой момент времени t обозначим через Ω , $|\Omega| = (k_1 + 1)(k_2 + 1)$. Пусть $\omega^{(1)}, \dots, \omega^{(m)}$ — состояния системы, где $m = (k_1 + 1)(k_2 + 1)$.

На основе сделанных предположений определим стохастическую игру с конечным множеством состояний и конечным множеством стратегий.

Время дискретно. Множество состояний в игре в каждый момент времени — $\Omega = \{\omega = (\omega_1, \omega_2) : \omega_i \in [0, k_i], i = 1, 2\}$. В состоянии ω множество стратегий игрока i есть A_i^ω :

$$A_i^\omega = \begin{cases} \{tr, w\}, & \text{если } \omega_i > 0, \\ \{w\}, & \text{если } \omega_i = 0, \end{cases} \quad (4.1.16)$$

где стратегия tr означает «послать» пакет, w — «ждать».

В состоянии $\omega = (\omega_1, \omega_2)$, где $\omega_1 = \omega_2 = 0$, функция выигрыша игрока i есть $u_i^\omega(a_1^\omega, a_2^\omega) = 0$ для любого $i = 1, 2$, где стратегии a_1^ω и a_2^ω есть w . Если $\omega = (\omega_1, \omega_2)$, $\omega_1 > 0$ и $\omega_2 = 0$, тогда функции выигрышей игроков заданы матрицей

$$\begin{pmatrix} (1 - c - d(\omega_1 - 1); 0) \\ (-d\omega_1; 0) \end{pmatrix}, \quad (4.1.17)$$

где игрок 1 выбирает строки (строка 1 соответствует стратегии tr , строка 2 — стратегии w), Игрок 2 выбирает столбцы. В случае, когда $\omega = (\omega_1, \omega_2)$, $\omega_1 = 0$

и $\omega_2 > 0$, функции игроков определены матрицей

$$\left(\begin{array}{cc} (0; 1 - c - d(\omega_2 - 1)) & (0; -d\omega_2) \end{array} \right). \quad (4.1.18)$$

Если $\omega = (\omega_1, \omega_2)$, $\omega_1 > 0$ и $\omega_2 > 0$, тогда тогда каждый игрок имеет две стратегии tr и w и функции выигрышей игроков в этом состоянии определены следующим образом:

$$\left(\begin{array}{cc} (-c - d\omega_1; -c - d\omega_2) & (1 - c - d(\omega_1 - 1); -d\omega_2) \\ (-d\omega_1; 1 - c - d(\omega_2 - 1)) & (-d\omega_1; -d\omega_2) \end{array} \right). \quad (4.1.19)$$

Предположим, что стратегии игроков в стохастической игре являются стационарными. Это предположение кажется естественным в модели передачи данных в сетях. Также предположим, что игрок не знает о состоянии буфера другого игрока, т. е. он не знает, сколько пакетов находится в буфере у другого игрока в настоящий момент времени. Пусть также игрок i использует одинаковые стратегии во всех состояниях (ω_i, ω_j) , где $\omega_i \in \{1, \dots, k_i - 1\}$, и, другую стратегию в состояниях (ω_i, ω_j) , где $\omega_i = k_i$. Следовательно, смешанная стратегия игрока i , которую обозначим через η_i есть (p_i^f, p_i^{nf}) , $i = 1, 2$, где $p_i^f \in [0, 1]$ — вероятность выбора стратегии tr в любом состоянии (ω_i, ω_j) , $j \neq i$, $j = 1, 2$, где $\omega_i = k_i$, $\omega_j \in [0, k_j]$ (вероятность пересылки пакета, когда буфер игрока i полный); $p_i^{nf} \in [0, 1]$ — вероятность выбора стратегии tr в любом состоянии (ω_i, ω_j) , $j \neq i$, $j = 1, 2$, где $\omega_i \in [1, k_i - 1]$, $\omega_j \in [0, k_j]$ (вероятность передачи пакета, если буфер игрока i не полон). Ситуация в стационарных стратегиях есть $(\eta_1, \eta_2) = ((p_1^f, p_1^{nf}), (p_2^f, p_2^{nf}))$. Обозначим через Ξ_i множество стационарных стратегий игрока i . Множество чистых стационарных стратегий игрока $i = 1, 2$ есть $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

Определим вероятности перехода $\pi(\omega''/\omega', \eta)$ в состояние $\omega'' \in \Omega$ из состояния $\omega' \in \Omega$, если ситуация η реализовалась. Вероятности $\pi(\omega''/\omega', \eta)$ представлены в табл. 4.9 и 4.10, где $\omega = (\omega_i, \omega_j)$, $i = 1, 2$, $j = 1, 2$, $i \neq j$ и $\eta = (\eta_i, \eta_j) = ((p_i^f, p_i^{nf}), (p_j^f, p_j^{nf}))$.

Таким образом, мы определили стохастическую игру G двух лиц в виде $\langle \Omega, \{A_i^\omega\}_{i=1,2;\omega \in \Omega}, \{u_i^\omega\}_{i=1,2;\omega \in \Omega}, \{\Xi_i\}_{i=1,2}, \{\pi(\omega''/\omega', \eta)\}_{\omega'' \in \Omega, \omega' \in \Omega, \eta \in \Xi_1 \times \Xi_2}, \delta \rangle$, где $\delta \in (0, 1)$ — одинаковый для всех игроков дисконт-фактор.

Таблица 4.9: Вероятности перехода

$\pi(\omega''/\omega', \eta)$	$\omega' = (\omega'_i, \omega'_j)$	$\omega'' = (\omega''_i, \omega''_j)$
$\nu_i \nu_j$	$\omega'_i = \omega'_j = 0$	$\omega''_i = \omega''_j = 1$
$\nu_i(1 - \nu_j)$	$\omega'_i = \omega'_j = 0$	$\omega''_i = 1, \omega''_j = 0$
$(1 - \nu_i)(1 - \nu_j)$	$\omega'_i = \omega'_j = 0$	$\omega''_i = \omega''_j = 0$
$(1 - \nu_i)(1 - \nu_j)p_j^{nf}$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 0,$ $\omega''_j = \omega'_j - 1$
$(1 - \nu_i)[p_j^{nf} \nu_j$ $+ (1 - p_j^{nf})(1 - \nu_j)]$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 0,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)(1 - p_j^{nf})\nu_j$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 0,$ $\omega''_j = \omega'_j + 1$
$\nu_i(1 - \nu_j)p_j^{nf}$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 1,$ $\omega''_j = \omega'_j - 1$
$\nu_i[p_j^{nf} \nu_j$ $+ (1 - p_j^{nf})(1 - \nu_j)]$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 1,$ $\omega''_j = \omega'_j$
$\nu_i(1 - p_j^{nf})\nu_j$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 1,$ $\omega''_j = \omega'_j + 1$
$(1 - \nu_i)(1 - \nu_j)p_i^{nf}(1 - p_j^{nf})$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = \omega'_i - 1,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)\nu_j p_i^{nf}(1 - p_j^{nf})$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = \omega'_i - 1,$ $\omega''_j = \omega'_j + 1$
$(1 - \nu_i)(1 - \nu_j)p_i^{nf} p_j^{nf}$ $+ \nu_i(1 - \nu_j)p_i^{nf}(1 - p_j^{nf})$ $+ (1 - \nu_i)\nu_j(1 - p_i^{nf})p_j^{nf}$ $+ (1 - \nu_i)(1 - \nu_j) \cdot$ $\cdot (1 - p_i^{nf})(1 - p_j^{nf})$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = \omega'_i,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)\nu_j(1 - p_i^{nf})(1 - p_j^{nf})$ $+ (1 - \nu_i)\nu_j p_i^{nf} p_j^{nf}$ $+ \nu_i \nu_j p_i^{nf}(1 - p_j^{nf})$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = \omega'_i,$ $\omega''_j = \omega'_j + 1$
$\nu_i \nu_j(1 - p_i^{nf})(1 - p_j^{nf})$ $+ \nu_i \nu_j p_i^{nf} p_j^{nf}$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = \omega'_i + 1,$ $\omega''_j = \omega'_j + 1$
$(1 - \nu_i)(1 - \nu_j)p_j^f$	$\omega'_i = 0,$ $\omega'_j = k_j$	$\omega''_i = 0,$ $\omega''_j = \omega'_j - 1$
$(1 - \nu_i)[1 - p_j^f(1 - \nu_j)]$	$\omega'_i = 0,$ $\omega'_j = k_j$	$\omega''_i = 0,$ $\omega''_j = \omega'_j$
$\nu_i(1 - \nu_j)p_j^f$	$\omega'_i = 0,$ $\omega'_j = k_j$	$\omega''_i = 1,$ $\omega''_j = \omega'_j - 1$

Таблица 4.10: Вероятности перехода (продолжение)

$\nu_i[1 - p_j^f(1 - \nu_j)]$	$\omega'_i = 0,$ $\omega'_j = k_j$	$\omega''_i = 1,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)p_i^f(1 - p_j^f)$	$\omega'_i = k_i,$ $\omega'_j = k_j$	$\omega''_i = \omega'_i - 1,$ $\omega''_j = \omega'_j$
$(1 - p_i^f)(1 - p_j^f)$ $+ \nu_i p_i^f(1 - p_j^f)$ $+ \nu_j(1 - p_i^f)p_j^f + p_i^f p_j^f$	$\omega'_i = k_i,$ $\omega'_j = k_j$	$\omega''_i = \omega'_i,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)p_i^{nf}(1 - p_j^f)$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j = k_j$	$\omega''_i = \omega'_i - 1,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)(1 - \nu_j)(1 - p_i^{nf})p_j^f$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j = k_j$	$\omega''_i = \omega'_i,$ $\omega''_j = \omega'_j - 1$
$\nu_i p_i^{nf} p_j^f$ $+ \nu_i(1 - p_i^{nf})(1 - p_j^f)$ $+ \nu_i \nu_j(1 - p_i^{nf})p_j^f$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j = k_j$	$\omega''_i = \omega'_i + 1,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)p_i^{nf} p_j^f$ $+ \nu_i p_i^{nf}(1 - p_j^f)$ $+ (1 - \nu_i)(1 - p_i^{nf})(1 - p_j^f)$ $+ (1 - \nu_i)\nu_j(1 - p_i^{nf})p_j^f$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j = k_j$	$\omega''_i = \omega'_i,$ $\omega''_j = \omega'_j$
$\nu_i(1 - \nu_j)(1 - p_i^{nf})p_j^f$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j = k_j$	$\omega''_i = \omega'_i + 1,$ $\omega''_j = \omega'_j - 1$

В качестве выигрыша игрока i в стохастической игре G будем рассматривать математическое ожидание выигрыша:

$$E_i(\eta) = \pi_0(\mathbb{I} - \delta\Pi(\eta))^{-1}u_i(\eta), \quad (4.1.20)$$

где π_0 — начальное распределение на множестве состояний игры, \mathbb{I} — единичная матрица размерности m , $\Pi(\eta)$ — матрица вероятностей перехода $\pi(\cdot/\cdot, \eta)$ размерности $m \times m$, элемент (l, n) которой есть вероятность перехода из состояния $\omega^{(l)}$ в состояние $\omega^{(n)}$, если ситуация η реализовалась.

При решении игры G используем два подхода (некооперативный и кооперативный). Рассмотрим равновесие по Нэшу в качестве принципа оптимальности при некооперативном подходе. В случае кооперативного варианта игры, будем искать кооперативное решение, при котором достигается максимум суммарного выигрыша игроков во всей игре G .

Как было показано ранее, в описанной выше игре G существуют как равновесие по Нэшу, так и кооперативное решение, поскольку множества стратегий всех игроков в любом состоянии ω конечны. Кооперативное решение η^* существует, поскольку является аргументом оптимизационной задачи, которая также имеет решение. Более того, эти решения достигаются в стационарных стратегиях.

Для того, чтобы оценить степень «анархии» в сети, вычислим так называемую цену анархии по формуле

$$PoA(G) = \frac{\sum_{i=1}^2 E_i(\eta^*)}{\min_{\eta \in NE(G)} \sum_{i=1}^2 E_i(\eta)}, \quad (4.1.21)$$

где $NE(G)$ — множество равновесий по Нэшу в игре G .

В качестве примера рассмотрим игру G с параметрами: $c = 0.2$, $d = 0.03$. Пусть игрок 1 имеет буфер меньшей емкости $k_1 = 2$ по сравнению с буфером игрока 2, емкость которого равна $k_2 = 4$. Оба игрока имеют дисконт-фактор 0.99. Вероятности появления пакета у игрока 1 и 2 равны $\nu_1 = 0.6$ и $\nu_2 = 0.2$ соответственно. В игре 15 состояний. Пусть игра начинается из состояния $(0, 0)$, при котором у вершин нет пакетов в буферах. Множество чистых стационарных стратегий любого игрока есть $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Смешанная стратегия i -го игрока — вектор (p_i^f, p_i^{nf}) , $i = 1, 2$. Матрица вероятностей перехода имеет размерность 15×15 , и это функция ситуации в стационарных стратегиях $\eta = (\eta_1, \eta_2)$. Для того чтобы найти кооперативное решение, необходимо посчитать суммарный выигрыш игроков во всей игре для каждой ситуаций в чистых стационарных стратегиях (всего таких ситуаций — 16) и найти максимальный суммарный выигрыш.

Для того чтобы найти равновесие по Нэшу в стохастической игре G , мы будем использовать алгоритм Лемке-Хоусона [105, 123, 198]. Для этого посчитаем матрицы ожидаемых выигрышей игроков 1 и 2 по формуле (4.1.20) для каждой ситуации в чистых стационарных стратегиях. Следовательно, получаем две

матрицы выигрышей \mathcal{A} и \mathcal{B} для игрока 1 и 2 соответственно:

$$\mathcal{A} = \begin{pmatrix} -5.8518 & -5.8518 & -5.8518 & -5.8518 \\ 47.5200 & -2.7638 & 5.0169 & -5.8518 \\ 43.7829 & 34.8935 & -12.9591 & -20.9107 \\ 47.5200 & 38.7075 & -7.5023 & -20.9767 \end{pmatrix}$$

$$\mathcal{B} = \begin{pmatrix} -10.6254 & 15.8400 & 5.5014 & 15.8400 \\ -10.6254 & 15.3433 & 3.7837 & 15.8400 \\ -10.6254 & -11.8717 & -26.0360 & -27.5207 \\ -10.6254 & -12.3383 & -21.0293 & -28.2416 \end{pmatrix}$$

Теперь используем эти матрицы в алгоритме Лемке-Хоусона.

В биматричной игре с матрицами выигрышей $(\mathcal{A}, \mathcal{B})$ существует четыре равновесия по Нэшу: (1) $\xi_1 = (0, 0.0619, 0, 0.93819)$, $\xi_2 = (1, 0, 0, 0)$ с выигрышами игрока 1 и 2, равными 47.5200 и -10.6254 соответственно; (2) $\xi_1 = (0, 0.9697, 0, 0.0303)$, $\xi_2 = (0, 0.2672, 0, 0.7328)$ с выигрышами игрока 1 и 2, равными -5.0266 и 14.5050 соответственно; (3) $\xi_1 = (0, 1, 0, 0)$, $\xi_2 = (0, 0, 0, 1)$ с выигрышами игрока 1 и 2, равными -5.8518 и 15.8400 соответственно; (4) $\xi_1 = (1, 0, 0, 0)$, $\xi_2 = (0, 0, 0, 1)$ с выигрышами игрока 1 и 2, равными -5.8518 и 15.8400 соответственно. «Наихудшее» равновесие по Нэшу, т. е. равновесие с наименьшим суммарным выигрышем игроков, есть $\xi_1 = (0, 0.9697, 0, 0.0303)$, $\xi_2 = (0, 0.2672, 0, 0.7328)$. Это равновесие в биматричной игре соответствует равновесию по Нэшу в стохастической игре G в стационарных стратегиях (η_1, η_2) . Смешанная стратегия η_1 предписывает выбирать чистую стратегию $(0, 1)$ с вероятностью 0.9697 и стратегию $(1, 1)$ с вероятностью 0.0303. Смешанная стратегия η_2 предписывает выбирать чистую стратегию $(0, 1)$ с вероятностью 0.2672 и стратегию $(1, 1)$ с вероятностью 0.7328. Оба игрока рандомизируют между своих двух стратегий $(0, 1)$ и $(1, 1)$, но игрок 1 с большей вероятностью 0.9697 выбирает стратегию $(0, 1)$ по сравнению с игроком 2, который выбирает стратегию $(0, 1)$ с вероятностью 0.2672. Суммарный ожидаемый выигрыш игроков в описанном равновесии по Нэшу равен 9.4784.

Кооперативное решение есть (η_1, η_2) , где $\eta_1 = (1, 1)$, $\eta_2 = (0, 0)$, с суммарным выигрышем игроков 36.8946. В кооперации игрок 1 посылает пакет в любом

состоянии (если он у него есть) независимо от того, полон или нет его буфер. при этом игрок 2 никогда не посылает пакеты. Цена анархии в игре G равна 3.8925.

§ 4.2 Устойчивые в динамике коалиционные структуры

§ 4.2.1 Постановка задачи. Определение вспомогательной стохастической игры

Рассмотрим кооперативную игру n лиц с коалиционной структурой π , определенную набором (N, v, π) , где N — конечное множество игроков, $|N| = n \geq 2$, вещественная функция $v : 2^N \mapsto \mathbb{R}$ — характеристическая функция, $\pi = \{B_1, \dots, B_m\}$ — коалиционная структура для множества N такая, что $B_1 \cup \dots \cup B_m = N$, $m \leq n$ и $B_i \cap B_\ell = \emptyset$ для всех $i, \ell = 1, \dots, m$, $i \neq \ell$. Обозначим множество всех коалиционных структур множества N через Π и коалиционную структуру, содержащую игрока i , через $B(i) \in \pi$.

Вектор $x^\pi = (x_1^\pi, \dots, x_n^\pi) \in \mathbb{R}^n$ называется распределением выигрыша (или вектором выигрышей) в игре (N, v, π) , если он удовлетворяет свойству эффективности, то есть, неравенство $\sum_{i \in B_\ell} x_i^\pi = v(B_\ell)$ справедливо для любой коалиции $B_\ell \in \pi$. Распределение выигрыша x^π является дележом в игре (N, v, π) , если для него выполняется свойство индивидуальной рациональности, то есть, $x_i^\pi \geq v(\{i\})$ для любого игрока $i \in N$. Решением игры (N, v, π) называется правило, которое ставит этой игре в соответствие распределение выигрыша³.

Построим стохастическую игру специального вида, определив ее элементы, включая состояния, стратегии игроков в каждом состоянии, функции выигрыша игроков в каждом состоянии, функцию перехода из одного состояния в другое, а также класс стратегий, в которых будем решать эту стохастическую игру. Время предполагается дискретным $t = 1, 2, \dots$, и каждый момент времени будем называть шагом игры.

³В классической теории кооперативных игр решением игры является дележ или подмножество множества дележей. При этом характеристическая функция обладает свойством супераддитивности. В этом параграфе мы считаем распределение выигрыша решением кооперативной игры (N, v, π) , поскольку мы не требуем выполнения свойства супераддитивности характеристической функции.

Множество состояний. Для заданного множества игроков N мы можем определить множество возможных коалиционных структур $\Pi = \{\pi_1, \dots, \pi_s\}$, которые игроки могут сформировать в игре, s — количество различных структур. Число s является числом Белла n -го порядка ⁴. Будем считать, что коалиционная структура определяет состояние стохастической игры, то есть Π — множество состояний.

Множество стратегий. Пусть на первом шаге кооперация игроков ограничена коалиционной структурой $\pi_j = \{B_1, \dots, B_m\} \in \Pi$, то есть, игровой процесс начинается из состояния π_j . В этом состоянии каждый игрок i выбирает чистую стратегию $a_i^{\pi_j}$ из конечного множества стратегий $A_i^{\pi_j}$. Множество $A_i^{\pi_j} \subset \Pi$ — подмножество коалиционных структур, которые могут сформироваться, если игрок i покинет коалицию $B(i)$ или останется в ней. Таким образом, $A_i^{\pi_j} = \{\pi_k \in \Pi : \pi_k = \{B(i) \setminus \{i\}, B_\ell \cup \{i\}, \pi_{-B(i) \cup B_\ell}\}\}$, где $B_\ell \in \pi_j \cup \emptyset$ и $\pi_{-B(i) \cup B_\ell} = \pi_j \setminus \{B(i) \cup B_\ell\}$ обозначает множество коалиций из π_j без коалиций $B(i)$ и B_ℓ . Очевидно, что $\pi_j \in A_i^{\pi_j}$ для любого игрока $i \in N$ и любой коалиционной структуры $\pi_j \in \Pi$. Смешанной стратегией игрока i в состоянии π_j является вектор $\xi_i^{\pi_j} = \{\xi_i(\pi | \pi_j)\}$ такой, что

$$\xi_i(\pi | \pi_j) \geq 0 \quad \text{и} \quad \sum_{\pi \in A_i^{\pi_j}} \xi_i(\pi | \pi_j) = 1 \quad \text{для всех } \pi \in A_i^{\pi_j}.$$

Элемент $\xi_i(\pi | \pi_j)$ — условная вероятность выбора коалиционной структуры π игроком i , если игра находится в структуре π_j на текущем шаге. Множество смешанных стратегий игрока i в состоянии π_j обозначим через $\Xi_i^{\pi_j}$. Пусть $\xi^{\pi_j} = (\xi_1^{\pi_j}, \dots, \xi_n^{\pi_j})$ — ситуация в смешанных стратегиях и $\Xi^{\pi_j} = \prod_{i \in N} \Xi_i^{\pi_j}$ — множество ситуаций в смешанных стратегиях.

Пример 1. Для игры трех лиц число Белла равно пяти, таким образом, может образоваться любая из пяти коалиционных структур из множества $\Pi = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\}$, т. е. в стохастической игре будет пять состояний. Здесь $\pi_1 = \{\{1\}, \{2\}, \{3\}\}$, $\pi_2 = \{\{1, 2\}, \{3\}\}$, $\pi_3 = \{\{1, 3\}, \{2\}\}$, $\pi_4 = \{\{1\}, \{2, 3\}\}$, $\pi_5 = \{\{1, 2, 3\}\}$. Пусть текущее состояние есть π_2 , тогда множества стратегий будут следующими: $A_1^{\pi_2} = \{\pi_1, \pi_2, \pi_3\}$, $A_2^{\pi_2} = \{\pi_1, \pi_2, \pi_4\}$ и

⁴Число Белла $B(n)$ порядка n вычисляется по рекуррентной формуле $B(n) = \sum_{k=0}^{n-1} \binom{n-1}{k} B(k)$ и $B(0) = 1$.

$A_3^{\pi_2} = \{\pi_2, \pi_5\}$. Например, игрок 1 имеет три чистые стратегии в состоянии π_2 : «отклониться в состояние π_1 », «остаться в состоянии π_2 », и «отклониться в состояние π_3 ». Смешанная стратегия игрока 1 в состоянии π_2 есть вектор вероятностей $\xi_1^{\pi_2} = (\xi_1(\pi_1|\pi_2), \xi_1(\pi_2|\pi_2), \xi_1(\pi_3|\pi_2))$.

Функции выигрыша. Для любого игрока $i \in N$ определим функцию выигрыша u_i^π в состоянии π как i -ую компоненту решения игры (N, v, π) для заданной коалиционной структуры π , то есть, $u_i^\pi = x_i^\pi$, $\pi \in \Pi$. Далее мы рассмотрим определенное решение игры (N, v, π) , а именно, в качестве решения игры будем рассматривать вектор Шепли для игры с коалиционной структурой или решение Аумана-Дрезе⁵ (подробнее см. [68]) $\phi^\pi = (\phi_1^\pi, \dots, \phi_n^\pi)$, элементы которого вычисляются по формуле:

$$\phi_i^\pi = \sum_{S \subseteq B(i); S \ni i} \frac{(|B(i)| - |S|)! (|S| - 1)!}{|B(i)|!} [v(S) - v(S \setminus \{i\})], \quad i \in N. \quad (4.2.1)$$

Таким образом, $u_i^\pi = \phi_i^\pi$ для любого игрока $i \in N$.

Функция перехода. В соответствии с определением стохастической игры необходимо задать функцию перехода из одного состояния в другое. Определим ее следующим образом. Вероятность перехода из состояния π_j в состояние π_k положительна только, если существует по крайней мере один игрок $i \in N$ такой, что $\pi_k \in A_i^{\pi_j}$. В противном случае, вероятность перехода равна нулю. Это означает, что с положительной вероятностью могут реализоваться только состояния из множества $A_1^{\pi_j} \cup \dots \cup A_n^{\pi_j}$. Эти состояния могут быть достигнуты из состояния π_j за один шаг по средствам выбора определенной стратегии игроком. Пусть в состоянии π_j реализуется ситуация ξ^{π_j} в смешанных стратегиях. Для данной ситуации ξ^{π_j} ненулевая вероятность $\lambda(\pi_k|\pi_j, \xi^{\pi_j})$ перехода из состояния π_j в состояние π_k определяется следующим образом. Предположим, что игроки используют свои смешанные стратегии независимо друг от друга,

⁵Можно рассмотреть другое решение кооперативной игры с заданной коалиционной структурой. В работе [155], элементы коалиционной структуры называются априорными множествами. Решение игры находится с использованием двухшаговой процедуры, вычисляя вектор Шепли на каждом шаге. На первом шаге значение максимальной коалиции распределяется между коалициями или элементами коалиционной структуры. На втором шаге полученные значения распределяются между игроками, членами той или иной коалиции. В нашем случае, значение Аумана-Дрезе распределяет значение каждого элемента коалиционной структуры, используя вектор Шепли.

и пусть $\lambda(\pi_k|\pi_j, \xi^{\pi_j})$ — вероятность «отклониться от структуры π_j на структуру π_k » одним игроком в то время как все остальные игроки играют стратегии «остаться в структуре π_j ». Получаем формулу для вычисления вероятностей перехода:

$$\lambda(\pi_k|\pi_j, \xi^{\pi_j}) = \begin{cases} \sum_{\substack{i \in N \\ \pi_k \in A_i^{\pi_j}}} \xi_i(\pi_k|\pi_j) \prod_{\ell \neq i} \xi_\ell(\pi_j|\pi_j), & \text{если } k \neq j, \quad \pi_k \in \bigcup_{i \in N} A_i^{\pi_j}, \\ 1 - \sum_{\ell \neq k} \lambda(\pi_\ell|\pi_j, \xi^{\pi_j}), & \text{если } k = j, \\ 0, & \text{если } \pi_k \notin \bigcup_{i \in N} A_i^{\pi_j}. \end{cases} \quad (4.2.2)$$

На следующем шаге игра переходит в новое состояние, только если ровно один игрок «отклоняется». В противном случае на следующем шаге игра остается в том же состоянии или в той же коалиционной структуре. В предложенном подходе важным моментом является предположение о том, что только один игрок должен отклониться, чтобы перевести игру в другое состояние. Это влияет на вычисление вероятностей перехода и, в последующем, на результат.

Заметим, что если мы рассматриваем только чистые стратегии игроков, то вероятности перехода, найденные по формуле (4.2.2), определяют детерминированное правило перехода из одного состояния в другое.

Замечание 4.5. *Стоит отметить, что можно предложить иные способы определения функции перехода из одного состояния в другое. Например, переход из состояния π_j в новое состояние $\pi_k \in A_1^{\pi_j} \cup \dots \cup A_n^{\pi_j}$ может произойти с положительной вероятностью, если реализуется одно из двух условий: (1) только один любой игрок i выбирает стратегию $\pi_k \in A_j^{\pi_j}$ в состоянии π_j в то время как все остальные игроки остаются в состоянии π_j , или (2) все игроки $t \in N$, у которых $A_t^{\pi_j} \ni \pi_k$ выбирают стратегию «отклониться от структуры π_j к структуре π_k », при этом все остальные игроки остаются в структуре π_j . В соответствии с определением этой функции перехода, веро-*

ятность $\lambda(\pi_k|\pi_j, \xi^{\pi_j})$ перехода из состояния π_j в состояние π_k равна:

$$\lambda(\pi_k|\pi_j, \xi^{\pi_j}) = \begin{cases} \sum_{\substack{i \in N \\ \pi_k \in A_i^{\pi_j}}} \xi_i(\pi_k|\pi_j) \prod_{\ell \neq i} \xi_\ell(\pi_j|\pi_j) + \\ + \prod_{\substack{i \in N \\ \pi_k \in A_i^{\pi_j}}} \xi_i(\pi_k|\pi_j) \cdot \prod_{\substack{\ell \in N \\ \pi_k \notin A_\ell^{\pi_j}}} \xi_\ell(\pi_j|\pi_j), & \text{если } k \neq j, \pi_k \in \bigcup_{i \in N} A_i^{\pi_j}, \\ 1 - \sum_{\ell \neq k} \lambda(\pi_\ell|\pi_j, \xi^{\pi_j}), & \text{если } k = j, \\ 0, & \text{если } \pi_k \notin \bigcup_{i \in N} A_i^{\pi_j}. \end{cases}$$

Более общее правило перехода рассмотрено в работе [116], где правило не учитывает стратегии игроков.

Стационарные стратегии. Для нахождения решения стохастической игры необходимо определить класс стратегий игроков. Будем предполагать, что игроки используют стационарные стратегии. Обозначим стационарную стратегию игрока $i \in N$ через $\eta_i = (\eta_i(\pi_1), \dots, \eta_i(\pi_s))$, где $\eta_i(\pi) = \xi_i^\pi \in \Xi_i^\pi$ для всех $\pi \in \Pi$. Стационарная стратегия η_i игрока i предписывает выбор определенной стратегии в каждом состоянии игры, т. е. предписывает вероятностное распределение $\xi_i^\pi \in \Xi_i^\pi$ на множестве чистых стратегий A_i^π для любого состояния $\pi \in \Pi$. Множество стационарных стратегий игрока i обозначим через H_i . Тогда $\eta = (\eta_1, \dots, \eta_n)$ — ситуация в стационарных стратегиях и $H = \prod_{i \in N} H_i$.

Матрица переходных вероятностей. Для заданной ситуации $\eta \in H$, функция перехода может быть записана в виде матрицы переходных вероятностей $\Lambda(\eta)$, которая является функцией ситуации η . Матрица переходных вероятностей имеет вид:

$$\Lambda(\eta) = \begin{matrix} & \pi_1 & \dots & \pi_j & \dots & \pi_s \\ \begin{matrix} \pi_1 \\ \vdots \\ \pi_j \\ \vdots \\ \pi_s \end{matrix} & \left(\begin{array}{cccccc} \lambda(\pi_1|\pi_1, \xi^{\pi_1}) & \dots & \lambda(\pi_j|\pi_1, \xi^{\pi_1}) & \dots & \lambda(\pi_s|\pi_1, \xi^{\pi_1}) \\ \vdots & & \vdots & & \vdots \\ \lambda(\pi_1|\pi_j, \xi^{\pi_j}) & \dots & \lambda(\pi_j|\pi_j, \xi^{\pi_j}) & \dots & \lambda(\pi_s|\pi_j, \xi^{\pi_j}) \\ \vdots & & \vdots & & \vdots \\ \lambda(\pi_1|\pi_s, \xi^{\pi_s}) & \dots & \lambda(\pi_j|\pi_s, \xi^{\pi_s}) & \dots & \lambda(\pi_s|\pi_s, \xi^{\pi_s}) \end{array} \right) \end{matrix}$$

Здесь элемент (j, k) есть $\lambda(\pi_k | \pi_j, \xi^{\pi_j})$ или вероятность перехода из состояния π_j в состояние π_k при условии, что реализуется ситуация ξ^{π_j} .

Теперь мы можем определить стохастическую игру.

Определение 4.4. *Стохастическая игра — это набор*

$$\Gamma = \langle N, \Pi, \{A_i, u_i\}_{i \in N}, \Lambda, \delta \rangle,$$

где $\delta \in (0, 1)$ — одинаковый для всех игроков дисконт-фактор.

В стохастической игре Γ в стационарных стратегиях $\eta_i \in H_i$, $i \in N$ дисконтированный выигрыш игрока i находится как решение рекуррентного уравнения (в векторной форме):

$$E_i(\eta) = u_i + \delta \Lambda(\eta) E_i(\eta) = \phi_i + \delta \Lambda(\eta) E_i(\eta),$$

или в эквивалентной форме

$$E_i(\eta) = (\mathbb{I}_s - \delta \Lambda(\eta))^{-1} \phi_i \quad (4.2.3)$$

где $E_i(\eta) = (E_i^{\pi_1}(\eta), \dots, E_i^{\pi_s}(\eta))^T$, $E_i^{\pi_j}(\eta)$ — математическое ожидание выигрыша игрока i в игре Γ , начинающейся из состояния π_j , $\phi_i = (\phi_i^{\pi_1}, \dots, \phi_i^{\pi_s})^T$ — вектор, элементы которого являются компонентами решений в состояниях, и \mathbb{I}_s — единичная матрица размерности s .

Определение 4.5. *Ситуация $\eta^* = (\eta_1^*, \dots, \eta_n^*) \in H$ называется равновесием по Нэшу, если для любого игрока $i \in N$ и любой стационарной стратегии $\eta_i \in H_i$ справедливо неравенство (в векторной форме):*

$$E_i(\eta_1^*, \dots, \eta_{i-1}^*, \eta_i^*, \eta_{i+1}^*, \dots, \eta_n^*) \geq E_i(\eta_1^*, \dots, \eta_{i-1}^*, \eta_i, \eta_{i+1}^*, \dots, \eta_n^*). \quad (4.2.4)$$

Существование равновесия по Нэшу в стохастической игре в стационарных стратегиях было доказано в работах [91, 217].

Теорема 4.5. *В стохастической игре Γ существует равновесие по Нэшу $\eta^* = (\eta_1^*, \dots, \eta_n^*)$, $\eta_i^* \in H_i$ в стационарных стратегиях.*

§ 4.2.2 D-устойчивые коалиционные структуры

Рассмотрим стохастическую игру Γ в стационарных стратегиях. Пусть ситуация $\eta^* = (\eta_1^*, \dots, \eta_n^*) \in H$ такая, что $\eta_i^*(\pi_j) = \xi_i^{\pi_j^*}$ для всех $\pi_j \in \Pi$ и любого игрока $i \in N$.

Определение 4.6. Коалиционная структура π_k называется *d-устойчивой*,⁶ если в игре Γ существует равновесие по Нэшу в стационарных стратегиях η^* такое, что $\lambda(\pi_k | \pi_k, \xi^{\pi_k^*}) = 1$, и для любой коалиционной структуры $\pi_j \in \Pi$, $\pi_j \neq \pi_k$, вероятность перехода из структуры π_j в структуру π_k за ℓ шагов положительна, где ℓ конечно.

Когда мы говорим о d-устойчивой коалиционной структуре, то мы подразумеваем, что эта структура устойчива относительно некоторого заранее оговоренного решения игры. Решение игры (N, v, π) , например, вектор Шепли, который взят в качестве решения, определяет функции выигрыша игроков во всех состояниях u_i^π , $\pi \in \Pi$ для игрока $i \in N$ в стохастической игре Γ специального вида. Для упрощения обозначений мы будем опускать фразу «относительно вектора Шепли», которая подразумевается каждый раз, когда мы говорим об устойчивости.

D-устойчивую коалиционную структуру можно считать «предельным» состоянием стохастического процесса, определяемого равновесием по Нэшу η^* (и матрицей вероятностей перехода $\Lambda(\eta^*)$). Таким образом, для любого начального состояния (начальной коалиционной структуры) ситуация равновесия по Нэшу может привести к одной из d-устойчивых коалиционных структур. Следует отметить, что мы не утверждаем единственности d-устойчивой коалиционной структуры.

Важной характеристикой d-устойчивой коалиционной структуры является ожидаемое число шагов, необходимых для того, чтобы игры достигла этой структуры из начальной структуры π_j . В терминологии цепей Маркова эта характеристика также является числом шагов до поглощения этим состоянием цепи Маркова, если начальным является состояние π_j . То есть, d-устойчивая ко-

⁶Буква d перед словом устойчивая обозначает dynamically.

алиционная структура — поглощающее состояние цепи Маркова, определенной матрицей $\Lambda(\eta^*)$. Цепь Маркова называется поглощающей, если существует хотя бы одно поглощающее состояние, и можно достичь хотя бы одно из поглощающих состояний из любого другого состояния за конечное число шагов. Следуя определению 4.6, мы можем утверждать, что если η^* — равновесие по Нэшу, и существует поглощающее состояние в цепи Маркова, генерируемой матрицей переходных вероятностей $\Lambda(\eta^*)$, то d -устойчивая коалиционная структура существует.

Для вычисления ожидаемого числа шагов до того, как стохастический процесс попадет в d -устойчивую коалиционную структуру, используем теорию цепей Маркова. Пусть поглощающая цепь Маркова с матрицей вероятностей перехода $\Lambda(\eta^*)$ имеет $s - p$ переходных состояний и m поглощающих состояний. Тогда матрица $\Lambda(\eta^*)$ представима в виде

$$\Lambda(\eta^*) = \begin{pmatrix} \Theta(\eta^*) & R(\eta^*) \\ \mathbf{0} & \mathbb{I}_p \end{pmatrix}$$

где $\Theta(\eta^*)$ — матрица размерности $(s - p) \times (s - p)$, $R(\eta^*)$ — матрица размерности $(s - p) \times p$ с ненулевыми элементами, $\mathbf{0}$ — нулевая матрица размерности $p \times (s - p)$, \mathbb{I}_p — единичная матрица размерности p . Когда игроки реализуют равновесие по Нэшу η^* , то ожидаемое число шагов, необходимое, чтобы процесс игры попал в d -устойчивую коалиционную структуру из всех переходных состояний определяется элементами вектора $\zeta'(\eta^*) = (\mathbb{I}_{s-p} - \Theta(\eta^*))^{-1} \mathbf{1}$, где матрица $(\mathbb{I}_{s-p} - \Theta)^{-1}$ — фундаментальная матрица для матрицы $\Lambda(\eta^*)$, и $\mathbf{1}$ — вектор размерности $(s - p)$, все элементы которого равны 1. Следуя определению 4.6, можно утверждать, что d -устойчивая коалиционная структура генерируется равновесием по Нэшу η^* таким, что $p = 1$ для матрицы $\Lambda(\eta^*)$. Добавляя нулевой элемент к вектору $\zeta'(\eta^*)$ на соответствующее место (на место, соответствующее номеру d -устойчивой коалиционной структуры), получаем вектор $\zeta(\eta^*)$ размерности s ожидаемого числа шагов до перехода игры в d -устойчивую коалиционную структуру.

§ 4.2.3 D-устойчивые коалиционные структуры для игры двух лиц

Рассмотрим игру двух лиц с характеристической функцией $v(\{1, 2\}) = v_{12}$, $v(\{1\}) = v_1$, $v(\{2\}) = v_2$. В стохастической игре Γ множество состояний есть $\Pi = \{\pi_1, \pi_2\}$, где $\pi_1 = \{\{1, 2\}\}$, $\pi_2 = \{\{1\}, \{2\}\}$. Пусть $\phi_1 = (\phi_1^{\pi_1}, \phi_1^{\pi_2})$, $\phi_2 = (\phi_2^{\pi_1}, \phi_2^{\pi_2})$, где $\phi_1^{\pi_1} = (v_{12} + v_1 - v_2)/2$, $\phi_1^{\pi_2} = v_1$, $\phi_2^{\pi_1} = (v_{12} - v_1 + v_2)/2$, $\phi_2^{\pi_2} = v_2$.

В этом параграфе мы покажем, что в любой игре двух лиц существует равновесие по Нэшу в чистых стационарных стратегиях, приводящее к d-устойчивой коалиционной структуре. Пусть $\eta_1 = (\eta_1(\pi_1), \eta_1(\pi_2))$ — стационарная стратегия игрока 1, $\eta_2 = (\eta_2(\pi_1), \eta_2(\pi_2))$ — стационарная стратегия игрока 2, где

$$\begin{aligned}\eta_i(\pi_1) &= (\xi_i(\pi_1|\pi_1), \xi_i(\pi_2|\pi_1)), \\ \eta_i(\pi_2) &= (\xi_i(\pi_1|\pi_2), \xi_i(\pi_2|\pi_2)), \quad i = 1, 2.\end{aligned}$$

Для ситуации η матрица вероятностей перехода $\Lambda(\eta)$ имеет вид:

$$\Lambda(\eta) = \begin{array}{cc} & \begin{array}{c} \pi_1 \\ \pi_2 \end{array} \\ \begin{array}{c} \pi_1 \\ \pi_2 \end{array} & \begin{pmatrix} 1 - \lambda(\pi_2|\pi_1, \xi^{\pi_1}) & \lambda(\pi_2|\pi_1, \xi^{\pi_1}) \\ \lambda(\pi_1|\pi_2, \xi^{\pi_2}) & 1 - \lambda(\pi_1|\pi_2, \xi^{\pi_2}) \end{pmatrix} \end{array}$$

где

$$\begin{aligned}\lambda(\pi_2|\pi_1, \xi^{\pi_1}) &= \xi_1(\pi_2|\pi_1)\xi_2(\pi_1|\pi_1) + \xi_2(\pi_2|\pi_1)\xi_1(\pi_1|\pi_1), \\ \lambda(\pi_1|\pi_2, \xi^{\pi_2}) &= \xi_1(\pi_1|\pi_2)\xi_2(\pi_2|\pi_2) + \xi_2(\pi_1|\pi_2)\xi_1(\pi_2|\pi_2).\end{aligned}$$

В чистых стационарных стратегиях матрица вероятностей перехода $\Lambda(\eta)$ принимает одну из четырех форм:

$$\Lambda_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad \Lambda_2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad \Lambda_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Lambda_4 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

1. Если $\Lambda(\eta) = \Lambda_1$, то коалиционная структура π_1 d-устойчива. Ожидаемое количество шагов до момента достижения d-устойчивой коалиционной структуры равно 0 и 1 для состояний π_1 и π_2 соответственно. Четыре ситуации в чистых стационарных стратегиях, которые дают матрицу вероятностей перехода формы Λ_1 следующие: $\{\eta_1 = ((1, 0), (0, 1)), \eta_2 =$

$((1, 0), (1, 0))\}, \{\eta_1 = ((1, 0), (1, 0)), \eta_2 = ((1, 0), (0, 1))\}, \{\eta_1 = ((0, 1), (0, 1)), \eta_2 = ((0, 1), (1, 0))\}, \{\eta_1 = ((0, 1), (1, 0)), \eta_2 = ((0, 1), (0, 1))\}$. При этом, выигрыши игроков в игре Γ равны

$$E_i(\eta) = (\mathbb{I}_2 - \delta\Lambda_1)^{-1}\phi_i = \begin{pmatrix} \phi_i^{\pi_1}/(1 - \delta) \\ v_i + \delta\phi_i^{\pi_1}/(1 - \delta) \end{pmatrix}, \quad i = 1, 2.$$

2. Если $\Lambda(\eta) = \Lambda_2$, коалиционная структура π_2 d-устойчива. Ожидаемое количество шагов до момента достижения d-устойчивой коалиционной структуры равно 1 и 0 для состояний π_1 и π_2 соответственно. Четыре ситуации в чистых стационарных стратегиях, которые дают матрицу вероятностей перехода вида Λ_2 следующие: $\{\eta_1 = ((0, 1), (0, 1)), \eta_2 = ((1, 0), (0, 1))\}, \{\eta_1 = ((1, 0), (0, 1)), \eta_2 = ((0, 1), (0, 1))\}, \{\eta_1 = ((0, 1), (1, 0)), \eta_2 = ((1, 0), (1, 0))\}, \{\eta_1 = ((1, 0), (1, 0)), \eta_2 = ((0, 1), (1, 0))\}$. Выигрыши игроков в игре Γ равны

$$E_i(\eta) = (\mathbb{I}_2 - \delta\Lambda_2)^{-1}\phi_i = \begin{pmatrix} \phi_i^{\pi_1} + \delta v_i/(1 - \delta) \\ v_i/(1 - \delta) \end{pmatrix}, \quad i = 1, 2.$$

3. Если $\Lambda(\eta) = \Lambda_3$, не существует d-устойчивых коалиционных структур, хотя есть четыре ситуации в чистых стационарных стратегиях, которые дают матрицу вероятностей перехода вида Λ_3 : $\{\eta_1 = ((1, 0), (0, 1)), \eta_2 = ((1, 0), (0, 1))\}, \{\eta_1 = ((1, 0), (1, 0)), \eta_2 = ((1, 0), (1, 0))\}, \{\eta_1 = ((0, 1), (0, 1)), \eta_2 = ((0, 1), (0, 1))\}, \{\eta_1 = ((0, 1), (1, 0)), \eta_2 = ((0, 1), (1, 0))\}$. Выигрыши игроков в игре Γ равны

$$E_i(\eta) = (\mathbb{I}_2 - \delta\Lambda_3)^{-1}\phi_i = \begin{pmatrix} \phi_i^{\pi_1}/(1 - \delta) \\ v_i/(1 - \delta) \end{pmatrix}, \quad i = 1, 2.$$

4. Если $\Lambda(\eta) = \Lambda_4$, не существует d-устойчивых коалиционных структур, хотя есть четыре ситуации в чистых стационарных стратегиях, которые дают матрицу вероятностей перехода вида Λ_4 : $\{\eta_1 = ((0, 1), (1, 0)), \eta_2 = ((1, 0), (0, 1))\}, \{\eta_1 = ((1, 0), (1, 0)), \eta_2 = ((0, 1), (0, 1))\}, \{\eta_1 = ((0, 1), (0, 1)), \eta_2 = ((1, 0), (1, 0))\}, \{\eta_1 = ((1, 0), (0, 1)), \eta_2 = ((0, 1), (1, 0))\}$. Выигрыши иг-

роков в игре Γ равны

$$E_i(\eta) = (\mathbb{I}_2 - \delta\Lambda_4)^{-1}\phi_i = \begin{pmatrix} (\phi_i^{\pi_1} + \delta v_i)/(1 - \delta^2) \\ (\delta\phi_i^{\pi_1} + v_i)/(1 - \delta^2) \end{pmatrix}, \quad i = 1, 2.$$

Теорема 4.6. *Для стохастической игры Γ двух лиц, в которой матрица вероятностей перехода определяется правилом (4.2.2), справедливы утверждения:*

1. *Если $v_{12} > v_1 + v_2$, коалиционная структура π_1 d -устойчива, и любая ситуация в стационарных стратегиях η такая, что $\Lambda(\eta) = \Lambda_1$, является равновесием по Нэшу.*
2. *Если $v_{12} < v_1 + v_2$, коалиционная структура π_2 d -устойчива, и любая ситуация в стационарных стратегиях η такая, что $\Lambda(\eta) = \Lambda_2$, является равновесием по Нэшу.*
3. *Если $v_{12} = v_1 + v_2$, то есть, характеристическая функция аддитивна, обе коалиционные структуры π_1 и π_2 d -устойчивы.*

Доказательство. Если следующие неравенства

$$\begin{aligned} \phi_i^{\pi_1}/(1 - \delta) &\geq \phi_i^{\pi_1} + \delta v_i/(1 - \delta), & v_i + \delta\phi_i^{\pi_1}/(1 - \delta) &\geq v_i/(1 - \delta), \\ \phi_i^{\pi_1}/(1 - \delta) &\geq \phi_i^{\pi_1}/(1 - \delta), & v_i + \delta\phi_i^{\pi_1}/(1 - \delta) &\geq v_i/(1 - \delta), \\ \phi_i^{\pi_1}/(1 - \delta) &\geq (\phi_i^{\pi_1} + \delta v_i)/(1 - \delta^2), & v_i + \delta\phi_i^{\pi_1}/(1 - \delta) &\geq (\delta\phi_i^{\pi_1} + v_i)/(1 - \delta^2), \end{aligned}$$

выполняются для любого $i = 1, 2$, то ситуация η такая, что $\Lambda(\eta) = \Lambda_1$, является равновесием по Нэшу. Эти неравенства эквивалентны следующим: $\phi_i^{\pi_1} \geq v_i$, $i = 1, 2$. Учитывая неравенство $v_{12} > v_1 + v_2$, получим $\phi_i^{\pi_1} > v_i$, $i = 1, 2$. Из определения d -устойчивой коалиционной структуры, мы можем сделать вывод о том, что коалиционная структура π_1 d -устойчива относительно вектора Шепли.

Если следующие неравенства выполняются для любого $i = 1, 2$

$$\begin{aligned} \phi_i^{\pi_1} + \delta v_i/(1 - \delta) &\geq \phi_i^{\pi_1}/(1 - \delta), & v_i/(1 - \delta) &\geq v_i + \delta\phi_i^{\pi_1}/(1 - \delta), \\ \phi_i^{\pi_1} + \delta v_i/(1 - \delta) &\geq \phi_i^{\pi_1}/(1 - \delta), & v_i/(1 - \delta) &\geq v_i/(1 - \delta), \\ \phi_i^{\pi_1} + \delta v_i/(1 - \delta) &\geq (\phi_i^{\pi_1} + \delta v_i)/(1 - \delta^2), & v_i/(1 - \delta) &\geq (\delta\phi_i^{\pi_1} + v_i)/(1 - \delta^2), \end{aligned}$$

то ситуация η такая, что $\Lambda(\eta) = \Lambda_2$, является равновесием по Нэшу. Эти неравенства эквивалентны следующим: $\phi_i^{\pi_1} \leq v_i$, $i = 1, 2$. Так как $v_{12} < v_1 + v_2$,

то мы получаем $\phi_i < v_i$, $i = 1, 2$. Из определения d-устойчивой коалиционной структуры, мы можем сделать вывод о том, что коалиционная структура π_2 d-устойчива относительно вектора Шепли.

Третье утверждение теоремы очевидно. \square

§ 4.2.4 Примеры нахождения устойчивых в динамике коалиционных структур в играх трех лиц

Рассмотрим четыре численных примера с супераддитивной и несупераддитивной функциями.

Пример 2. Рассмотрим игру трех лиц с несупераддитивной характеристической функцией: $v(\{1, 2, 3\}) = v(\{1, 3\}) = 5$, $v(\{1, 2\}) = v(\{2, 3\}) = 6$, $v(\{1\}) = 2$, $v(\{2\}) = 3$, $v(\{3\}) = 1$. В игре возможно пять коалиционных структур $\Pi = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\}$. Коалиционные структуры и соответствующие им решения, вычисленные по формуле (4.2.1), представлены в табл. 4.11.

Дисконт-фактор δ равен 0.95. Стационарная стратегия игрока $i \in N$ в игре Γ — ситуация $\eta_i = (\xi_i^{\pi_1}, \xi_i^{\pi_2}, \xi_i^{\pi_3}, \xi_i^{\pi_4}, \xi_i^{\pi_5})$ где

$$\begin{aligned} \eta_1 &= ((\xi_1(\pi_1|\pi_1), \xi_1(\pi_2|\pi_1), \xi_1(\pi_3|\pi_1)), (\xi_1(\pi_1|\pi_2), \xi_1(\pi_2|\pi_2), \xi_1(\pi_3|\pi_2)), \\ &\quad (\xi_1(\pi_1|\pi_3), \xi_1(\pi_2|\pi_3), \xi_1(\pi_3|\pi_3)), (\xi_1(\pi_4|\pi_4), \xi_1(\pi_5|\pi_4)), (\xi_1(\pi_4|\pi_5), \xi_1(\pi_5|\pi_5))), \\ \eta_2 &= ((\xi_2(\pi_1|\pi_1), \xi_2(\pi_2|\pi_1), \xi_2(\pi_4|\pi_1)), (\xi_2(\pi_1|\pi_2), \xi_2(\pi_2|\pi_2), \xi_2(\pi_4|\pi_2)), \\ &\quad (\xi_2(\pi_3|\pi_3), \xi_2(\pi_5|\pi_3)), (\xi_2(\pi_1|\pi_4), \xi_2(\pi_2|\pi_4), \xi_2(\pi_4|\pi_4)), (\xi_2(\pi_3|\pi_5), \xi_2(\pi_5|\pi_5))), \\ \eta_3 &= ((\xi_3(\pi_1|\pi_1), \xi_3(\pi_3|\pi_1), \xi_3(\pi_4|\pi_1)), (\xi_3(\pi_2|\pi_2), \xi_3(\pi_5|\pi_2)), \\ &\quad (\xi_3(\pi_1|\pi_3), \xi_3(\pi_3|\pi_3), \xi_3(\pi_4|\pi_3)), (\xi_3(\pi_1|\pi_4), \xi_3(\pi_3|\pi_4), \xi_3(\pi_4|\pi_4)), \\ &\quad (\xi_3(\pi_2|\pi_5), \xi_3(\pi_5|\pi_5))). \end{aligned}$$

Каждый игрок имеет 108 чистых стратегий, следовательно, в игре существует 1 259 712 ситуаций в чистых стационарных стратегиях.

Элементы $\lambda(\pi_k|\pi_j, \xi^{\pi_j})$ матрицы вероятностей перехода $\Lambda(\eta)$ определены с помощью правила (4.2.2). Например, первая строка матрицы $\Lambda(\eta)$ состоит

из элементов:

$$\begin{aligned}\lambda(\pi_1|\pi_1, \xi^{\pi_1}) &= 1 - \lambda(\pi_2|\pi_1, \xi^{\pi_1}) - \lambda(\pi_3|\pi_1, \xi^{\pi_1}) - \lambda(\pi_4|\pi_1, \xi^{\pi_1}), \\ \lambda(\pi_2|\pi_1, \xi^{\pi_1}) &= [\xi_1(\pi_2|\pi_1)\xi_2(\pi_1|\pi_1) + \xi_2(\pi_2|\pi_1)\xi_1(\pi_1|\pi_1)] \cdot \xi_3(\pi_1|\pi_1), \\ \lambda(\pi_3|\pi_1, \xi^{\pi_1}) &= [\xi_1(\pi_3|\pi_1)\xi_3(\pi_1|\pi_1) + \xi_3(\pi_3|\pi_1)\xi_1(\pi_1|\pi_1)] \cdot \xi_2(\pi_1|\pi_1), \\ \lambda(\pi_4|\pi_1, \xi^{\pi_1}) &= [\xi_2(\pi_4|\pi_1)\xi_3(\pi_1|\pi_1) + \xi_3(\pi_4|\pi_1)\xi_2(\pi_1|\pi_1)] \cdot \xi_1(\pi_1|\pi_1), \\ \lambda(\pi_5|\pi_1, \xi^{\pi_1}) &= 0.\end{aligned}$$

В игре Γ существует 3 125 различных матриц вероятностей перехода $\Lambda(\eta)$. Вычисления показывают, что в игре есть 3 001 равновесие по Нэшу в чистых стационарных стратегиях, и все эти равновесия формируют 11 различных матриц $\Lambda(\eta^*)$. Число матриц, генерирующих d -устойчивые коалиционные структуры, равно 5 (результаты вычислений представлены в табл. 4.12). При анализе матриц $\Lambda(\eta^*)$, получаем, что коалиционные структуры π_2, π_3, π_4 d -устойчивы.

Таблица 4.11: Вектор Шепли для примера 2.

π	ϕ_1^π	ϕ_2^π	ϕ_3^π
$\pi_1 = \{\{1\}, \{2\}, \{3\}\}$	2	3	1
$\pi_2 = \{\{1, 2\}, \{3\}\}$	2.5	3.5	1
$\pi_3 = \{\{1, 3\}, \{2\}\}$	3	3	2
$\pi_4 = \{\{1\}, \{2, 3\}\}$	2	4	2
$\pi_5 = \{\{1, 2, 3\}\}$	1.5	2.5	1

В табл. 4.12 представлены результаты вычислений для примера 2, а, именно, d -устойчивые коалиционные структуры (первый столбец); количество соответствующих равновесий по Нэшу в чистых стационарных стратегиях, которые генерируют матрицы вероятностей перехода, и сами матрицы (второй столбец); ожидаемые выигрыши игроков в игре Γ и вектора ожидаемого количества шагов до того, как процесс игры попадет в d -устойчивую коалиционную структуру (последний столбец).

Пример 3. Теперь рассмотрим игру трех лиц с супераддитивной характеристической функцией: $v(\{1, 2, 3\}) = 10$, $v(\{1, 3\}) = 5$, $v(\{1, 2\}) = v(\{2, 3\}) =$

Таблица 4.12: Результаты вычислений для Примера 2.

π	Число равновесий	$\Lambda_1(\eta^*)$	$E_1(\eta^*)$	$E_2(\eta^*)$	$E_3(\eta^*)$	$\zeta(\eta^*)$
π_2	2	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 48.122 \\ 50.000 \\ 49.550 \\ 48.550 \\ 49.000 \end{pmatrix}$	$\begin{pmatrix} 69.072 \\ 70.000 \\ 68.550 \\ 69.550 \\ 69.000 \end{pmatrix}$	$\begin{pmatrix} 20.950 \\ 20.000 \\ 21.000 \\ 21.000 \\ 20.000 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 0 \\ 2 \\ 2 \\ 1 \end{pmatrix}$
		$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 49.072 \\ 50.000 \\ 49.550 \\ 48.550 \\ 49.000 \end{pmatrix}$	$\begin{pmatrix} 68.122 \\ 70.000 \\ 68.550 \\ 69.550 \\ 69.000 \end{pmatrix}$	$\begin{pmatrix} 20.950 \\ 20.000 \\ 21.000 \\ 21.000 \\ 20.000 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 0 \\ 2 \\ 2 \\ 1 \end{pmatrix}$
		$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 48.122 \\ 50.000 \\ 49.550 \\ 48.550 \\ 49.000 \end{pmatrix}$	$\begin{pmatrix} 60.000 \\ 60.500 \\ 60.000 \\ 60.976 \\ 59.975 \end{pmatrix}$	$\begin{pmatrix} 39.000 \\ 39.000 \\ 40.000 \\ 38.147 \\ 38.050 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 3 \\ 2 \end{pmatrix}$
		$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 56.268 \\ 59.500 \\ 60.000 \\ 57.124 \\ 58.025 \end{pmatrix}$	$\begin{pmatrix} 60.927 \\ 60.500 \\ 60.000 \\ 60.976 \\ 59.975 \end{pmatrix}$	$\begin{pmatrix} 37.240 \\ 39.000 \\ 40.000 \\ 38.147 \\ 38.050 \end{pmatrix}$	$\begin{pmatrix} 4 \\ 1 \\ 0 \\ 3 \\ 2 \end{pmatrix}$
		$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 40.000 \\ 40.500 \\ 40.976 \\ 40.000 \\ 39.975 \end{pmatrix}$	$\begin{pmatrix} 79.000 \\ 79.500 \\ 77.124 \\ 80.000 \\ 78.025 \end{pmatrix}$	$\begin{pmatrix} 39.000 \\ 39.000 \\ 38.147 \\ 40.000 \\ 38.050 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \\ 3 \\ 0 \\ 2 \end{pmatrix}$

6, $v(\{1\}) = 2$, $v(\{2\}) = 3$, $v(\{3\}) = 1$. Коалиционные структуры и соответствующие им решения, вычисленные по формуле (4.2.1), представлены в табл. 4.13.

Пусть дисконт-фактор δ равен 0.95. Ситуация η в стационарных стратегиях и матрица вероятностей перехода $\Lambda(\eta)$ определены в предыдущем примере. В игре Γ возможно 3 125 различных матриц вероятностей перехода $\Lambda(\eta)$. Вычисления показывают, что имеется 3 043 равновесия по Нэшу в чистых стационарных стратегиях и все эти равновесия приводят к

26 различным матрицам $\Lambda(\eta^*)$. Количество матриц, которые генерируют d -устойчивые коалиционные структуры равно 19 (см. табл. 4.14–4.16). Проанализировав матрицы $\Lambda(\eta^*)$, можно утверждать, что существует единственная d -устойчивая коалиционная структура π_5 .

Таблица 4.13: Вектор Шепли для примера 3.

π	ϕ_1^π	ϕ_2^π	ϕ_3^π
$\pi_1 = \{\{1\}, \{2\}, \{3\}\}$	2	3	1
$\pi_2 = \{\{1, 2\}, \{3\}\}$	2.5	3.5	1
$\pi_3 = \{\{1, 3\}, \{2\}\}$	3	3	2
$\pi_4 = \{\{1\}, \{2, 3\}\}$	2	4	2
$\pi_5 = \{\{1, 2, 3\}\}$	$3\frac{1}{6}$	$4\frac{1}{6}$	$2\frac{2}{3}$

В таблице 4.14 представлены результаты вычислений для примера 3, а именно: d -устойчивые коалиционные структуры (первый столбец); количество соответствующих равновесий по Нэшу в чистых стационарных стратегиях, которые генерируют матрицы вероятностей перехода и сами матрицы; ожидаемые выигрыши игроков в игре Γ и вектора, элементы которого равны ожидаемым количеством шагов до того, как процесс игры попадет в d -устойчивую коалиционную структуру π_5 (последний столбец).

Таблица 4.14: Результаты вычислений для примера 3.

Число равновесий	$\Lambda_1(\eta^*)$	$E_1(\eta^*)$	$E_2(\eta^*)$	$E_3(\eta^*)$	$\zeta(\eta^*)$			
1	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 61.558 \\ 61.480 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 82.508 \\ 81.383 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 50.482 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \\ 3 \\ 1 \\ 0 \end{pmatrix}$			
	2	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.457 \\ 62.667 \\ 61.457 \\ 61.533 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 81.407 \\ 82.667 \\ 81.407 \\ 82.533 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 49.529 \\ 51.667 \\ 50.529 \\ 51.083 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 1 \\ 3 \\ 2 \\ 0 \end{pmatrix}$		
		1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 62.008 \\ 60.363 \\ 63.167 \\ 60.908 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 81.058 \\ 80.455 \\ 82.167 \\ 81.005 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 48.958 \\ 52.667 \\ 50.482 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 3 \\ 1 \\ 3 \\ 0 \end{pmatrix}$	
			2	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.908 \\ 62.667 \\ 63.167 \\ 62.008 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.955 \\ 82.667 \\ 82.167 \\ 82.058 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 50.432 \\ 51.667 \\ 52.667 \\ 52.033 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 0 \end{pmatrix}$
				1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.955 \\ 62.667 \\ 62.058 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.908 \\ 82.667 \\ 82.008 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 50.432 \\ 51.667 \\ 52.033 \\ 52.667 \\ 53.333 \end{pmatrix}$

Таблица 4.15: Продолжение таблицы 4.14

Число равновесий	$\Lambda_1(\eta^*)$	$E_1(\eta^*)$	$E_2(\eta^*)$	$E_3(\eta^*)$	$\zeta(\eta^*)$			
1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.955 \\ 61.455 \\ 62.058 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.908 \\ 81.408 \\ 82.008 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 50.432 \\ 50.432 \\ 52.033 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 0 \end{pmatrix}$			
	1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 62.008 \\ 62.508 \\ 63.167 \\ 61.383 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 81.058 \\ 81.558 \\ 82.167 \\ 81.480 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 52.667 \\ 50.482 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \\ 1 \\ 3 \\ 0 \end{pmatrix}$		
		1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.407 \\ 62.667 \\ 62.533 \\ 61.407 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.457 \\ 82.667 \\ 81.533 \\ 81.457 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 49.529 \\ 51.667 \\ 51.083 \\ 50.529 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 1 \\ 2 \\ 3 \\ 0 \end{pmatrix}$	
			1	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 62.508 \\ 63.167 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 81.558 \\ 82.167 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 52.667 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \\ 1 \\ 1 \\ 0 \end{pmatrix}$
				2	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 60.455 \\ 61.005 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 80.363 \\ 80.908 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 48.958 \\ 50.482 \\ 52.667 \\ 53.333 \end{pmatrix}$
2					$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 62.008 \\ 61.558 \\ 63.167 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 81.058 \\ 82.508 \\ 82.167 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 52.667 \\ 52.667 \\ 53.333 \end{pmatrix}$
	1				$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 61.558 \\ 63.167 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 82.508 \\ 82.167 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 52.667 \\ 52.667 \\ 53.333 \end{pmatrix}$

Таблица 4.16: Продолжение таблицы 4.14

Число равновесий	$\Lambda_1(\eta^*)$	$E_1(\eta^*)$	$E_2(\eta^*)$	$E_3(\eta^*)$	$\zeta(\eta^*)$			
1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.406 \\ 61.558 \\ 61.480 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.314 \\ 82.508 \\ 81.383 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 48.958 \\ 51.033 \\ 50.482 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 4 \\ 2 \\ 3 \\ 1 \\ 0 \end{pmatrix}$			
	2	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.908 \\ 61.408 \\ 63.167 \\ 62.008 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.955 \\ 81.455 \\ 82.167 \\ 82.058 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 50.432 \\ 50.432 \\ 52.667 \\ 52.033 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 3 \\ 1 \\ 2 \\ 0 \end{pmatrix}$		
		1	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 62.667 \\ 62.058 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 82.667 \\ 82.008 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.667 \\ 52.033 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}$	
			1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 62.008 \\ 62.508 \\ 63.167 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 81.058 \\ 81.558 \\ 82.167 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 52.667 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \\ 1 \\ 1 \\ 0 \end{pmatrix}$
				1	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.336 \\ 62.667 \\ 62.533 \\ 61.407 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.384 \\ 82.667 \\ 81.533 \\ 81.457 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 49.003 \\ 51.667 \\ 51.083 \\ 50.529 \\ 53.333 \end{pmatrix}$
1					$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.314 \\ 62.508 \\ 63.167 \\ 61.383 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.406 \\ 81.558 \\ 82.167 \\ 81.480 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 48.958 \\ 51.033 \\ 52.667 \\ 50.482 \\ 53.333 \end{pmatrix}$
	1				$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 61.455 \\ 62.058 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 81.408 \\ 82.008 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 50.432 \\ 52.033 \\ 52.667 \\ 53.333 \end{pmatrix}$

§ 4.3 Стохастическая игра «Дилемма заключенного» с неполной информацией о дисконтирующих факторах

§ 4.3.1 Модель

Рассмотрим стохастическую бесконечно повторяющуюся игру «Дилемма заключенного» с двумя состояниями $\{s_1, s_2\}$ и двумя игроками $\{i, j\}$. Время предполагается дискретным и в любой момент времени $t = 1, 2, \dots$, может реализоваться одно из двух состояний. В каждом состоянии задается одновременная игра типа Дилемма заключенного. Множество стратегий любого игрока ν в каждом состоянии s есть $\mathbf{X} = \{C, D\}$. Стратегия C соответствует действию «кооперироваться», стратегия D — «отклониться». Функция выигрыша любого игрока $\nu \in \{i, j\}$ в состоянии s есть $u^s : \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$:

$$\begin{array}{cc}
 & \begin{array}{c} j \\ C \quad D \end{array} \\
 \begin{array}{c} i \\ C \\ D \end{array} & \begin{array}{|c|c|} \hline a^s; a^s & c^s; b^s \\ \hline b^s; c^s & d^s; d^s \\ \hline \end{array}
 \end{array} \tag{4.3.1}$$

Игроки получают одинаковые выигрыши, но они зависят от состояния. Сделаем стандартное предположение о выигрышах в игре Дилемма заключенного:

$$b^s > a^s > d^s > c^s,$$

and

$$2a^s > b^s + c^s$$

для любого $s \in \{s_1, s_2\}$. Второе неравенство говорит о том, что игроки, совместно максимизируя суммарный выигрыш, будут разыгрывать ситуацию (C, C) , а не (D, C) или (C, D) в любом состоянии. Это условие обычно не приводится, но мы будем предполагать, что игроки в кооперации будут разыгрывать (C, C) на каждом шаге [101]. Также предположим, что все игроки знают текущее состояние и историю этого шага.

Следующее предположение гарантирует, что в одношаговой игре индивидуальное отклонение от ситуации (C, C) более прибыльно в состоянии $s = s_2$, чем в состоянии $s = s_1$, это условие отражено в следующем предположении.

Предположение 4.1. Пусть $b^{s_2} - a^{s_2} \geq b^{s_1} - a^{s_1}$.

В качестве начального состояния может реализоваться любое из состояний из множества $\{s_1, s_2\}$. Переход из одного состояния в другое не зависит от шага и реализованной ситуации. Вероятности того, что игра перейдет из состояния s в состояние s_1 (s_2) равны π^s ($1 - \pi^s$) соответственно. Переход из одного состояния в другое может интерпретироваться как «шок».

§ 4.3.2 Случай с известными дисконтирующими факторами

Сначала рассмотрим игру, когда дисконтирующие факторы игроков являются публичной информацией, или они известны всем игрокам. Этот анализ будет использован ниже для определения классов игроков в соответствии с их «степенью терпимости». Более того, ситуации, являющиеся равновесиями в игре с полной информацией будут использованы для построения решения в двухшаговой игре как часть равновесия, реализуемая во второй фазе байесовской игры, когда дисконтирующие факторы игроков становятся известными противнику (см. § 4.3.3).

Ситуации в игре.

В этом параграфе рассмотрим некоторые ситуации в стратегиях поведения, которые опишем подробнее ниже, а также найдем условия, при которых эти ситуации будут являться ситуациями абсолютного равновесия по Нэшу. Будем рассматривать только ситуации в чистых стратегиях. Обозначим ситуацию в стратегиях поведения через $\sigma = (\sigma_\nu : \nu \in \{i, j\})$, где σ_ν — стратегия игрока ν , определяющая его поведение в любой момент времени и в любом состоянии в зависимости от истории этого момента.

Ограничимся рассмотрением трех ситуаций в стратегиях поведения: 1) игроки выбирают стратегию D в любом состоянии в любой момент времени («некооперативная ситуация» σ_n), 2) игроки выбирают стратегию C в состоянии s_1 и стратегию D в состоянии s_2 в любой момент времени и переключаются на разыгрывание стратегии D в любом состоянии, если в истории наблюдается отклонение игрока от описанного поведения («полукооперативная ситуация» σ_{sc}), 3) игроки выбирают стратегию C в любой момент времени и в любом состоянии

и переключаются на разыгрывание стратегии D , если наблюдается отклонение от описанного поведения в истории («кооперативная ситуация» σ_c).

Дадим строгие математические определения ситуаций, используемых в данной модели. Обозначим ситуацию в стратегиях поведения через

$$\sigma = (\sigma_\nu : \nu \in \{i, j\}). \quad (4.3.2)$$

где $\sigma_\nu = \{\sigma_{\nu,t}^s\}_{t=1}^\infty$ и $\sigma_{\nu,t+1}^s : H(t) \rightarrow \mathbf{X}$ — стратегия игрока ν в момент времени $t + 1$ и состоянии $s \in \{s_1, s_2\}$; $H(t) = ((s(1), x(1)), \dots, (s(t), x(t)))$ — история момента времени t , где $s(t)$ — состояние в момент времени t и $x(t)$ — ситуация, разыгрываемая в состоянии $s(t)$.

Определение 4.7. *Некооперативная стратегия игрока ν , обозначенная через $\sigma_{n,\nu} = \{\sigma_{\nu,t}^s\}_{t=1,\dots,\infty}^{s=s_1,s_2}$ такова, что $\sigma_{\nu,t+1}^s(H(t)) = D$ для $s = s_1, s_2, t = 1, \dots, \infty$ и любой истории $H(t)$.*

Назовем ситуацию $\sigma_n = (\sigma_{n,\nu} : \nu \in \{i, j\})$ некооперативной.

Определение 4.8. *Полукооперативная стратегия игрока ν , обозначенная через $\sigma_{sc,\nu} = \{\sigma_{\nu,t}^s\}_{t=1,\dots,\infty}^{s=s_1,s_2}$ такова, что*

$$\sigma_{\nu,t+1}^s(H(t)) = \begin{cases} C, & \text{если } s = s_1 \text{ и } H(t) = H_{sc}(t), \\ D, & \text{в противном случае,} \end{cases}$$

где $H_{sc}(t)$ — история момента времени t , содержащая только элементы $(s_1, (C, C))$ и $(s_2, (D, D))$.

Назовем ситуацию $\sigma_{sc} = (\sigma_{sc,\nu} : \nu \in \{i, j\})$ полукооперативной, по определению которой игроки выбирают стратегию C в состоянии s_1 и стратегию D в состоянии s_2 , если не было отклонения от истории $H_{sc}(t)$. В противном случае игроки переключаются на стратегию D в любом состоянии до бесконечности.

Определение 4.9. *Кооперативная стратегия игрока ν , обозначенная через $\sigma_{c,\nu} = \{\sigma_{\nu,t}^s\}_{t=1,\dots,\infty}^{s=s_1,s_2}$ такова, что*

$$\sigma_{\nu,t+1}^s(H(t)) = \begin{cases} C, & \text{если } H(t) = H_c(t), \\ D, & \text{в противном случае,} \end{cases}$$

где $H_{\mathbf{c}}(t) = ((s(1), (C, C)), \dots, (s(t), (C, C)))$ — история момента времени t , в соответствии с этой стратегией игроки выбирают C во все моменты времени до $t + 1$.

Назовем ситуацию $\sigma_{\mathbf{c}} = (\sigma_{\mathbf{c}, \nu} : \nu \in \{i, j\})$ кооперативной, она предписывает игрокам выбирать C в любой период $t + 1$, если история шага не содержит отклонений от кооперации в предыдущие моменты времени. Если игрок наблюдает отклонение от ситуации (C, C) в истории, то он начинает играть стратегию D до бесконечности.

Ожидаемые выигрыши.

Пусть δ_{ν} — дисконтирующий фактор выигрышей игрока ν . Найдем условия, при которых ситуации в чистых стратегиях, определенные выше, являются ситуациями абсолютных равновесий. Ситуация равновесия по Нэшу является ситуацией абсолютного равновесия, если для любого момента времени и любого состояния вектор усеченных стратегий образует равновесие по Нэшу в соответствующей подыгре.

Дисконтированный выигрыш игрока в ситуации σ есть

$$V(\sigma, \delta_{\nu}) = \sum_{t=1}^{\infty} \delta_{\nu}^{t-1} \Pi^{t-1} U_{\nu, t}^s(\sigma),$$

где $V(\sigma, \delta_{\nu}) = (V^{s_1}(\sigma, \delta_{\nu}), V^{s_2}(\sigma, \delta_{\nu}))'$, $U_{\nu, t}^s(\sigma) = (u_{\nu, t}^{s_1}(\sigma), u_{\nu, t}^{s_2}(\sigma))'$, операция $()'$ означает транспонирование, $u_{\nu, t}^s(\sigma)$ — выигрыш игрока ν в момент времени t и состоянии s при реализации ситуации σ и

$$\Pi = \begin{pmatrix} \pi^{s_1} & 1 - \pi^{s_1} \\ \pi^{s_2} & 1 - \pi^{s_2} \end{pmatrix}.$$

Обозначим дисконтированные выигрыши игрока ν в равновесиях $\sigma_{\mathbf{n}}$, $\sigma_{\mathbf{sc}}$ и $\sigma_{\mathbf{c}}$ через $V_{\mathbf{n}}^s(\delta_{\nu})$, $V_{\mathbf{sc}}^s(\delta_{\nu})$ и $V_{\mathbf{c}}^s(\delta_{\nu})$ соответственно, где индексы \mathbf{n} , \mathbf{sc} и \mathbf{c} означают некооперативное (“non-cooperative”), полукооперативное (“semi-cooperative”) и кооперативное (“cooperative”) равновесия, а индекс $s \in \{s_1, s_2\}$ указывает на состояние в игре в первый момент времени.

Лемма 4.1. *Для любого $\delta_{\nu} \in (0, 1)$ справедливо неравенство $V_{\mathbf{c}}(\delta_{\nu}) > V_{\mathbf{sc}}(\delta_{\nu}) > V_{\mathbf{n}}(\delta_{\nu})$.*

Доказательство. Вычислим ожидаемые выигрыши игроков. Для удобства введем обозначения:

$$\tilde{\Pi}(\delta_\nu) \equiv \frac{1}{(1 - \delta_\nu)(1 - \delta_\nu(\pi^{s_1} - \pi^{s_2}))} \begin{pmatrix} 1 - \delta_\nu(1 - \pi^{s_2}) & \delta_\nu(1 - \pi^{s_1}) \\ \delta_\nu\pi^{s_2} & 1 - \delta_\nu\pi^{s_1} \end{pmatrix}.$$

Нетрудно вычислить дисконтированный выигрыш игрока в равновесиях $\sigma_{\mathbf{n}}$, $\sigma_{\mathbf{c}}$ и $\sigma_{\mathbf{sc}}$:

1. Дисконтированный выигрыш игрока ν в равновесии $\sigma_{\mathbf{n}}$ равен

$$V_{\mathbf{n}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{n}}^{s_1}(\delta_\nu) \\ V_{\mathbf{n}}^{s_2}(\delta_\nu) \end{pmatrix} = \tilde{\Pi}(\delta_\nu) \begin{pmatrix} d^{s_1} \\ d^{s_2} \end{pmatrix}. \quad (4.3.3)$$

2. Дисконтированный выигрыш игрока ν в равновесии $\sigma_{\mathbf{c}}$ равен

$$V_{\mathbf{c}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{c}}^{s_1}(\delta_\nu) \\ V_{\mathbf{c}}^{s_2}(\delta_\nu) \end{pmatrix} = \tilde{\Pi}(\delta_\nu) \begin{pmatrix} a^{s_1} \\ a^{s_2} \end{pmatrix}. \quad (4.3.4)$$

3. Дисконтированный выигрыш игрока ν в равновесии $\sigma_{\mathbf{sc}}$ есть

$$V_{\mathbf{sc}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{sc}}^{s_1}(\delta_\nu) \\ V_{\mathbf{sc}}^{s_2}(\delta_\nu) \end{pmatrix} = \tilde{\Pi}(\delta_\nu) \begin{pmatrix} a^{s_1} \\ d^{s_2} \end{pmatrix}. \quad (4.3.5)$$

Эти формулы получены с учетом определений ситуаций.

Дисконтированный выигрыш игрока ν в равновесии $\sigma_{\mathbf{n}}$ есть

$$V_{\mathbf{n}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{n}}^{s_1}(\delta_\nu) \\ V_{\mathbf{n}}^{s_2}(\delta_\nu) \end{pmatrix} = \begin{pmatrix} d_{s_1} + \delta_\nu \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_\nu) \\ d_{s_2} + \delta_\nu \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_\nu) \end{pmatrix},$$

или в векторной форме:

$$V_{\mathbf{n}}(\delta_\nu) = \begin{pmatrix} d_{s_1} \\ d_{s_2} \end{pmatrix} + \delta_\nu \Pi V_{\mathbf{n}}(\delta_\nu).$$

Из этого уравнения следует, что

$$V_{\mathbf{n}}(\delta_\nu) = (\mathbb{I} - \delta_\nu \Pi)^{-1} \begin{pmatrix} d_{s_1} \\ d_{s_2} \end{pmatrix},$$

где \mathbb{I} — единичная матрица размерности 2×2 . Обозначим $(\mathbb{I} - \delta_\nu \Pi)^{-1}$ через $\tilde{\Pi}(\delta_\nu)$ и получаем требуемое. Обратная матрица $(\mathbb{I} - \delta_\nu \Pi)^{-1}$ существует для любого $\delta_\nu \in (0, 1)$.

Теперь вычислим дисконтированный выигрыш игрока ν в равновесии $\sigma_{\mathbf{c}}$, получим

$$V_{\mathbf{c}}(\delta_{\nu}) = \begin{pmatrix} V_{\mathbf{c}}^{s_1}(\delta_{\nu}) \\ V_{\mathbf{c}}^{s_2}(\delta_{\nu}) \end{pmatrix} = \begin{pmatrix} a^{s_1} \\ a^{s_2} \end{pmatrix} + \delta_{\nu} \begin{pmatrix} \pi^{s_1} & 1 - \pi^{s_1} \\ \pi^{s_2} & 1 - \pi^{s_2} \end{pmatrix} \begin{pmatrix} V_{\mathbf{c}}^{s_1}(\delta_{\nu}) \\ V_{\mathbf{c}}^{s_2}(\delta_{\nu}) \end{pmatrix}.$$

Переписав это уравнение в векторной форме, получим выражение (4.3.4).

Теперь вычислим дисконтированный выигрыш игрока ν в равновесии $\sigma_{\mathbf{sc}}$:

$$V_{\mathbf{sc}}(\delta_{\nu}) = \begin{pmatrix} V_{\mathbf{sc}}^{s_1}(\delta_{\nu}) \\ V_{\mathbf{sc}}^{s_2}(\delta_{\nu}) \end{pmatrix} = \begin{pmatrix} a^{s_1} \\ d^{s_2} \end{pmatrix} + \delta_{\nu} \begin{pmatrix} \pi^{s_1} & 1 - \pi^{s_1} \\ \pi^{s_2} & 1 - \pi^{s_2} \end{pmatrix} \begin{pmatrix} V_{\mathbf{sc}}^{s_1}(\delta_{\nu}) \\ V_{\mathbf{sc}}^{s_2}(\delta_{\nu}) \end{pmatrix}.$$

Переписав это уравнение в векторной форме, получим выражение (4.3.5).

Заметим, что в выражениях (4.3.3), (4.3.4) и (4.3.5) матрица $\tilde{\Pi}(\delta_{\nu})$ одинакова.

Результат леммы следует из выражений ожидаемых выигрышей игроков $V_{\mathbf{n}}(\delta_{\nu})$, $V_{\mathbf{c}}(\delta_{\nu})$ и $V_{\mathbf{sc}}(\delta_{\nu})$, полученных выше, и того, что $a^{s_2} > d^{s_2}$ и $a^{s_1} > d^{s_1}$. \square

Теперь определим допустимые значения δ , для которых каждая из вышеописанных ситуаций является ситуацией абсолютного равновесия по Нэшу. Введем обозначения:

$$\delta_1^* = \frac{\Delta_1 - \pi^{s_1}(b^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_1} - a^{s_1})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1})}, \quad (4.3.6)$$

$$\delta_2^* = \frac{\Delta_2 - \pi^{s_2}(a^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) - \pi^{s_1}(b^{s_2} - a^{s_2})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2})} \quad (4.3.7)$$

где $\delta_1^* \in (0, 1)$, $\delta_2^* \in (0, 1)$, $\mathbf{p}^s = (\pi^s, 1 - \pi^s)$ для любого $s \in \{s_1, s_2\}$ и

$$\begin{aligned} \Delta_1 &= ((\pi^{s_1}(b^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_1} - a^{s_1}))^2 \\ &\quad + 4(b^{s_1} - a^{s_1})(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1}))^{1/2}, \\ \Delta_2 &= ((\pi^{s_2}(a^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) + \pi^{s_1}(b^{s_2} - a^{s_2}))^2 \\ &\quad + 4(b^{s_2} - a^{s_2})(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2}))^{1/2}. \end{aligned}$$

Следующая теорема дает условия, при которых выше упомянутые ситуации являются ситуациями абсолютного равновесия по Нэшу (subgame perfect Nash equilibrium или SPNE).

Теорема 4.7. Пусть выполнено предположение 4.1 и игрок $\nu \in \{i, j\}$ имеет дисконтирующий фактор δ_{ν} , что является известным другим игрокам. Кооперативная ситуация является ситуацией абсолютного равновесия по Нэшу

тогда и только тогда, когда $\delta_\nu \geq \delta_2^*$ для любого $\nu \in \{i, j\}$. Полукооперативная ситуация является ситуацией абсолютного равновесия по Нэшу тогда и только тогда, когда $\delta_\nu \geq \delta_1^*$ для любого $\nu \in \{i, j\}$. Некооперативная ситуация является ситуацией абсолютного равновесия по Нэшу для любых $\delta_\nu \in (0, 1)$, $\nu \in \{i, j\}$.

Доказательство. Докажем, что каждая из ситуаций является ситуацией абсолютного равновесия по Нэшу при выполнении тех или иных условий, пользуясь принципом невыгодности отклонения в одношаговой игре. Сначала рассмотрим некооперативную ситуацию. Очевидно, что ситуация $\sigma_{\mathbf{n}}$ является ситуацией абсолютного равновесия по Нэшу для любого δ , потому что эта ситуация предписывает игрокам выбирать стратегии, равновесные по Нэшу на каждом шаге игры в любом состоянии и отклонение в одношаговой игре не выгодно ни одному из игроков.

Далее рассмотрим полукооперативную ситуацию. Выигрыш игрока в этой ситуации можно найти в утверждении леммы 4.1. Требуется доказать, что если

$$\delta_\nu \geq \frac{\Delta_1 - \pi^{s_1}(b^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_1} - a^{s_1})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1})},$$

где

$$\begin{aligned} \Delta_1 = & ((\pi^{s_1}(b^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_1} - a^{s_1}))^2 \\ & + 4(b^{s_1} - a^{s_1})(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1}))^{0.5}, \end{aligned}$$

то $\sigma_{\mathbf{sc}}$ — абсолютное равновесие по Нэшу.

Если наблюдается отклонение, то со следующего шага отклонившийся игрок будет наказан другими игроками получением выигрыша при равновесии по Нэшу (то есть, d^{s_1} в состоянии s_1 и d^{s_2} в состоянии s_2). Естественно, что отклонение в состоянии s_2 не принесет игроку прибыль. Ситуация $\sigma_{\mathbf{sc}}$ — абсолютное равновесие по Нэшу, если у игрока нет выгоды от отклонения в состоянии s_1 :

$$a^{s_1} + \delta_\nu \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_\nu) \geq b^{s_1} + \delta_\nu \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_\nu).$$

Подставим выражения для $V_{\mathbf{sc}}(\delta_\nu)$ и $V_{\mathbf{n}}(\delta_\nu)$ из леммы 4.1, перегруппируем и

получим условие:

$$f_1(\delta_\nu) = \delta_\nu^2(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1}) + \delta_\nu \{ \pi^{s_1}(b^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_1} - a^{s_1}) \} - (b^{s_1} - a^{s_1}) \geq 0.$$

Функция $f_1(\delta_\nu)$ — квадратичная и удовлетворяет условиям:

1. $f_1(0) \leq 0$,
2. $\lim_{\delta \rightarrow 1} f_1(\delta_\nu) > 0$.

Следовательно, существует единственное решение $\delta_\nu = \delta_1^*$ уравнения $f_1(\delta_\nu) = 0$ для $\delta_\nu \in (0, 1)$, где δ_1^* определена равенством (4.3.6).

Решение неравенства $f_1(\delta_\nu) \geq 0$ есть $\delta_\nu \in [\delta_1^*, 1)$. Следовательно, для любых $\delta_\nu \in [\delta_1^*, 1)$ ситуация σ_{sc} — абсолютное равновесие по Нэшу.

Теперь рассмотрим кооперативную ситуацию. Выигрыш игрока в этой ситуации получен в лемме 4.1. Требуется доказать, что для любого дисконтирующего фактора

$$\delta_\nu \geq \frac{\Delta_2 - \pi^{s_2}(a^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) - \pi^{s_1}(b^{s_2} - a^{s_2})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2})},$$

где

$$\Delta_2 = ((\pi^{s_2}(a^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) + \pi^{s_1}(b^{s_2} - a^{s_2}))^2 + 4(b^{s_2} - a^{s_2})(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2}))^{0.5},$$

ситуация σ_c — абсолютное равновесие по Нэшу. Если наблюдается отклонение, то со следующего за отклонением шага отклонившийся игрок наказывается получением выигрыша в равновесии по Нэшу во всех последующих реализовавшихся состояниях (d^{s_1} в состоянии s_1 и d^{s_2} в состоянии s_2). Ситуация σ_c — абсолютное равновесие по Нэшу, если для любого игрока $\nu \in \{i, j\}$ не существует выгодного отклонения ни в каком состоянии. Игрок ν не отклоняется в состоянии s_1 , если

$$a^{s_1} + \delta \mathbf{p}^{s_1} V_c(\delta_\nu) \geq b^{s_1} + \delta_\nu \mathbf{p}^{s_1} V_n(\delta_\nu),$$

и в состоянии s_2 , если

$$a^{s_2} + \delta_\nu \mathbf{p}^{s_2} V_c(\delta_\nu) \geq b^{s_2} + \delta_\nu \mathbf{p}^{s_2} V_n(\delta_\nu).$$

Подставим выражения $V_{\mathbf{c}}(\delta_\nu)$ и $V_{\mathbf{n}}(\delta_\nu)$ из леммы 4.1 и перегруппируем слагаемые, получим:

$$f_3(\delta_\nu) = \delta_\nu^2(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1}) + \delta_\nu \{ \pi^{s_1}(b^{s_1} - d^{s_1}) + (1 - \pi^{s_1})(a^{s_2} - d^{s_2}) + (1 - \pi^{s_2})(b^{s_1} - a^{s_1}) \} - (b^{s_1} - a^{s_1}) \geq 0,$$

$$f_2(\delta_\nu) = \delta_\nu^2(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2}) + \delta_\nu \{ \pi^{s_2}(a^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) + \pi^{s_1}(b^{s_2} - a^{s_2}) \} - (b^{s_2} - a^{s_2}) \geq 0.$$

Здесь $f_2(\delta_\nu)$ и $f_3(\delta_\nu)$ — квадратичные функции, удовлетворяющие следующим условиям, используя 4.1:

1. $0 \geq f_3(0) \geq f_2(0)$,
2. $\lim_{\delta \rightarrow 1} f_3(\delta_\nu) = \lim_{\delta \rightarrow 1} f_2(\delta_\nu) > 0$.

Следовательно, существует единственное решение $\delta_\nu = \delta_3^*$ уравнения $f_3(\delta_\nu) = 0$ для $\delta_\nu \in (0, 1)$ и единственное решение $\delta_\nu = \delta_2^*$ уравнения $f_2(\delta_\nu) = 0$ для $\delta_\nu \in (0, 1)$, где δ_2^* из формулы (4.3.7). Учитывая, что обе функции $f_2(\delta_\nu)$ и $f_3(\delta_\nu)$ неубывающие, нетрудно доказать, что решение системы неравенств

$$\begin{cases} f_3(\delta_\nu) \geq 0, \\ f_2(\delta_\nu) \geq 0, \end{cases}$$

есть $\delta_\nu \in [\delta_2^*, 1)$. Следовательно, для любого $\delta_\nu \in [\delta_2^*, 1)$ кооперативная ситуация $\sigma_{\mathbf{c}}$ — абсолютное равновесие по Нэшу. \square

В зависимости от параметров модели, получается или $\delta_1^* < \delta_2^*$, или $\delta_1^* \geq \delta_2^*$. Если $\delta_1^* < \delta_2^*$, выделим три различных комбинации дисконтирующих факторов игроков i и j :

- $\delta_i \in (0, \delta_1^*)$, $\delta_j \in (0, 1)$: только ситуация $\sigma_{\mathbf{n}}$ (из рассмотренных выше) — абсолютное равновесие по Нэшу;
- $\delta_i \in [\delta_1^*, \delta_2^*)$, $\delta_j \in [\delta_1^*, 1)$: и $\sigma_{\mathbf{n}}$, и $\sigma_{\mathbf{sc}}$ — абсолютные равновесия по Нэшу, но $V_{\mathbf{sc}}(\delta_\nu) > V_{\mathbf{n}}(\delta_\nu)$ по лемме 4.1;

- $\delta_i \in [\delta_1^*, 1)$, $\delta_j \in [\delta_1^*, 1)$: ситуации $\sigma_{\mathbf{n}}$, σ_{sc} , $\sigma_{\mathbf{c}}$ — абсолютные равновесия по Нэшу. Опять же по лемме 4.1, выигрыши игроков при кооперативной ситуации наибольшие.

Если $\delta_1^* \geq \delta_2^*$, интервал $(\delta_1^*, 1]$, для которого полукооперативная ситуация — абсолютное равновесие по Нэшу, входит в интервал $(\delta_2^*, 1]$, где также кооперативная ситуация — абсолютное равновесие по Нэшу. Тогда по лемме 4.1, полукооперативная ситуация не выгодна для разыгрывания, поскольку выигрыши игроков при σ_{sc} не меньше, чем при $\sigma_{\mathbf{c}}$. В этом случае анализ модели схож с анализом, проведенным в [101], где в интервале $(0, \delta_2^*)$, только $\sigma_{\mathbf{n}}$ — абсолютное равновесие по Нэшу (единственная ситуация из множества $\{\sigma_{\mathbf{n}}, \sigma_{\text{sc}}, \sigma_{\mathbf{c}}\}$), тогда как в интервале $(\delta_2^*, 1)$ абсолютное равновесие с наибольшими выигрышами $\sigma_{\mathbf{c}}$.

⁷ Далее будем рассматривать первый случай и предположим:

Предположение 4.2. Пусть $\delta_2^* > \delta_1^*$.

Классы игроков.

Следуя теореме 4.7 и предположению 4.2, определим классы игроков в соответствии с их дисконтирующими факторами.

Определение 4.10. Игрок ν принадлежит классу ℓ_ν , где $\ell_\nu \in \{I, M, P\}$:

1. I («нетерпеливый» или “*impatient*”) с дисконтирующим фактором $\delta_\nu = \delta_I$, удовлетворяющем условию $\delta_I \in (0, \delta_1^*)$;
2. M («умеренный» или “*mildly patient*”) с дисконтирующим фактором $\delta_\nu = \delta_M$, удовлетворяющем условию $\delta_M \in [\delta_1^*, \delta_2^*)$;
3. P («терпеливый» или “*patient*”) с дисконтирующим фактором $\delta_\nu = \delta_P$, удовлетворяющем условию $\delta_P \in [\delta_2^*, 1)$.

⁷Рассмотрим три комбинации дисконтирующих факторов игроков i и j :

- $\delta_i \in (0, \delta_2^*)$, $\delta_j \in (0, 1)$: только $\sigma_{\mathbf{n}}$ — абсолютное равновесие по Нэшу.
- $\delta_i \in [\delta_2^*, \delta_1^*)$, $\delta_j \in [\delta_2^*, 1)$: и $\sigma_{\mathbf{n}}$, и $\sigma_{\mathbf{c}}$ — абсолютные равновесия по Нэшу. Но при кооперативном равновесии выигрыши игроков больше, чем при полукооперативном.
- $\delta_i \in [\delta_1^*, 1)$, $\delta_j \in [\delta_1^*, 1)$: ситуации $\sigma_{\mathbf{n}}$, σ_{sc} , $\sigma_{\mathbf{c}}$ — абсолютные равновесия по Нэшу. Равновесие с наибольшими выигрышами — $\sigma_{\mathbf{c}}$.

Кооперативная ситуация — абсолютное равновесие по Нэшу тогда и только тогда, когда оба игрока класса P . В противном случае, игрок, не принадлежащий классу P , отклонится от кооперативной ситуации, потому как отклонение выгодно для него. Полукооперативная ситуация — абсолютное равновесие по Нэшу, когда 1) оба игрока класса P , 2) оба игрока класса M или 3) один игрок класса P , а другой — класса M . Если по крайней мере один из игроков класса I , то ни кооперативная, ни полукооперативная ситуация не являются абсолютным равновесием по Нэшу.

§ 4.3.3 Двухфазная игра с неизвестными дисконтирующими факторами игроков

Теперь мы предположим, что дисконтирующий фактор игрока являются неизвестными для другого игрока. Возможно, в игре существует несколько классов равновесий, но мы остановимся на анализе некоторых равновесий в игре, состоящей из двух фаз [101].

Первая фаза игры называется *обучающей*, при ней дисконтирующие факторы игроков неизвестны для оппонентов, и игроки пытаются узнать класс игрока-оппонента. На этой фазе игры стратегии игроков являются марковскими, а именно, она зависит от ожиданий игрока о классе игрока-оппонента и не зависит от истории игры. Вторая фаза игры — *кооперативная*, в которой игроки реализуют стратегии поведения из множества $\{\sigma_{c,\nu}, \sigma_{sc,\nu}, \sigma_{n,\nu}\}$. Стратегия игрока во второй фазе определяется по информации о классе игрока-оппонента в последний момент времени обучающей фазы игры.

Обучающая фаза: стратегии и правило обновления информации.

В этом разделе опишем, по каким правилам происходит процесс изменения информации о классе другого игрока. В момент времени t , игрок имеет информацию о том, что другой игрок класса P с вероятностью α_t , класса M с вероятностью β_t и класса I с вероятностью $\gamma_t = 1 - \alpha_t - \beta_t$. Определим стратегии для каждого $t = 1, \dots, T$: игрок выбирает стратегию в момент времени t на основе вероятностей (ожиданий) α_t и β_t . Предположим, что начальные ожидания о классе другого игрока в момент времени 1 даны и известны игрокам,

т. е. $\alpha_1 \in (0, 1)$, $\beta_1 \in (0, 1)$ известны и $\alpha_1 + \beta_1 \in (0, 1)$.

Обозначим стратегию игрока ν в обучающей фазе игры через ψ_ν и множество стратегий игрока ν в обучающей фазе через $\Psi(\ell_\nu)$, это множество зависит от класса игрока $\ell_\nu \in \{I, M, P\}$. Множество марковских стратегий игрока класса P есть $\Psi(P) = \{q_t^s, t = 1, \dots, T, s = s_1, s_2\}$, где $q_t^s : [0, 1] \times [0, 1] \rightarrow [0, 1]$ функция α_t и β_t , которая определяет вероятность выбора стратегии C в состоянии s в момент времени t . Множество марковских стратегий игрока класса M есть $\Psi(M) = \{r_t^s, t = 1, \dots, T, s = s_1, s_2\}$. Заметим, что $r_t^{s_1} : [0, 1] \times [0, 1] \rightarrow [0, 1]$ — функция α_t и β_t , которая определяет вероятность выбора стратегии C в состоянии s_1 в момент времени t . В состоянии s_2 , $r_t^{s_2} = 0$ для любого t , т. е. стратегия $r_t^{s_2}$ предписывает ив любой момент времени. Определим множество стратегий игрока ν класса $\ell_\nu = I$. Множество марковских стратегий игрока класса I в обучающей фазе игры есть $\Psi(I) = \{z_t^s, t = 1, \dots, T, s = s_1, s_2\}$, где $z_t^s : [0, 1] \times [0, 1] \rightarrow 0$ для любого момента времени t и любого состояния s . Стратегия z_t^s предписывает игроку класса I выбирать стратегию D в любом состоянии $s = s_1, s_2$ с вероятностью 1 в любой момент времени.

Стратегия игрока ν в обучающей фазе игры есть $\psi(\ell_\nu)$ и она определяет вероятность выбора стратегии C в любой момент времени t и любом состоянии s . Она зависит от класса игрока ℓ_ν и равна:

$$\psi(\ell_\nu) = \begin{cases} q_t^s \in [0, 1], & \text{если } \ell_\nu = P, \quad s = s_1, s_2; \\ r_t^s \in [0, 1], & \text{если } \ell_\nu = M, \quad s = s_1; \\ r_t^s = 0, & \text{если } \ell_\nu = M, \quad s = s_2; \\ z_t^s = 0, & \text{если } \ell_\nu = I, \quad s = s_1, s_2. \end{cases}$$

По определению стратегий игроков в обучающей фазе можно утверждать, что если игрок выбирает стратегию C в состоянии s_2 , его класс раскрывается другому игроку как класс P , поскольку только игроки этого класса выбирают стратегию C в состоянии s_2 с положительной вероятностью. Также если игрок выбирает стратегию C в состоянии s_1 , то он идентифицирует себя как игрок класса P или M , потому как только игроки этих двух классов выбирают стратегию C в состоянии s_1 с положительной вероятностью.

Используем байесовское правило для пересчета вероятностей α_t и β_t во времени, $t = 2, \dots, T + 1$. Правило определения вероятностей для момента $t + 1$ зависит от состояния s , реализовавшегося с период t . Если состояние s реализуется в момент времени t , то в следующий момент времени $t + 1$, вероятность того, что оппонент — игрок класса P есть

$$\alpha_{t+1} = \begin{cases} \alpha_{t+1}^{s_1}, & \text{если } s = s_1, \\ \alpha_{t+1}^{s_2}, & \text{если } s = s_2. \end{cases}$$

Такое же правило применяется к β_{t+1} и γ_{t+1} .

Сначала рассмотрим правило изменения ожиданий (вероятностей) для состояния $s = s_1$. Если игрок выбирает стратегию C в момент t , он идентифицируется как игрок класса⁸

$$\begin{cases} I & \text{с вер. } \gamma_{t+1}^{s_1} = 0; \\ P & \text{с вер. } \alpha_{t+1}^{s_1} = \frac{\alpha_t q_t^{s_1}}{\alpha_t q_t^{s_1} + \beta_t r_t^{s_1}}; \\ M & \text{с вер. } \beta_{t+1}^{s_1} = 1 - \alpha_{t+1}^{s_1} = \frac{\beta_t r_t^{s_1}}{\alpha_t q_t^{s_1} + \beta_t r_t^{s_1}}. \end{cases} \quad (4.3.8)$$

Если игрок выбирает стратегию D в состоянии $s = s_1$, он идентифицируется как игрок класса⁹

$$\begin{cases} I & \text{с вер. } \gamma_{t+1}^{s_1} = 1 - \alpha_{t+1}^{s_1} - \beta_{t+1}^{s_1} = \frac{1 - \alpha_t - \beta_t}{1 - \alpha_t q_t^{s_1} - \beta_t r_t^{s_1}}; \\ P & \text{с вер. } \alpha_{t+1}^{s_1} = \frac{\alpha_t(1 - q_t^{s_1})}{1 - \alpha_t q_t^{s_1} - \beta_t r_t^{s_1}}; \\ M & \text{с вер. } \beta_{t+1}^{s_1} = 1 - \alpha_{t+1}^{s_1} = \frac{\beta_t(1 - r_t^{s_1})}{1 - \alpha_t q_t^{s_1} - \beta_t r_t^{s_1}}. \end{cases} \quad (4.3.9)$$

Далее рассмотрим правило изменения ожиданий для состояния s_2 . Как уже говорилось, если игрок выбирает стратегию C , то он идентифицируется как игрок класса P с вероятностью 1. Если же игрок выбирает стратегию D , он

⁸Вероятности $\alpha_{t+1}^{s_1}$ и $\beta_{t+1}^{s_1}$ определены, если $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} \neq 0$. Если $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} = 0$, невозможно наблюдать стратегию C в состоянии $s = s_1$.

⁹Вероятности $\alpha_{t+1}^{s_1}$, $\beta_{t+1}^{s_1}$ и $\gamma_{t+1}^{s_1}$ определены, если $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} \neq 1$. Если $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} = 1$, невозможно наблюдать стратегию D в состоянии $s = s_1$.

идентифицируется как игрок класса¹⁰

$$\left\{ \begin{array}{l} I \quad \text{с вер. } \gamma_{t+1}^{s_2} = 1 - \alpha_{t+1}^{s_2} - \beta_{t+1}^{s_2} = \frac{1 - \alpha_t - \beta_t}{1 - \alpha_t q_t^{s_2}}; \\ P \quad \text{с вер. } \alpha_{t+1}^{s_2} = \frac{\alpha_t(1 - q_t^{s_2})}{1 - \alpha_t q_t^{s_2}}; \\ M \quad \text{с вер. } \beta_{t+1}^{s_2} = \frac{\beta_t}{1 - \alpha_t q_t^{s_2}}. \end{array} \right. \quad (4.3.10)$$

Кооперативная фаза.

Теперь определим стратегии игроков в кооперативной фазе игры. Стратегия игрока ν в кооперативной фазе — это функция, зависящая от класса игрока и его ожиданий о классе оппонента, имеющих в конце обучающей фазы игры, с областью значений во множестве $\{\sigma_{\mathbf{c},\nu}, \sigma_{\mathbf{sc},\nu}, \sigma_{\mathbf{n},\nu}\}$. Пусть T — последний период обучающей фазы, который предполагается конечным.¹¹

В конце обучающей фазы, ожидания о том, что оппонент — игрок класса P или M есть α_{T+1} и β_{T+1} соответственно. Тогда стратегия игрока ν — функция класса ℓ_ν и ожиданий $\alpha_{T+1}, \beta_{T+1}$ такая, что

$$\sigma_\nu(\ell_\nu, \alpha_{T+1}, \beta_{T+1}) = \left\{ \begin{array}{l} \sigma_{\mathbf{c},\nu}, \quad \text{если } \ell_\nu = P, \quad \alpha_{T+1} = 1, \\ \sigma_{\mathbf{sc},\nu}, \quad \text{если } \{\ell_\nu = P, \quad \alpha_{T+1} + \beta_{T+1} = 1, \quad \alpha_{T+1} \neq 1\} \\ \quad \text{or } \{\ell_\nu = M, \quad \alpha_{T+1} + \beta_{T+1} = 1\}, \\ \sigma_{\mathbf{n},\nu}, \quad \text{если } \{\ell_\nu \in \{P, M\}, \quad \alpha_{T+1} + \beta_{T+1} < 1\} \\ \quad \text{or } \{\ell_\nu = I\}. \end{array} \right. \quad (4.3.11)$$

Из правила (4.3.11) следует, что нетерпеливый игрок всегда выбирает стратегию $\sigma_{\mathbf{n},\nu}$, умеренно терпеливый игрок выбирает полукооперативную стратегию $\sigma_{\mathbf{sc},\nu}$, если в соответствии с его ожиданиями оппонент не является нетерпеливым игроком. Он выбирает некооперативную стратегию $\sigma_{\mathbf{n},\nu}$ в противном случае. Терпеливый игрок выбирает кооперативную стратегию $\sigma_{\mathbf{c},\nu}$ только в случае, когда согласно его ожиданиям оппонент — терпеливый игрок с вероятностью единица. Он выбирает полукооперативную стратегию $\sigma_{\mathbf{sc},\nu}$, если согласно его

¹⁰Вероятности $\alpha_{t+1}^{s_2}, \beta_{t+1}^{s_2}$ и $\gamma_{t+1}^{s_2}$ определены, если $\alpha_t q_t^{s_2} \neq 1$. Если $\alpha_t q_t^{s_2} = 1$ ($\alpha_t = q_t^{s_2} = 1$), невозможно наблюдать стратегию D в состоянии $s = s_2$.

¹¹Здесь мы не рассматриваем случай бесконечной обучающей фазы.

ожиданиям оппонент не является нетерпеливым игроком, но его ожидание, что оппонент — игрок класса P не равны единице. В противном случае, терпеливый игрок играет некооперативную стратегию $\sigma_{\mathbf{n},\nu}$.

Выигрыши и равновесие.

Выигрыш игрока $\nu = i, j$ чей класс есть $\ell_\nu \in \{I, M, P\}$ равен сумме выигрышей в двух фазах игры, и это функция класса игрока, начальных ожиданий и стратегий игроков i и j в обучающей фазе:¹²

$$\begin{aligned} \Phi_\nu(\psi_i, \psi_j, \alpha_1, \beta_1 | \ell_\nu) &= \sum_{t=1}^T \delta_\nu^{t-1} \Pi^{t-1} U_\nu(\psi_{i,t}, \psi_{j,t}) \\ &+ \delta_\nu^T \Pi^T V(\delta_\nu, (\sigma_i(\ell_i, \alpha_{T+1}, \beta_{T+1}), \sigma_j(\ell_j, \alpha_{T+1}, \beta_{T+1}))), \end{aligned} \quad (4.3.12)$$

где $\psi_\nu \in \Psi(\ell_\nu)$. Первое слагаемое в правой части (4.3.12) — выигрыш в обучающей фазе, второе слагаемое — выигрыш игрока в кооперативной фазе. В формуле (4.3.12), $\sigma_i(\ell_i, \alpha_{T+1}, \beta_{T+1})$ и $\sigma_j(\ell_j, \alpha_{T+1}, \beta_{T+1})$ определены правилом (4.3.11).

Определим теперь концепцию равновесия, используемую для решения игры. Под равновесием здесь будем понимать абсолютное байесовское равновесие [94]. В абсолютном байесовском равновесии для каждого момента времени $t = 1, \dots, T$ и любой истории этого момента, усеченные стратегии игроков образуют байесовское равновесие в соответствующей подыгре. Дадим определение этого равновесия.

Определение 4.11. *Набор стратегий или ситуация $\psi^* = (\psi_i^*, \psi_j^*)$ называется абсолютным байесовским равновесием, если для любого игрока $\nu \in \{i, j\}$, любого класса $\ell_\nu \in \{M, P\}$, любых начальных ожиданий $\alpha_1 \in (0, 1)$, $\beta_1 \in (0, 1)$, $\alpha_1 + \beta_1 \in (0, 1)$ и любой стратегии $\psi_\nu \in \Psi(\ell_\nu)$, имеет место неравенство*

$$\Phi_\nu(\psi^*, \alpha_1, \beta_1 | \ell_\nu) \geq \Phi_\nu((\psi_\nu, \psi_{-\nu}^*), \alpha_1, \beta_1 | \ell_\nu). \quad (4.3.13)$$

Множество с стратегий игрока в двухфазовой игре состоит из стратегии в обучающей и кооперативной фазах. Концепция равновесия близка к марковскому абсолютному байесовскому равновесию (Markov Perfect Bayesian Equilibrium

¹²Опустим стратегии игроков в кооперативной фазе аргумент функции, поскольку они определены единственным образом правилом (4.3.11) по заданным стратегиям ψ_i, ψ_j .

или МРВЕ) со следующей модификацией. В нашем случае стратегия любого игрока является марковской только в обучающей фазе игры, когда классы игроков неизвестны. В кооперативной фазе игры стратегии игроков являются стратегиями поведения из множества $\{\sigma_{c,\nu}, \sigma_{sc,\nu}, \sigma_{n,\nu}\}$, как определено формулой (4.3.11). Для того чтобы избежать путаницы, используем назовем равновесие частично марковским абсолютным байесовским равновесием (Partially Markov Perfect Bayesian Equilibrium¹³ или РМРВЕ [101]).

§ 4.3.4 Двухфазовая игра с кооперативным и полukoоперативным равновесиями

В этом параграфе опишем равновесия, согласно которым во второй фазе игры разыгрывается кооперативное или полukoоперативное равновесия (в зависимости от классов игроков, участвующих в игре). Правило выбора стратегии в кооперативной фазе определено равенством (4.3.11). Сначала рассмотрим случай, когда продолжительность обучающей фазы равна одному периоду. Хотя это ограничение кажется достаточно строгим, этот случай позволяет понять некоторые свойства равновесий, которые будут найдены во втором случае, когда продолжительность обучающей фазы не ограничена.

Обучающая фаза продолжительностью в один период.

В этом разделе ограничим продолжительность обучающей фазы одним периодом. Это упрощение делается для того, чтобы на простом случае показать роль ожиданий игроков в решении поставленной задачи. Похожие результаты будут получены и в случае, когда мы опустим ограничение в один период для продолжительности обучающей фазы, но анализ и условия для равновесия будут более сложными. Тем не менее, заданная изначально продолжительность обучающей фазы может быть обусловлена внешними факторами, например, один обучающий период может иметь большую стоимость и игроки заинтересованы в наискорейшем окончании обучающей фазы.

Найдем условия, при которых частично марковское абсолютное байесовское равновесие существует. Сформулируем теоремы, сгруппировав равновесия по

¹³Равновесие названо “частично” марковским, потому что в кооперативной фазе стратегии игроков не марковские, а являются стратегиями поведения как только закончивается обучающая фаза.

типу равновесий, реализуемых в кооперативной фазе. Для удобства введем константы A_1, A_2, A_3 и A_4 , которые определены ниже в доказательстве теоремы.

Теорема 4.8. *Рассмотрим двухфазную игру Дилемма заключенного с обучающей фазой продолжительностью в один период. Пусть выполнены предположения 4.1 и 4.2, тогда существуют следующие равновесия (РМРВЕ):*

1. Если начальное состояние есть $s = s_1$:

1.i $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ — частично марковское абсолютное байесовское равновесие при $\alpha_1 \in [A_1; A_2]$.

1.ii $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ — частично марковское абсолютное байесовское равновесие при $\alpha_1 + \beta_1 \geq A_3$.

2. Если начальное состояние есть $s = s_2$:

2.i $(q_1^{s_2}) = 1$ — частично марковское абсолютное байесовское равновесие при $\alpha_1 \geq A_4$.

2.ii $(q_1^{s_2}) = \frac{A_4}{\alpha_1}$ — частично марковское абсолютное байесовское равновесие при $\alpha_1 \geq A_4$.

Доказательство. Для доказательства теоремы введем в рассмотрение еще одну ситуацию в стратегиях поведения. Эта ситуация может реализоваться в случае отклонения одного из игроков. После этого приступим к доказательству теоремы.

Ситуация отклонения

Определение 4.12. “Ситуацию отклонения” обозначим через $\sigma_{\mathbf{d}} = (\sigma_{\mathbf{d},i}, \sigma_{\mathbf{c},j})$, где $\sigma_{\mathbf{d},i} = \{\sigma_{i,t}^s\}_{t=1,\dots,\infty}^{s=s_1,s_2}$ такой что

$$\sigma_{i,t+1}^s(H(t)) = \begin{cases} C, & \text{если } s = s_1 \text{ и } H(t) = H_{\mathbf{c}}(t) \\ D, & \text{если } s = s_1 \text{ и } H(t) \neq H_{\mathbf{c}}(t) \\ D, & \text{если } s = s_2 \end{cases}$$

При этой ситуации игрок j играет стратегию $\sigma_{\mathbf{c},j}$ согласно определению 4.9, тогда как игрок i играет стратегию $\sigma_{\mathbf{d},i}$. Этот набор стратегий может реализоваться, если игрок j имеет ожидание, что оппонент i класса P и будет играть кооперативную стратегию, а он на самом деле будет отклоняться в состоянии $s = s_2$. Когда игрок j увидит отклонение игрока i от кооперативной ситуации, он переключится на стратегию D , которую будет выбирать во все последующие моменты времени согласно стратегии $\sigma_{\mathbf{c},j}$.

Обозначим через $V_{\mathbf{d}}(\delta_i)$ ожидаемый выигрыш отклоняющегося игрока i в ситуации $\sigma_{\mathbf{d}}$. Посчитаем ожидаемый выигрыш $V_{\mathbf{d},i}(\delta_i)$ отклоняющегося игрока i по формуле

$$V_{\mathbf{d}}(\delta_i) = \begin{pmatrix} V_{\mathbf{d}}^{s_1}(\delta_i) \\ V_{\mathbf{d}}^{s_2}(\delta_i) \end{pmatrix},$$

где $V_{\mathbf{d}}^s(\delta_i)$ — выигрыш игрока i в подыгре, начинающейся из состояния s . Если подыгра начинается из состояния s_1 , игрок i получает

$$V_{\mathbf{d}}^{s_1}(\delta_i) = a^{s_1} + \delta(\pi^{s_1} V_{\mathbf{d},i}^{s_1}(\delta_i) + (1 - \pi^{s_1}) V_{\mathbf{d},i}^{s_2}(\delta_i)).$$

Если подыгра начинается из состояния s_2 , игрок i отклоняется i и получает b^{s_2} , после этого он будет наказан разыгрыванием ситуации (D, D) в любом состоянии, начиная со следующего периода и до бесконечности. Его выигрыш равен

$$V_{\mathbf{d}}^{s_2}(\delta_i) = b^{s_2} + \delta \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_i).$$

Из этих двух выражений получаем

$$V_{\mathbf{d}}(\delta_i) = \begin{pmatrix} \frac{1}{1 - \delta_i \pi^{s_1}} [a^{s_1} + \delta_i (1 - \pi^{s_1}) (b^{s_2} + \delta_i \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_i))] \\ b^{s_2} + \delta \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_i) \end{pmatrix}.$$

Начальное состояние $s = s_1$. Ситуация $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$.

Состояние $s = s_1$. Начнем с игрока типа P . Если он не отклоняется от ситуации $(1, 0)$, он получает выигрыш

$$\alpha_1 (a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{c}}(\delta_P)) + (1 - \alpha_1) (c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)). \quad (4.3.14)$$

Если он отклоняется от состояния $(1, 0)$ на стратегию $q_1^{s_1} = 0$, он получает ожидаемый выигрыш:

$$\alpha_1 (b^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)) + (1 - \alpha_1) (d^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)). \quad (4.3.15)$$

Отклонение не приносит прибыли, если выигрыш (4.3.14) больше или равен выигрышу (4.3.15) при условии, что $\delta_P \geq \delta_2^*$.

Теперь рассмотрим игрока типа M . Его выигрыш в ситуации $(1, 0)$ равен

$$\alpha_1(b^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)) + (1 - \alpha_1)(d^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)). \quad (4.3.16)$$

Если он отклоняется от ситуации $(1, 0)$ и играет стратегию $r_1^{s_1} = 1$, он получает

$$\alpha_1(a^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{d}}(\delta_M)) + (1 - \alpha_1)(c^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)), \quad (4.3.17)$$

где $V_{\mathbf{d}}(\delta_M)$ — выигрыш игрока типа M , когда он кооперируется в состоянии $s = s_1$ и отклоняется в состоянии $s = s_2$, что является прибыльным, исходя из условий на дисконтирующий фактор. Отклонение не приносит прибыли, если выигрыш (4.3.16) больше или равен выигрышу (4.3.17) при условии, что $\delta_1^* \leq \delta_M \leq \delta_2^*$.

Если игра начинается из состояния $s = s_1$, то ситуация $(1, 0)$ является частично марковским абсолютным байесовским равновесием, если одна из следующих систем имеет решение:

$$\begin{cases} \alpha_1 \geq \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_P \mathbf{P}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))} \\ d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \leq 0 \end{cases},$$

или

$$\begin{cases} \alpha_1 \geq \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_P \mathbf{P}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))} \\ d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) > 0 \\ \alpha_1 \leq \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M))} \end{cases}.$$

Теперь необходимо проверить выполнение неравенства:

$$d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)). \quad (4.3.18)$$

Сначала рассмотрим разность $\delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M))$. По лемме 4.1, получаем

$$\delta_M \mathbf{P}^{s_1} V_{\mathbf{d}}(\delta_M) > \delta_M \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_M)$$

или в эквивалентном виде

$$\delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) > \delta_M \mathbf{P}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \quad (4.3.19)$$

Учитывая, что

$$\delta_M \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \geq b^{s_1} - a^{s_1},$$

получаем

$$d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \geq d^{s_1} - c^{s_1}.$$

Таким образом, получаем, что (4.3.18) положительно.

Упрощая системы и вычислив $\delta_P \geq \delta_2^*$, мы получаем условие:

$$\alpha_1 \in [A_1, A_2], \quad (4.3.20)$$

где

$$A_1 \equiv \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_P \mathbf{P}^{s_1}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))},$$

и

$$A_2 \equiv \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}.$$

Начальное состояние $s = s_1$. **Ситуация** $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$.

Начнем с игрока типа P . Если он не отклоняется от ситуации $(1, 1)$, то получает выигрыш

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_P)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_P)). \quad (4.3.21)$$

Если он отклоняется от ситуации $(1, 1)$ и реализует стратегию $q_1 = 0$, он получает

$$(\alpha_1 + \beta_1)(b^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_P)) + (1 - \alpha_1 - \beta_1)(d^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_P)). \quad (4.3.22)$$

Отклонение не будет прибыльным, если выигрыш (4.3.21) больше или равен выигрышу (4.3.22), учитывая ограничение $\delta_P \geq \delta_2^*$.

Далее рассмотрим игрока типа M . Его выигрыш в ситуации $(1, 1)$ равен

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_M)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)). \quad (4.3.23)$$

Если он отклоняется от ситуации $(1, 1)$ на стратегию $r_1^{s_1} = 0$, он получает:

$$(\alpha_1 + \beta_1)(b^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)) + (1 - \alpha_1 - \beta_1)(d^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)). \quad (4.3.24)$$

Отклонение не будет прибыльным, если выигрыш (4.3.23) больше или равен выигрышу (4.3.24), учитывая ограничение $\delta_1^* \leq \delta_M \leq \delta_2^*$.

Если игра начинается в состоянии s_1 , ситуация $(1, 1)$ является частично марковским абсолютным байесовским равновесием, если система:

$$\begin{cases} (\alpha_1 + \beta_1) [a^{s_1} - b^{s_1} - c^{s_1} + d^{s_1} + \delta_P \mathbf{P}^{s_1}(V_{\text{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))] \geq d^{s_1} - c^{s_1}, \\ (\alpha_1 + \beta_1) [a^{s_1} - b^{s_1} - c^{s_1} + d^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\text{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))] \geq d^{s_1} - c^{s_1}. \end{cases}$$

имеет решение.

Так как $\delta_M < \delta_P$, эта система эквивалентна неравенству:

$$\alpha_1 + \beta_1 \geq \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\text{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}. \quad (4.3.25)$$

Введем обозначение:

$$A_3 \equiv \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\text{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}.$$

Начальное состояние $s = s_2$. Ситуация $(q_1^{s_2}) = (1)$.

Если игра начинается из состояния s_2 , ситуация $(q_1^{s_2}) = (1)$ является частично марковским абсолютным байесовским равновесием, если следующее неравенство имеет решение:

$$\alpha_1 [a^{s_2} - b^{s_2} - c^{s_2} + d^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))] \geq d^{s_2} - c^{s_2}.$$

Так как $\delta_P \geq \delta_2^*$, тогда $\delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P)) \geq b^{s_2} - a^{s_2}$, что

$$\alpha_1 \geq \frac{d^{s_2} - c^{s_2}}{a^{s_2} - b^{s_2} - c^{s_2} + d^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}. \quad (4.3.26)$$

Введем обозначение:

$$A_4 \equiv \frac{d^{s_2} - c^{s_2}}{a^{s_2} - b^{s_2} - c^{s_2} + d^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}.$$

Начальное состояние $s = s_2$. Ситуация $(q_1^{s_2}), q_1^{s_2} \in (0, 1)$.

Если игрок типа P не отклоняется от ситуации $(q_1^{s_2}) = (q_1)$, он получает выигрыш

$$\begin{aligned} & \alpha_1 \{ (q_1)^2 (a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1 - q_1) q_1 (b^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) \\ & + (1 - q_1) q_1 (c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + (1 - q_1)^2 (d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) \} \\ & + (1 - \alpha_1) \{ q_1 (c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + (1 - q_1) (d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) \}. \end{aligned} \quad (4.3.27)$$

Если он отклоняется от ситуации (q_1) на чистую стратегию $q_1 = 0$, он получает выигрыш

$$\begin{aligned} & \alpha_1 \{q_1(b^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + (1 - q_1)(d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P))\} \\ & + (1 - \alpha_1)(d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)). \end{aligned} \quad (4.3.28)$$

Если он отклоняется от ситуации (q_1) на чистую стратегию $q_1 = 1$, он получает выигрыш

$$\begin{aligned} & \alpha_1 \{q_1(a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1 - q_1)(c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P))\} \\ & + (1 - \alpha_1)(c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)). \end{aligned} \quad (4.3.29)$$

Отклонение невыгодно, если выигрыш (4.3.27) больше или равен выигрышей (4.3.28) и (4.3.29), учитывая ограничение $\delta_P \geq \delta_2^*$. Это имеет место, если

$$\alpha_1 q_1 = \frac{d^{s_2} - c^{s_2}}{d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}. \quad (4.3.30)$$

Следовательно, стратегия

$$q_1 = q^* \equiv \frac{d^{s_2} - c^{s_2}}{\alpha_1 (d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P)))}$$

является оптимальной, если

$$\alpha_1 \geq \frac{d^{s_2} - c^{s_2}}{d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}$$

или $\alpha_1 \geq A_4$. □

На рисунках 4.8 и 4.9 изображены области РМРВЕ для начальных состояний s_1 и s_2 соответственно, на плоскости, осями которой являются ожидания (α_s, β_s) . Правило пересчета ожиданий (4.3.8) и (4.3.9) помогают понять, будет ли реализована та или иная ситуация в кооперативной фазе. Предположим, например, что игра начинается из состояния s_1 и реализуется ситуация $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$. Если наблюдается стратегия C , то вероятности пересчитываются и становятся следующими: $\alpha_2 = 1, \beta_2 = \gamma_2 = 0$. Таким образом, за один период возможно распознать класс другого игрока, если он является игроком класса P . Рассмотрим другой случай, когда встретились два игрока класса P , тогда равновесие $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ приводит к кооперативному равновесию σ_c

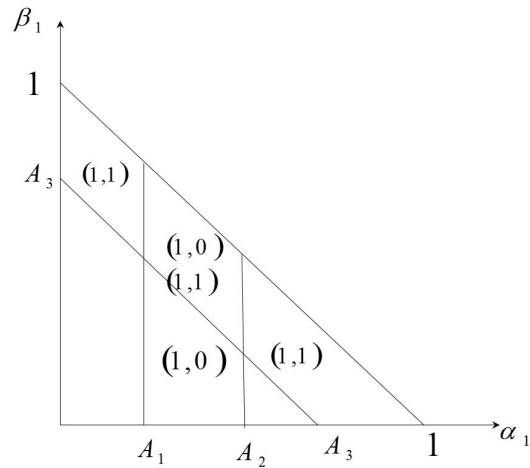


Рис. 4.8: Равновесия для начального состояния $s = s_1$.

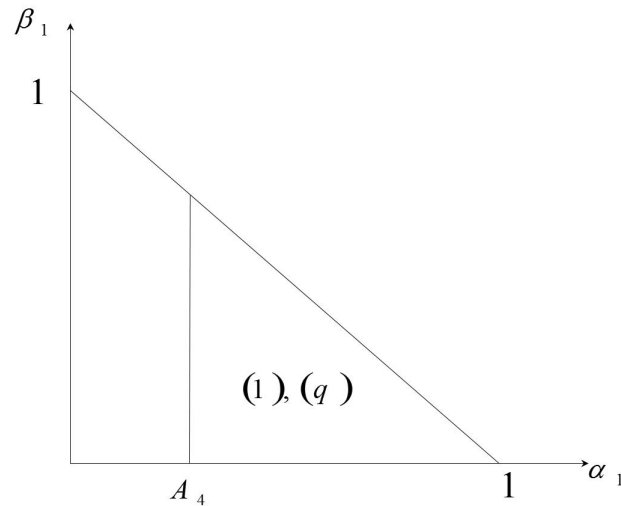


Рис. 4.9: Равновесия для начального состояния $s = s_2$.

в кооперативной фазе игры. Но если хотя бы один из игроков не является игроком класса P , то в кооперативной фазе будет реализована некооперативное равновесие σ_n . Это происходит, потому что «умеренно терпеливый» и «нетерпеливый» игроки используют одинаковые стратегии, и терпеливый игрок не может распознать «умеренно терпеливого» игрока во время обучающей фазы игры, поэтому полukoоперативное равновесие никогда не будет разыграно в кооперативной фазе.

Если игра начинается из состояния s_1 , и реализуется ситуация $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$, то есть, игроки классов M и P кооперируются с вероятностью 1, ожидания о типе другого игрока при наблюдении стратегии C становятся следующи-

ми:

$$\alpha_2 = \frac{\alpha_1}{\alpha_1 + \beta_1}, \beta_2 = \frac{\beta_1}{\alpha_1 + \beta_1}, \gamma_2 = 0.$$

В этом случае вероятности того, что другой игрок является игроком класса P или M , положительны. Поэтому равновесие приводит к полукооперативному равновесию σ_{sc} во второй фазе игры в соответствии с (4.3.11), что позволяет игрокам кооперироваться во всех будущих состояниях s_1 и играть некооперативно во всех будущих состояниях s_2 . Этот результат объясняется тем, что игрок не может распознать с вероятностью 1, является ли его оппонент класса P или M .

Если игра начинается из состояния s_2 , то игрок класса M не может быть распознан, поскольку он всегда отклоняется в этом состоянии. Поэтому ожидание β_1 не играет никакой роли в определении ситуации равновесия. Тем не менее, игрок класса P может быть идентифицирован с вероятностью 1, и два игрока класса P , встретившись в игре, будут реализовывать кооперативное равновесие σ_c в кооперативной фазе.

Сравним выигрыши игроков в равновесиях в зависимости от параметров игры в случае, если существуют одновременно два равновесия. Это позволит игрокам выбирать равновесие в случае, если возникает такая задача.

Следствие 4.1. *Если двухфазная стохастическая игра начинается из состояния $s = s_1$, при этом начальные ожидания α_1, β_1 удовлетворяют условиям: $\alpha_1 \in [A_1, A_2]$ и $\alpha_1 + \beta_1 \in [A_3, 1]$. Тогда выигрыш игрока класса M в равновесии $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ не меньше, чем его выигрыш в равновесии $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$.*

Если игра начинается из состояния $s = s_2$ и при этом $\alpha_1 \geq A_4$, тогда выигрыш игрока класса P в равновесии $q_1^{s_2} = 1$ не меньше, чем его выигрыш в равновесии $q_1^{s_2} = q^$.*

Доказательство. Выигрыш игрока класса P в ситуации $(1, 0)$ равен

$$\alpha_1(a^{s_1} + \delta_P \mathbf{P}^{s_1} V_c(\delta_P)) + (1 - \alpha_1)(c^{s_1} + \delta_P \mathbf{P}^{s_1} V_n(\delta_P)),$$

и в ситуации $(1, 1)$ равен

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_P \mathbf{P}^{s_1} V_{sc}(\delta_P)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_P \mathbf{P}^{s_1} V_n(\delta_P)).$$

Выигрыш игрока класса P в ситуации $(1, 1)$ не меньше, чем его выигрыш в ситуации $(1, 0)$, если

$$\beta_1(c^{s_1} - a^{s_1} - \delta_P \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))) + \alpha_1 \delta_P \mathbf{P}^{s_1}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P)) \leq 0,$$

или

$$\frac{\beta_1}{\alpha_1} \geq \frac{\delta_P \mathbf{P}^{s_1}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P))}{a^{s_1} - c^{s_1} + \delta_P \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}.$$

Выигрыш игрока класса M в ситуации $(1, 0)$ равен

$$\alpha_1(b^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)) + (1 - \alpha_1)(d^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M))$$

и в ситуации $(1, 1)$ равен

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_M)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)).$$

Выигрыш игрока класса M в ситуации $(1, 1)$ не меньше, чем его выигрыш в ситуации $(1, 0)$, если

$$\begin{aligned} \alpha_1(b^{s_1} - a^{s_1} + c^{s_1} - d^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{n}}(\delta_M) - V_{\mathbf{sc}}(\delta_M))) \\ + \beta_1(c^{s_1} - a^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{n}}(\delta_M) - V_{\mathbf{sc}}(\delta_M))) + d^{s_1} - c^{s_1} \leq 0. \end{aligned}$$

или

$$(\alpha_1 + \beta_1)(d^{s_1} - b^{s_1} + a^{s_1} - c^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))) \geq \beta_1(d^{s_1} - b^{s_1}) + (d^{s_1} - c^{s_1}). \quad (4.3.31)$$

Учитывая, что $\alpha_1 + \beta_1 \geq A_3$, мы можем утверждать, что

$$(\alpha_1 + \beta_1)(d^{s_1} - b^{s_1} + a^{s_1} - c^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))) \geq d^{s_1} - c^{s_1}.$$

Последнее неравенство гарантирует, что неравенство (4.3.31) выполнено, потому что $d^{s_1} - b^{s_1} < 0$.

Теперь рассмотрим начальное состояние $s = s_2$. Выигрыш игрока класса P в равновесии $(q_1^{s_2}) = 1$ равен

$$\alpha_1(a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1 - \alpha_1)(c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)).$$

Выигрыш игрока класса P в равновесии $(q_1^{s_2}) = q^*$, где $q^* = A_4/\alpha_1$ (см. теорему 4.8) равен

$$\begin{aligned} \alpha_1 q q (a^{s_2} - b^{s_2} + d^{s_2} - c^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))) + \alpha_1 q (b^{s_2} - d^{s_2}) + q (c^{s_2} - d^{s_2}) \\ + d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P) = d^{s_2} + \alpha_1 q (b^{s_2} - d^{s_2}) + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P). \end{aligned}$$

Выигрыш игрока класса P в ситуации $(q_1^{s_2}) = 1$ не меньше, чем его выигрыш в ситуации $(q_1^{s_2}) = q$, если

$$\begin{aligned} & \alpha_1(a^{s_2} - c^{s_2} + \delta_P \mathbf{P}^{s_2}(V_c(\delta_P) - V_n(\delta_P))) \\ & \geq (d^{s_2} - c^{s_2}) \left[1 + \frac{b^{s_2} - d^{s_2}}{d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2}(V_c(\delta_P) - V_n(\delta_P))} \right], \end{aligned}$$

что всегда выполнено для $\alpha_1 \geq A_4$.

Выигрыш игрока класса M в ситуации $(q_1^{s_2}) = 1$ равен

$$\alpha_1(b^{s_2} + \delta_M \mathbf{P}^{s_2} V_n(\delta_M)) + (1 - \alpha_1)(d^{s_2} + \delta_M \mathbf{P}^{s_2} V_n(\delta_M)),$$

и в ситуации $(q_1^{s_2}) = q$ равен

$$\alpha_1(q(b^{s_2} + \delta_M \mathbf{P}^{s_2} V_n(\delta_M)) + (1 - q)(d^{s_2} + \delta_M \mathbf{P}^{s_2} V_n(\delta_M))) + (1 - \alpha_1)(d^{s_2} + \delta_M \mathbf{P}^{s_2} V_n(\delta_M)).$$

Выигрыш игрока класса M в ситуации $(q_1^{s_2}) = 1$ не меньше выигрыша в ситуации $(q_1^{s_2}) = q$, если

$$\alpha_1(b^{s_2} - qb^{s_2} - (1 - q)d^{s_2}) \geq 0,$$

что всегда выполнено, поскольку $b^{s_2} > d^{s_2}$. \square

Следствие 4.1 позволят сделать вывод о том, что равновесию $(1, 1)$ может быть отдано предпочтение в случае существования нескольких равновесий в игре, начинающейся из состояния s_1 . А также равновесие в чистых стратегиях предпочтительнее равновесия в смешанных стратегиях, если игра начинается из состояния s_2 . В частности, для начального состояния s_1 , этот результат интуитивно предполагает, что когда ожидания того, что другой игрок класса P или M , близки, то маловероятно достичь полной кооперации (во всех состояниях второй фазы игры). И результатом будет как раз полукооперативное равновесие в кооперативной фазе.

Следующая теорема позволяет провести анализ выигрышей игроков в равновесии как функций ожиданий.

Теорема 4.9. *Выигрыши игроков классов P и M в равновесиях — возрастающие функции по параметру α_1 . Выигрыши игроков классов P и M в равновесии $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ — возрастающие функции параметра β_1 .*

Доказательство. Рассмотрим выигрыши игроков классов P and M как функции параметра α_1 . Следуя Теореме 4.8, существует три равновесия:

1. Равновесие $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$: выигрыш игрока класса P равен

$$\alpha_1(a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{c}}(\delta_P)) + (1 - \alpha_1)(c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)).$$

Это линейная функция по параметру α_1 с коэффициентом $a^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))$, который положителен, потому что $a^{s_1} > c^{s_1}$ и $V_{\mathbf{c}}(\delta) > V_{\mathbf{n}}(\delta)$ для всех $\delta \in (0, 1)$.

Выигрыш игрока класса M равен

$$\alpha_1(b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)) + (1 - \alpha_1)(d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)).$$

Это также линейная функция по параметру α_1 с коэффициентом $b^{s_1} - d^{s_1}$, который положителен для любого $\delta \in (0, 1)$.

2. Равновесие $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$: рассмотрим сначала игрока класса P . Его выигрыш равен

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)).$$

Это линейная функция по параметру α_1 с коэффициентом $a^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))$, который положителен, потому что $a^{s_1} > c^{s_1}$ и $V_{\mathbf{sc}}(\delta) > V_{\mathbf{n}}(\delta)$ для любого $\delta \in (0, 1)$.

Перейдем к игроку класса M . Его выигрыш равен

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)).$$

Это линейная функция по параметру α_1 с коэффициентом $a^{s_1} - c^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))$, который положителен, потому что $a^{s_1} > c^{s_1}$ и $V_{\mathbf{sc}}(\delta) > V_{\mathbf{n}}(\delta)$ для любого $\delta \in (0, 1)$.

Производные выигрышей игроков классов P и M по параметру β_1 равны соответствующим производным по параметру α_1 . Следовательно, выигрыши также являются возрастающими функциями по параметру β_1 .

3. Равновесие $(q_1) = (1)$ при начальном состоянии $s = s_2$: выигрыш игрока класса P равен

$$\alpha_1(a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1 - \alpha_1)(c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P))$$

Это линейная функция по параметру α_1 с коэффициентом $a^{s_2} - c^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))$, который положителен, потому что $a^{s_2} > c^{s_2}$ и $V_{\mathbf{c}}(\delta) > V_{\mathbf{n}}(\delta)$ для любого $\delta \in (0, 1)$.

4. Равновесие $(q_1) = q^*$ при начальном состоянии $s = s_2$. Выигрыш игрока класса P равен

$$q^* = \frac{d^{s_2} - c^{s_2}}{\alpha_1(d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P)))}.$$

Выигрыш игрока класса P равен

$$\begin{aligned} & \alpha_1 \{q_1(a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1 - q_1)q_1(b^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) \\ & + (1 - q_1)q_1(c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + (1 - q_1)(d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P))\} \\ & + (1 - \alpha_1) \{q_1(c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + (1 - q_1)(d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P))\}. \end{aligned}$$

Подставляя q^* в выигрыш и найдя производную по параметру α_1 , получаем производную, равную

$$\frac{A_4^2}{2\alpha_1^2}(b^{s_2} + c^{s_2} + s_2 \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + \frac{A_4}{2\alpha_1^2}(d^{s_2} - c^{s_2}),$$

которая положительна, потому что $d^{s_2} > c^{s_2}$.

□

Теорема 4.9 вместе с Теоремой 4.8 позволяют сделать любопытный вывод: даже сильное ожидание того, что другой игрок является игроком класса P не приводит к кооперативному равновесию (полной кооперации) во второй фазе игры. Это можно легко заметить, взглянув на рис. 4.8. Причина в следующем: большие значения α_1 дают больше мотивации для игрока класса M имитировать «терпеливость», что будет означать, что он будет играть как игрок класса P для того, чтобы другой игрок, будучи игроком класса P во второй фазе игры стал реализовывать кооперативную стратегию. Тогда, если во второй фазе

будет игрок будет разыгрывать кооперативную стратегию, и в некоторый период реализуется состояние s_2 , то игрок класса M отклонится от кооперативной стратегии (ему это выгодно), тем самым обманув игрока класса P . Учитывая, что игроки знают об этом эффекте «поддельного терпения», то реализуется полукооперативное равновесие во второй фазе игры: кооперация в состоянии s_1 и отсутствие кооперации в состоянии s_2 .

§ 4.3.5 Двухфазная игра с обучающей фазой непредписанной продолжительности

В этом параграфе мы рассмотрим случай, когда продолжительность обучающей фазы не фиксирована. Мы приведем условия существования нескольких равновесий и покажем, что эффект имитации «терпимости» также наблюдается в этом случае. Мы ограничимся рассмотрением равновесий, для которых кооперативная фаза игры наступает как можно скорее.

Начальное состояние $s = s_1$.

Естественно предположить следующую структуру равновесий, отвечающих упомянутым требованиям. В первый момент времени игроки классов P и M выбирают стратегию C , чтобы точно узнать, принимает ли участие в игре игрок класса I . Далее, если игроков класса I нет в игре, то игровой процесс переходит в состояние s_1 или s_2 , в котором игроки классов P и M выбирают разные стратегии, чтобы тем самым идентифицировать свои классы с вероятностью единица. Теперь найдем условия, при которых описанные стратегии образуют РМРВЕ. Далее рассмотрим ситуации, в соответствии с которыми игроки классов P и M при начальном состоянии s_1 используют различные стратегии в момент времени $t = 1$.

Предположим, что в первый момент времени игроки классов P и M выбирают стратегии $q_1^{s_1} = r_1^{s_1} = 1$. Во второй момент времени и в состояниях

- $s = s_1$, стратегии следующие: $q_2^{s_1} = 1, r_2^{s_1} = 0$;
- $s = s_2$, стратегиях игрока класса P есть $q_2^{s_2} = 1$.

Используя такие стратегии, классы игроков идентифицируются не позднее

второго момента времени. Следующая теорема дает условия на начальные ожидания игроков, при которых описанные стратегии образуют частично марковское абсолютное байесовское равновесие. Для упрощения записи введем в рассмотрение коэффициенты A_5 , A_6 , A_7 и A_8 , которые определены в доказательстве теоремы, при этом $A_5, A_7 > 0$.

Теорема 4.10. Пусть начальное состояние игры $s = s_1$ и выполнены следующие условия:

$$\left\{ \begin{array}{l} i. \quad A_5\alpha_1 + A_6\beta_1 \geq d^{s_1} - c^{s_1}, \\ ii. \quad A_7\alpha_1 + A_8\beta_1 \geq d^{s_1} - c^{s_1}, \\ iii. \quad \frac{\beta_1}{\alpha_1} \leq \min_{s=s_1, s_2} \left\{ \frac{\delta_P \mathbf{P}^s (V_c(\delta_P) - V_{sc}(\delta_P)) - (b^s - a^s)}{d^s - c^s} \right\}, \\ iv. \quad \frac{\beta_1}{\alpha_1} \geq \frac{\delta_M \mathbf{P}^{s_1} (V_d(\delta_M) - V_{sc}(\delta_M)) - (b^{s_1} - a^{s_1})}{d^{s_1} - c^{s_1}}, \end{array} \right.$$

тогда следующие стратегии образуют частично марковское абсолютное байесовское равновесие:

$$P : (q_1^{s_1}, q_2^{s_1}, q_2^{s_2}) = (1, 1, 1), \quad M : (r_1^{s_1}, r_2^{s_1}) = (1, 0).$$

В кооперативной фазе игры разыгрывается кооперативное равновесие, если в игре принимают участие два игрока класса P , и полужооперативное равновесие, если — два игрока класса M , или один игрок класса P и другой игрок класса M .

Доказательство. Момент времени 1, состояние s_1 .

Если игрок класса P придерживается стратегии $q_1^{s_1} = 1$, его выигрыш будет

$$\begin{aligned} & \alpha_1 [a^{s_1} + \delta_P \mathbf{P}^{s_1} V_c(\delta_P)] \\ & + \beta_1 [a^{s_1} + \delta_P \pi^{s_1} (c^{s_1} + \delta_P \mathbf{P}^{s_1} V_{sc}(\delta_P)) + \delta_P (1 - \pi^{s_1}) (c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{sc}(\delta_P))] \\ & + (1 - \alpha_1 - \beta_1) [c^{s_1} + \delta_P \mathbf{P}^{s_1} V_n(\delta_P)]. \end{aligned}$$

Если он отклоняется на стратегию $q_1^{s_1} = 0$, его класс будет идентифицирован как I и его выигрыш составит

$$\alpha_1 [b^{s_1} + \delta_P \mathbf{P}^{s_1} V_n(\delta_P)] + \beta_1 [b^{s_1} + \delta_P \mathbf{P}^{s_1} V_n(\delta_P)] + (1 - \alpha_1 - \beta_1) [d^{s_1} + \delta_P \mathbf{P}^{s_1} V_n(\delta_P)].$$

Учитывая, что $\mathbf{p}^s = (\pi^s, 1 - \pi^s)$, отклонение игрока класса P в период 1 не прибыльно, если

$$\begin{aligned} & \alpha_1 [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1} (V_c(\delta_P) - V_n(\delta_P))] \\ & + \beta_1 [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \pi^{s_1} (c^{s_1} - a^{s_1}) \\ & + \delta_P (1 - \pi^{s_1}) (c^{s_2} - d^{s_2}) + \delta_P \mathbf{p}^{s_1} (V_{sc}(\delta_P) - V_n(\delta_P))] \\ & \geq d^{s_1} - c^{s_1}. \end{aligned}$$

Введем обозначения:

$$\begin{aligned} A_5 & \equiv [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1} (V_c(\delta_P) - V_n(\delta_P))], \\ A_6 & \equiv [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \pi^{s_1} (c^{s_1} - a^{s_1}) + \delta_P (1 - \pi^{s_1}) (c^{s_2} - d^{s_2}) \\ & + \delta_P \mathbf{p}^{s_1} (V_{sc}(\delta_P) - V_n(\delta_P))]. \end{aligned}$$

Заметим, что $A_5 > 0$, учитывая, что $d^{s_1} - c^{s_1} > 0$ и теорему 4.7 и лемму 4.1, получаем

$$\delta_P > \frac{a^{s_1} - b^{s_1}}{\mathbf{p}^{s_1} (V_c(\delta_P) - V_n(\delta_P))}.$$

Если игрок M следует стратегии $r_1^{s_1} = 1$, его выигрыш будет равен

$$\begin{aligned} & \alpha_1 [a^{s_1} + \delta_M \pi^{s_1} (b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{sc}(\delta_M)) + \delta_M (1 - \pi^{s_1}) (b^{s_2} + \delta_M \mathbf{p}^{s_2} V_{sc}(\delta_M))] \\ & + \beta_1 [a^{s_1} + \delta_M \pi^{s_1} (d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{sc}(\delta_P)) + \delta_M (1 - \pi^{s_1}) (d^{s_2} + \delta_M \mathbf{p}^{s_2} V_{sc}(\delta_M))] \\ & + (1 - \alpha_1 - \beta_1) [c^{s_1} + \delta_M \mathbf{p}^{s_1} V_n(\delta_M)]. \end{aligned}$$

Если он отклоняется на стратегию $r_1^{s_1} = 0$, его класс будет идентифицирован как класс I , его выигрыш будет равен

$$\alpha_1 [b^{s_1} + \delta_M \mathbf{p}^{s_1} V_n(\delta_M)] + \beta_1 [b^{s_1} + \delta_M \mathbf{p}^{s_1} V_n(\delta_M)] + (1 - \alpha_1 - \beta_1) [d^{s_1} + \delta_M \mathbf{p}^{s_1} V_n(\delta_M)].$$

Отклонение игрока M в момент времени 1 невыгодно, если

$$\begin{aligned} & \alpha_1 [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{sc}(\delta_M) - V_n(\delta_M))] \\ & + \delta_M \pi^{s_1} (b^{s_1} - a^{s_1}) + \delta_M (1 - \pi^{s_1}) (b^{s_2} - d^{s_2})] \\ & + \beta_1 [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{sc}(\delta_M) - V_n(\delta_M))] \\ & + \delta_M \pi^{s_1} (d^{s_1} - a^{s_1})] \geq d^{s_1} - c^{s_1}. \end{aligned}$$

Введем обозначения:

$$A_7 \equiv [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) + \delta_M \pi^{s_1}(b^{s_1} - a^{s_1}) + \delta_M(1 - \pi^{s_1})(b^{s_2} - d^{s_2})],$$

$$A_8 \equiv [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) + \delta_M \pi^{s_1}(d^{s_1} - a^{s_1})].$$

Заметим, что $A_7 > 0$, учитывая, что $b^s > a^s > d^s > c^s$ и теорему 4.7 и лемму 4.1, получаем

$$\delta_M > \frac{a^{s_1} - b^{s_1}}{\mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}.$$

Момент времени 2. Состояние s_1 .

Если в период 1 классы игроков не были распознаны, т. е. только стратегии C были выбраны игроками, то обучающая фаза продолжается и ожидания станут следующими:

$$\alpha_2 = \frac{\alpha_1}{\alpha_1 + \beta_1}, \quad \beta_2 = \frac{\beta_1}{\alpha_1 + \beta_1}.$$

Если игрок P выбирает стратегию $q_2^{s_1} = 1$, его выигрыш равен

$$\alpha_2 [a^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{c}}(\delta_P)] + \beta_2 [c^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_P)].$$

Если он отклоняется на стратегию $q_2^{s_1} = 0$, его класс идентифицируется как M и его выигрыш будет равен

$$\alpha_2 [b^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_P)] + \beta_2 [d^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_P)].$$

Отклонение игрока P в момент времени 2 и состоянии s_1 , не выгодно, если

$$\alpha_2 [a^{s_1} - b^{s_1} + \delta_P \mathbf{P}^{s_1}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P))] + \beta_2 [c^{s_1} - d^{s_1}] \geq 0,$$

учитывая выражения α_2 и β_2 , получаем условие

$$\frac{\beta_1}{\alpha_1} \leq \frac{\delta_P \mathbf{P}^{s_1}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P)) - (b^{s_1} - a^{s_1})}{d^{s_1} - c^{s_1}}.$$

Если игрок M использует стратегию $r_2^{s_1} = 0$, его выигрыш равен

$$\alpha_2 [b^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_M)] + \beta_2 [d^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_M)].$$

Если он отклоняется на стратегию $r_2^{s_1} = 1$, его класс идентифицируется как P и его выигрыш будет равен

$$\alpha_2 [a^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{d}}(\delta_M)] + \beta_2 [c^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_M)].$$

Отклонение игрока M в момент времени 2 и состоянии s_1 , не выгодно, если

$$\alpha_2 [b^{s_1} - a^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\text{sc}}(\delta_M) - V_{\text{d}}(\delta_M))] + \beta_2 [d^{s_1} - c^{s_1}] \geq 0,$$

учитывая выражения α_2 и β_2 , получаем условие

$$\frac{\beta_1}{\alpha_1} \geq \frac{\delta_M \mathbf{P}^{s_1} (V_{\text{d}}(\delta_M) - V_{\text{sc}}(\delta_M)) - (b^{s_1} - a^{s_1})}{d^{s_1} - c^{s_1}}.$$

Момент времени 2. Состояние s_2 .

Если в период 2 игрок P использует стратегию $q_2^{s_2} = 1$, его выигрыш равен

$$\alpha_2 [a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\text{c}}(\delta_P)] + \beta_2 [c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\text{sc}}(\delta_P)].$$

Если он отклоняется на стратегию $q_2^{s_2} = 0$, его класс идентифицируется как M и его выигрыш будет равен

$$\alpha_2 [b^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\text{sc}}(\delta_P)] + \beta_2 [d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\text{sc}}(\delta_P)].$$

Отклонение игрока P в момент времени 2 и состоянии s_2 , не выгодно, если

$$\alpha_2 [a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2} (V_{\text{c}}(\delta_P) - V_{\text{sc}}(\delta_P))] + \beta_2 [c^{s_2} - d^{s_2}] \geq 0.$$

Учитывая выражения α_2 и β_2 , получаем условие

$$\beta_1 \leq \frac{\delta_P \mathbf{P}^{s_2} (V_{\text{c}}(\delta_P) - V_{\text{sc}}(\delta_P)) - (b^{s_2} - a^{s_2})}{d^{s_2} - c^{s_2}} \alpha_1.$$

Объединяя все условия, получим утверждение теоремы. \square

Равновесие, описанное в теореме 4.10 показывает эффект «имитации терпеливости» в первый момент времени, когда игрок класса M , кооперируясь, не раскрывает своего класса. Но далее в момент времени 2, игрок класса M может отклоняться в состоянии s_1 и, поступая таким образом, он вскрывает свой класс, и обучающая фаза игры заканчивается тем, что в кооперативной фазе будет разыгрываться полукооперативное равновесие. Интуитивно понятно, что эффект «имитации терпеливости» откладывает наступление кооперативной фазы. Предположим, что игрок класса M продолжает играть $r_t^{s_1} = 1$ во все периоды t , пока не сменится состояние. В этом случае обучающая фаза не закончится, пока не наступит состояние s_2 .

В отличие от случая, когда обучающая фаза длится один период, условия теоремы 4.10 сложнее интерпретировать. Тем не менее, можно посмотреть на значения α_1 в условиях *i.* и *ii.* Они положительны, что доказано в теореме, и предполагает, что увеличение α_1 увеличивает вероятность выполнения обоих условий. Другими словами, чем больше ожидание того, что оппонент — игрок класса P , тем больше эффект «имитации терпеливости», и тем менее вероятно реализация полной кооперации.

Начальное состояние $s = s_2$.

В состоянии s_2 игрок класса M всегда выбирает стратегию D . Исследуем ситуацию, в соответствии с которой игрок класса P выбирает стратегию C , т. е. $q_1^{s_2} = 1$. Следовательно, если по крайней мере один из двух игроков выбирает стратегию C в период 1, обучающая фаза заканчивается и кооперативная фаза начинается с момента времени 2. В соответствии с правилом (4.3.11), если оба игрока выбрали C , тогда в кооперативной фазе игроки реализуют кооперативное равновесие. Если же один из игроков выбирает C , а другой выбирает D , тогда игроки реализуют некооперативное равновесие, следуя (4.3.11).

Если оба игрока выбрали D , тогда обучающая фаза игры продолжается и в период 2 и игроки идентифицируются как M или I со следующими ожиданиями:

$$\alpha_2 = 0, \quad \beta_2 = \frac{\beta_1}{1 - \alpha_1}, \quad \gamma_2 = \frac{1 - \alpha_1 - \beta_1}{1 - \alpha_1}.$$

Заметим, что игроки таких классов продолжают играть D пока не наступит состояние s_1 , поскольку стратегии игроков в обучающей фазе марковские. Рассмотрим теперь случай, когда состояние s_1 реализуется в момент времени 2. Игрок класса I продолжит играть D . Но игрок класса M будет выбирать стратегию C (стратегия $r_2^{s_1} = 1$) или D (стратегия $r_2^{s_1} = 0$). Если он использует стратегию $r_2^{s_1} = 0$, его ожидания не меняются и стратегия $r_t^{s_1}$ будет равна 0 до бесконечности из-за марковского свойства стратегии.¹⁴ Теперь найдем условия, при которых стратегия $r_2^{s_1} = 1$ — часть равновесия РМРВЕ.

Теорема 4.11. *Пусть игра начинается из состояния $s = s_2$. Предположим,*

¹⁴Следовательно, эта стратегия игрока класса M — часть равновесия РМРВЕ с бесконечной обучающей фазой.

что

$$\begin{cases} \beta_1 \geq \frac{(d^{s_1} - c^{s_1})(1 - \alpha_1)}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}, \\ \beta_1 \leq \frac{\alpha_1 [d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))] - (d^{s_2} - c^{s_2})}{\delta_P \mathbf{P}^{s_2}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}, \end{cases}$$

тогда следующие стратегии образуют частично марковское абсолютное байесовское равновесие:

- стратегия игрока класса P : $q_1^{s_2} = 1$. Обучающая фаза закончится в момент $t = 1$, если по крайней мере один игрок является игроком типа P .
- стратегия игрока класса M : $r_k^{s_2} = 0$ с периода $k = 1$ пока не наступит $s = s_2$; $r_t^{s_1} = 1$, пока $s = s_1$ в момент $t > 1$. Обучающая фаза закончится в момент времени t , если по крайней мере один из игроков класса M .

Доказательство. Отклонение на стратегию $r_2^{s_1} = 0$ не выгодно, если

$$\begin{aligned} & \beta_2(a^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_M)) + \gamma_2(c^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)) \\ & \geq \beta_2(b^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)) + \gamma_2(d^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)). \end{aligned}$$

Учитывая, что $\beta_2 = \frac{\beta_1}{1 - \alpha_1}$ и $\gamma_2 = \frac{1 - \alpha_1 - \beta_1}{1 - \alpha_1}$, получаем

$$\beta_1 \geq \frac{(d^{s_1} - c^{s_1})(1 - \alpha_1)}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}.$$

□

§ 4.3.6 Некооперативные равновесия

Рассмотрим теперь частично марковские абсолютные байесовские равновесия, приводящие к некооперативному равновесию $\sigma_{\mathbf{n}}$ во второй фазе игры согласно правилу (4.3.11). Также как и в предыдущем случае, сначала рассмотрим обучающую фазу продолжительностью в один период.

Теорема 4.12. Пусть продолжительность обучающей фазы равна $T = 1$, тогда наборы стратегий $(q_1^{s_1}, r_1^{s_1}) = (0, 0)$ и $(q_1^{s_2}) = (0)$ — частично марковские абсолютные байесовские равновесия для игры с начальными состояниями $s = s_1$ и $s = s_2$ соответственно.

Доказательство. Сначала рассмотрим начальное состояние $s = s_1$ и ситуацию $(q_1^{s_1}; r_1^{s_1}) = (0, 0)$. Игрок класса P получает выигрыш

$$d^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P). \quad (4.3.32)$$

если он не отклоняется от ситуации $(0, 0)$.

Если он отклоняется от ситуации $(0, 0)$ на стратегию $q_1^{s_1} = 1$, он получает:

$$c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P). \quad (4.3.33)$$

Заметим, что выигрыш (4.3.32) всегда больше или равен выигрыша (4.3.33), поскольку $d^{s_1} \geq c^{s_1}$. Отклонение игрока класса M также не является прибыльным.

Теперь рассмотрим начальное состояние s_2 и ситуацию $(q_1^{s_2}) = (0)$. Игрок класса P получает выигрыш

$$d^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_P). \quad (4.3.34)$$

если он не отклоняется от ситуации (0) .

Если он отклоняется от ситуации (0) на стратегию $q_1^{s_2} = 1$, он получает:

$$c^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_P). \quad (4.3.35)$$

Заметим, что выигрыш (4.3.34) всегда больше или равен выигрыша (4.3.35), поскольку $d^{s_2} \geq c^{s_2}$. Следовательно, ситуация $(q_1^{s_1}, r_1^{s_1}) = (0, 0)$ является частично марковским абсолютным байесовским равновесием в двухфазной игре с начальным состоянием $s = s_1$, и ситуация $(q_1^{s_2}) = (0)$ является частично марковским абсолютным байесовским равновесием в двухфазной игре с начальным состоянием $s = s_2$. \square

Теперь перейдем к случаю, когда продолжительность обучающей фазы не фиксирована. Пусть начальным состоянием будет s_2 . Рассмотрим стратегию $q_1^{s_2} = 0$ игрока P в момент времени 1 в состоянии $s = s_2$. В этом случае все игроки играют стратегию D и после этого периода ожидания игроков не меняются, т. е. $\alpha_2 = \alpha_1$, $\beta_2 = \beta_1$, $\gamma_2 = \gamma_1$. Если в последующие периоды будет появляться только состояние $s = s_2$, то стратегия игрока класса P будет $q_t^{s_2} = 0$ из-за марковского свойства стратегий. Ожидания могут измениться только, если

состояние $s = s_1$ появится в игре. Пусть состояние s_1 реализуется в момент времени $t > 1$. Если в этом состоянии игроки выберут стратегии $q_1^{s_1} = 0$ и $r_1^{s_1} = 0$, тогда ожидания игроков опять не изменятся: $\alpha_t = \alpha_1$, $\beta_t = \beta_1$. Следовательно, по марковскому свойству стратегии по индукции получаем $q_k^{s_1} = r_k^{s_1} = 0$ для любого k . Эти стратегии определяют частично марковское абсолютное байесовское равновесие с бесконечной обучающей фазой, когда игроки всегда играют стратегию D в любом состоянии. Существование частично марковского абсолютного байесовского равновесия можно доказать и для случая, когда игра начинается из состояния s_1 и игроки выбирают стратегию D в этом состоянии и в первый раз появившемся состоянии s_2 также выбирают D . Таким образом, доказана теорема.

Теорема 4.13. *Для любых начальных вероятностей $\alpha_1 > 0$, $\beta_1 > 0$ таких, что $\alpha_1 + \beta_1 < 1$, существует частично марковское абсолютное байесовское равновесие, согласно которому стратегии игроков для любого начального состояния s_1 или s_2 следующие: $q_t^{s_1} = r_t^{s_1} = 0$ и $q_t^{s_2} = 0$, $t = 1, 2, \dots$ (игроки всех классов выбирают стратегию D в любом состоянии в любой момент времени). В этом случае обучающая фаза имеет бесконечную продолжительность.*

§ 4.3.7 Численный пример

Рассмотрим игру с двумя состояниями, представленными биматричными играми

$$s = s_1 : \begin{array}{cc} & C & D \\ C & (11,3), (11,3) & (3,12) \\ D & (12,3) & (5,5) \end{array} \quad s = s_2 : \begin{array}{cc} & C & D \\ C & (11,11) & (3,16) \\ D & (16,3) & (10,10) \end{array}$$

для которой предположение 4.1 выполнено. Пусть вероятности перехода из состояния s_1 в состояния s_1 и s_2 есть $\pi_1 = 0.8$ и $1 - \pi_1 = 0.2$ соответственно, и вероятности перехода из состояния s_2 в состояния s_1 и s_2 есть $\pi_2 = 0.1$ и $1 - \pi_2 = 0.9$ соответственно. Предположение 4.2 также справедливо, и $\delta_1^* \approx 0.1245$, $\delta_2^* \approx 0.7089$. Пусть игроки классов M и P имеют дисконтирующие факторы $\delta_M = 0.7$ и $\delta_P = 0.9$ соответственно.

Нач. состояние	$s = s_1$		$s = s_2$
РМРВЕ	$(q_1^{s_1}, r_1^{s_1}) = (1, 0)$	$(q_1^{s_1}, r_1^{s_1}) = (1, 1)$	$(q_1^{s_2}) = (1)$
Игрок P	$72.3243 + 39.2162\alpha_1$	$72.3243 + 34.3514(\alpha_1 + \beta_1)$	$80.8378 + 29.8919\alpha_1$
Игрок M	$21.2418 + 7\alpha_1$	$19.2418 + 17.2353(\alpha_1 + \beta_1)$	$31.0458 + 6\alpha_1$

Таблица 4.17: Выигрыши в равновесиях для различных начальных состояний.

Обучающая фаза продолжительностью в один период.

Параметры A_1 , A_2 , A_3 и A_4 равны

$$A_1 = 0.0621, \quad A_2 = 0.1781, \quad A_3 = 0.1954, \quad A_4 = 0.2930.$$

По теореме 4.8 получаем

- Если игра начинается из состояния $s = s_1$:

– $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ — РМРВЕ, если $\alpha_1 \in [A_1, A_2] = [0.0621, 0.1781]$.

– $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ — РМРВЕ, если $\alpha_1 + \beta_1 \geq A_3 = 0.1954$.

- Если игра начинается из состояния $s = s_2$:

– $(q_1^{s_2}) = (1)$ и $(q_1^{s_2}) = (q_1^{s_2*})$, где $q_1^{s_2*} = \frac{0.2930}{\alpha_1}$ — РМРВЕ, если $\alpha_1 \geq A_4 = 0.2930$.

Из теоремы 4.12 следует, что ситуации $(q_1^{s_1}, r_1^{s_1}) = (0, 0)$ и $(q_1^{s_2}) = (0)$ — РМРВЕ для начальных состояний s_1 и s_2 соответственно (с бесконечной обучающей фазой).

Таблица 4.17 содержит ожидаемые выигрыши игроков классов P и M , вычисленные при соответствующих равновесиях. Показано, что оба игрока имеют большие ожидаемые выигрыши в игре, начинающейся из состояния s_2 , чем при равновесии $(1, 0)$ с начальным состоянием s_1 . Но выигрыши игроков в равновесии $(1, 1)$, при котором стратегии игроков в кооперативной фазе образуют полукооперативное равновесие, могут быть больше, чем при равновесии $(1, 0)$, согласно которому стратегии игроков в кооперативной фазе образуют кооперативное равновесие. Это может произойти, если начальное ожидание того, что другой игрок класса M достаточно большое. Это продемонстрировано на рис. 4.10, случай (1).



(1) Игра начинается из состояния $s = s_1$.

(2) Игра начинается из состояния $s = s_2$.

Рис. 4.10: Равновесия.

Область, закрашенная желтым цветом, — область параметров α_1 и β_1 , при которых набор стратегий $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ образует РМРВЕ. Голубая область — область α_1 и β_1 , для которых набор стратегий $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ образует РМРВЕ. Область над синей прямой — область параметров, для которых ожидаемый выигрыш игрока класса P в равновесии $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ больше выигрыша в равновесии $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$. Область над красной прямой — область параметров, для которых ожидаемый выигрыш игрока класса M в равновесии $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ больше выигрыша в равновесии $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$. Область параметров, для которых согласно теореме 4.8 набор стратегий $q_1^{s_2} = 1$ образует РМРВЕ в игре, начинающейся из состояния $s = s_2$, изображена на рис. 4.10, случай (2) (оранжевая область).

Из теоремы 4.9 следует, что выигрыши игроков классов P и M в любом равновесии — линейные возрастающие функции вероятности α_1 .

Обучающая фаза непредписанной продолжительности.

Рассмотрим случай, когда продолжительность обучающей фазы не ограничена. Если игра начинается из состояния $s = s_1$, условия теоремы 4.10 образуют систему

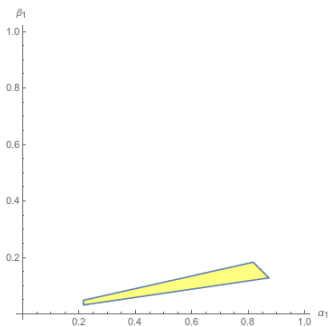
$$\left\{ \begin{array}{l} 32.2162\alpha_1 + 20.1154\beta_1 \geq 2, \\ 11.4673\alpha_1 + 6.70729\beta_1 \geq 2, \\ \frac{\beta_1}{\alpha_1} \leq \min \{2.08243, 0.223938\}, \\ \frac{\beta_1}{\alpha_1} \geq 0.145989, \end{array} \right.$$

при выполнении которой набор стратегий $q_1^{s_1} = 1, r_1^{s_1} = 1$ в период 1 и $q_2^{s_2} = 1$ в период 2 образуют РМРВЕ. Область параметров (α_1, β_1) , для которой система

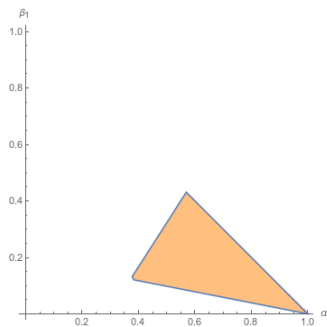
выполнена (желтая область) изображена на рис. 4.11, случай (1). можем отметить эффект «имитации терпеливости», который не позволяет достичь равновесия с полукооперативным и кооперативным равновесиями для низких значений отношения β_1/α_1 . Чем меньше это отношение, тем больше ожидание, что оппонент — игрок класса P и тем больше выгоды у игрока класса M «имитировать терпеливость».

Если игра начинается из состояния $s = s_2$, условия теоремы 4.11 записаны в системе

$$\begin{cases} \frac{1 - \alpha_1 - \beta_1}{\beta_1} \leq 4.11765, \\ 23.8919\alpha_1 - 15.3243\beta_1 \geq 7. \end{cases}$$



(1) Игра начинается из состояния $s = s_1$.



(2) Игра начинается из состояния $s = s_2$.

Рис. 4.11: Равновесия.

Область параметров, для которых набор стратегий $q_1^{s_2} = 1$ и $r_t^{s_1} = 1$ для любого $t > 1$ в обучающей фазе согласно теореме 4.11 является РМРВЕ в игре с начальным состоянием $s = s_2$, изображена на рис. 4.11, случай (2) (оранжевая область).

Заключения.

Когда обучающая фаза игры имеет продолжительность в один период, то игрок, который имеет выгоду отклониться от кооперации, имеет повод «имитировать терпеливость» в определенном состоянии. Такое поведение игрока ожидаемо, и вероятность такого поведения увеличивается вместе с увеличением ожидания того, что оппонент является терпеливым игроком. В случае, когда продолжительность обучающей фазы не фиксирована, эффект «имитации терпеливости» также наблюдается. Учитывая существование множества равновесий, в работе рассматриваются только равновесия в чистых стратегиях и те,

при которых кооперативная фаза игры наступает как можно быстрее.

Интересным продолжением работы может быть использование других стратегий при нахождении равновесий. В работе рассматривались стратегии наказания в кооперативной фазе игры. Этот класс стратегий используется в народных теорем при доказательстве существования абсолютного равновесия по Нэшу с кооперативными выигрышами. Этот класс стратегий естественно использовать в случае неполной информации о дисконтирующем факторе оппонента [130]. Можно также построить равновесия, используя другие стратегии наказания, так называемые “tit-for-tat” стратегии [70], согласно которым в каждый момент времени игрок выбирает стратегию, которую выбирал его оппонент на предыдущем шаге. Набор таких стратегий не будет абсолютным равновесием. Также можно построить равновесие в стратегиях наказания с ограниченным числом периодов наказания отклонившегося игрока.

Заключение

Модели конфликтно-управляемых систем часто нуждаются в учете случайных факторов, влияющих на оптимальное поведение участников системы и на их выигрыши. Удобным инструментом при моделировании такого рода систем являются стохастические игры. В работе предложен метод построения кооперативного варианта игры в случае, когда игровой процесс развивается на конечном графе (Глава 1), а также, когда игра имеет бесконечную продолжительность (Глава 2). Впервые понятие стохастической игры двух лиц бесконечной продолжительности введено Л. Шепли в 1953 году [210], что дало толчок к развитию этого раздела теории игр. В диссертационной работе акцент делается на построение кооперативной игры, в случае, когда изначально задана некооперативная стохастическая игра, а также на исследование свойств кооперативных решений, которые игроки реализуют в игре при кооперативном поведении. Основные принципы устойчивой кооперации для динамических игр сформулированы Л. А. Петросяном в цикле работ (из ранних работ см. [35, 43]). Основными принципами можно считать динамическую устойчивость и сильную динамическую устойчивость, стратегическую устойчивость, а также условие защиты от иррационального поведения. В диссертационной работе условие динамической устойчивости кооперативного решения формулируется как условие позиционной состоятельности, чтобы подчеркнуть значимость «позиции» или вершины графа, на котором происходит игра, а также состояние этой вершины.

Глава 3 посвящена изучению свойств кооперативных решений в динамических играх, разыгрываемых на деревьях событий. В этом случае на стохастический процесс не влияют стратегии игроков. Этот класс игр широко используется на практике. Деревья событий позволяют построить возможные сценарии, которые могут реализоваться, начиная с основного события, и используются для

определения и анализа последовательности событий, которые возможно произойдут. Для этого класса игр также получены условия, при которых выполняются свойства позиционной состоятельности, стратегической устойчивости и существования приближенного абсолютного равновесия по Нэшу.

Глава 4 содержит прикладные модели, построенные с использованием стохастических игр, а именно: три модели передачи данных в беспроводных сетях различных топологий в том числе с неполной информацией, модель определения коалиционной структуры, устойчивой в динамике, которая находится в результате решения специальным образом построенной стохастической игры, и, наконец, модель двухфазной игры с неполной информацией о дисконтирующих факторах игроков, для которой найдено оптимальное поведение игроков, позволяющих гарантировать кооперацию в игре.

Основные результаты работы:

1. Предложен метод построения кооперативной стохастической игры по заданной изначально некооперативной игре в случае конечной продолжительности в форме α -характеристической функции и метод регуляризации кооперативных решений для данного вида стохастических игр на основе построения процедуры распределения дележа с целью получения позиционно состоятельных решений. Для случая неотрицательных функций выигрыша игроков предложен способ построение «новой» процедуры распределения дележа с неотрицательными компонентами и получены условия позиционной состоятельности решений для данного класса игр. Для кооперативных стохастических игр конечной продолжительности сформулирован принцип сильной позиционной состоятельности дележей из s -ядра и получено достаточное условие его выполнения.
2. Предложен метод построения кооперативной стохастической игры по заданной изначально некооперативной стохастической игре в случае бесконечной продолжительности и конечного множества состояний игры. При построении характеристической функции использовался минимаксный подход. Сформулированы принципы устойчивой кооперации для данного вида стохастических игр. Для построения позиционно состоятельного коопера-

тивного решения предложен метод регуляризации игры с использованием процедуры распределения дележа. Получены условия выполнения принципа стратегической устойчивости кооперативного решения в стохастических играх такого вида, а именно, доказано существование ситуации равновесия по Нэшу с выигрышами, равными компонентам дележа. Также найдены условия существования абсолютного равновесия по Нэшу и сильного трансферабельного равновесия с кооперативными выигрышами. Сформулирован принцип защиты кооперативного решения от иррационального поведения для стохастических игр бесконечной продолжительности и конечного множества состояний и получены достаточные условия его выполнения.

3. Предложен метод построения позиционно состоятельного s -ядра на основе процедуры распределения дележа в кооперативных стохастических играх, разыгрываемых на деревьях событий. Получены условия существования абсолютного ε -равновесия в стохастических играх такого вида с выигрышами игроков, равными выигрышам при реализации кооперативного решения и условия существования абсолютного ε -равновесия в регуляризованной стохастической игре, заданной на дереве событий, с выигрышами игроков, равными соответствующим компонентам дележа, на основе которого была проведена регуляризация игры. Предложен метод построения позиционно состоятельного вектора Шепли в стохастических играх случайной продолжительности, разыгрываемых на деревьях событий. Все теоретические результаты проиллюстрированы на одной модели охраны окружающей среды.
4. Исследованы стохастические игры, разыгрываемые на бинарных деревьях событий с линейным уравнением состояния и идентичными игроками. В этом случае найдены равновесие по Нэшу и кооперативное решение или набор стратегий, максимизирующий суммарный ожидаемый выигрыш игроков. Эти наборы стратегий получены в явном виде. Получена формула для вычисления цены анархии для стохастических игр данного вида.
5. Представлены теоретико-игровые модели передачи данных в беспроводных

сетях нескольких конфигураций, в том числе, одна модель — для случая полной и неполной информации. В качестве инструмента моделирования предложено использовать стохастические игры. Для трех моделей найдены равновесия по Нэшу и кооперативные решения, вычислена так называемая стоимость отказа от кооперации, сделаны выводы о необходимости координации действий участников сети. Для одной модели передачи данных найден дележ суммарного выигрыша при реализации кооперативного решения, проведена его регуляризация для удовлетворения принципа позиционной состоятельности, а также выполнена проверка принципа стратегической устойчивости.

6. Построена модель нахождения устойчивых в динамике (d -устойчивых) коалиционных структур с помощью решения специальным образом определенной стохастической игры. Проведено численное моделирование решения игры трех лиц, найдены все d -устойчивые коалиционные структуры.
7. Построена модель двухфазной стохастической байесовской игры двух лиц типа «Дилемма заключенного», в которой имеется два состояния. В случае полной информации о дисконтирующем факторе другого игрока найдены ситуации абсолютного равновесия, в которых игроки 1) кооперируются в обоих состояниях, 2) кооперируются только в одном состоянии, 3) никогда не кооперируются. В случае отсутствия полной информации о дисконтирующем факторе другого игрока строится двухфазная игра, в первой фазе которой обновляется информация о типе игрока, а во второй — реализуются одно из равновесий. В работе получены условия существования байесовских равновесий в случае обучающей фазы продолжительностью в один и более периодов, приводящие к полной или частичной кооперации игроков во второй фазе игры. Проведен сравнительный анализ выигрышей игроков в этих равновесиях.

Литература

- [1] Баранова Е. М. Многошаговые стохастические игры со случайным временем окончания // Современные методы теории краевых задач. Материалы Воронежской весенней математической школы «Понтрягинские чтения – XV». г. Воронеж, 2004, стр. 22–23.
- [2] Баранова Е. М. Условие Д. В. К. Янга для стохастических кооперативных игр // Тезисы докладов Международного рабочего совещания «Задачи оптимальной остановки и стохастического управления», г. Петрозаводск, 2005, стр. 8.
- [3] Беллман Р. Динамическое программирование. М.: Изд-во иностр. лит-ры, 1960, 400 с.
- [4] Берж К. Теория графов и её применения. М.: Изд-во иностр. лит-ры, 1962, 320 с.
- [5] Болтянский В. Г., Гамкредидзе Р. В., Понтрягин Л. С. К теории оптимальных процессов // Доклады академии наук, 1956, Т. 110, № 1, стр. 7–10.
- [6] Буре В. М., Парилина Е. М. Стохастические модели передачи данных в сетях с различными топологиями // Управление большими системами, 2017. Выпуск 68. стр. 6–29.
- [7] Буре В. М., Парилина Е. М. Теория вероятностей и математическая статистика. СПб.: Изд-во «Лань», 2013.

- [8] Буре В. М., Парилина Е. М. Игра «Множественный доступ» с неполной информацией // Математическая теория игр и ее приложения, 2017. Т. 9. Вып. 4, стр. 3–17.
- [9] Васин А. А., Морозов В. В. Теория игр и модели математической экономики. М.: МАКС Пресс, 2005.
- [10] Вилкас Э. Й. Оптимальность в играх и решениях. М.: Наука, 1990, 256 с.
- [11] Воробьев Н. Н. Устойчивые ситуации в коалиционных играх // Доклады АН СССР, 1960, Т. 131, с. 493–495.
- [12] Воробьев Н. Н. Коалиционные игры // Теория вероятности и её применение, 1967, Т. 12, вып. 2, с. 289–306.
- [13] Воробьев Н. Н. Теория игр для экономистов-кибернетиков. М.: Наука, 1985, 272 с.
- [14] Громова Е. В., Петросян Л. А. Об одном способе построения характеристической функции в кооперативных дифференциальных играх // Математическая теория игр и её приложения, 2015, т. 7, в. 4. С. 19–39.
- [15] Громова Е. В., Петросян Л. А. Сильно динамически устойчивое кооперативное решение в одной дифференциальной игре управления вредными выбросами // Управление большими системами, 2015, т. 55, с. 140–159.
- [16] Захаров В. В. О регуляризации и динамической устойчивости решений иерархических дифференциальных игр // Вестник Ленинградского университета, 1988, сер. 1, вып. 2, № 8, 8 стр.
- [17] Захаров В. В., Петросян Л. А. Математические модели в экологии. — СПб.: Изд-во СПбГУ, 1997.
- [18] Захаров В. В., Щегряев А. Н. Устойчивая кооперация в динамических задачах маршрутизации транспорта // МТИиП, 2012, 4 (2), стр. 39–56.

- [19] Колокольцов В. Н., Малафеев О. А. Математическое моделирование многоагентных систем конкуренции и кооперации. Теория игр для всех. — СПб.: Изд-во «Лань», втор. изд., 2012, 624 с.
- [20] Корниенко Е. А. Решения кооперативных динамических игр. Диссертация на соиск. уч. ст. канд. физ.-мат. н.: 01.01.09 // СПб.: С.-Петербург. гос. ун-т, 2003.
- [21] Кузютин Д. В. Динамическая устойчивость решений для некоторых классов игр с неполной информацией. Диссертация на соиск. уч. ст. канд. физ.-мат. н.: 01.01.09 // СПб.: С.-Петербург. гос. ун-т, 1993.
- [22] Мазалов В. В. Математическая теория игр и приложения. — СПб.: Санкт-Петербург-Москва-Краснодар, Изд-во «Лань», втор. изд., 2016, 446 с.
- [23] Малафеев О. А. О существовании ситуации ε -равновесия в динамических играх с зависимыми движениями. Ж. вычисл. матем. и матем. физ., 1974, 14:1, стр. 88–98.
- [24] Малафеев О. А. О существовании ситуаций ε -равновесия в смешанных стратегиях в бескоалиционных дифференциальных играх // Дифференц. уравнения, 1979, т. 15, № 4, стр. 609–613.
- [25] Марковкин М. В. Линейно-квадратичные кооперативные дифференциальные игры. Диссертация на соиск. уч. ст. канд. физ.-мат. н.: 01.01.09 // СПб.: С.-Петербург. гос. ун-т, 2006.
- [26] Наумова Н. И. Вектор Шепли и его обобщения. Учебное пособие. — СПб.: Изд-во ВВМ, 2017.
- [27] Парилина Е. М. Кооперативная стохастическая игра с конечным числом игровых элементов // Тезисы докладов международного конгресса «Нелинейный динамический анализ-2007» — СПб.: Изд-во СПбГУ, 2007, стр. 337.
- [28] Парилина Е. М. Кооперативная игра передачи данных в беспроводной сети // Математическая теория игр и ее приложения, 2009, т. 1, № 4, с. 93–110.

- [29] Парилина Е. М. Устойчивая кооперация в стохастических играх // Математическая теория игр и ее приложения, 2010, т. 2, № 3, с. 21–40.
- [30] Парилина Е. М. Стохастические игры в сетевых задачах // Сборник тезисов конф. «Устойчивость и процессы управления» г. СПб, 2010, стр. 170–171.
- [31] Парилина Е. М. Кооперативная игра передачи данных в беспроводной сети // Управление большими системами, 2010, т. 31-1, с. 191–209.
- [32] Парилина Е. М. Кооперативные стохастические игры, учитывающие отношение игроков к риску // Материалы международной конф. «Математика, экономика, менеджмент: 100 лет со дня рождения Л.В. Канторовича», г. Санкт-Петербург, 2012, с. 114.
- [33] Парилина Е. М. Стратегическая устойчивость одноточечных принципов оптимальности в кооперативных стохастических играх // Математическая теория игр и ее приложения, 2014, т. 6, № 1, с. 56–72.
- [34] Парилина Е. М., Петросян Л. А. Сильно позиционно состоятельное с-ядро в стохастических играх // Математическая теория игр и ее приложения, 2017, т. 9, в. 2, с. 39–61.
- [35] Петросян Л. А. Устойчивость решений в дифференциальных играх со многими участниками // Вестник Ленинградского университета. Серия 1: математика, механика, астрономия. 1977. Вып. 19. С. 46-52.
- [36] Петросян Л. А. Построение сильно-динамически устойчивых решений в кооперативных дифференциальных играх // Вестник Ленинградского университета. Серия 1: математика, механика, астрономия. 1992. № 2. С. 33-38.
- [37] Петросян Л. А. Сильно динамически устойчивые принципы оптимальности в многокритериальных задачах оптимального управления // Техническая кибернетика. 1993. № 1. С. 169-174.

- [38] Петросян Л. А. Полукооперативные игры // Вестник Санкт-Петербургского Университета. Серия 1: Математика, механика, астрономия, 1998. № 2, с. 57–63.
- [39] Петросян Л. А., Баранова Е. М. Стохастические игры со случайной продолжительностью // Труды XXXIV научной конференции аспирантов и студентов «Процессы управления и устойчивость», г. СПб, 2003. С. 456–462.
- [40] Петросян Л. А., Баранова Е. М. Кооперативные стохастические игры // Тезисы докладов Международного семинара «Теория управления и теория обобщенных решений уравнений Гамильтона-Якоби», г. Екатеринбург, 2005, стр. 33–35.
- [41] Петросян Л. А., Баранова Е. М. Кооперативные стохастические игры в стационарных стратегиях // Сборник трудов международной конференции «Устойчивость и процессы управления», г. СПб, 2005. Т. 1, стр. 495–503.
- [42] Петросян Л. А., Баранова Е. М., Шевкопляс Е. В. Многошаговые кооперативные игры со случайной продолжительностью // Труды института математики и механики УрО РАН. 2004. Т. 10. № 2. С. 116–130.
- [43] Петросян Л. А., Данилов Н. А. Устойчивые решения неантагонистических дифференциальных игр с транзитивными выигрышами // Вестник ЛГУ. 1979. №1. С. 46–54.
- [44] Петросян Л. А., Зенкевич Н. А. Принципы устойчивой кооперации // Математическая теория игр и ее приложения, 2009. Т. 1, № 1, с. 106–123.
- [45] Петросян Л. А., Зенкевич Н. А., Шевкопляс Е. В. Теория игр. Санкт-Петербург: БХВ – Петербург, 2012. 480 с.
- [46] Петросян Л. А., Кузютин Д. В. Игры в развернутой форме: оптимальность и устойчивость. — СПб: Изд-во С.-Петербургского университета. 2000.

- [47] Петросян Л. А., Мурзов Н. В. Теоретико-игровые задачи механики // Литовский математический сборник, 1966, Т. 6, с. 423–432.
- [48] Петросян Л. А., Седаков А. А., Сюрин А. Н. Многошаговые игры с коалиционной структурой // Вестник Санкт-Петербургского университета. Серия 10: Прикладная математика, информатика, процессы управления, 2006, № 4, с. 97–110.
- [49] Петросян Л. А., Шевкопляс Е. В. Кооперативные дифференциальные игры со случайной продолжительностью // Вестник Санкт-Петербургского Университета. Серия 1: Математика, механика, астрономия, 2000, № 4, с. 14–18.
- [50] Петросян О. Л., Громова Е. В., Погожев С. В. О сильно динамически устойчивом подмножестве C -ядра в кооперативных дифференциальных играх с предписанной продолжительностью // Математическая теория игр и ее приложения. 2016. Т. 8. № 4. С. 79-106.
- [51] Печерский С. Л., Яновская Е. Б. Кооперативные игры: решения и аксиомы. СПб.: Изд-во Европ. унив-та в С.-Петербурге, 2004, 459 с.
- [52] Седаков А. А. О сильной динамической устойчивости c -ядра // Математическая теория игр и ее приложения. 2015. Т. 7. № 2. С. 69-84.
- [53] Тур А. В. Линейно-квадратичные стохастические дискретные игры со случайной продолжительностью // Математическая теория игр и ее приложения. 2014. Т. 6. № 3. С. 76–92.
- [54] Тур А. В. Кооперация в дискретных линейно-квадратичных играх. Диссертация на соиск. уч. ст. канд. физ.-мат. н.: 01.01.09 // СПб.: С.-Петерб. гос. ун-т, 2015.
- [55] Шевкопляс Е. В. Устойчивая кооперация в дифференциальных играх со случайной продолжительностью // Математическая теория игр и ее приложения, 2010, Т. 2, № 3, с. 162–190.

- [56] Чистяков С. В. Программные итерации и универсальные ε -оптимальные стратегии в позиционной дифференциальной игре, Докл. АН СССР, 1991, 319:6, стр. 1333–1335.
- [57] Чиркова Ю. В. Цена анархии в игре баланса загрузки системы обслуживания с тремя машинами // Математическая Теория Игр и ее Приложения, 2014, т. 6, в. 4, стр. 85–96.
- [58] Algaba, E., Bilbao, J.M., van den Brink, R., Jiménez-Losada, A. Axiomatizations of the Shapley value for cooperative games on antimatroids // Mathematical Methods of Operations Research, 2003, 57, 49–65.
- [59] Altman E., Avrachenkov K., Garnaev A. A jamming game in wireless networks with transmission cost // Lecture Notes in Computer Science, 2007, 4465 LNCS, pp. 1-12.
- [60] Altman E., Avrachenkov K., Garnaev A. Jamming in wireless networks under uncertainty // Mobile Networks and Applications, 2011, 16 (2), pp. 246–254.
- [61] Altman E., Avrachenkov K., Miller G., Prabhu B. Discrete power control: Cooperative and non-cooperative optimization, Proceedings - IEEE INFOCOM, 2007, art. no. 4215595, pp. 37–45.
- [62] E. Altman, D. Barman, R. El Azouzi and T. Jimenez, “A game theoretic approach for delay minimization in slotted ALOHA,” 2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577), Vol.7, pp. 3999–4003, 2004.
- [63] Amir R. Stochastic games in economics: The lattice-theoretic approach. In A. Neyman and S. Sorin (eds.) // Stochastic Games and Applications, NATO Science Series C, Mathematical and Physical Sciences, 2003, vol. 570, pp. 443–453.
- [64] Anshelevich E., Dasgupta A., Kleinberg J., Tardos, E., Wexler, T. and Roughgarden, T. The price of stability for network design with fair cost

allocation // In 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2004, pp. 59–73.

- [65] Avrachenkov K., Cottatellucci L., Maggi L. Cooperative Markov decision processes: Time consistency, greedy players satisfaction, and cooperation maintenance // International Journal of Game Theory, 2013, 42 (1), pp. 239–262.
- [66] Aumann R. J. Acceptable points in general cooperative n-person games. // In: Tucker, A. W., Luce, R. D. (eds.) Contributions to the Theory of Games IV. 1959. Princeton: Princeton University Press.
- [67] Aumann R. J. Subjectivity and correlation in randomized strategies // Journal of Mathematical Economics, 1974, Vol. 1, pp. 67–96.
- [68] Aumann R. J., Dreze J. H. Cooperative games with coalition structure // International Journal of Game Theory, 1974, Vol. 3, pp. 217–237.
- [69] Aumann R. J., Peleg B., Von Neumann-Morgenstern Solutions to Cooperative Games without Side Payments // Bulletin of the American Mathematical Society, 1960, vol. 66, pp. 173–179.
- [70] Axelrod R., Hamilton W. D. The evolution of cooperation // Science, 1981, vol. 211, pp. 1390–1396.
- [71] Baranova E. M. The Condition for Keeping Cooperation in Stochastic Cooperative Games // Proceedings of the Russian-Finnish Graduate School Seminar “Dynamic Games and Multicriteria Optimization”, edited by V.V. Mazalov, Petrozavodsk, 2006, p. 54–58.
- [72] Baranova E. M., Petrosjan L. A. Cooperative Stochastic Games in Stationary Strategies // Game Theory and Applications, 2006, Vol. 11, pp. 1–7.
- [73] Basar T., Zhu Q. Price of Anarchy, Information, and Cooperation in Differential Games // Dynamic Games and Applications, 2011, vol. 1(1), pp. 50–73.

- [74] Bazenkov N. I. Double best response dynamics in topology formation game for ad hoc networks // Automation and Remote Control. 2015. 76 (2). Pp. 323–335.
- [75] Bellman R. Dynamic programming. Princeton: Princeton University Press, New Jersey, 1957.
- [76] Bogomolnaia A., Jackson M. O. The stability of hedonic coalition structures // Games and Economic Behavior, 2002, 38 (2), pp. 201–230.
- [77] Breton M., Haurie A., Filar J. A. On the computation of equilibria in discounted stochastic dynamic games // Journal of Economic Dynamics and Control, 1986, Vol. 10, Is. 1–2, pp. 33–36.
- [78] Breton M., L'Écuyer P. Approximate solutions to continuous stochastic games. In: Hämäläinen R.P., Ehtamo H.K. (eds) Differential Games — Developments in Modelling and Computation. Lecture Notes in Control and Information Sciences, 1991, vol. 156, pp. 257-264. Springer, Berlin, Heidelberg.
- [79] Bure V., Parilina E. Stochastic models of Data Transmission with Specific Network Topologies // Collected abstracts of papers presented on the 11th International Conference Game Theory and Management Ed. L.A. Petrosyan and N.A. Zenkevich.- SPb.: Saint Petersburg State University, 2017, p. 125.
- [80] Buttyan L., Hubaux J.-P. Security and Cooperation in Wireless Networks: Thwarting Malicious and Selfish Behavior in the Age of Ubiquitous Computing // Cambridge University Press New York. - 2007. - NY, USA.
- [81] Chander P., Tulkens H. The Core of an Economy with Multilateral Environmental Externalities // International Journal of Game Theory, 1997, 23, 379–401.
- [82] Chirkova Yu. V. Price of anarchy in machine load balancing game // Automation and Remote Control, 2015, Vol. 76, Is. 10. Pp. 1849–1864.

- [83] Chistyakov S., Petrosyan L. Strong Strategic Support of Cooperative Solutions in Differential Games // Contributions to game theory and management, 2011, Vol. IV, p. 105–111.
- [84] Cole R., Dodis Y., Roughgarden T. How much can taxes help selfish routing? // Journal of Computer and System Sciences, 2006, 72, pp. 444–467.
- [85] Correa J. R., Schulz A. S., Stier Moses N. E. Selfish Routing in Capacitated Networks // Mathematics of Operations Research, 2004, 29(4), pp. 961–976.
- [86] Correa J. R., Schulz A. S., Stier Moses N. E. Fast, Fair and Efficient Flows in Networks // Operations Research, 2007, 55(2), pp. 215–225.
- [87] Driessen T., Muto S., Nakayama M. A cooperative Game of Information Trading: the Core, the Nucleolus and the Kernel // ZOR Methods and Models of Operations Research, 1992, vol. 36, no 1, pp 55–72.
- [88] Dutta P., 1995. A folk theorem for stochastic games. // Journal of Economic Theory, 1995, Vol. 66, pp. 1–32.
- [89] Filar J., Petrosjan L. Dynamic Cooperative Games // International Game Theory Review, 2000, 2:1, 47–65.
- [90] Filar J., Vrieze K. Competitive Markov Decision Processes. N.Y.: Springer-Verlag New York, 1997.
- [91] Fink A. M. Equilibrium in a stochastic n -person game // Journal of Science of the Hiroshima University, 1964, A-I 28, pp. 89–93.
- [92] Flesh J., Kuipers J., Mashiah-Yaakovi A., Shmaya E., Shoenmakers, G., Solan, E., Vrieze, K. Nonexistence of subgame-perfect-equilibrium in perfect information games with infinite horizon // International Journal of Game Theory, 2014, vol. 43, pp. 945–951.
- [93] Flesh J., Predtethinski A. On refinements of subgame perfect ϵ -equilibrium // International Journal of Game Theory, 2015, DOI 10.1007/s00182-015-0468-8.

- [94] Fudenberg D., Tirole J. 1991. Perfect Bayesian equilibrium and sequential equilibrium // *Journal of Economic Theory*, vol. 53, no. 2, pp. 236–260.
- [95] Fudenberg D., Yamamoto Y. Repeated games where the payoffs and monitoring structure are unknown // *Econometrica*, 2010, vol. 78, pp. 1673–1710.
- [96] Gene T., Reynolds S. S., Sen S. Dynamic Oligopolistic Games Under Uncertainty: A Stochastic Programming Approach. // *Journal of Economic Dynamics & Control*, 2007, 31, 55–80.
- [97] Gene T., Sen S. An Analysis of Capacity and Price Trajectories for the Ontario Electricity Market Using Dynamic Nash Equilibrium Under Uncertainty. // *Energy Economics*, 2008, vol. 30, p. 173–191.
- [98] Germain M. Toint P., Tulkens H., de Zeeuw A. Transfers to Sustain Dynamic Core-Theoretic Cooperation in International Stock Pollutant Control // *Journal of Economic Dynamics & Control*, 2003, Vol. 28 (1), pp. 79–99.
- [99] Gillies D. B. Solutions to general non-zero-sum games. In Tucker, A. W.; Luce, R. D. *Contributions to the Theory of Games IV*. (Annals of Mathematics Studies 40). Princeton: Princeton University Press. 1959, pp. 47–85.
- [100] González-Díaz, J., Sánchez-Rodríguez, E. Towards an axiomatization of the core-center // *European Journal of Operational Research*, Vol. 195, Is. 2, 2009, Pp. 449–459.
- [101] Harrington Jr. J. E., Zhao W. Signaling and tacit collusion in an infinitely repeated Prisoners' Dilemma // *Mathematical Social Sciences*, 2012, 64, p. 277–289.
- [102] Hart S., Kurz M. Endogenous formation of coalitions // *Econometrica*, 1983, 51 (4), pp. 1047–1064.
- [103] Haurie A., Krawczyk J. B., Zaccour G. *Games and Dynamic Games*, Singapore: Scientific World, 2012.

- [104] Haurie A., Zaccour G., Smeers Y. (1990). Stochastic Equilibrium Programming for Dynamic Oligopolistic Markets // Journal of Optimization Theory and Applications 66(2), 243–253.
- [105] P. Jean-Jacques Herings and R. Peeters, “Homotopy methods to compute equilibria in game theory,” *Econ Theory*, vol. 42, pp. 119–156, 2010.
- [106] Hörner J., Rosenberg D., Solan E., Vieille N. On a Markov game with one-sided information // *Operations Research*, 2010, vol. 58 (4 PART 2), pp. 1107–1115.
- [107] Hörner J., Sugaya T., Takahashi S., Vieille N. Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \rightarrow 1$ and a Folk Theorem // *Econometrica*, 2011, vol. 79, pp. 1277–1318.
- [108] Hörner J., Takahashi S., Vieille N. Truthful Equilibria in Dynamic Bayesian Games // *Econometrica*, 2015, vol. 79, pp. 1277–1318.
- [109] Jørgensen S., Martín-Herrán G., Zaccour G. Dynamic Games in the Economics and Management of Pollution // *Environmental Modeling and Assessment*, 2010, Vol. 15, pp. 433–467.
- [110] Jászkiewicz A., Nowak A. S. Stationary almost markov perfect equilibria in discounted stochastic games // *Mathematics of Operations Research*, 2016, vol. 41 (2), pp. 430–441.
- [111] Kalai E. Bounded rationality and strategic complexity in repeated games // *Game theory and applications*, Pro. Int. Conf., Columbus/OH (USA), p. 131–157.
- [112] Kamhoua C., Pissinou N. Mitigating selfish misbehavior in multi-hop networks using stochastic game theory // In *Proceedings of the 2010 IEEE 35th Conference on Local Computer Networks (LCN '10)*. IEEE Computer Society, Washington, DC, USA. - 2010. - Pp. 232-235.
- [113] Kohlberg E. On the Nucleolus of a Characteristic Function Game // *SIAM Journal of Applied Mathematics*, 1971, vol. 20 , pp. 62–66.

- [114] Kohlberg E. The Nucleolus as a Solution to a Minimization Problem // SIAM Journal of Applied Mathematics, 1972, vol. 23, no. 1, pp. 34–39.
- [115] Kohlberg E., Neyman A., 2015. The cooperative solution of stochastic games. Harvard Business School. Working Paper, No. 15-071, March 2015.
- [116] Konishi H., Ray D. Coalition formation as a dynamic process // Journal of Economic Theory, 2003, 110, p. 1–41.
- [117] Koutsoupias E., Papadimitriou C. H. Worst-case equilibria // In Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS), 1999, pp. 404–413.
- [118] Kozlovskaya N., Zenkevich N. Stable cooperation under environmental constraints, International Game Theory Review, 2010, 12 (4), pp. 453–470.
- [119] Kreps V., Domansky V. Repeated games with asymmetric information modeling financial markets with two risky assets // RAIRO Recherche Operationnelle, 2013, vol. 47 (3), pp. 251–272.
- [120] Kuhn H. W. Extensive Games // Proceedings of National Academy of Sciences of the USA, 1950, vol. 36, pp. 570–576.
- [121] Kuhn H. W. Extensive Games and the Problem of Information // Annals of Mathematics Studies, 1953, no. 28, pp. 193–216.
- [122] Lehrer E., Scarsini M. On the Core of Dynamic Cooperative Games // Dynamic games and applications, 2013, vol. 3, p. 359–373.
- [123] Lemke C. E., Howson J. T. Equilibrium Points of Bimatrix Games // J Soc Indust Appl Math, vol. 12, pp. 413–423, 1964.
- [124] Liu Y., Garnaev A., Trappe W. Maintaining throughput network connectivity in ad hoc networks // Proceedings of ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing. - 2016. Art. no. 7472905. Pp. 6380-6384.

- [125] Long N. V. Applications of dynamic games to global and transboundary environmental issues: a review of the literature. // *Strategic Behavior and the Environment*, 2012, 2(1), pp. 1-59.
- [126] WOLFRAM MATHEMATICA // <https://www.wolfram.com/mathematica/>
- [127] Maggi L., Avrachenkov K., Cottatellucci L. Stochastic games for cooperative network routing and epidemic spread, 2011, IEEE International Conference on Communications, art. no. 5963532.
- [128] Mailath G. J., Postlewaite A., Samuelson L. Contemporaneous perfect epsilon-equilibria // *Games and Economic Behavior*, 2005, Vol. 53, pp. 126–140.
- [129] Mailath G. J., Samuelson L. *Repeated Games and Reputations: Long-Run Relationships*. Oxford: Oxford University Press, 2006.
- [130] Maor C., Solan E. Cooperation under incomplete information on the discount factors // *International Journal of Game Theory*, 2015, Vol. 44, pp. 321–346.
- [131] S. Marbán, P. van de Ven, P. Borm, and H. Hamers “ALOHA networks: a game-theoretic approach,” *Math Meth Oper Res*, vol. 78, is. 2, pp. 221–242, 2013.
- [132] Marín-Solano J., Shevkoplyas E. V. Non-constant discounting and differential games with random time horizon, *Automatica*, 2011, Vol. 47, Is. 12, p. 2626–2638.
- [133] Markovkin M. V., D. W. K. Yeung’s Condition for Linear Quadratic Differential Games, In: *Dynamic Games and Their Applications*, eds. L. A. Petrosyan, A. Y. Garnaev, St. Petersburg State University, St. Petersburg, 2006, 207–216.
- [134] Maschler M., Solan E., Zamir S. *Game Theory*. Cambridge University Press, 2013.
- [135] Maplesoft. <http://www.maplesoft.com/>
- [136] Mathworks, 2017, <https://se.mathworks.com>

- [137] Mazalov V. V., Rettieva A. N. Fish wars and cooperation maintenance // *Ecological Modelling*, 2010, vol. 221 (12), pp. 1545–1553.
- [138] Mazalov V. V., Rettieva A. N. Fish wars with many players // *International Game Theory Review*, 2010, 12 (4), pp. 385–405.
- [139] Meinhardt H.I. Graphical Extensions of the Mathematica Package TU Games // <http://library.wolfram.com/infocenter/MathSource/5709/TuGamesView3D.pdf>.
- [140] Mertens J-F, Neyman A. Minimax Theorems for Undiscounted Stochastic Games // *Game Theory and Mathematical Economics*, 1981, p. 83–87.
- [141] Mertens J. F., Neyman A., Stochastic Games. // *International Journal of Game Theory*, 1981, Vol. 10, pp. 53–66.
- [142] Mertens J-F., Parthasarathy T. P. Equilibria for discounted stochastic games // In: Neyman A, Sorin S, eds., *Stochastic Games and Applications*, Kluwer Academic Publishers, 2003, pp. 131–172.
- [143] Mertens J. F., Zamir S. Formulation of Bayesian analysis for games with incomplete information, *International Journal of Game Theory*, 1985, Vol. 14, No. 1, pp. 1–29.
- [144] Montero M. On the nucleolus as a Power Index // *Homo Oeconomicus*, 2005, vol. 22(4), pp. 551–567.
- [145] Naumova N. Solidary Solutions to Games with Restricted Cooperation // *Contributions to Game Theory and Management*, vol. 6, 2013, pp. 316–337.
- [146] von Neumann J., Morgenstern O. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.
- [147] Neyman A. Existence of Optimal Strategies in Markov Games with Incomplete Information // *International Journal of Game Theory*, 2008, 37(4), p. 581–596.
- [148] Neyman A. Stochastic Games with Short-Stage Duration // *Dynamic Games and Applications*, 2013, Vol.3, p. 236–278.

- [149] Neyman A., Sorin, S. Stochastic Games and Applications. Dordrecht: Kluwer Academic Press, 2003.
- [150] Njilla L.Y., Pissinou N. Dynamics of data delivery in mobile ad-hoc networks: A bargaining game approach // Proceedings of 2015 IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2015 - art. no. 7208634. - 2015. - Pp. 98-103.
- [151] Novikov D.A. Games and networks // Automation and Remote Control. - 2014. - 75 (6). - Pp. 1145-1154.
- [152] Nowak A. S. Existence of equilibrium stationary strategies in discounted noncooperative stochastic games with uncountable state space // Journal of Optimization Theory and Applications, 1985, 45 (4), pp. 591–602.
- [153] Nowak A. S. Sensitive equilibria for ergodic stochastic games with countable state spaces // Mathematical Methods of Operations Research, 1999, vol. 50 (1), pp. 65–76.
- [154] Nowak A. S., Radzik T. A solidarity value for n-person transferable utility games // International Journal of Game Theory, 1994, 23 (1), pp. 43–48.
- [155] Owen G. Values of games with a priori unions // In: Henn, R., Moeschlin, O. (eds.) Essays in mathematical economics and game theory, Springer-Verlag, Berlin, 1977, p. 76–88.
- [156] Parilina E. M. Subgame Consistency of Optimality Principles in Cooperative Stochastic Games // Collected abstracts of papers presented on International Conference Game Theory and Management, SPb, 2007, p. 104.
- [157] Parilina E. M. Game-theoretic Approach to Resource Sharing Management // Proceedings of The Second International Conference on Game Theory and Applications — World Academic Union (World Academic Press), 2007, pp. 49-52.

- [158] Parilina E. M. Subgame Consistency of Shapley Value in Cooperative Data Transmission Game in Wireless Network // Contributions to Game Theory and Management, SPb, 2008, Vol. 1, p. 381–994.
- [159] Parilina E. Principles of Stable Cooperation in Stochastic Games // Collected abstracts of papers presented on GTM 2011, SPb, 2011., стр. 176–177.
- [160] Parilina E. M. Principles of Stable Cooperation in Stochastic Games // Contributions to game theory and management, Vol. 5, 2012. P. 243–256.
- [161] Parilina E. M. Cooperation in Dynamic Games with Stochastic Payoffs // Конструктивный негладкий анализ и смежные вопросы. Тезисы докладов международной конференции, Санкт-Петербург, 2012, p. 131–132.
- [162] Parilina E. M. Dynamic Scenarios of Data Transmission in Wireless Networks // Расширенные тезисы докладов международного семинара «Сетевые игры и менеджмент», Петрозаводск, 2012, стр. 51–53.
- [163] Parilina E. M. Stable cooperation in stochastic games // Automation and Remote Control. 2015. Vol. 76. Is. 6, Pp. 1111–1122.
- [164] Parilina E. Strategic Support of the Shapley Value in Stochastic Games // Contributions to Game Theory and Management, 2016. Vol. IX, p. 246–265.
- [165] Parilina E. A Survey on Cooperative Stochastic Games with Finite and Infinite Duration // Contributions to Game Theory and Management, Vol. 11, 2018, pp. 129–196.
- [166] Parilina E. M., Petrosyan L. A. Strongly Subgame-Consistent Core in Stochastic Games // Automation and Remote Control, Vol. 79, Is. 8, pp. 1515–1527.
- [167] Parilina E., Sedakov A. Stable coalition structure as solution of stochastic game // Collected abstracts of papers presented on 8th International Conference GTM — 2014. SPb: Graduate School of Management SpbU, 2014, p.140–141.

- [168] Parilina E., Sedakov A. Stochastic Approach for Determining Stable Coalition Structure // *International Game Theory Review*, 2015. Vol. 17, № 4, p. 155009-1–155009-22.
- [169] Parilina E., Sedakov A., Zaccour G. Price of Anarchy in a Linear-State Stochastic Dynamic Game // *European Journal of Operational Research*, 2017. Vol. 258, no. 2, p. 790–800.
- [170] Parilina E., Tampieri A. Stability and Cooperative Solution in Stochastic Games // *Theory and Decision*, <https://doi.org/10.1007/s11238-017-9619-7>.
- [171] Parilina E., Zaccour G. Node-Consistent Core for Games Played over Event Trees // *Automatica*. 2015. Vol. 53. P. 304–311.
- [172] Parilina E., Zaccour G. Approximated cooperative equilibria for games played over event trees // *Operations Research Letters*, 2015, Vol. 43 (5), p. 507–513.
- [173] Parilina E., Zaccour G. Strategic Support of Node-Consistent Cooperative Outcomes in Dynamic Games Played over Event Trees // *International Game Theory Review*, 2016. Vol. 18, № 2, art.no. 1640002, 16 pages.
- [174] Parilina E., Zaccour G. Node-Consistent Shapley Value for Games Played over Event Trees with Random Terminal Time // *Journal of Optimization Theory and Applications*, 2017. Volume 175, Issue 1, pp 236–254.
- [175] Patrone F., Dinar A., Moretti S. Application of Stochastic Cooperative Games in Water Resources // "Frontiers in Water Resource Economics Edited by R. Goetz and D. Berga (Kluwer Academic Publishers), Series: Natural Resource Management and Policy, Vol. 29:XXI + 275, 2006.
- [176] Peleg B., Sudhölter P. Introduction to the theory of cooperative games. Berlin, Heidelberg, New York: Springer Science + Business Media B. V. 328 p. (Theory and Decision Library Series C, Vol. 34).
- [177] Peski M. Repeated games with incomplete information and discounting // *Theoretical Economics*, 2014, Vol. 9, p. 651–694.

- [178] Petrosjan L. A. *Cooperative Stochastic Games* // Advances in Dynamic Games. Annals of the ISDG. Application to Economics, Engineering and Environmental Management, ed. by A. Haurie, S. Muto, L. A. Petrosjan, T.E.S. Raghavan. 2006. P. 139-146.
- [179] Petrosjan L. A., Baranova E. M. Cooperative Stochastic Games in Stationary Strategies // Proceedings of the Fifth International ISDG Workshop International Society of Dynamic Games, Segovia (Spain), 2005, p. 225–234.
- [180] Petrosjan L. A., Grauer L. V. Strong Nash equilibrium in multistage games // International Game Theory Review, 2002, Vol. 4, № 2, p. 255–264.
- [181] Petrosjan L., Mamkina S. Dynamic games with coalitional structures // International Game Theory Review, 2006, vol. 8 (2), pp. 295–307.
- [182] Petrosjan L., Zaccour G. Time-consistent Shapley value allocation of pollution cost reduction // Journal of Economic Dynamics & Control, 2003, 27, pp. 381–398.
- [183] Petrosjan L. A., Zenkevich N. A., Conditions for Sustainable Cooperation // Automation and Remote Control, 2015, Vol. 76(10), pp. 1894–1904.
- [184] Petrosjan L. A., Zenkevich N. A. Game Theory, Singapore-London-New York: World Scientific Publishing Co, 2016, 552 p.
- [185] Petrosyan L., Chistyakov S. Strategic support of Cooperative Solutions in 2-Person Differential Games with Dependent Motions // Contributions to Game Theory and Management, 2013. Vol. 6, p. 388–394.
- [186] Petrosyan L., Sedakov A. One-way flow two-stage network games // Вестник Санкт-Петербургского университета. Серия 10: прикладная математика, информатика, процессы управления, 2014, № 4, с. 72–81.
- [187] Petrosyan L., Sedakov A. Strategic support of cooperation in dynamic games on networks // "Stability and Control Processes" in Memory of V.I. Zubov (SCP), 2015 International Conference, 2015, pp. 256–260.

- [188] Petrosyan L., Sedakov A. The Subgame-Consistent Shapley Value for Dynamic Network Games with Shock // *Dynamic Games and Applications*, 2016, Vol. 6, No 4, pp. 520–537.
- [189] Pineau P.-O., Murto P. An Oligopolistic Investment Model of the Finnish Electricity Market // *Annals of Operations Research*, 2003, vol. 121, p. 123–148.
- [190] Pineau P.-O., Rasata H., Zaccour G. Impact of some Parameters on Investments in Oligopolistic Electricity Markets // *European Journal of Operational Research*, 2011, vol. 213 (1), p. 180–195.
- [191] Predtetchinski A. The strong sequential core for stationary cooperative games // *Games and Economic Behavior*, 2007, vol. 61, p. 50–66.
- [192] Radner R. Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives // *Journal of Economic Theory*, 1980, vol. 22, p. 136–154.
- [193] Reactor Safety Study. An Assessment of Accident Risks in U. S. Commercial Nuclear Power Plants. Executive Summary: Main Report. [PWR and BWR]. United States: 1975. Web. doi:10.2172/7134131.
- [194] Reddy P. V., Shevkoplyas E., Zaccour G. Time-Consistent Shapley Value for Games Played over Event Trees // *Automatica*, 2013, 49 (6), p. 1521–1527.
- [195] Reddy P. V., Zaccour G. A friendly computable characteristic function // *Mathematical Social Sciences*, 2016, Vol. 82, pp. 18–25.
- [196] Perakis G., Roels G. The price of anarchy in supply chains: Quantifying the efficiency of price-only contracts // *Management Science*, 2007, 53 (8), pp. 1249–1268.
- [197] Raghavan T. E. S. A stochastic game model of tax evasion. In: *Advances in dynamic games*, Birkhauser Boston, Boston, MA, 2006, pp. 397–420.

- [198] Raghavan T. E. S., Filar J. A. Algorithms for Stochastic Games — A Survey // ZOR — Methods and Models of Operations Research, vol. 35, pp. 437–472, 1991.
- [199] Rettieva A. N. Stable coalition structure in bioresource management problem // Ecological Modelling, 2012, 235-236, pp. 102–118.
- [200] Rosen J. B. Existence and Uniqueness of Equilibrium Points for Concave n -Person Games // Econometrica, 1965, Vol. 33 (3). pp. 520–534.
- [201] Roughgarden T. Selfish Routing and the Price of Anarchy, MIT Press, 2005.
- [202] Roughgarden T. The price of anarchy in games of incomplete information // ACM Transactions on Economics and Computation 3 (1), Article 6 (March 2015).
- [203] Roughgarden T., Schoppmann, F. Local smoothness and the price of anarchy in splittable congestion games // Journal of Economic Theory, 2015, 156, pp. 317–342.
- [204] Roughgarden T., Tardos, E. How Bad is Selfish Routing? // Journal of the ACM, 2002, 49(2), pp. 236–259.
- [205] Sagduyu Y. E., Ephremides A. A game-theoretic look at simple relay channel // Wireless Networks. - 2006. - Vol. 12, No. 5. - Pp. 545–560.
- [206] Schmeidler D. The Nucleolus of a Characteristic Function game // SIAM Journal of Applied Mathematics, 1969, vol. 17, pp. 1163–1170.
- [207] Schneider R. Convex bodies: the Brunn–Minkowski theory, Cambridge Univ. Press, 1993.
- [208] Selten R. Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetraägheit // Zeitschrift für die gesamte Staatswissenschaft, 1965, Vol. 121, pp. 301–24 and 667–89.
- [209] Selten R. Reexamination of the perfectness concept for equilibrium points in extensive games // International Journal of Game Theory, 1975, 4, pp. 25–55.

- [210] Shapley L.S. Stochastic Games // Proceedings of National Academy of Sciences of the USA, 1953, vol. 39, pp. 1095–1100.
- [211] Shapley L.S. A Value for n-person Games. In H.W. Kuhn, A.W. Tucker, eds. Contributions to the theory of games II, Princeton: Princeton Univ. Press, 1953, pp. 307–317.
- [212] Solan E. Discounted Stochastic Games // Mathematics of Operations Research, 1998, vol. 23, pp. 1010–1021.
- [213] Solan E. Stochastic Games, 2009, in Encyclopedia of Database Systems, Springer.
- [214] Solan E., Vieille N. Correlated Equilibrium in Stochastic Games // Games and Economic Behavior, 2002, vol. 38, pp. 362–399.
- [215] Solan E., Vieille N. Deterministic multi-player Dynkin games // Journal of Mathematical Economics, 2003, vol. 39, p. 911–929.
- [216] Solan E., Vieille N. Stochastic games // Proceedings of the National Academy of Sciences of the United States of America, 2015, Vol. 112 (45), pp. 13743–13746.
- [217] Takahashi M., Stochastic games with infinitely many strategies // Journal of Science of the Hiroshima University, 1964, A-I 28, pp. 95–99.
- [218] <http://mmiras.webs.uvigo.es/TUGlab/>
- [219] Tur A. V. The irrational behavior proof condition for linear-quadratic discrete-time dynamic games with nontransferable payoffs // Contributions to Game Theory and Management, 7 (2014), 384–392.
- [220] Vieille N. Solvable states in n-player stochastic games // SIAM Journal on Control and Optimization, 2000, vol. 38 (6), pp. 1794–1804.
- [221] Vieille N. Chapter 48 Stochastic games: Recent results // Handbook of Game Theory with Economic Applications, 2002, 3, pp. 1833–1850.

- [222] Xu N., Veinott Jr. A., Sequential stochastic core of a cooperative stochastic programming game. *Operations Research Letters*, 2013, vol. 41, p. 430–435.
- [223] Yeung D.W.K., 2006. An irrational-behavior-proof condition in cooperative differential games // *International Game Theory Review*, 2006, Vol. 8, pp. 739–744.
- [224] Yeung D.W.K., Petrosyan L. *Cooperative Stochastic Differential Games*. Springer, New York, NY, 2005
- [225] Yeung D.W.K., Petrosyan L.A. Subgame Consistent Cooperative Solution of Dynamic Games with Random Horizon // *Journal of Optimization Theory and Applications*, 2011, 150 (1), pp. 78–97.
- [226] Zaccour G. *Théorie des jeux et marchés énergétiques: marché européen de gaz naturel et échanges d'électricité*, Ph.D. Thesis, HEC Montréal, 1987.
- [227] Zaccour G. Time Consistency in Cooperative Differential Games: A Tutorial // *INFOR*, 2008, 46 (1), p. 81–92.
- [228] Zakharov V., Dementieva M. Multistage cooperative games and problem of time consistency // *Int. Game Theory Rev.* 2004, vol. 06, is. 01, pp. 1–14.
- [229] Zyskowski M., Zhu Q. Price and Variance of Anarchy in Mean-Variance Cost Density-Shaping Stochastic Differential Games, 52nd IEEE Conference on Decision and Control, December 10-13, 2013, Florence, Italy.

SAINT PETERSBURG STATE UNIVERSITY

Printed as a manuscript

Parilina Elena Mikhailovna

**SOLUTIONS OF COOPERATIVE STOCHASTIC GAMES
WITH TRANSFERABLE PAYOFFS**

Volume 2

Specialization 01.01.09 —

Discrete mathematics and mathematical cybernetics

Thesis submitted in conformity with the requirements
for the degree of doctor of physico-mathematical sciences

Scientific consultant:

Doctor of physico-mathematical sciences,

Professor Petrosyan Leon Aganesovich

Saint Petersburg

2018

Contents

Introduction	6
Chapter 1. Cooperative Stochastic Games	
with Finite Duration	26
§ 1.1 Non-cooperative stochastic games with finite duration	26
§ 1.2 Main functional equations for stochastic games with finite duration ..	29
§ 1.3 Cooperative stochastic games with finite duration	30
§ 1.4 Shapley value, core and nucleolus	35
§ 1.5 Imputation distribution procedure in stochastic games with finite duration	37
§ 1.6 Subgame consistency of solution of cooperative stochastic game with finite duration	39
§ 1.7 Nonnegative components of imputation distribution procedure in stochastic games with finite duration. Regularization of imputations	43
§ 1.8 Regularization of the Shapley value and the core in stochastic games with finite duration	47
§ 1.9 Strongly subgame consistency of the core in stochastic games with finite duration	62
Chapter 2. Cooperative Stochastic Games with Infinite Duration ..	72
§ 2.1 Noncooperative stochastic games with infinite duration	72
§ 2.2 Most functional equations for stochastic games with infinite duration and stationary strategies	75
§ 2.3 Cooperative stochastic game with infinite duration	77

§ 2.4 Principles of stable cooperation in stochastic games with infinite duration	80
§ 2.4.1 Subgame consistency of optimality principle in stochastic games with infinite duration	81
§ 2.4.2 Strategic stability of cooperative solution in stochastic games with infinite duration	85
§ 2.4.3 Irrational-behaviour-proof in stochastic games with infinite duration	91
§ 2.5 Existence of stable cooperative solution in stochastic games with infinite duration	94
§ 2.6 Strong transferable equilibrium in stochastic games with infinite duration	100
§ 2.7 Stochastic game with one absorbing state	102
§ 2.8 Strongly subgame consistency of the core in stochastic games with infinite duration	106
Chapter 3. Dynamic Games Played over Event Trees	111
§ 3.1 Definitions of a dynamic game played over event tree	111
§ 3.2 Node consistency of the core in dynamic games played over event trees	115
§ 3.3 Node-consistent core in the game of pollution control	121
§ 3.4 Subgame-perfect ε -equilibrium in dynamic games played over event trees	126
§ 3.5 Subgame perfect ε -equilibrium in the game of pollution control	132
§ 3.6 Strategic support of a cooperative solution in dynamic games played over event trees	138
§ 3.7 Subgame perfect ε -equilibrium using imputation distribution procedure in a game of pollution control	143
§ 3.8 Dynamic games played over binary event trees with linear state equations	146
§ 3.8.1 Binary event tree with homogeneous players	148

§ 3.8.2 S -adapted Nash equilibrium in dynamic games played over event trees	149
§ 3.8.3 Joint optimization in dynamic games played over binary event trees	154
§ 3.8.4 Price of anarchy in dynamic games played over binary event trees	159
§ 3.9 Price of Anarchy in an Example in Pollution Control.....	161
§ 3.9.1 Model	161
§ 3.9.2 Results	163
§ 3.9.3 Interaction of Results.....	169
§ 3.10 Node-consistent Shapley Value in a game with random termination time	171
§ 3.10.1 Model	171
§ 3.10.2 Node-consistent Shapley value in dynamic games played over event trees	175
§ 3.10.3 Necessary conditions for the S -adapted equilibrium.....	180
§ 3.10.4 Game-theoretical model of pollution control with random termination time	182
Chapter 4. Applications of Stochastic Games.....	187
§ 4.1 Models of data transmission in wireless networks.....	187
§ 4.1.1 “Forwarder’s dilemma” game.....	187
§ 4.1.2 “Joint package transmission” game.....	194
§ 4.1.3 “Multiple access game”.....	197
§ 4.1.4 “Multiple access game” with the lack of information.....	203
§ 4.1.5 “Forward and transmit” game.....	209
§ 4.1.6 Stochastic game of data transmission in the presence of buffers of finite capacity	219
§ 4.2 Dynamically stable coalition structures as solutions of stochastic game.....	226
§ 4.2.1 Basic definitions. An auxiliary stochastic game	226
§ 4.2.2 D-stable coalition structures	230

§ 4.2.3 D-stable coalition structures in two-person games	232
§ 4.2.4 D-stable coalition structures in three-person games. Simulation study	235
§ 4.3 A stochastic Prisoners' Dilemma with incomplete information on the discount factors	241
§ 4.3.1 Model	241
§ 4.3.2 Publicly known discount factors	242
§ 4.3.3 Two-phase game with unknown discount factors	250
§ 4.3.4 Two-phase game with cooperative and semi-cooperative outcomes	254
§ 4.3.5 Two-phase game with endogenous learning phase	266
§ 4.3.6 Non-cooperative outcomes	271
§ 4.3.7 Numerical example	273
Conclusions	277
Bibliography	280

Introduction

Relevance of thesis topic

The thesis is devoted to studying the methods of constructing cooperative behaviour in stochastic conflict-controlled dynamic systems. The part of dynamic game theory which takes into account uncertainty and randomness, is actual in the modeling of real conflict processes. As players' total payoff in cooperation is not less than their payoff with individually rational behaviour, the problem of constructing a cooperative version of initially given non-cooperative stochastic game is actual. When solving cooperative games played in dynamics, it is necessary to take into account the special features of implementation of the initially chosen cooperative solutions, in particular, the changes in the environmental conditions or "nature" as well as the changes in the behavior of conflict participants, whose interests may change and damage the realization of joint arrangements and break up the cooperative agreement. So the actual problem is to develop the principles of stable cooperation in conflict-controlled dynamic systems with finite and infinite durations including the systems taking into account the stochastic nature of dynamics.

When modeling the possible scenarios of development in technological systems of various complexity and business processes, an analysis of event trees is widely used. The event trees represent a stochastic process with initially given probabilities that are not affected by the players' actions. The construction of time-consistent cooperative solutions in conflict-controlled stochastic systems defined over event trees is an actual task for the above-mentioned areas. Especially in such systems, the problem of implementing strategic support of the cooperative solutions is also a field of applied interest. The principle of strategic support allows to protect the cooperative agreement from deviations of the conflict participants who wish to increase their payoffs.

Another actual task of the thesis is to develop the models of data transmission in the networks with various configurations using the theory of stochastic games. The solutions of stochastic games, both non-cooperative and cooperative, allow one to estimate the necessity of coordinating strategies of participants in the network. In most cases, coordination of strategies of the network participants leads to a significant increase of its capacity, and thus, allows to reduce the total costs of network participants. In the thesis, we consider an important problem of calculating the estimation of the cost of cooperation rejection. We also introduce the calculations of this estimation in the networks of various configurations.

The application of the theory of stochastic games to the problem of determining stable coalition structures in dynamics is also relevant. The coalition structure is a partition of a set of players into disjoint subsets. This approach allows to take into account the implementation of the process of players' transitions from one structure to another and find the coalition structure from which none of the players will deviate individually. Dynamically stable or d -stable coalition structure is found as a solution of a stochastic game.

Overview of the results in this area

The starting point of stochastic game theory is a paper of L. Shapley [210] in which he introduced the definition of a stochastic game and proved the existence of a value of zero-sum game with infinite duration. We may notice that some years before, H. W. Kuhn determined a positional game [120, 121]. The main equations used to solve stochastic games are the modified Bellman's equations [3, 75] for dynamic game problems. Extension of Shapley's results in case of stochastic games with n players was obtained in papers [91, 217], where the existence of the Nash equilibrium in stationary strategies in stochastic games with compact set of strategies and finite set of states was proved. Thereafter, a large number of publications were devoted to prove the existence of the Nash equilibrium in stochastic games in various classes of strategies and also to study stochastic games with incomplete information, asymmetric players, and stochastic games with a special structure. I would like to mention the contribution made by N. Vieille [214, 220], J.-F. Mertens [141], A. Neyman [141, 147, 148], A. S. Nowak [152, 153, 154], E. Solan [106, 212, 214],

A. Jáskiewicz [110] to the theory of stochastic games. One should also mention the books [140, 149] and surveys of the results in stochastic games [213, 216, 221]. The computational problems in finding the equilibria in stochastic games are discussed and solved in [77, 78, 105].

A method for constructing a cooperative version of a stochastic game implemented on a finite tree graph was first proposed by L. A. Petrosyan in the paper [178], in which the problem of the subgame consistency of the Shapley value was formulated and the problem of regularization of a subgame-inconsistent Shapley value was solved for a given class of games. Afterwards the method of constructing a cooperative game version was generalized to the class of stochastic games with infinite duration in the paper of E. M. Parilina and L. A. Petrosyan [72]. Then cooperative stochastic games with infinite duration and a finite set of strategies were studied in the papers [115, 163, 170]. In the thesis the principles of stable cooperation or sustainability of cooperative solutions were implemented in stochastic games. These principles were formulated for dynamic and differential games by L. A. Petrosyan and N. A. Zenkevich in the paper [44].

The first principle or the property of dynamic consistency (subgame consistency) of cooperative solutions was formulated by L. A. Petrosyan [35] for differential games. The mechanism for determining payments to the players to regularize dynamic-inconsistent cooperative solutions by defining the so-called imputation distribution procedure was proposed by L. A. Petrosyan and V. V. Danilov [43]. The problem of constructing dynamic-consistent cooperative solutions was also studied in the works of E. V. Shevkopyas for differential games with a random duration [49], D. W. K. Yeung and L. A. Petrosyan for dynamic games with a random duration [225], L. A. Petrosyan and A. A. Sedakov for network games [186, 188], V. V. Zakharov [16] and E. A. Kornienko [20] for hierarchical games, M. V. Markovkin for linear-quadratic differential games [25], A. V. Tur for linear-quadratic dynamic games [54], D. V. Kuzyutin for dynamic games with incomplete information [21], N. V. Kozlovskaya and N. A. Zenkevich [118] and L. A. Petrosyan and G. Zaccour [182], V. V. Zakharov and L. A. Petrosyan [17] in the field of environmental problems, and V. V. Mazalov and A. V. Rettieva in the field of bioresource management [137, 138]. The problem

of stable cooperation in dynamic transport routing games was studied in [18]. The detailed description of conditions of dynamic consistency of the solutions can be found in the book [224] and various approaches to this problem in survey [227].

The second principle of stable cooperation in dynamic and differential games, namely, strategic stability or strategic support of a cooperative solution, is formulated by L. A. Petrosyan in [38]. The principle occurred to be relevant and then was adapted for various classes of differential and dynamic games [55, 180, 185, 187].

The third principle of stable cooperation, irrational-behavior-proof condition, was formulated by D. W. K. Yeung in [223] and later applied for linear-quadratic games in the papers [53, 133]. The conditions of stable cooperation in Markov processes when the cooperation between players is allowed including the condition of protection against irrational behavior are formulated in the paper [65].

The condition of dynamic consistency in case when the cooperative solution is a set was generalized by L. A. Petrosyan in [37] and called strongly dynamic consistency. Recently, this condition has been investigated in different classes of games (see, for example, [15, 52, 83]).

In the thesis, we also investigate a class of dynamic games played over event trees, in which players' strategies do not affect the transition probabilities. The analysis of event trees became widespread in simulating possible incidents in nuclear reactors [193]. The class of dynamic games played over event trees was first introduced in the works of G. Zaccour [226] and A. Haurie et al. [104] for studying non-cooperative equilibria on the European natural gas market. A detailed description of this class of games can be also found in the book of Haurie et al. [103]. Stochastic games played over event trees are used in modeling unregulated electricity markets and forecasting equilibrium investments in the technology of different generations [96, 97, 189, 190]. In the paper [194], a node-consistent Shapley value was constructed for this class of games. In the thesis, a node-consistent core is constructed for the games played over event trees. The papers [81, 89, 98, 122, 228] are devoted to dynamic consistency of the core in deterministic dynamic games. In [191, 222], the consistency of the core in stochastic games of special forms is examined. The idea of constructing ε -equilibrium used in the thesis was proposed in [192] for finitely repeated games.

The existence of ε -equilibria in dynamic and differential games is proved in the papers [23, 24]. A method of constructing ε -optimal strategies in differential games is proposed in [56]. A comparison of ε -equilibrium and subgame ε -equilibrium in repeated games is described in [111]. The papers [92, 93, 215] should be noted among the works devoted to the construction of ε -equilibria.

Considering dynamic games played over event trees in the paper the price of anarchy (*price of anarchy* or PoA) is calculated. This term is first proposed in [117]. It is a measure of difference in players' payoffs in coordinating their actions and in its absence. In telecommunication networks with complex configuration or large dimension, it is not easy to calculate the exact value of the price of anarchy, so the task of finding its boundaries is actual [57, 82, 84, 196]. A measure similar to the price of anarchy, called the *price of stability* or PoS, is defined in [64]. Recently, a large number of works by T. Roughgarden and other authors are devoted to game-theoretic simulation of data transfer, in which the price of the network anarchy is calculated [85, 86, 201, 202, 203, 204]. The result of calculating the price of anarchy in differential games is presented in [73, 229]. Game-theoretic simulation of data transmission using stochastic games is used for both wireless and other types of telecommunications networks [74, 80, 112, 150, 151, 205]. In this research area we note the works of K. E. Avrachenkov, A. Yu. Garnaev, E. Altman, and others [59, 60, 61, 124, 127].

As an application of stochastic games we present a model of finding a dynamically stable coalition structure. The generalization of cooperative solutions in case of existing a coalition structure is obtained in [68], and further in [155]. Different static approaches to the definition of stable coalition structures have been proposed in the papers [76, 102]. In the thesis, a dynamic approach close to the one described in [116] is modeled. The idea of similarity between the absorbing states of Markov chains and stable coalition structures is used. We note the works on dynamic games in which the possibilities of formation of a coalition structure (which is not just the grand coalition) by the players is assumed [48, 181, 199].

The theory of stochastic games allows to create more realistic models in the field of taxation [197], environmental problems [175] and economics [63]. Game

theoretical applied models are presented in [9, 45]. A simplified model of a stochastic game, repeated game, is often used in applications. In particular, the problem of stimulating cooperation and finding optimal behavior with incomplete or asymmetric information is widely studied [95, 101, 107, 108, 119, 129, 177].

The research object of the thesis is stochastic conflict systems with finite and infinite duration, as well as the conflict systems, the implementation of which takes place over the so-called event trees, in which the stochastic process does not depend on behavior of the participants of the conflict. **The subject of the research** are the methods of constructing cooperative solutions which are consistent in stochastic conflict-controlled systems.

The goal of thesis

The goal of the thesis is to create a theory of cooperative stochastic games with finite and infinite duration with discounted payoffs, as well as the development of the theory of dynamic games played over event trees, in which the stochastic process does not depend on the players' strategies. The aim of the work is also the verification of the principles of stable cooperation, namely, subgame consistency, strongly subgame consistency, strategic stability and irrational-behavior-proof condition of classical cooperative solutions in stochastic games. Another goal is the development of methods for regularization of solutions to construct subgame-consistent solutions based on imputation distribution procedures proposed in the work. The aim of the thesis is also the construction of applied game-theoretic models using the theory of stochastic games.

Main tasks

One of the tasks of the thesis is the development of a method of determining cooperative versions of stochastic games with finite and infinite duration with finite set of game states. In the work we also state the problem of constructing a method for regularization of cooperative solutions for such stochastic games based on the imputation distribution procedure in order to obtain subgame-consistent cooperative solutions. In case of non-negative payoff functions, it is suggested that the payments to the players in the realized vertices (states) are should be non-negative. Therefore, we formulate the problem of constructing a non-negative distribution procedure

satisfying the principle of subgame consistency of a solution for stochastic games with finite duration defined on graphs. For the class of stochastic games with finite duration, the problem of constructing a strongly subgame-consistent core is also formulated.

When studying stochastic games with infinite duration, one of the problems is the regularization of cooperative solutions by constructing subgame-consistent imputation distribution procedure belonging to these solutions. The implementation of a cooperative solution in dynamics requires the verification of satisfaction of the principle of strategic stability, which guarantees the existence of the Nash equilibrium (or, in some cases, subgame perfect Nash equilibrium) with payoffs equal to the components of the imputation. In order to “protect” the dynamic realization of a cooperative solution from deviation, we formulate a problem of finding the conditions under which any player will receive the total payoff that is not less than the guaranteed payoff, even if at some stage the cooperative agreement is broken up, and in the rest of the game, the players will play individually (irrational-behavior-proof condition). One of the tasks formulated in the work for a class of stochastic games with infinite duration is to construct a strong transferable equilibrium with players’ payoffs equal to the corresponding components of the imputation, as well as to construct a strongly subgame-consistent core.

The next task of the thesis is to determine a subgame-consistent core based on the imputation distribution procedure in a cooperative stochastic game played over the event tree. Since stochastic games played over event trees have finite duration, it is impossible to guarantee the existence of the Nash equilibrium with cooperative payoffs in most games. Therefore, for these stochastic games the problem of constructing an approximate equilibrium or subgame-perfect ε -equilibrium with cooperative payoffs is stated. Moreover, the minimum value of ε depending on the parameters of the game is found. In the thesis we formulate the problem of constructing a subgame perfect ε -equilibrium in a regularized stochastic game according to which the players’ payoffs are equal to the corresponding components of the imputation, on the basis of which the regularization has been made.

When modeling real dynamic processes with event trees, binary trees are often

used. In these trees each non-terminal node has two edges. In the thesis, we formulate the task of studying stochastic games played over binary event trees with symmetric players. In this case, we find the Nash equilibrium and a cooperative strategy profile maximizing the total expected payoff of the players and obtain the formula of calculating the price of anarchy for stochastic games of this class, which allows to estimate how much the total players' payoff under cooperation is greater than their payoff with the Nash equilibrium.

Another task of the thesis is the development of game-theoretic models of data transmission in wireless networks with different configurations using the theory of stochastic games, as well as finding the Nash equilibria and cooperative solutions, the calculation of the cost of cooperation rejection to explain the necessity of coordination of the network participants' actions. For one data transmission model, we formulate a problem of verifying subgame consistency and strategic stability of a cooperative solution and constructing regularization of subgame-inconsistent solutions.

One more task of the thesis is to define a game-theoretic model of finding “stable in dynamics”, or d -stable coalition structures using the theory of stochastic games. For this purpose, we provide a method of constructing a stochastic game. For this game, we find the Nash equilibria satisfying the conditions that allow to determine d -stable coalition structures. The aim is to perform a numerical simulation to solve a three-person game and find all d -stable coalition structures.

In the work, we also formulate a task to build a two-phase two-person stochastic game “Prisoner’s dilemma” with incomplete information in which there are two states. The incompleteness of information is in the absence of information about the value of discount factor of the other player. The task is to find Bayesian equilibria in this game.

Scientific novelty

In the thesis, a method of determining a cooperative version of a stochastic game based on the initially defined non-cooperative stochastic game with finite and infinite duration is proposed. The mechanisms of constructing cooperative solutions satisfying the principles of stable cooperation in the above-mentioned classes of

stochastic games, as well as in dynamic games played over event trees are examined in detail in the thesis. The imputation distribution procedure is applied to define subgame-consistent solutions. For cooperative stochastic games with finite and infinite duration with a finite set of states, the sufficient conditions of strongly subgame consistency of the core are obtained. All these methods and mechanisms are first applied to the above-mentioned classes of games. For stochastic games with infinite duration, the conditions of existence of cooperative solutions satisfying the principles of strategic stability and irrational-behavior-proof are obtained.

The methods of supporting cooperation in dynamic games played over event trees are also studied in detail. Therefore, the conditions of subgame consistency of the core, existence of subgame perfect ε -equilibrium with cooperative payoffs when redistribution of payoffs does not take place, and the condition of strategic stability of the cooperative solution when redistribution of players' payoffs takes place are obtained. All these results are first obtained in the works of the author. We may notice a new approach of comparing the players' payoffs in a cooperative and non-cooperative behaviour using the price of anarchy. This approach is borrowed from data transmission models and first applied to dynamic games played over event trees. For dynamic games played over binary event trees with linear state equations and symmetric players, the price of anarchy is calculated in an explicit form.

The methods of modeling data transmission in wireless networks with some configurations have already been considered in the literature before. We first proposed an approach based on the cooperative version of the game modeling data transmission using a stochastic game. We find cooperative solutions and propose their regularization in case of its subgame inconsistency. The regularization allows to change the scheme of payments to network participants in order to increase their payoffs and network capacity. In the work, we proposed an original method of finding d -stable coalition structures (structures that are stable in dynamics), which requires solving a stochastic game constructed in a special way according to the initially given values of a characteristic function. In the thesis, we propose an approach to finding the optimal behavior of the participants in stochastic game "Prisoner's dilemma" with two states when the discounting factor of the other player is unknown. In this

case, a two-phase game is constructed, in the first phase of which players try to recognize the other player's class, and in the second phase they implement a fully cooperative (cooperation in both states), semi-cooperative (cooperation only in one state) or non-cooperative equilibrium.

Research methods

In the thesis we use the methods of dynamic game theory (construction of multi-stage games, subgames, optimality principles, dynamic or subgame consistency, the imputation distribution procedure), cooperative game theory (imputations, coalition structures, stability of coalition structures), optimal control theory (dynamic programming, Pontryagin's maximum principle), optimization and probability theories (distributions of random variables, stochastic processes, Markov chains).

Theoretical and practical significance

The theoretical results obtained in the work are related to the field of dynamic games. Their significance is in the development of the theory of cooperative stochastic games for different ways of defining the dynamics of the game and stochastic process, as well as in the development of stable mechanisms of cooperation and methods of construction of subgame-consistent and strongly subgame-consistent cooperative solutions, as well as their strategic stability for the classes of games being studied.

The practical significance of the work is determined by the range of applications of stochastic games: in mathematical modeling of economic, social conflict-controlled processes, in solving problems of insurance and environmental sciences. Therefore, the area of applications of the obtained results can be estimated by the described quite wide area of applications of stochastic games, in which cooperation of players is meaningful. The theoretical results of Chapter 2 are demonstrated by an economic example. In particular, the problem of constructing a stable cooperative solution in a competitive environment has been solved. The theoretical results of Chapter 3 are illustrated by the examples of the problems in environmental economics, when several countries participating in the game produce emissions and pollute the joint area. The results show that the cooperative behavior of the countries leads to a joint reduction of costs and decreases the total environmental pollution.

Chapter 4 of the thesis is devoted to the construction of applied models using the

theory of stochastic games. In particular, in this chapter several dynamic models of data transmission in wireless networks with different configurations have been proposed. For each model we make a conclusion if cooperation is reasonable and the cooperative control of data transmission increases the network capacity.

The research conducted in the thesis was supported by the following grants: Russian Foundation for Basic Research (RFBR) No. 06-01-39005 “Mathematical analysis of conflict and cooperation” (2006–2007), RFBR No. 09-01-00334 “Optimization in multi-agent wireless networks” (2009–2010), RFBR No. 10-01-91160 “Dynamic network games” (2010–2011), RFBR No. 13-01-91160 “Cooperation in network games” (2013–2014), RFBR No. 14-01-31141 “Stability of coalition structures” (2014–2015), RFBR No. 16-01-00713 “Cooperation with structural and information restrictions” (2016–2017), RFBR travel grant No. 08-01-09236 (2008); grant of Saint Petersburg State University No. 9.0.189.2010 “Mathematical modeling multi-agent dynamic controlled systems” (2010–2012), grant of Saint Petersburg State University (SPSU) No. 9.38.245.2014 “Optimality principles in dynamic and differential games with prescribed and changing coalition structure” (2014–2016), travel grants, SPSU (2011, 2012, 2013, 2014, 2015, 2018); Russian Science Foundation No. 17-11-01079 “Optimal behaviour in conflict-controlled systems” (2017–present); financial support of research center GERAD (2013, 2015, 2018).

The results of the research were used and are used now in the following courses of Saint Petersburg State University including “Game theory and operations research”, “Methods and models of operations research”, “Dynamic games”, “Risk management in projects, finance and logistics”, and soon.

Veracity of the results obtained in the thesis is confirmed by the correct mathematical proofs.

Brief description

The thesis consists of Introduction, four Chapters and Conclusion, includes 304 pages (in Russian version — 329 pages), 42 tables and 24 figures. The list of references contains 229 items.

The first chapter of the thesis is devoted to studying cooperative stochastic games played over finite tree graphs. The main purpose of this chapter is to construct a

cooperative game based on a non-cooperative stochastic game, the development of methods of verification of cooperative solutions of the games on subgame consistency and strongly subgame consistency, as well as regularization of subgame-inconsistent solutions. In § 1.1 the definition of a non-cooperative stochastic game is given, the dynamics of the game and the class of strategies in which the solution will be found is described. The basic functional equations for calculating the expectations of players' payoffs in the game and its subgames are presented in § 1.2. A cooperative version of a stochastic game in the characteristic function form is constructed in § 1.2. This section also discusses the approaches to defining characteristic functions. Preference is given to the so-called α -characteristic function. Bellman's equations are presented to determine the values of this function for all possible coalitions. The definition of the imputation in a stochastic game and its any subgame is given. The solution of a cooperative stochastic game and its subgame is introduced. Section 1.4 contains definitions of the solutions of cooperative games used in the work, including the Shapley value, the core, the nucleolus. The definition of the the imputation distribution procedure for a stochastic game with finite duration is given in § 1.5, which also contains Bellman's functional equations for calculating the expected sum of players' payoffs in accordance with the imputation distribution procedure and proof of lemma, which makes possible to construct the imputation distribution procedure. In § 1.6 the problem of subgame inconsistency of the solutions of a cooperative stochastic game is described and a definition of a subgame-consistent imputation is introduced. A method of constructing a subgame-consistent cooperative solution is also provided in this paragraph. In case of non-negative payoff functions in the initially given game and with the requirements of the non-negative payments to players according to the imputation distribution procedure, we introduce a method of constructing a new imputation distribution procedure in § 1.7. Here we also define a new imputation in a cooperative game with a new characteristic function which satisfies the properties of subgame consistency and non-negativity of the corresponding distribution procedure. The theoretical results are specified in case the Shapley value and the core are chosen as solutions of the cooperative stochastic game in § 1.8. In particular, they are regularized in such a way that the regularized Shapley

value and the core satisfy the principle of subgame consistency. Two examples of finding subgame-consistent solutions (the Shapley value in the first example and the core in the second example) are presented in §1.8. In case when the solution of the game is not a single imputation, but it is a set, the problem of constructing a strongly subgame-consistent solution becomes actual. The problem of strongly subgame consistency of the core for cooperative stochastic games with finite duration is formulated in §1.9. Sufficient conditions of strongly subgame consistency of the core are introduced and proved. The numerical example of constructing a strongly subgame-consistent distribution procedure of the imputation belonging to the core is given.

The second chapter is devoted to studying the principles of stable cooperation in stochastic games with infinite duration and a finite set of states. A stochastic game, first introduced in [210] is a basis for constructing a cooperative version of a game. In §2.1 we describe the model of a non-cooperative stochastic game, classes of players' strategies and methods for calculating their payoffs. In case when a stochastic game is considered in the class of stationary strategies, the players' payoffs can be calculated in an explicit form for the given strategy profiles and these formulas can be found in §2.2. A cooperative version of a stochastic game with infinite duration and a finite set of states is constructed in §2.3. The value of the characteristic function for the given coalition of players is defined as the value of a zero-sum game, where the considered coalition acts as a maximizing player, and the coalition consisting of the players who are not included in the first coalition acts as a minimizing player. The existence of the value of such a game is guaranteed by Shapley's theorem [210]. In §2.3, we present formulas of calculating the values of a characteristic function in the game and its any subgame for all possible coalitions. In §2.4, the basic principles of stable cooperation in stochastic games of this class are defined. These principles were first formulated as a whole in work [44]. The principle of subgame consistency, which is a generalization of the principle of dynamic consistency, is presented in §2.4.1, where the concept of the subgame-consistent distribution of an imputation is introduced and the principle of subgame consistency of the imputation and corresponding distribution procedure is defined. The formula for calculating the

components of subgame-consistent imputation distribution procedure in an explicit form is given in § 2.4.1, where we also propose a method of regularization of the initially given noncooperative game, in which the cooperative solution (initially chosen imputation) is already subgame-consistent. The principle of strategic stability of a cooperative solution is formulated in § 2.4.2, namely, the behaviour strategy profile is defined which consists of two “regimes”: (i) cooperative and (ii) punishment, allowing to punish the player who deviates from the cooperative behavior. The theorem on the existence of the Nash equilibrium with the players’ payoffs equal to the components of a previously chosen imputation is proved. Moreover, in this paragraph the existence of a subgame perfect Nash equilibrium with the same payoffs in a stochastic game is proved. Irrational-behavior-proof condition is given in § 2.4.3, where a sufficient condition for this principle is obtained. The problem of existence of a stable cooperative solution satisfying all three principles of stable cooperation is discussed in § 2.5. In this section we also consider a stochastic dynamic model of competition between asymmetric firms, find a cooperative solution of a stochastic game, and verify if the principles of stable cooperation are satisfied. In § 2.6 the extension of a principle of strategic stability is introduced, namely, the existence of a strong transferable equilibrium in a regularized game with cooperative payoffs is proved. The principle of strong subgame consistency for stochastic games with infinite duration and a finite set of states is formulated in § 2.8. Section 2.8 contains an example of constructing a cooperative solution in a stochastic game with one absorbing state, as well as a verification of the principles of stable cooperation for such a game. As a result, the system of restrictions on the discount factor of the players is derived. The restrictions allow to realize stable cooperation in this game.

The third chapter studies games played over event trees, in which the stochastic process does not depend on the players’ strategies. For this class of games, the state equation is given. A cooperative version of the initially defined non-cooperative game is constructed. The theoretical results of Chapter 3 are illustrated with several examples from the field of environmental economics. In § 3.1 the model of a stochastic game played over event tree is specified and the class of \mathcal{S} -adapted strategies is defined. The basic equations to find the Nash equilibrium and the co-

operative profile, maximizing the total expected players' payoff are also introduced in §3.1. Paragraph §3.2 contains a definition of a characteristic function based on γ -approach, according to which the value of function for coalition S is the total payoff of the players belonging to this coalition in the Nash equilibrium, when the players from S maximize their payoff, and all other players maximize their individual payoffs. The core is chosen as a solution of a stochastic game. The problem of node consistency (subgame consistency) of the core is formulated, and a method of players' payoffs redistribution is proposed to guarantee node consistency of the chosen imputation from the core in case when the cores of all subgames on the cooperative trajectory are non-empty. An example of constructing a node-consistent core in one game of environmental problems is considered in §3.3. For a game with three periods a node-consistent core based on the nucleoli is derived. The subgame perfect ε -equilibrium is constructed for a non-regularized stochastic game played over event tree in §3.4. In defined approximate equilibrium, the players payoffs are equal to the corresponding payoffs in case the players implement cooperative strategies. The examples of constructing subgame perfect ε -equilibrium in a problem of environmental economics with four and eleven periods and binary trees are presented in §3.5. For the game with eleven time periods, a numerical analysis of the minimum value of ε as a function of the discount factor is performed. Assuming that players maximize their total payoff and conduct regularization of the payoffs according to a node-consistent imputation distribution procedure, the subgame perfect ε -equilibrium is constructed and the minimum value of ε is found in §3.6. The numerical comparison of the minimum values of ε in subgame perfect ε -equilibria in a cooperative game with 6 periods without and with regularization of players' payoffs is presented in §3.7. The case of a binary event tree with linear-quadratic payoff functions is studied in detail in §3.8. In this paragraph it is also assumed the symmetry of the players. For this stochastic game, the players' strategies forming S -adapted Nash equilibrium are obtained in an explicit form (see §3.8.2), the strategies that form a cooperative profile can also be found in §3.8.3. The formula for calculating the price of anarchy in the described game is given in §3.8.4. The theoretical results of §3.8 are presented with a game of environmental economics.

Stochastic games with random duration played over event trees, for which a method of constructing a node-consistent Shapley value is proposed, and also the necessary conditions for the existence of S -adapted Nash equilibrium are provided in § 3.10. The theoretical results of Section § 3.10 are presented by an example in § 3.10.4.

Chapter 4 is devoted to game-theoretic modeling using the theory of stochastic games as well as solutions of applied stochastic games. Paragraph § 4.1 contains four models of data transmission in wireless networks. The “Forwarder’s dilemma” game (see § 4.1.1) describes a conflict system consisting of four vertices, two of which represent players who make decisions on forwarding a package appeared at a given node. The process of data transmission is modeled by a stochastic game with four states and solved in stationary strategies, the Nash equilibria and a cooperative strategy profile are found. The comparison of players’ payoffs assuming cooperation of the players or its absence is made by calculating a “cost of cooperation rejection”. For a numerical example of “Forwarder’s dilemma” game, the conclusion if there is a need to coordinate the players’ actions in order to increase the system capacity is made. The similar calculations are performed for “Joint package transmission” game (see § 4.1.2) and “Multiple access game” (see § 4.1.3). For “Joint package transmission” game the results are obtained in a general form, and for “Multiple access game” the results are obtained in case of a uniform initial distribution over the set of states. In § 4.1.5 data transmission in a network of a special configuration is modeled as a cooperative game, the Shapley value is found as a solution of the game, an imputation distribution procedure is defined, the subgame consistency and the strategic stability of the Shapley value are verified. The model of data transmission in the network when nodes have buffers of finite capacities is provided in § 4.1.6. A dynamic model of finding stable coalition structures as solutions of a stochastic game constructed in a special way is proposed in § 4.2. A method of constructing an auxiliary stochastic game is given in § 4.2.1. The numerical modeling of three-person game with two cases: (i) superadditive and (ii) non-superadditive characteristic functions is presented in § 4.2.4. In § 4.3 a stochastic Bayesian model of infinitely repeated Prisoner’s dilemma game with two players and two possible states is proposed. The model description is given in § 4.3.1 where the conditions for which a “cooperative”

strategy profile (cooperation in both states), “semi-cooperative” strategy profile (cooperation only in one state) are subgame perfect equilibria are obtained. Depending on the discount factors, there are three classes of players, that are “patient”, “mildly patient” and “impatient”. In § 4.3.3 the model with unknown discount factors of the players and two phases (learning and cooperation) is proposed, the players’ strategies in both phases are described, and a partially Markov perfect Bayesian equilibrium is defined. The existence of the equilibria according to which the players’ strategies in the second phase form cooperative or semi-cooperative equilibria is proved in § 4.3.4. The learning phase lasts one period in this case. In this paragraph the comparison of the players’ payoffs in equilibria is provided. The case when the learning phase has endogenous duration is considered in § 4.3.5, for which the conditions of the existence of the equilibria with cooperative and semi-cooperative outcomes are obtained. The theoretical results of § 4.3 are demonstrated by a numerical example in § 4.3.7.

The conclusion contains a brief description of the results obtained in the work.

Results submitted for defense

1. A method of constructing a cooperative stochastic game based on the initially given noncooperative stochastic game with finite duration as well as infinite duration with finite set of states.
2. A method of constructing subgame-consistent solution based on the imputation distribution procedure in cooperative stochastic games with finite and infinite duration, and also in dynamic games played over the event trees.
3. Regularization of cooperative solutions in stochastic games with finite duration and non-negative payoff functions of the players as well as the construction of a “new” subgame-consistent imputation distribution procedure with non-negative components, and conditions of subgame consistency for this class of games.
4. A method of constructing a strongly subgame-consistent core in cooperative stochastic games with finite and infinite duration.
5. The conditions for implementing the principles of stable cooperation including subgame consistency, strategic stability of a cooperative solution and irrational-

behavior-proof condition for stochastic games with infinite duration and a finite set of states.

6. The condition of existence of a strong transferable equilibrium with players' payoffs equal to the corresponding components of the imputation in stochastic games with infinite duration and a finite set of states.
7. The conditions of existence of subgame-perfect ε -equilibrium in dynamic games played over event trees with players' payoffs equal to their payoffs when a cooperative strategy profile is implemented, and a condition of the strategic stability of a cooperative solution for this class of games.
8. Construction of the Nash equilibrium and cooperative solution in dynamic games played over binary event trees with linear state equations and symmetric players, as well as a method of calculating the price of anarchy in this class of games.
9. A method of constructing a cooperative solution in dynamic data transmission for wireless networks with different configurations, and a method of regularization of a stochastic game of data transmission to obtain a subgame-consistent cooperative solution.
10. A method of finding d -stable (stable in dynamics) coalition structures by solving a specially constructed stochastic game when there are no restrictions on the set of admissible coalitions.
11. The conditions of existence of Bayesian equilibria in a two-phase stochastic "Prisoner's dilemma" game with two states which leading to full or partial cooperation of the players in the second phase of the game.

Verification of results

The main results of the thesis were presented at the International Conferences "Game Theory and Management" (Saint Petersburg, 2007, 2010, 2011, 2012, 2013, 2014, 2016, 2018); International meeting "Optimal stopping problems and stochastic control" (Petrozavodsk, 2005); the International Conferences "Stability and control processes" (Saint Petersburg, 2005, 2010, 2015); the International Conference

“Constructive non-smooth analysis and related topics” (Saint Petersburg, 2012); the International Conference “Control Theory and Theory of Generalized Solutions of Hamilton-Jacobi Equations” (Ekaterinburg, 2005); Russian Voronezh Spring Mathematical School “Pontryagin’s Readings — XV” (Voronezh, 2004); the seminar of the Russian-Finnish school for postgraduate students “Dynamic games and multi-criteria optimization” (Petrozavodsk, 2006); the International conference “Non-linear dynamic analysis” (Saint Petersburg, 2007); the International conference “Game theory and applications” (Qingdao, 2007); the International Conference “Spanish-Italian-Netherlands Game Theory Meeting (SING)” (Wroclaw, 2008); Graduate School Seminar “Citations as impacts of research — who reads our papers?” (Stockholm, 2007), Tel Aviv International Workshop on Game Theory (Tel Aviv, 2011); the XVI conference on Dynamics, Economic Growth, and International Trade — DEGIT-XVI (Saint Petersburg, 2011); the Fourth Workshop on Dynamic Games in Management Science (Padua, 2012); the International Conference “Mathematics, Economic Sciences, and Management: the Centenary of L.V. Kantorovich” (Saint-Petersburg, 2012); the International Seminar “Networking Games and Management” (Petrozavodsk, 2012); the 15th, 16th, 18th International Symposia on Dynamic Games and Applications (Byšice, 2012; Amsterdam, 2014; Grenoble, 2018); the 9th and 10th International ISDG Workshops (Barcelona, 2013; Glasgow, 2015); European Meeting on Game Theory “SING11-GTM2015” (Saint Petersburg, 2015); XIV International Conference “Stability and Oscillations of Nonlinear Control Systems” (Pyatnitskiy’s Conference, Moscow, 2018); at the seminars of Department of mathematical game theory and statistical decisions of Saint Petersburg State University, Institute of Applied Mathematical Research of Karelian Research Centre, Russian Academy of Sciences (Petrozavodsk, 2006, 2018); Department of dynamic systems of N. N. Krasovskii Institute of Mathematics and Mechanics of the Ural Branch of the Russian Academy of Sciences (2018); Institute of Control Sciences V. A. Trapeznikov of Russian Academy of Sciences (2018); University of Padova (2013); Qingdao University (2018); Research Center GERAD (Montreal, 2013, 2015, 2018).

Publications

The author has published more than 85 scientific works, 39 of them are the

scientific publications on the topic of thesis research including one chapter in the monograph and 15 papers published in well-known peer-reviewed scientific journals (Mathematical Game Theory and its Applications, Large-Scale Systems Control, Proceedings of Institute of Mathematics and Mechanics Ural Branch of the Russian Academy of Sciences, European Journal of Operational Research, Automatica, Operations Research Letters, Theory and Decision, International Game Theory Review, Automation and Remote Control) from the list of journals recommended by High Certifying Commission (VAK) and equivalent journals. All main results submitted for defense are obtained by the author independently.

Acknowledgments

The author expresses her deep appreciation and thanks to Doctor of Physical and Mathematical Sciences, Professor Leon Aganesovich Petrosyan for the topic proposed for scientific research, evolution of scientific interests in mathematical game theory, long-term and fruitful collaboration, careful attention to the work and comprehensive assistance. The author is also sincerely grateful to Professor Georges Zaccour for his support, kind attention, interesting discussions and joint work. The author thanks the parents and family for their patience to scientific work, help and support in all situations.

Chapter 1

Cooperative Stochastic Games with Finite Duration

§ 1.1 Non-cooperative stochastic games with finite duration

We define a finite stochastic game played on a graph. Let $\Psi = (Z, L)$ be a finite graph of a tree structure (see [4]), where Z is the set of vertices of the graph, and $L : Z \rightarrow Z$ is a point-set mapping defined on the set Z , with values in the set of the subsets of Z . The vertex z_0 is the initial vertex of the tree graph Ψ . We denote the terminal vertices of graph Ψ by $Z^T \subset Z$, that is, the vertices z for which $L(z) = \emptyset$. The finite tree graph with initial vertex z_0 is denoted by $\Psi(z_0)$.

Let at each vertex $z \in Z$ of the graph $\Psi(z_0)$ the normal form game of n players

$$\Gamma(z) = \langle N, A_1^z, \dots, A_n^z, K_1^z, \dots, K_n^z \rangle,$$

be given, and $N = \{1, 2, \dots, n\}$ is a finite set of players, the same for all vertices $z \in Z$; A_i^z is a finite set of actions of player $i \in N$, $K_i^z(a_1^z, \dots, a_n^z) : \prod_{j \in N} A_j^z \rightarrow \mathbb{R}$ is a payoff function of player i , $a_i^z \in A_i^z$. The collection of actions $a^z = (a_1^z, \dots, a_n^z)$, $a_i^z \in A_i^z$, $i \in N$, is called an action profile in the game $\Gamma(z)$. And $a^z \in A^z = \prod_{i \in N} A_i^z$, A^z is the set of action profiles in game $\Gamma(z)$.

For each vertex $z \in Z$ we define the transition probabilities to the vertices $y \in L(z)$ of the graph $\Psi(z_0)$ following vertex z . These probabilities depend on the action profile a^z realized in the game $\Gamma(z)$. Thus, for each vertex $z \in Z$ we define a function $p(\cdot | z, a^z) : A^z \rightarrow \Delta(L(z))$, where $\Delta(L(z))$ is a probability distribution

over set $L(z)$:

$$\begin{aligned} p(y|z, a^z) &\geq 0, \\ \sum_{y \in L(z)} p(y|z, a^z) &= 1 \end{aligned}$$

for any action profile $a^z \in A^z$. The value $p(y|z, a^z)$ is the probability that game $\Gamma(y)$, $y \in L(z)$, will be played after game $\Gamma(z)$ if the action profile $a^z = (a_1^z, \dots, a_n^z)$ has been realized in $\Gamma(z)$.

We also suppose that the duration of the game is random which values are $0, 1, \dots, l$, and l is the length of the game (by the length of the game we mean the number of stages in the game of maximal possible path). The probabilities q_k of the event that the game will end at stage k are given. Notice that $0 \leq q_k \leq 1$, $k = 0, \dots, l-1$, $q_l = 1$, stage k at vertex $z \in Z$ in a stochastic game with random duration is determined from the condition: $z \in (L(z_0))^k$.

Remark 1.1. Notice that the probabilities q_k , $k = 0, \dots, l$ are conditional probabilities and may be used to determine probability distribution of the game duration. In case when all paths in graph $\Psi(z_0)$ have the same length l , the discrete distribution of the game duration, determined by the conditional probabilities q_k , is presented in Table 1.1, in which P_k is the probability that the game will end at stage k .

Table 1.1: Probability distribution of the game duration.

k	P_k
0	q_0
1	$(1 - q_0)q_1$
2	$(1 - q_0)(1 - q_1)q_2$
\vdots	\vdots
l	$(1 - q_0)(1 - q_1) \cdot \dots \cdot (1 - q_{l-1})$

Definition 1.1. Stochastic game with random duration $G(z_0)$, where z_0 is an initial vertex of a tree graph $\Psi(z_0)$, is a quintuple

$$G(z_0) = \langle N, \Psi(z_0), \{\Gamma(z)\}_{z \in Z}, \{q_k\}_{k=0}^l, \{p(\cdot|z, a^z)\}_{z \in Z \setminus Z^T, a^z \in A^z} \rangle. \quad (1.1.1)$$

From the definition of a stochastic game with random duration it is clear that the transitions from some vertices of the graph $\Psi(z_0)$ to the others, as well as the final stage of the game are random.

Stochastic game with random duration $G(z_0)$ is played in the following way:

1. At vertex z_0 of the graph $\Psi(z_0)$, a normal-form game $\Gamma(z_0)$ is played. Suppose that in this game action profile $a^{z_0} \in A^{z_0}$ is realized by the players. Each player $i \in N$ receives a payoff $K_i^{z_0}(a^{z_0})$. The stochastic game $G(z_0)$ either terminates with probability q_0 , $0 \leq q_0 \leq 1$, or continues with probability $1 - q_0$ and transmits to the vertex $y \in L(z_0)$ of the graph $\Psi(z_0)$ with probability $p(y|z_0, a^{z_0})$, depending on the action profile a^{z_0} realized in the game $\Gamma(z_0)$.
2. Suppose that at stage k the game process is at the vertex $z_k \in Z$, at which the normal-form game $\Gamma(z_k)$ is given. Let the action profile $a^{z_k} \in A^{z_k}$ is realized in this game. Each player $i \in N$ receives a payoff $K_i^{z_k}(a^{z_k})$. Stochastic game either ends with probability q_k , $0 \leq q_k \leq 1$, or continues with probability $1 - q_k$ and transits to the vertex $z_{k+1} \in L(z_k)$ with probability $p(z_{k+1}|z_k, a^{z_k})$ depending on the action profile a^{z_k} realized in game $\Gamma(z_k)$. In case when the set $L(z_k)$ is empty, the game terminates at the vertex z_k with probability 1.
3. The stochastic game continues until the terminal vertex is reached or it may end according to realizations of probabilities q_0, \dots, q_l .

We denote by $G(z_k)$ the subgame (see [120, 121]) of the game $G(z_0)$ starting at the vertex $z_k \in Z$ of graph $\Psi(z_0)$ (starting with game $\Gamma(z_k)$), which is also a stochastic game with random duration. Subgame $G(z_k)$ is defined on the subgraph $\Psi(z_k)$ with the set of vertices $Z(z_k) \subset Z$ and is given by the quintuple

$$G(z_k) = \langle N, \Psi(z_k), \{\Gamma(z)\}_{z \in Z'}, \{q_s\}_{s=k}^l, \{p(\cdot|z, a^z)\}_{z \in Z(z_k) \setminus Z^T, a^z \in A^z} \rangle.$$

To solve the game you need to determine the set of players' strategies. We denote by $\varphi_i : Z \times H \rightarrow \prod_{z \in Z} \Delta(A_i^z)$ the behavior strategy of player i in game $G(z_0)$, where H is a set of histories of all vertices. By the history of a vertex z we mean the set of action profiles realised on a subpath from z_0 to z . $\Delta(A_i^z)$ is the set of mixed actions of the player i at the vertex $z \in Z$. The strategy profile in stochastic game $G(z_0)$ is a collection of the players' strategies given by $\varphi = (\varphi_i : i \in N)$. Denote by Σ_i the set of behavior strategies of player i in the stochastic game $G(z_0)$, then $\Sigma = \prod_{i \in N} \Sigma_i$ is the set of behavior strategy profiles in game $G(z_0)$. Obviously, the

restriction of strategy φ_i on subgraph $\Psi(z_k)$ of graph $\Psi(z_0)$ is a strategy in subgame $G(z_k)$. Denote this restriction of a strategy by $\varphi_i^{z_k}$.

§ 1.2 Main functional equations for stochastic games with finite duration

Assume that players implement strategies φ_i , $i \in N$, in stochastic game $G(z_0)$. Define the payoff of player i as a mathematical expectation of his payoff relative to a random variable equal to the game duration, and for the realized path $z_0, z_1 \in L(z_0)$, $z_2 \in L(z_1), \dots, z_l \in L(z_{l-1}), L(z_l) = \emptyset$, we obtain

$$E_i(z_0, \varphi) = \sum_{k=0}^l P_k \sum_{j=0}^k K_i^{z_j}(a^{z_j}) = \sum_{k=0}^l q_k \left(\prod_{j=0}^{k-1} (1 - q_j) \right) \left(\sum_{m=0}^k K_i^{z_m}(a^{z_m}) \right),$$

where $a^{z_0} \in A^{z_0}, a^{z_1} \in A^{z_1}, \dots, a^{z_l} \in A^{z_l}$ is a sequence of realized action profiles when players implement strategies $(\varphi_i : i \in N)$.

Since transitions from the vertices to the following vertices are stochastic, we consider mathematical expectation of the player's payoff relative to random transitions from vertices to the following vertices as a player's payoff in the stochastic game. The mathematical expectation $E_i(z_0, \varphi)$ of player i 's payoff in the game satisfies the functional equation

$$\begin{aligned} E_i(z_0, \varphi) &= q_0 K_i^{z_0}(a^{z_0}) + (1 - q_0) \left(K_i^{z_0}(a^{z_0}) + \sum_{y \in L(z_0)} p(y|z_0, a^{z_0}) E_i(y, \varphi^y) \right) \\ &= K_i^{z_0}(a^{z_0}) + (1 - q_0) \sum_{y \in L(z_0)} p(y|z_0, a^{z_0}) E_i(y, \varphi^y), \end{aligned} \tag{1.2.1}$$

where $E_i(y, \varphi^y)$ is the mathematical expectation of player i 's payoff in the subgame $G(y)$ starting at the vertex $y \in L(z_0)$ of graph $\Psi(z_0)$.

Assume that $z \in (L(z_0))^k$, that is, the game process enters vertex $z \in Z$ at stage k , then the mathematical expectation of player i 's payoff in the subgame $G(z)$

satisfies the functional equation

$$\begin{aligned} E_i(z, \varphi^z) &= q_k K_i^z(a^z) + (1 - q_k) \left(K_i^z(a^z) + \sum_{y \in L(z)} p(y|z, a^z) E_i(y, \varphi^y) \right) \\ &= K_i^z(a^z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, x^z) E_i(y, \varphi^y). \end{aligned}$$

To define a cooperative version of the game, it is necessary to determine a cooperative strategy profile realisation of which determines the so-called “cooperative path” (one of the cooperative paths, if there are several ones), that is, the path that maximizes the total players’ payoff. In the case of stochastic games, this is a subtree with the given transition probabilities, at which the maximum of the mathematical expectation of the total players’ payoff in the whole game is achieved. However, the maximal mathematical expectation of the total players’ payoffs in mixed strategies is equal to the maximum mathematical expectation of the summarized players’ payoffs in pure strategies. Therefore, we can restrict ourselves and consider the class of pure strategies to find cooperative strategies in the stochastic game.

§ 1.3 Cooperative stochastic games with finite duration

Denote by $\bar{\varphi} = (\bar{\varphi}_1, \dots, \bar{\varphi}_n)$ the pure strategy profile in game $G(z_0)$ which maximizes the total mathematical expectation of the players’ payoffs:

$$V(N, z_0) = \max_{\varphi \in \Sigma} \left[\sum_{i \in N} E_i(z_0, \varphi) \right] = \sum_{i \in N} E_i(z_0, \bar{\varphi}).$$

We call this strategy profile as a cooperative one. Let strategy profile $\bar{\varphi}$ be such that $\bar{\varphi}_i(z) = \bar{a}_i^z$, $i \in N$, $z \in Z$. We can determine the cooperative strategy profile for any subgame $G(z)$, $z \in Z$, starting with normal-form game $\Gamma(z)$.

We construct a cooperative version of a stochastic game on the basis of a non-cooperative stochastic game with random duration $G(z_0)$ described above. For this purpose it is necessary to define the characteristic function for any subset S (*coalition*) of the set of players N . The characteristic function calculated for the subgame $G(z)$, $z \in Z$, is denoted by $V(S, z)$, where $S \subset N$.

Characteristic function $V(S, z)$ shows which total payoff can be obtained by the players joining into coalition S . There are different approaches to defining the

characteristic function that determines the cooperative game on the basis of a non-cooperative one. We introduce some of these approaches:

1. α -approach. In this case, $V(S, z)$ is the *maxmin* value of the zero-sum game between coalitions S and $N \setminus S$. Moreover, the maxmin is found in the pure strategies of coalition S . This approach can be described as “pessimistic”, since $V(S, z)$ is equal to maximal total payoff of coalition S which coalition S can obtain if coalition $N \setminus S$ acts against S . This approach was proposed in the book of Neumann and Morgenstern [146].
2. β -approach. Following this approach, $V(S, z)$ is the *minmax* value of the zero-sum game G_S between coalitions S and $N \setminus S$. Moreover, the minimax is found in pure strategies. This approach can be considered as “optimistic”. Comparison of α - and β -approaches can be found in [69].
3. The value of game G_S . In this case, value $V(S, z)$ is equal to the value of zero-sum game G_S game between coalitions S and $N \setminus S$. Moreover, this value always exists in mixed strategies, while it is equal to the maxmin and minimax of G_S . In case the minmax and maxmin are found in mixed strategies, the values of α - and β -characteristic functions coincide.
4. γ -approach. According to this approach, $V(S, z)$ is equal to the payoff of coalition S in the Nash equilibrium, when all the players who do not belong to coalition S play individually. This approach was proposed in the paper [81].
5. δ -approach. Value $V(S, z)$ is equal to the maximum payoff of coalition S in the strategy profile when the players who do not belong to coalition S adopt the Nash equilibrium strategies optimal in the n -person game when all players act individually. This approach was proposed in the paper [182] and further considered in detail in the paper [195].
6. ζ -approach. In this case, $V(S, z)$ is equal to the payoff of coalition S in the strategy profile when the players from coalition S use strategies that maximize the total payoff of coalition N , and the players who do not belong to coalition S minimize the total payoff of the players from coalition S (see [14]).

In this chapter we will use the so-called α -approach and assume that the “power” of coalition S is equal to the maxmin value of a two-person zero-sum stochastic game G_S between coalitions S and $N \setminus S$. This approach was used in the paper [178], in which for the first time a cooperative stochastic game was constructed on the basis of a non-cooperative one and the problem of subgame inconsistency of the Shapley value is considered.

We determine the values of the characteristic function. First we consider the case when $S = N$ and find the maximum of the total payoff of the coalition N in stochastic game $G(z_0)$. For this purpose, we write Bellman’s equation (see [75] or [3]) for the maximum sum of the mathematical expectations of players’ payoffs:

$$\begin{aligned} V(N, z_0) &= \max_{\substack{a_i^{z_0} \in A_i^{z_0} \\ i \in N}} \left[\sum_{i \in N} K_i^{z_0}(a^{z_0}) + (1 - q_0) \sum_{y \in L(z_0)} p(y|z_0, a^{z_0}) V(N, y) \right] \quad (1.3.1) \\ &= \sum_{i \in N} K_i^{z_0}(\bar{a}^{z_0}) + (1 - q_0) \sum_{y \in L(z_0)} p(y|z_0, \bar{a}^{z_0}) V(N, y) \end{aligned}$$

with boundary condition

$$V(N, z) = \max_{\substack{a_i^z \in A_i^z \\ i \in N}} \sum_{i \in N} K_i^z(a^z), \quad z \in \{z : L(z) = \emptyset\}. \quad (1.3.2)$$

Later on in this chapter, we suppose that $z \in (L(z_0))^k$.

For subgame $G(z)$, $z \in Z$, the equation (1.3.1) with the boundary condition (1.3.2) takes the form:

$$\begin{aligned} V(N, z) &= \max_{\substack{a_i^z \in A_i^z \\ i \in N}} \left[\sum_{i \in N} K_i^z(a^z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, a^z) V(N, y) \right] \quad (1.3.3) \\ &= \sum_{i \in N} K_i^z(\bar{a}^z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(N, y) \end{aligned}$$

with boundary condition

$$V(N, z) = \max_{\substack{a_i^z \in A_i^z \\ i \in N}} \sum_{i \in N} K_i^z(a^z), \quad z \in \{z : L(z) = \emptyset\}. \quad (1.3.4)$$

Strategy profile $(\bar{\varphi}_i : i \in N)$ in stochastic game $G(z_0)$ generates the probability distributions over set Z of the vertices of graph $\Psi(z_0)$.

Definition 1.2. A subgraph of graph $\Psi(z_0)$, which consists of the vertices $z \in Z$ of the graph $\Psi(z_0)$, having positive realization probabilities, generated by the cooperative strategy profile $\bar{\varphi}$, is called a cooperative subtree and denoted by $\bar{\Psi}(z_0)$.

Obviously, subgraph $\bar{\Psi}(z_0)$ is a finite tree graph. The set of vertices in graph $\bar{\Psi}(z_0)$ is denoted by $CZ \subseteq Z$.

Let $S \subset N$, $S \neq N$. For each vertex $z \in CZ$ we define the auxiliary zero-sum game denoted by $G_S(z)$. It is a zero-sum game between coalition $S \subset N$ acting as a maximizing player, and coalition $N \setminus S$ acting as a minimizing player. In this case, the payoff of coalition S is calculated as the sum of the payoffs of the players belonging to coalition S . Then, the value of characteristic function $V(S, z)$ is given by the lower value of zero-sum game $G_S(z)$ in pure strategies (similar to the lower value of the matrix game) ¹.

Function $V(S, z)$, $z \in CZ$, satisfies the following functional equation

$$V(S, z) = \max_{a_S^z \in A_S^z} \min_{a_{N \setminus S}^z \in A_{N \setminus S}^z} \left[\sum_{i \in S} K_i^z(a_S^z, a_{N \setminus S}^z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, (a_S^z, a_{N \setminus S}^z)) V(S, y) \right] \quad (1.3.5)$$

with boundary condition

$$V(S, z) = \max_{a_S^z \in A_S^z} \min_{a_{N \setminus S}^z \in A_{N \setminus S}^z} \sum_{i \in S} K_i^z(a_S^z, a_{N \setminus S}^z), \quad z \in \{z : L(z) = \emptyset\}, \quad (1.3.6)$$

where $a_S^z = (a_i^z : i \in S)$ is an action of coalition S ; $A_S^z = \prod_{i \in S} A_i^z$ is the action set of coalition S ; $a_{N \setminus S}^z = (a_j^z : j \in N \setminus S)$ is an action of coalition $N \setminus S$; $A_{N \setminus S}^z = \prod_{j \in N \setminus S} A_j^z$ is the action set of coalition $N \setminus S$.

For all $z \in CZ$ it is natural to suppose that

$$V(\emptyset, z) = 0. \quad (1.3.7)$$

Thus, for each subgame $G(z)$, $z \in CZ$, we have determined the characteristic function $V(S, z)$, $S \subset N$. The characteristic function $V(S, z)$ is determined by the

¹In this chapter we use the α -approach for construction of the characteristic function, proposed by Neumann and Morgenstern [146].

equation (1.3.3) with the boundary condition (1.3.4), and also the equation (1.3.5) with the boundary condition (1.3.6) and equation (1.3.7).

The characteristic function $V(S, z)$ defined by formulas (1.3.3) – (1.3.7) is super-additive with S , i. e., for any vertex $z \in CZ$ and any coalitions $S, P \subset N$, $S \cap P = \emptyset$, the inequality

$$V(S \cup P, z) \geq V(S, z) + V(P, z).$$

holds. The proof of this fact can be found, e. g., in the book [51].

Definition 1.3. *A cooperative stochastic game with random duration $\bar{G}(z_0)$ constructed on the basis of non-cooperative stochastic game $G(z_0)$, is a tuple $\langle N, V(S, z_0) \rangle$, where $V(S, z_0)$ is a characteristic function defined by equation (1.3.3) with boundary condition (1.3.4) for coalition N , by equation (1.3.5) with boundary condition (1.3.6) for coalition $S \subset N$, $S \neq N$, $S \neq \emptyset$, and by formula (1.3.7) for coalition $S = \emptyset$.*

Definition 1.4. *An imputation in cooperative stochastic game $\bar{G}(z_0)$ is vector $\xi(z_0) = (\xi_1(z_0), \dots, \xi_n(z_0))$, satisfying two properties:*

1. *Efficiency:* $\sum_{i \in N} \xi_i(z_0) = V(N, z_0)$;
2. *Individual rationality:* $\xi_i(z_0) \geq V(\{i\}, z_0)$ for any $i \in N$.

The imputation set (see [10, 11, 12, 13, 51]) of cooperative stochastic game $\bar{G}(z_0)$ is denoted by $I(z_0)$.

Definition 1.5. *A solution of cooperative stochastic game $\bar{G}(z_0)$ is a subset $C(z_0)$ of the imputation set $I(z_0)$.*

The solutions of a cooperative game can be conventionally divided into single-valued and multi-valued ones. The well-known single-valued solutions are the Shapley value [211], the nucleolus [206]. The most well-known multi-valued solution is the core [99]. Suppose that solution $C(z_0)$ of cooperative stochastic game $\bar{G}(z_0)$ is non-empty.

Definition 1.6. *A cooperative stochastic subgame $\bar{G}(z)$, $z \in Z$, of game $\bar{G}(z_0)$, constructed on the basis of non-cooperative stochastic subgame $G(z)$, is a pair $\langle N,$*

$V(S, z)$, where $V(S, z)$ is a characteristic function defined by equation (1.3.3) with boundary condition (1.3.4) for coalition N , by equation (1.3.5) with boundary condition (1.3.6) for coalition $S \subset N$, $S \neq N$, $S \neq \emptyset$, and by formula (1.3.7) for coalition $S = \emptyset$.

Determine the imputation, the imputation set and the solution for any cooperative subgame $\bar{G}(z)$, $z \in Z \setminus z_0$.

Definition 1.7. An imputation in cooperative stochastic subgame $\bar{G}(z)$ is vector $\xi(z) = (\xi_1(z), \dots, \xi_n(z))$, satisfying two properties:

1. $\sum_{i \in N} \xi_i(z) = V(N, z)$;
2. $\xi_i(z) \geq V(\{i\}, z)$ for any $i \in N$.

The set of imputations of cooperative stochastic subgame $\bar{G}(z)$ is denoted by $I(z)$.

Definition 1.8. The solution of a cooperative stochastic subgame $\bar{G}(z)$ is a subset $C(z)$ of the set of imputations $I(z)$.

Suppose that solution $C(z)$ of any cooperative subgame $\bar{G}(z)$ is non-empty for any $z \in CZ$.

§ 1.4 Shapley value, core and nucleolus

In this section we define some solutions which will be used further in the work. The Shapley value of a cooperative stochastic game or subgame $\bar{G}(z)$, $z \in CZ$, is a vector $Sh(z) = (Sh_1(z), \dots, Sh_n(z))$, and $Sh_i(z)$, $i \in N$, is calculated by the formula

$$Sh_i(z) = \sum_{\substack{S \subset N \\ S \ni i}} \frac{(|S| - 1)! (n - |S|)!}{n!} (V(S, z) - V(S \setminus \{i\}, z)),$$

where $|S|$ is the cardinality of S . The definition of the Shapley value is introduced in Shapley's paper [211].

A core of a cooperative stochastic game or subgame $\bar{G}(z)$, $z \in CZ$, is a set denoted by $CO(z)$, and it is the set

$$CO(z) = \left\{ \xi(z) \in I(z) : \sum_{i \in S} \xi_i(z) \geq V(S, z) \text{ for any } S \subset N \right\}. \quad (1.4.1)$$

For the cooperative stochastic game or subgame $\bar{G}(z)$ and any vector $\xi(z) \in I(z)$, by $\theta(\xi(z))$ we denote the vector of excesses values $e(S, \xi(z)) = V(S, z) - \sum_{i \in S} \xi_i(z)$ located in a descending order:

$$\theta(\xi(z)) = (e(S_1, \xi(z)), e(S_2, \xi(z)), \dots, e(S_{2^n-1}, \xi(z))),$$

where coalitions are numbered so that $e(S_1, \xi(z)) \geq e(S_2, \xi(z)) \geq \dots \geq e(S_{2^n-1}, \xi(z))$.

On the set of excesses $\{\theta(\xi(z)) : \xi(z) \in I(z)\}$ we consider the lexicographic ordering \succ_{lex} :

$$\theta(\xi(z)) \succ_{lex} \theta(\psi(z)) \iff \exists l \in \{1, \dots, 2^n\},$$

such that

$$\begin{cases} \theta_k(\xi(z)) = \theta_k(\psi(z)), & \text{for all } k = 1, \dots, l-1; \\ \theta_l(\xi(z)) > \theta_l(\psi(z)), \end{cases}$$

where $\psi(z) \in I(z)$.

The nucleolus of a cooperative stochastic game or subgame $\bar{G}(z)$, $z \in CZ$, is a subset of the imputation set on which $\min_{\substack{\theta(\xi(z)) \\ \xi(z) \in I(z)}} \succ_{lex}$ is reached. The definition of the nucleolus is first introduced by Schmeidler [206]. The definition and some useful theorems and lemmas about the properties of the nucleolus may be found in [51, 87, 113]. The works [114, 144] are devoted to the calculation of the nucleolus which is a unique vector.

If $C(z_0)$ is the solution of cooperative stochastic game $\bar{G}(z_0)$, then later on in the work by solution $C(z)$ of cooperative subgame $\bar{G}(z)$ we mean a solution constructed according to the same “rules” as $C(z_0)$. For example, if $C(z_0)$ is the Shapley value in cooperative stochastic game $\bar{G}(z_0)$, then $C(z)$ is the Shapley value in subgame $\bar{G}(z)$, $z \in CZ$.

Here we assume that the players choose some fixed subset of the imputation set which contains the imputations satisfying “optimal” properties, i. e., the players forming coalition N , are going to follow some “rule” distributing the payoffs of coalition N throughout the game process. Set $C(z)$ may consist of a single imputation, if, e. g., the players have decided to use the Shapley value or the nucleolus, or it may be empty (e. g., the core may be empty). The solution of the game or subgame

$\bar{G}(z)$ can be any other imputations from the classical “static” cooperative theory, such as von Neumann-Morgenstern solution (or the so-called stable set), the kernel, \mathcal{M} -stable sets (see [51]).

Further in the work we will suppose that $C(z)$ is a nonempty subset of set $I(z)$ for any $z \in CZ$, that is, for each vertex $z \in CZ$ there exists at least one imputation

$$\xi(z) = (\xi_1(z), \dots, \xi_n(z)) \in C(z) \subset I(z).$$

§ 1.5 Imputation distribution procedure in stochastic games with finite duration

In this section we introduce the definition of an imputation distribution procedure [43] of the cooperative stochastic game solution, which has been chosen by the players. The imputation distribution procedure determines the payments to the players at each vertex of the cooperative subtree $\bar{\Psi}(z_0)$.

Denote by a^{z_0, z_1, \dots, z_l} a vector of profiles $a^{z_0}, a^{z_1}, \dots, a^{z_l}$ realised on a path z_0, z_1, \dots, z_l in a stochastic game, where a^{z_i} is the action profile in game $\Gamma(z_i)$, $z_i \in L(z_{i-1})$, $i = 1, \dots, l$.

Consider any vertex $z \in CZ$, $z \in (L(z_0))^k$, of the cooperative subtree. Each player receives payments implementing a cooperative strategy profile². Let at the vertex $z \in CZ$ the payment to player $i \in N$ be $\beta_i(z)$. In any cooperative subgame $\bar{G}(z)$, the player can calculate the sum of the payments with \bar{a}^{z, \dots, z_l} along the path z, \dots, z_l . We denote by $B_i(z)$ the mathematical expectation of the sum of such payments, calculated along the path segment z, \dots, z_l in cooperative subgame $\bar{G}(z)$. The value $B_i(z)$ satisfies the following functional equation:

$$B_i(z) = \beta_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) B_i(y) \quad (1.5.1)$$

with boundary condition

$$B_i(z) = \beta_i(z) \text{ for } z \in \{z : L(z) = \emptyset\}. \quad (1.5.2)$$

²Obviously, all $z, \dots, z_l \in CZ$, since CZ is the set of vertices of the cooperative subtree, and the strategy profile $\bar{\varphi}$ is determined.

Now we define the distribution procedure of the imputation belonging to cooperative solution $C(z_0)$, chosen by the players at the beginning of the game.

Definition 1.9. *Let $\xi(z_0)$ be a vector $(\xi_1(z_0), \dots, \xi_n(z_0)) \in C(z_0)$. The set of vectors $\{\beta(z) = (\beta_1(z), \dots, \beta_n(z)) : z \in CZ\}$ is called a distribution procedure of the imputation $\xi(z_0)$ if the following conditions are satisfied:*

1. *For each vertex $z \in CZ$:*

$$\sum_{i \in N} \beta_i(z) = \sum_{i \in N} K_i^z(\bar{a}^z).$$

2. *The components $\xi_i(z_0)$, $i \in N$, of imputation $\xi(z_0)$ coincide with the mathematical expectation of the corresponding components of the imputation distribution procedure with respect to the probability distribution of transitions and the termination time of the game, i. e., $\xi_i(z_0) = B_i(z_0)$, where $B_i(z_0)$ satisfies the functional equation (1.5.1) with the boundary condition (1.5.2).*

For each cooperative subgame $\bar{G}(z)$, $z \in CZ$, we write the functional equation for the components $\xi_i(z)$ of the imputation $\xi(z) \in C(z) \subset I(z)$ of type (1.5.1) and define the values $\gamma_i(z)$ from equation:

$$\xi_i(z) = \gamma_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \xi_i(y), \quad (1.5.3)$$

where $\xi(y) = (\xi_i(y) : i \in N)$ is an imputation belonging to the solution $C(y)$ of the cooperative subgame $\bar{G}(y)$. The boundary condition for $\gamma_i(z)$ is as follows:

$$\gamma_i(z) = \xi_i(z) \text{ for } z \in \{z : L(z) = \emptyset\}. \quad (1.5.4)$$

Lemma 1.1. *The vector $\gamma(z) = (\gamma_i(z) : i \in N)$ given by equation (1.5.3) with boundary condition (1.5.4) is an imputation distribution procedure.*

Proof. It is obvious that for terminal vertices and the vertices at which the probability of the game end equals one, the equality (1.5.4) holds, and properties 1 and 2 of Definition 1.9 are satisfied.

Now we prove that these conditions are satisfied for the remaining vertices of the cooperative subtree. From (1.5.3) we express the values $\gamma_i(z)$ and summing them

up over $i \in N$, and obtain

$$\sum_{i \in N} \gamma_i(z) = \sum_{i \in N} \xi_i(z) - (1 - q_k) \sum_{i \in N} \left(\sum_{y \in L(z)} p(y|z, \bar{a}^z) \xi_i(y) \right). \quad (1.5.5)$$

As we have

$$\begin{aligned} \xi(z) &= (\xi_i(z) : i \in N) \in C(z) \subset I(z), \\ \xi(y) &= (\xi_i(y) : i \in N) \in C(y) \subset I(y), \end{aligned}$$

then from (1.5.5) we obtain:

$$\sum_{i \in N} \gamma_i(z) = V(N, z) - (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(N, y). \quad (1.5.6)$$

From (1.5.6) and (1.3.3) it follows that $\sum_{i \in N} \gamma_i(z) = \sum_{i \in N} K_i^z(\bar{a}^z)$ for action profile $\bar{a}^z = (\bar{a}_i^z : i \in N)$, which has been realized in game $\Gamma(z)$ when the players used a cooperative strategy profile $\bar{\varphi}$. Therefore, $\gamma_i(z)$ satisfies Condition 1 from Definition 1.9.

Now we verify if Condition 2 of Definition 1.9 is satisfied. Specifically, we find the mathematical expectation of the sums $\gamma_i(z)$, defined by formula (1.5.3), along the vertices of the cooperative subtree. For the vertices $z \in \{z : L(z) = \emptyset \text{ or } q_k = 1\}$, Condition 2 is satisfied. Continue with the vertices of the cooperative subtree, from which the vertices mentioned above are reached with one stage. For these vertices, we obtain the equality:

$$\begin{aligned} B_i(z_l) &= \xi_i(z_l) - (1 - q_l) \sum_{y \in L(z_l)} p(y|z_l, \bar{a}^{z_l}) \xi_i(y) + (1 - q_l) \sum_{y \in L(z_l)} p(y|z_l, \bar{a}^{z_l}) \gamma_i(y) \\ &= \xi_i(z_l), \end{aligned}$$

because $\xi_i(y) = \gamma_i(y)$. Following from the terminal vertices to the initial one, we prove condition 2 of Definition 1.9. \square

§ 1.6 Subgame consistency of solution of cooperative stochastic game with finite duration

Before the game starts, players come to an agreement about cooperation, i. e., they agree to maximize the mathematical expectation of the total payoff of coalition N

and expect to receive the imputation $\xi(z_0) \in C(z_0)$. The game process takes place along the vertices of cooperative subtree $\bar{\Psi}(z_0)$. But since the stochastic structure of the game implies uncertainty in realization of vertices of the cooperative subtree, then moving along a certain path, that is, along the vertices of the cooperative subtree, does not yet ensure the support of cooperation. Indeed, players moving along the cooperative path get into cooperative subgames with the current initial states in which the same player may have different opportunities. Conditions of a conflict and players' opportunities involved in the conflict change over time. And it will be natural to require maintenance of the optimality principle or "approach" in the choice of solutions of cooperative subgames. But once at vertex $z \in CZ$, the sum of the remaining payments to player i may not be equal to the i th component of the imputation from solution $C(z)$ of a cooperative subgame $\bar{G}(z)$. Therefore, at vertex $z \in CZ$ player i may think whether it is worth keeping the cooperative agreement to act "jointly optimally" proposed before the game starts. Thus, player i may wish to deviate from the cooperative strategy profile. If this deviation is beneficial for at least one player, it means subgame inconsistency of imputation $\xi(z_0) \in C(z_0)$.

Definition 1.10. *An imputation $\xi(z_0) \in C(z_0)$ is called subgame-consistent in cooperative stochastic game $\bar{G}(z_0)$ if for each vertex $z \in CZ \cap (L(z_0))^k$ there exists the imputation distribution procedure $\beta(z) = (\beta_i(z) : i \in N)$ such that*

$$\xi_i(z) = \beta_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \xi_i(y), \quad (1.6.1)$$

and

$$\xi_i(z) = \beta_i(z), \quad z \in \{z : L(z) = \emptyset\}, \quad (1.6.2)$$

where $\xi(y) = (\xi_i(y) : i \in N)$ is an imputation belonging to solution $C(y)$ of cooperative subgame $\bar{G}(y)$.

Remark 1.2. *If $C(z_0)$ consists of more than one imputation, then the choice of imputation $\xi(z_0)$ is indefinite. If players have chosen a certain imputation $\xi(z_0) \in C(z_0)$ and decided to verify if it is subgame consistent, first it is necessary to verify condition (1.6.1) for vertex z_0 . This means to verify if there exists the imputation distribution procedure $\beta(z_0) = (\beta_i(z_0) : i \in N)$ satisfying condition (1.6.1) for*

imputation $\xi(y) \in C(y)$, where $y \in L(z_0)$. Obviously, there is indefiniteness in the choice of imputation $\xi(y) \in C(y)$, which in its turn should also be subgame-consistent in cooperative subgame $\bar{G}(y)$. This means that condition (1.6.1) should be satisfied for imputation $\xi(y) \in C(y)$. From Definition 1.10 it follows that this condition should be satisfied for all z from the set of vertices of the cooperative subtree.

Definition 1.11. We say that cooperative stochastic game $\bar{G}(z_0)$ has subgame-consistent solution $C(z_0)$ if all imputations $\xi(z_0) \in C(z_0)$ are subgame consistent.

Obviously, if the payments to the players are made at the vertices of the cooperative subtree in accordance with the initially defined payoff functions, it is impossible in general to achieve subgame consistency of the cooperative solution. This may lead to the breakup of the cooperative agreement. Therefore, the problem of finding a scheme or procedure of payments to the players at the vertices of the cooperative subtree in order to satisfy the property of subgame consistency of a cooperative solution becomes natural. For this we need to find such an imputation distribution procedure $(\beta_i(z) : i \in N)$ for all vertices $z \in CZ$, for which the conditions (1.6.1) and (1.6.2) will be satisfied.

Theorem 1.1. Let in the cooperative stochastic game $\bar{G}(z_0)$ and any subgame the cooperative solutions $C(z_0)$ and $C(z)$, $z \in CZ$, be nonempty. If for any $\xi(z) = (\xi_i(z) : i \in N) \in C(z)$ the imputation distribution procedure is defined by formula

$$\beta_i(z) = \xi_i(z) - (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \xi_i(y) \quad (1.6.3)$$

for any $z \in CZ$, $z \notin \{z : L(z) = \emptyset\}$, where $\xi(y) = (\xi_i(y) : i \in N) \in C(y)$, and by formula (1.6.2) for any $z \in \{z : L(z) = \emptyset\}$, then cooperative solution $C(z_0)$ is subgame-consistent.

Proof. To prove subgame consistency of the cooperative solution $C(z_0)$, it is required to prove that for each vector $\xi(z_0) \in C(z_0)$ conditions (1.6.1) and (1.6.2) are satisfied.

From Lemma 1.1 it follows that the payments, determined by formulas (1.6.3) and (1.6.2), are the components of the imputation distribution procedure. Condition

(1.6.1) follows from (1.6.3) taking into account that $\xi(y) = (\xi_i(y) : i \in N)$ belongs to the cooperative solution of subgame $\bar{G}(y)$. \square

The proposed method of implementing the imputation has an important property: at each vertex of the cooperative path, players are guided by the same “optimality principle” (property of subgame consistency) and, in this sense, have no reasons for interruption of the previously adopted cooperative agreement and deviation from the cooperative strategy profile. The sum of payments to the players at each vertex of the cooperative subtree is also equal to the sum of the payoffs received by the players at that vertex (condition 1 of Definition 1.9 of an imputation distribution procedure). The latter condition may be called a condition of attainability of the payments, since players redistribute the sum which they obtain in the game and do not take any funds outside.

Notice that Definition 1.10 does not require the nonnegativity of functions $\beta_i(z)$, $z \in CZ$. All imputations belonging to solution $C(z)$ will be subgame-consistent if solution is such that $C(z) \neq \emptyset$ for all vertices $z \in CZ$. This is possible if the payments to the players are not made according to their initially defined payoff functions in games along the cooperative path, but according to the imputation distribution procedure $\beta(z) = (\beta_1(z), \dots, \beta_n(z))$ defined by (1.6.1), (1.6.2) for all $z \in CZ$, where $\beta_i(z)$ is the payment to player i at the vertex $z \in CZ$. Moreover, the mathematical expectation of all payments to player i coincides with the mathematical expectation of the i th component of the imputation belonging to the solution chosen by the players. It follows from Theorem 1.1. Thus, players can agree on getting negative payments at some vertices to ensure that the cooperation is supported throughout the game in order to guarantee receiving the components of initially selected imputation $\xi(z_0)$ belonging to solution $C(z_0)$ of cooperative stochastic game $\bar{G}(z_0)$.

§ 1.7 Nonnegative components of imputation distribution procedure in stochastic games with finite duration.

Regularization of imputations

In this section, we consider the case when for any player $i \in N$ payoff function is non-negative: $K_i^z(a^z) \geq 0$ for all $z \in CZ$ and any a^z . Assume that the players are interested in receiving non-negative payments at each vertex of the cooperative subtree and at the same time they want to guarantee subgame consistency of the cooperative solution. In case when non-negativity of $\beta_i(z)$ cannot be guaranteed for all vertices $z \in CZ$, one can construct a new subgame-consistent solution based on the solution initially chosen by the players from set $C(z_0)$. We present how this is done when set $C(z_0) \subset I(z_0)$ is considered as a solution. Notice that this procedure can be applied to the imputations well-known in the classical “static” cooperative game theory (the core, the nucleolus, the Shapley value).

For each vertex $z \in CZ$ define new imputation distribution procedure by

$$\beta_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}_1^z, \dots, \bar{a}_n^z)}{V(N, z)} \xi_i(z), \quad (1.7.1)$$

where $\xi(z) = (\xi_1(z), \dots, \xi_n(z)) \in C(z)$, and $\bar{a}^z = (\bar{a}_1^z, \dots, \bar{a}_n^z)$ is the realization of the cooperative strategy profile $\bar{\varphi} = (\bar{\varphi}_1(\cdot), \dots, \bar{\varphi}_n(\cdot))$ at vertex $z \in CZ$ maximizing the sum of mathematical expectations of the players' payoffs in stochastic game $\bar{G}(z_0)$, $V(N, z)$ is the value of characteristic function of coalition N calculated for cooperative subgame $\bar{G}(z)$.

As $K_i^z(a^z) \geq 0$ for any vertex $z \in CZ$ and player $i \in N$, then $\beta_i(z) \geq 0$ for any vertex $z \in CZ$. Taking into account equation (1.7.1) and equity $\sum_{i \in N} \xi_i(z) = V(N, z)$, we obtain that the current payment $\beta_i(z)$ to player i in game $\Gamma(z)$ should be proportional to the i th component of the imputation $\xi(z) \in C(z)$ in cooperative subgame $\bar{G}(z)$ of stochastic game $\bar{G}(z_0)$.

Determine a new imputation for cooperative subgame $\bar{G}(z)$, where $z \in CZ$, and $z \in (L(z_0))^k$ on the basis of the “old” imputation $\xi(z)$ as a solution of the functional

equation

$$\hat{\xi}_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{\xi}_i(y) \quad (1.7.2)$$

with boundary condition

$$\hat{\xi}_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \xi_i(z) = \xi_i(z) \quad (1.7.3)$$

for $z \in \{z : L(z) = \emptyset\}$.

Construct a new characteristic function $\hat{V}(S, z)$ for any cooperative subgame $\bar{G}(z)$ for all $z \in CZ$ using functional equation

$$\hat{V}(S, z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} V(S, z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{V}(S, y) \quad (1.7.4)$$

with boundary condition

$$\hat{V}(S, z) = V(S, z) \text{ for } z \in \{z : L(z) = \emptyset\}. \quad (1.7.5)$$

Functions $\hat{V}(S, z)$ and $V(S, z)$ are superadditive, and $\hat{V}(N, z) = V(N, z)$ because $\hat{V}(N, z)$ and $V(N, z)$ satisfy the functional equation (1.3.3) with boundary condition (1.3.4).

For all vertices $z \in CZ$ and all imputations $\xi(z) \in C(z)$, we compute the regularized imputations $\hat{\xi}(z)$ and define the set of solutions $\hat{C}(z)$ as follows:

$$\hat{C}(z) = \left\{ \hat{\xi}(z) : \hat{\xi}_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{\xi}_i(y), \quad (1.7.6) \right. \\ \left. \hat{\xi}_i(z) = \xi_i(z) \text{ for } z \in \{z : L(z) = \emptyset\}, \xi(z) \in C(z) \right\}.$$

Definition 1.12. *The set $\hat{C}(z_0)$ defined by formula (1.7.6), is called a regularized solution of cooperative stochastic game $\bar{G}(z_0)$.*

Therefore, players have an opportunity to regularize the solution chosen at the beginning of the game so that at each vertex of stochastic game $\bar{G}(z_0)$ “new” solution $\hat{C}(z_0)$ is subgame-consistent and the payment to the players are non-negative. But the imputation belonging to the new regularized solution $\hat{C}(z_0)$, generally speaking,

will not be an imputation for cooperative game with characteristic function $V(S, z_0)$, defined by (1.3.5) and (1.3.6). It will be an imputation for a cooperative stochastic game with a new characteristic function $\hat{V}(S, z_0)$ defined by formulas (1.7.4), (1.7.5).

Theorem 1.2. *An imputation $\hat{\xi}(z) = (\hat{\xi}_1(z), \dots, \hat{\xi}_n(z))$, defined by formula (1.7.2) with boundary condition (1.7.3), is subgame consistent imputation in cooperative game $\langle N, \hat{V} \rangle$ where characteristic function $\hat{V}(S, z)$ is defined by functional equation (1.7.4) with boundary condition (1.7.5). Moreover, IDP defined by (1.7.1) is non-negative for any $i \in N, z \in CZ$.*

Proof. Subgame consistency follows from the method of construction of a “new” imputation $\hat{\xi}(z)$. The non-negativity of component

$$\frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \hat{\xi}_i(z),$$

is obvious because

$$K_i^z(a_1^z, \dots, a_n^z) \geq 0$$

for all $z \in Z$ and each player $i \in N$.

Now we prove that $\hat{\xi}(z) = (\hat{\xi}_1(z), \dots, \hat{\xi}_n(z))$ satisfies the properties of an imputation in cooperative game with characteristic function $\hat{V}(S, z)$, which is given by the functional equation (1.7.4) with boundary condition (1.7.5). To do this, for any player $i \in N$ and vertex $z \in CZ$, it is necessary to prove satisfaction of two properties:

1. $\sum_{i \in N} \hat{\xi}_i(z) = \hat{V}(N, z)$,
2. $\hat{\xi}_i(z) \geq \hat{V}(\{i\}, z)$.

The first property is obviously satisfied for vertices $z \in \{z : L(z) = \emptyset\}$ and $z \in CZ$. Now prove these properties for vertices $z \in \{z : L(z) \ni y \text{ and } L(y) = \emptyset\}$

and such that $z \in CZ$:

$$\begin{aligned}
\sum_{i \in N} \hat{\xi}_i(z) &= \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \sum_{i \in N} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} \left(p(y|z, \bar{a}^z) \sum_{i \in N} \hat{\xi}_i(y) \right) \\
&= \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} V(N, z) + (1 - q_k) \sum_{y \in L(z)} \left(p(y|z, \bar{a}^z) \hat{V}(N, z) \right) \\
&= V(N, z) = \hat{V}(N, z),
\end{aligned}$$

because $y \in \{y : L(y) = \emptyset\}$.

The second property is also obviously satisfied for the vertices $z \in \{z : L(z) = \emptyset\}$. We show that $\hat{\xi}_i(z) - \hat{V}(\{i\}, z) \geq 0$ for the vertices $z \in \{z : L(z) \ni y \text{ and } L(y) = \emptyset\}$, using formulas (1.7.2) и (1.7.4):

$$\begin{aligned}
\hat{\xi}_i(z) - \hat{V}(\{i\}, z) &= \frac{\sum_{i \in N} K_i(\bar{a}^z)}{V(N, z)} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{\xi}_i(y) \\
&\quad - \left\{ \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} V(\{i\}, z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{V}(\{i\}, y) \right\} \\
&= \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} (\xi_i(z) - V(\{i\}, z)) \\
&\quad + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) (\hat{\xi}_i(y) - \hat{V}(\{i\}, y)) \geq 0.
\end{aligned}$$

The first term is non-negative since $\xi(z)$ is an imputation of cooperative subgame $\bar{G}(z)$, and the second term is non-negative, because $y \in \{y : L(y) = \emptyset\}$. We prove recursively for the previous vertices $z \in CZ$ and proceeding until vertex z_0 . \square

It is important to know in what relation the set $\hat{C}(z)$ which is a regularized solution defined by the formula (1.7.6), and the set $\tilde{C}(z)$ which is the solution found for the cooperative subgame $\bar{G}(z)$ with the characteristic function $\hat{V}(S, z)$ (i. e., the solution constructed using the same rules as the solution $C(z) \subset I(z)$ for the cooperative subgame $\bar{G}(z)$). In the next paragraph we find the sets $\tilde{C}(z)$ and $\hat{C}(z)$ for the cooperative stochastic subgame $\bar{G}(z)$ when the solutions of stochastic game $\bar{G}(z_0)$ are the Shapley value and the core.

§ 1.8 Regularization of the Shapley value and the core in stochastic games with finite duration

We start with the case when players choose the single-point optimality principle—Shapley value—as a cooperative solution. The Shapley value calculated in cooperative stochastic game $\bar{G}(z_0)$, is denoted by $Sh(z_0) = (Sh_i(z_0) : i \in N)$, and in cooperative subgame $\bar{G}(z)$, $z \in CZ$, by $Sh(z) = (Sh_i(z) : i \in N)$.

Define the regularized Shapley value in cooperative subgame $\bar{G}(z)$, where $z \in CZ$ and $z \in (L(z_0))^k$ based on the Shapley value of the initially given game as a solution of the functional equation

$$\hat{Sh}_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} Sh_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{Sh}_i(y) \quad (1.8.1)$$

with boundary condition

$$\hat{Sh}_i(z) = Sh_i(z) \quad (1.8.2)$$

for $z \in \{z : L(z) = \emptyset\}$.

The following theorem holds.

Theorem 1.3. *Vector satisfying functional equation (1.8.1) with boundary condition (1.8.2), is subgame-consistent and it is the Shapley value of the cooperative subgame $\langle N, \hat{V}(\cdot, z) \rangle$, $z \in CZ$, of stochastic game $\langle N, \hat{V}(\cdot, z_0) \rangle$, where the values of characteristic function $\hat{V}(\cdot, z)$ are calculated by formulas (1.7.4) and (1.7.5).*

Remark 1.3. *Theorem 1.3 provides the relation between sets $\tilde{C}(z)$ and $\hat{C}(z)$, which are mentioned at the end of the previous paragraph. If the Shapley value is chosen as a solution of the stochastic game $\bar{G}(z_0)$, then $\tilde{C}(z) = \hat{C}(z)$ for any $z \in CZ$. Therefore, we may reformulate Theorem 1.2 in the following way.*

Theorem 1.4. *Vector satisfying equation (1.8.1) with boundary condition (1.8.2), is subgame-consistent, and $\hat{Sh}(z_0) = \hat{C}(z_0) = \tilde{C}(z_0)$, where $\hat{C}(z_0)$ is a regularized solution satisfying equation (1.7.6), and $\tilde{C}(z_0)$ is the Shapley value of cooperative stochastic game $\langle N, \hat{V}(\cdot, z_0) \rangle$ with characteristic function given by formulas (1.7.4), (1.7.5).*

Proof. The fact that the vector satisfying functional equation (1.8.1) with initial condition (1.8.2) is subgame-consistent, follows from Theorem 1.1 which is formulated for a general case, i. e., for any solution $C(z)$.

Calculate the Shapley value of cooperative stochastic game $\langle N, \hat{V}(\cdot, z) \rangle$, $z \in CZ$, with regularized characteristic function given by formulas (1.7.4), (1.7.5):

$$\hat{S}h_i(z) = \sum_{\substack{S \subset N \\ S \ni i}} \frac{(|S| - 1)! (n - |S|)!}{n!} \left(\hat{V}(S, z) - \hat{V}(S \setminus \{i\}, z) \right).$$

Rewrite (1.7.4) for coalition $S \setminus \{i\}$ and obtain

$$\begin{aligned} \hat{V}(S \setminus \{i\}, z) &= \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} V(S \setminus \{i\}, z) \\ &\quad + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{V}(S \setminus \{i\}, y). \end{aligned} \quad (1.8.3)$$

Subtracting (1.8.3) from (1.7.4), multiplying by $\frac{(|S|-1)!(n-|S|)!}{n!}$ and summing up over all possible coalitions $S \subset N$ such that $S \ni i$, we obtain

$$\begin{aligned} &\sum_{\substack{S \subset N \\ S \ni i}} \frac{(|S| - 1)! (n - |S|)!}{n!} \left[\hat{V}(S, z) - \hat{V}(S \setminus \{i\}, z) \right] \quad (1.8.4) \\ &= \left\{ \sum_{\substack{S \subset N \\ S \ni i}} \frac{(|S| - 1)! (n - |S|)!}{n!} [V(S, z) - V(S \setminus \{i\}, z)] \right\} \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \\ &\quad + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \times \\ &\quad \times \left\{ \sum_{\substack{S \subset N \\ S \ni i}} \frac{(|S| - 1)! (n - |S|)!}{n!} \left[\hat{V}(S, y) - \hat{V}(S \setminus \{i\}, y) \right] \right\} \\ &= Sh_i(z) \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{S}h_i(y). \end{aligned}$$

The result of the theorem follows from (1.8.4) and (1.8.1). \square

Now we assume that the core is chosen as a solution of cooperative stochastic game $\bar{G}(z_0)$. As before, we suppose that $CO(z) \neq \emptyset$ for any vertex $z \in CZ$. We

also assume that $CO(z_0)$ is not subgame-consistent, i. e., there exists at least one imputation $\xi(z_0) \in CO(z_0)$ for which the condition of subgame consistency is not satisfied.

Definition 1.13. *The regularized core of stochastic game $\bar{G}(z_0)$ is the set:*

$$\begin{aligned} \widehat{CO}(z_0) = & \left\{ \hat{\xi}(z_0) : \hat{\xi}_i(z_0) = \frac{\sum_{i \in N} K_i^{z_0}(\bar{a}^{z_0})}{V(N, z_0)} \xi_i(z_0) + (1 - q_0) \sum_{y \in L(z_0)} (p(y|z_0, \bar{a}^{z_0}) \hat{\xi}_i(y)), \right. \\ & \left. \hat{\xi}_i(z_0) = \xi_i(z_0), z_0 \in \{z : L(z) = \emptyset\}, \xi(z_0) \in CO(z_0) \right\}. \end{aligned} \quad (1.8.5)$$

Definition 1.14. *The regularized core of cooperative subgame $\bar{G}(z)$ is the set*

$$\begin{aligned} \widehat{CO}(z) = & \left\{ \hat{\xi}(z) : \hat{\xi}_i(z) = \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \hat{\xi}_i(y) \right. \\ & \left. \hat{\xi}_i(z) = \xi_i(z) \text{ for } z \in \{z : L(z) = \emptyset\}, \xi(z) \in CO(z) \right\}. \end{aligned} \quad (1.8.6)$$

Denote by $\widetilde{CO}(z)$ the core calculated for cooperative subgame $\langle N, \hat{V}(\cdot, z) \rangle$, $z \in CZ$, with characteristic function $\hat{V}(S, z)$, defined by formulas (1.7.4), (1.7.5). We prove the theorem providing the relation between $\widetilde{CO}(z)$ and $\widehat{CO}(z)$.

Theorem 1.5. *The regularized core defined by formula (1.8.5) is a subgame-consistent solution. Moreover, $\widehat{CO}(z_0) \subset \widetilde{CO}(z_0)$, where $\widetilde{CO}(z_0)$ is the core of cooperative stochastic game $\langle N, \hat{V}(\cdot, z) \rangle$ with characteristic function defined by formulas (1.7.4), (1.7.5).*

Proof. Subgame consistency of the core follows from Theorem 1.1. To prove that $\widehat{CO}(z_0) \subset \widetilde{CO}(z_0)$, we need to prove that any imputation $\hat{\xi}(z_0) \in \widehat{CO}(z_0)$ belongs to set $\widetilde{CO}(z_0)$, which is equivalent to the following: for any $\hat{\xi}(z) \in \widehat{CO}(z)$, $z \in CZ$ and $S \subset N$ the inequality

$$\sum_{i \in S} \hat{\xi}_i(z) \geq \hat{V}(S, z) \quad (1.8.7)$$

is true.

The proof is obvious for the vertices $z \in \{z : L(z) = \emptyset\}$. Now we prove this

inequality for the vertices $z \in \{z : L(z) \ni y \text{ and } L(y) = \emptyset\}$:

$$\begin{aligned} \sum_{i \in S} \hat{\xi}_i(z) &= \frac{\sum_{i \in N} K_i^z(\bar{a}^z)}{V(N, z)} \sum_{i \in S} \xi_i(z) + (1 - q_k) \sum_{y \in L(z)} \left(p(y|z, \bar{a}^z) \sum_{i \in S} \hat{\xi}_i(y) \right) \geq \\ &\geq \hat{V}(S, z), \end{aligned}$$

which is true because $y \in \{z : L(z) = \emptyset\}$ and $\sum_{i \in S} \xi_i(z) \geq V(S, z)$, as $\xi(z)$ is the imputation belonging to the core $CO(z)$.

The following part of the proof is made for the next vertices up to the initial vertex z_0 like in the proof of Theorem 1.1. \square

Now we consider examples how to construct and regularize the solution in cooperative stochastic games defined on finite graphs.

Example 1.1 [42] Consider stochastic game $G(z_0)$ defined on graph $\Psi(z_0)$ which is represented on Fig. 1.1.

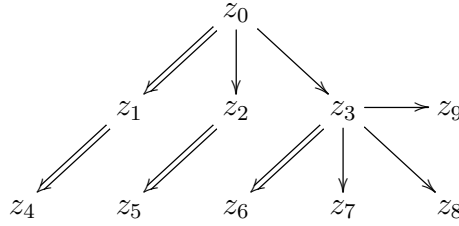


Figure 1.1: Graph for Example 1.1

The set of vertices of graph $\Psi(z_0)$ is $Z = \{z_0, \dots, z_9\}$. The set of players is $N = \{1, 2\}$. In each vertex of graph $\Psi(z_0)$ two-player normal-form game $\Gamma(z)$, $z \in Z$, is given and the payoffs in these games are the following:

$$\begin{aligned} \Gamma(z_0) &: \begin{pmatrix} (5, 5) & (0, 8) \\ (8, 0) & (1, 1) \end{pmatrix}, & \Gamma(z_2) &: \begin{pmatrix} (3, 0) & (6, 4) \\ (5, 6) & (2, 2) \end{pmatrix}, \\ \Gamma(z_3) &: \begin{pmatrix} (1, 11) & (4, 2) \\ (1, 3) & (1, 1) \end{pmatrix}, & \Gamma(z_7) &: \begin{pmatrix} (1, 1) & (0, 2) \\ (2, 0) & (1, 2) \end{pmatrix}, \end{aligned}$$

$$\Gamma(z_8) : \begin{pmatrix} (5, 5) & (6, 1) \\ (1, 6) & (6, 6) \end{pmatrix}, \quad \Gamma(z_9) : \begin{pmatrix} (4, 2) & (3, 4) \\ (5, 6) & (1, 5) \end{pmatrix},$$

$$\Gamma(z_1), \Gamma(z_4), \Gamma(z_5), \Gamma(z_6) : \begin{pmatrix} (0, 0) & (1, 0) \\ (1, 0) & (0, 1) \end{pmatrix}.$$

To determine non-cooperative stochastic game $G(z_0)$ we need to define transition probabilities and probabilities of the game duration. First, define the transition probabilities. If in game $\Gamma(z_0)$ the action profile $(2, 2)$ is realised, then the game process $G(z_0)$ transits to the vertex z_2 with probability $1/3$ and to the vertex z_3 with probability $2/3$. If any other action profile different from $(2, 2)$ is realised (arrow \implies means the deterministic transition), then the transition to vertex z_1 is performed. At vertices z_1, z_2 when any action profile is played, the game process transits to vertices z_4 and z_5 respectively. If in the game $\Gamma(z_3)$ the action profile $(2, 2)$ is played, then the game process transits to vertices z_8 and z_9 with equal probabilities $1/2$. And if in the game $\Gamma(z_3)$ the action profile $(2, 1)$ is played, the transition to vertex z_7 is made with probability 1. The deterministic transition (with probability 1) is made from other action profiles to vertex z_6 .

Let probabilities q_k that the game ends at stage k be given:

$$q_1 = \frac{1}{8}, \quad q_2 = 0, \quad q_3 = 1.$$

Let the Shapley value be chosen as a cooperative solution of the game. For two-player game, it is calculated by formulas:

$$Sh_1(z) = V(\{1\}, z) + \frac{V(\{1, 2\}, z) - V(\{1\}, z) - V(\{2\}, z)}{2},$$

$$Sh_2(z) = V(\{2\}, z) + \frac{V(\{1, 2\}, z) - V(\{1\}, z) - V(\{2\}, z)}{2},$$

where $V(\{1\}, z)$ and $V(\{2\}, z)$ are the values of characteristic function for the sub-game beginning at vertex z calculated for coalitions $\{1\}$ and $\{2\}$ respectively.

The above described sets and values determine stochastic game with random duration $G(z_0)$ (see (1.1.1)).

We start to find the solution of the cooperative game from the terminal vertices of the graph, i. e., the vertices from which it is impossible to transmit to any

other vertices of the graph. First, calculate $V(\{1\}, z_9)$ and $V(\{2\}, z_9)$ as maximum guaranteed players' payoffs in the game $\Gamma(z_9)$ using formula (1.3.6):

$$V(\{1\}, z_9) = 3, \quad V(\{2\}, z_9) = 4, \quad V(\{1, 2\}, z_9) = 11.$$

Then, we may calculate the Shapley value of subgame $\bar{G}(z_9)$ starting from game $\Gamma(z_9)$:

$$Sh_1(z_9) = 5, \quad Sh_2(z_9) = 6.$$

We make the similar calculations for the subgames starting from the games $\Gamma(z_4)$, $\Gamma(z_5)$, $\Gamma(z_6)$, $\Gamma(z_7)$ and $\Gamma(z_8)$ using formula (1.3.6) while they are realised at the vertices belonging to set $\{z : L(z) = \emptyset\}$. The values of characteristic functions for these subgames and corresponding Shapley values are given in the Table 1.2.

Table 1.2: Characteristic functions and the Shapley values in subgames $\bar{G}(z)$, $z \in \{z_4, z_5, z_6, z_7, z_8, z_9\}$.

Vertex z	$V(\{1\}, z)$	$V(\{2\}, z)$	$V(\{1, 2\}, z)$	$Sh_1(z)$	$Sh_2(z)$
z_4	0	0	1	1/2	1/2
z_5	0	0	1	1/2	1/2
z_6	0	0	1	1/2	1/2
z_7	1	2	3	1	2
z_8	5	5	12	6	6
z_9	3	4	11	5	6

Now consider the vertices from the set $\{z : (L(z))^2 = \emptyset\}$. We start from vertex z_3 . As the game process transits to the other vertices of the graph, we need to transform the payoff matrix of the game to calculate the Shapley value of cooperative subgame $\bar{G}(z_3)$. With action profile (2,2) the mathematical expectations of the players' payoffs are found in the following way:

- for Player 1:

$$1 + (1 - q_2) \left(\frac{1}{2}V(\{1\}, z_8) + \frac{1}{2}V(\{1\}, z_9) \right) = 5,$$

- for Player 2:

$$1 + (1 - q_2) \left(\frac{1}{2}V(\{2\}, z_8) + \frac{1}{2}V(\{2\}, z_9) \right) = 5.5.$$

With action profile (2,1) they are

- for Player 1:

$$1 + (1 - q_2)V(\{1\}, z_7) = 2,$$

- for Player 2:

$$1 + (1 - q_2)V(\{2\}, z_7) = 3.$$

Similarly, with action profile (1,1) the mathematical expectations of the players' payoffs are

- for Player 1:

$$1 + (1 - q_2)V(\{1\}, z_6) = 1,$$

- for Player 2:

$$11 + (1 - q_2)V(\{2\}, z_6) = 11;$$

and with action profile (1,2) the mathematical expectations of the players' payoffs are

- for Player 1:

$$4 + (1 - q_2)V(\{1\}, z_6) = 4,$$

- for Player 2:

$$2 + (1 - q_2)V(\{2\}, z_6) = 2.$$

Then bi-matrix game written for the calculations of the values of characteristic functions $V(\{1\}, z_3)$ and $V(\{2\}, z_3)$ looks like

$$\begin{pmatrix} (1, 11) & (4, 2) \\ (2, 5) & (5, 5.5) \end{pmatrix}.$$

The values of characteristic function of cooperative subgame $\bar{G}(z_3)$ of the game $\bar{G}(z_0)$ for coalitions $\{1\}$, $\{2\}$ are

$$V(\{1\}, z_3) = 2, \quad V(\{2\}, z_3) = 5.$$

To calculate $V(\{1, 2\}, z_3)$ we use formula (1.3.5) and obtain bi-matrix game:

$$\begin{pmatrix} 12 + (1 - q_2)V(\{1, 2\}, z_6) & 6 + (1 - q_2)V(\{1, 2\}, z_6) \\ 4 + (1 - q_2)V(\{1, 2\}, z_7) & 2 + (1 - q_2)(0.5V(\{1, 2\}, z_8) + 0.5V(\{1, 2\}, z_9)) \end{pmatrix}$$

or in numeric form:

$$\begin{pmatrix} 13 & 7 \\ 7 & 13.5 \end{pmatrix}.$$

Therefore,

$$\begin{aligned} V(\{1, 2\}, z_3) &= 13, 5, \\ Sh_1(z_3) &= 5.25, \quad Sh_2(z_3) = 8.25. \end{aligned}$$

We make similar calculations for the cooperative subgame $\bar{G}(z_1)$:

$$\begin{aligned} V(\{1\}, z_1) &= 0, \quad V(\{2\}, z_1) = 0, \quad V(\{1, 2\}, z_1) = 2, \\ Sh_1(z_1) &= Sh_2(z_1) = 1, \end{aligned}$$

and for subgame $\bar{G}(z_2)$:

$$\begin{aligned} V(\{1\}, z_2) &= 3, \quad V(\{2\}, z_2) = 2, \quad V(\{1, 2\}, z_2) = 12, \\ Sh_1(z_2) &= 6.5, \quad Sh_2(z_2) = 5.5. \end{aligned}$$

For cooperative stochastic game $\bar{G}(z_0)$, bi-matrix game for the calculation of the values of characteristic function for coalitions $\{1\}$, $\{2\}$ can be found by formula (1.3.5). With action profile (2,2) the mathematical expectations of the players' payoffs are

- for Player 1:

$$1 + (1 - q_1) \left(\frac{1}{3}V(\{1\}, z_2) + \frac{2}{3}V(\{1\}, z_3) \right) = 3\frac{1}{24},$$

- for Player 2:

$$1 + (1 - q_1) \left(\frac{1}{3}V(\{2\}, z_2) + \frac{2}{3}V(\{2\}, z_3) \right) = 4\frac{1}{2}.$$

With action profile (2,1) the mathematical expectations of the players' payoffs are

- for Player 1:

$$8 + (1 - q_1)V(\{1\}, z_1) = 8,$$

- for Player 2:

$$0 + (1 - q_1)V(\{2\}, z_1) = 0.$$

Similarly, with action profile (1,1) the mathematical expectations of the players' payoffs are

- for Player 1:

$$5 + (1 - q_1)V(\{1\}, z_1) = 5,$$

- for Player 2:

$$5 + (1 - q_1)V(\{2\}, z_1) = 5.$$

With action profile (1,2) the mathematical expectations of the players' payoffs are

- for Player 1:

$$0 + (1 - q_1)V(\{1\}, z_1) = 0,$$

- for Player 2:

$$8 + (1 - q_1)V(\{2\}, z_1) = 8.$$

Finally, we obtain bi-matrix:

$$\begin{pmatrix} (5, 5) & (0, 8) \\ (8, 0) & (3\frac{1}{24}, 4\frac{1}{2}) \end{pmatrix},$$

$$V(\{1\}, z_0) = 3\frac{1}{24}, V(\{2\}, z_0) = 4\frac{1}{2}.$$

To calculate $V(\{1, 2\}, z_0)$ we form matrix game using formula (1.3.5):

$$\begin{pmatrix} 10 + (1 - q_1)V(\{1, 2\}, z_1) & 8 + (1 - q_1)V(\{1, 2\}, z_1) \\ 8 + (1 - q_1)V(\{1, 2\}, z_1) & 2 + (1 - q_1)(\frac{1}{3}V(\{1, 2\}, z_2) + \frac{2}{3}V(\{1, 2\}, z_3)) \end{pmatrix}$$

or in a numeric form:

$$\begin{pmatrix} 11\frac{3}{4} & 9\frac{3}{4} \\ 9\frac{3}{4} & 13\frac{3}{8} \end{pmatrix}.$$

We obtain:

$$V(\{1, 2\}, z_0) = 13\frac{3}{8}, \quad Sh_1(z_0) = 5\frac{23}{24}, \quad Sh_2(z_0) = 7\frac{5}{12}.$$

The set of vertices forming the cooperative subtree consists of the vertices $z_0, z_2, z_3, z_5, z_8, z_9$.

Now we verify if the imputation distribution procedure is non-negative. Calculations show that it is negative at vertex z_3 . This follows from equation (1.6.1), in which the vertex z_3 is used:

$$\begin{aligned} Sh_1(z_3) &= \beta_1(z_3) + (1 - q_1) \left[\frac{1}{2} \cdot Sh_1(z_8) + \frac{1}{2} \cdot Sh_1(z_9) \right], \\ 5.25 &= \beta_1(z_3) + (1 - 0) \left[\frac{1}{2} \cdot 5 + \frac{1}{2} \cdot 6 \right], \\ \beta_1(z_3) &= -0.25. \end{aligned}$$

As $\beta_1(z_3)$ is negative, we make the regularization of the Shapley value to construct a «new» non-negative Shapley value.

Determine the new Shapley value for the vertices of the cooperative subtree with vertices $z_0, z_2, z_3, z_5, z_8, z_9$ by formulas (1.8.1) and (1.8.2):

$$\begin{aligned} \hat{Sh}_1(z_5) &= 0.5, \hat{Sh}_1(z_8) = 6, \hat{Sh}_1(z_9) = 5, \\ \hat{Sh}_2(z_5) &= 0.5, \hat{Sh}_2(z_8) = 6, \hat{Sh}_2(z_9) = 6, \\ \hat{Sh}_1(z_2) &= \frac{11}{12} \cdot 6.5 + \frac{1}{2} = 6\frac{11}{24}, \\ \hat{Sh}_2(z_2) &= \frac{11}{12} \cdot 5.5 + \frac{1}{2} = 5\frac{13}{24}, \\ \hat{Sh}_1(z_3) &= \frac{2}{13.5} \cdot 5.25 + \left[\frac{1}{2} \cdot 5 + \frac{1}{2} \cdot 6 \right] = 6\frac{5}{18}, \\ \hat{Sh}_2(z_3) &= \frac{2}{13.5} \cdot 8.25 + \left[\frac{1}{2} \cdot 6 + \frac{1}{2} \cdot 6 \right] = 7\frac{2}{9}, \\ \hat{Sh}_1(z_0) &= \frac{2}{13\frac{3}{8}} \cdot 5\frac{23}{24} + \left(1 - \frac{1}{8} \right) \left[\frac{1}{3} \cdot 6\frac{11}{24} + \frac{2}{3} \cdot 6\frac{5}{18} \right] = 6\frac{80741}{184896} \approx 6.437, \\ \hat{Sh}_2(z_0) &= \frac{2}{13\frac{3}{8}} \cdot 7\frac{5}{12} + \left(1 - \frac{1}{8} \right) \left[\frac{1}{3} \cdot 5\frac{13}{24} + \frac{2}{3} \cdot 7\frac{2}{9} \right] = 6\frac{173491}{184896} \approx 6.938. \end{aligned}$$

The «new» vector is the Shapley value of the cooperative game with characteristic function defined by formulas (1.7.4), (1.7.5). It is subgame-consistent which follows from Theorem 1.4. Moreover, the components of IDP are non-negative.

For the game and subgames starting from vertices z_2, z_3, z_5, z_8 and z_9 , the new characteristic functions are presented in Table 1.3.

Remark 1.4. *The nucleolus may be chosen as a solution of the cooperative game [206]. Notice that the nucleolus a unique vector, so there are no problems with the*

Table 1.3: «New» characteristic functions.

Vertex z	$\hat{V}(\{1\}, z)$	$\hat{V}(\{2\}, z)$	$\hat{V}(\{1, 2\}, z)$
z_0	3.763	4.265	13.375
z_2	2.750	1.833	12.000
z_3	4.296	5.574	13.500
z_5	0.000	0.000	1.000
z_8	5.000	5.000	12.000
z_9	3.000	4.000	11.000

choice of an imputation from the imputation set. And the nucleolus belongs to the core when the latter is non-empty.

Example 1.2. Consider stochastic game $G(z_0)$ defined on graph $\Psi(z_0)$ which is presented on Fig. 1.3. The set of vertices of graph $\Psi(z_0)$ is $Z = \{z_0, \dots, z_5\}$. The

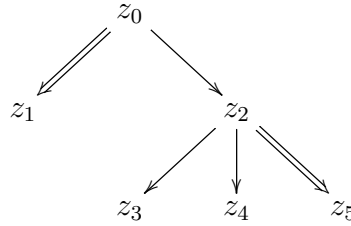


Figure 1.2: Graph of Example 1.2.

set of players is $N = \{1, 2, 3\}$. At each vertex three-player normal-form game $\Gamma(z)$, $z \in Z$, is given. The payoff matrices are the following:

$$\begin{aligned} \Gamma(z_0), \Gamma(z_2) &: \left(\begin{pmatrix} (1, 1, 1) & (2, 2, 0) \\ (2, 2, 0) & (0, 0, 3) \end{pmatrix} \begin{pmatrix} (1, 1, 2) & (2, 2, 1) \\ (1, 3, 1) & (3, 0, 1) \end{pmatrix} \right), \\ \Gamma(z_1), \Gamma(z_5) &: \left(\begin{pmatrix} (2, 0, 1) & (1, 0, 1) \\ (3, 1, 2) & (2, 2, 2) \end{pmatrix} \begin{pmatrix} (2, 2, 1) & (1, 1, 3) \\ (2, 1, 1) & (2, 1, 2) \end{pmatrix} \right), \\ \Gamma(z_3) &: \left(\begin{pmatrix} (1, 1, 1) & (2, 2, 2) \\ (3, 2, 0) & (1, 4, 1) \end{pmatrix} \begin{pmatrix} (2, 0, 1) & (2, 1, 1) \\ (4, 0, 1) & (0, 4, 1) \end{pmatrix} \right), \\ \Gamma(z_4) &: \left(\begin{pmatrix} (2, 1, 0) & (2, 1, 3) \\ (3, 1, 2) & (3, 6, 4) \end{pmatrix} \begin{pmatrix} (4, 5, 0) & (0, 5, 4) \\ (2, 8, 0) & (0, 8, 2) \end{pmatrix} \right). \end{aligned}$$

In each game defined above, Player 1 chooses a row, Player 2 chooses a column and Player 3 chooses a matrix.

First, we define the transition probabilities. If in game $\Gamma(z_0)$ action profile $(1,1,1)$ is played, then game process transits to vertex z_1 with probability $1/3$ and to vertex z_2 with probability $2/3$. Otherwise, if any action profile different from $(1,1,1)$ is played (arrow \implies means the deterministic transition), then the game process transits to vertex z_1 . If action profile $(2,1,2)$ is realised at vertex z_2 , transitions to the vertex z_3 and z_4 are made with probabilities $1/3$, $2/3$ respectively. If any other action profile different from $(2,1,2)$ is realised, game process transits to vertex z_5 with probability 1.

The probabilities q_k that stochastic game $G(z_0)$ ends at stage k are given:

$$q_1 = 0.5, \quad q_2 = 0, \quad q_3 = 1.$$

Let players choose the Shapley value as a solution of the game. We start solving the game with the vertices of the graph which belong to the set $\{z : L(z) = \emptyset\}$. We calculate the values of characteristic function and the Shapley value for subgame $\bar{G}(z_3)$. Similar calculations are made for the vertices z_1 , z_5 , z_4 , and then for the vertices z_2 and z_0 using formula (1.3.5). The results are presented in Tables 1.4 and 1.5.

Table 1.4: Characteristic functions for subgames $\bar{G}(z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

z	$V(\{1\}, z)$	$V(\{2\}, z)$	$V(\{3\}, z)$	$V(\{1, 2\}, z)$	$V(\{2, 3\}, z)$	$V(\{1, 3\}, z)$	$V(\{1, 2, 3\}, z)$
z_0	2	1	$3/2$	$11/2$	$9/2$	2	$83/9$
z_1	2	0	1	3	4	3	6
z_2	3	1	$4/3$	7	6	7	$47/3$
z_3	1	1	1	4	4	3	6
z_4	0	1	0	8	9	5	13
z_5	2	0	1	3	4	3	6

The set of vertices of a cooperative subtree is $CZ = \{z_0, z_1, z_2, z_3, z_4\}$. Let regularize the Shapley value

$$Sh(z_0) = \left(3\frac{31}{54}, 2\frac{89}{108}, 2\frac{89}{108} \right)$$

and verify if the imputation distribution procedure of the Shapley value is non-negative. For this, we find values $\beta_i(z)$ for vertices $z_0 \in CZ$ and $z_2 \in CZ$ using

Table 1.5: The Shapley values in subgames $\bar{G}(z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

z	$Sh_1(z)$	$Sh_2(z)$	$Sh_3(z)$
z_0	193/54	305/108	305/108
z_1	8/3	7/6	13/6
z_2	37/6	14/3	29/6
z_3	11/6	7/3	11/6
z_4	10/3	35/6	23/6
z_5	8/3	7/6	13/6

formula (1.6.3) and verify if imputation distribution procedure $\beta_i(z)$ is non-negative:

$$\beta_i(z_2) = Sh_i(z_2) - (1 - q_2) \left(\frac{1}{3} Sh_i(z_3) + \frac{2}{3} Sh_i(z_4) \right),$$

$$\beta_i(z_0) = Sh_i(z_0) - (1 - q_1) \left(\frac{1}{3} Sh_i(z_1) + \frac{2}{3} Sh_i(z_2) \right).$$

We obtain

$$\beta_1(z_2) = 3\frac{1}{3}, \quad \beta_2(z_2) = 0, \quad \beta_3(z_2) = 1\frac{2}{3};$$

$$\beta_1(z_0) = 1\frac{2}{27}, \quad \beta_2(z_0) = 1\frac{2}{27}, \quad \beta_3(z_0) = \frac{23}{27}.$$

For $z \in \{z_0, z_2\}$ the following conditions: $\beta_i(z) \geq 0$ and $\sum_{i \in N} \beta_i(z_2) = 5$, $\sum_{i \in N} \beta_i(z_0) = 3$ are satisfied.

In all vertices of the cooperative subtree, conditions of subgame consistency and non-negativity of the Shapley value are satisfied. Therefore, we state that the Shapley value is subgame-consistent imputation in game $\bar{G}(z_0)$.

Now we repeat calculations assuming that players adopt the nucleolus as a solution of the game $\bar{G}(z_0)$. The nucleolus was initially proposed by David Schmeidler [206]. For the calculation of the nucleolus, one may use Matlab [136] and program TUGlab [218], written for calculation in cooperative game theory, or Mathematica [126] and program TUGames [139] written for the same tasks.

The characteristic function is calculated above. The nucleolus of the subgame $\bar{G}(z)$, $z \in CZ$, is denoted by $n(z) = (n_i(z) : i \in N)$.

We calculate the nucleoli for all subgames of the game $\bar{G}(z_0)$. The results are presented in Table 1.6.

Now we verify the subgame consistency of the nucleolus using formula (1.6.1)

Table 1.6: The nucleoli in subgames $\bar{G}(z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

z	$n_1(z)$	$n_2(z)$	$n_3(z)$
z_0	$3\frac{5}{9}$	$3\frac{1}{18}$	$2\frac{11}{18}$
z_1	$2\frac{1}{2}$	$1\frac{1}{4}$	$2\frac{1}{4}$
z_2	$6\frac{1}{3}$	$4\frac{1}{2}$	$4\frac{2}{3}$
z_3	$1\frac{2}{3}$	$2\frac{2}{3}$	$1\frac{2}{3}$
z_4	$2\frac{2}{3}$	$6\frac{2}{3}$	$3\frac{2}{3}$
z_5	$2\frac{1}{2}$	$1\frac{1}{4}$	$2\frac{1}{4}$

and calculate $\beta_i(z_2)$ for vertex z_2 by formula:

$$n_i(z_2) = \beta_i(z_2) + (1 - q_2)(p(z_3|z_2, \bar{a}^{z_2})n_i(z_3) + p(z_4|z_2, \bar{a}^{z_2})n_i(z_4)).$$

We obtain:

$$\beta_1(z_2) = 4, \quad \beta_2(z_2) = -\frac{5}{6}, \quad \beta_3(z_2) = 1\frac{5}{6}.$$

The nucleolus of the cooperative stochastic game $\bar{G}(z_0)$ is subgame-consistent if the payments to the players are made by IDP. The imputation distribution procedure is not non-negative, because $\beta_2(z_2) < 0$. We make the regularization of the nucleolus.

Calculate «new» nucleolus for each vertex $z \in CZ$ by formula (1.7.2) with initial condition (1.7.3). For vertex z_2 , we use the following formula:

$$\hat{n}_i(z_2) = \frac{\sum_{i \in N} K_i^{z_2}(\bar{a}^{z_2})}{V(N, z_2)} n_i(z_2) + (1 - q_2) \sum_{y \in L(z_2)} p(y|z_2, \bar{a}^{z_2}) \hat{n}_i(y),$$

for vertex z_0 , we use formula:

$$\hat{n}_i(z_0) = \frac{\sum_{i \in N} K_i^{z_0}(\bar{a}^{z_0})}{V(N, z_0)} n_i(z_0) + (1 - q_1) \sum_{y \in L(z_0)} p(y|z_0, \bar{a}^{z_0}) \hat{n}_i(y).$$

«New» nucleoli for the vertices of set CZ are given in Table 1.7.

Calculate characteristic function $\hat{V}(S, z)$ for each vertex $z \in CZ$ by formulas (1.7.4) and (1.7.5). Moreover, $\hat{V}(S, z_3) = V(S, z_3)$, $\hat{V}(S, z_4) = V(S, z_4)$, $\hat{V}(S, z_1) = \hat{V}(S, z_5) = V(S, z_1) = V(S, z_5)$. For the calculation of $\hat{V}(S, z_2)$ we use formula:

$$\hat{V}(S, z_2) = \frac{\sum_{i \in N} K_i^{z_2}(\bar{a}^{z_2})}{V(N, z_2)} V(S, z_2) + (1 - q_2) \sum_{y \in L(z_2)} p(y|z_2, \bar{a}^{z_2}) \hat{V}(S, y),$$

Table 1.7: The nucleoli in subgames $\bar{G}(z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

z	$\hat{n}_1(z)$	$\hat{n}_2(z)$	$\hat{n}_3(z)$
z_0	$3 \frac{3487}{140436}$	$3 \frac{128867}{280872}$	$2 \frac{69149}{93624}$
z_1	$2 \frac{1}{2}$	$1 \frac{1}{4}$	$2 \frac{1}{4}$
z_2	$4 \frac{50}{141}$	$6 \frac{217}{282}$	$4 \frac{51}{94}$
z_3	$1 \frac{2}{3}$	$2 \frac{2}{3}$	$1 \frac{2}{3}$
z_4	$2 \frac{2}{3}$	$6 \frac{2}{3}$	$3 \frac{2}{3}$
z_5	$2 \frac{1}{2}$	$1 \frac{1}{4}$	$2 \frac{1}{4}$

and for $\hat{V}(S, z_0)$:

$$\hat{V}(S, z_0) = \frac{\sum_{i \in N} K_i^{z_0}(\bar{a}^{z_0})}{V(N, z_0)} V(S, z_0) + (1 - q_1) \left(p(z_1|z_0, \bar{a}^{z_0}) \hat{V}(S, z_1) + p(z_2|z_0, \bar{a}^{z_0}) \hat{V}(S, z_2) \right).$$

The values of function $\hat{V}(S, \cdot)$ are given in Table 1.8.

Table 1.8: Characteristic function $\hat{V}(S, z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

z	$\hat{V}(\{1\}, z)$	$\hat{V}(\{2\}, z)$	$\hat{V}(\{3\}, z)$	$\hat{V}(\{1, 2\}, z)$	$\hat{V}(\{2, 3\}, z)$	$\hat{V}(\{1, 3\}, z)$	$\hat{V}(\{1, 2, 3\}, z)$
z_0	$1 \frac{245}{249}$	$\frac{164}{249}$	$1 \frac{74}{747}$	$4 \frac{155}{249}$	$3 \frac{80}{83}$	$4 \frac{52}{83}$	$9 \frac{2}{9}$
z_1	2	0	1	3	4	3	6
z_2	$1 \frac{41}{141}$	$1 \frac{15}{47}$	$\frac{107}{141}$	$8 \frac{127}{141}$	$9 \frac{35}{141}$	$6 \frac{80}{141}$	$15 \frac{2}{3}$
z_3	1	1	1	4	4	3	6
z_4	0	1	0	8	9	5	13
z_5	2	0	1	3	4	3	6

Notice that the “new” nucleolus $\hat{n}(z_2)$ of subgame $\bar{G}(z_2)$ belongs to the imputation set with characteristic function $\hat{V}(S, z_2)$ (the nucleolus $\hat{n}(z_2)$ also belongs to the set $I(z_2)$, which is not true in general), but it is not the nucleolus of the cooperative game with initially defined characteristic function. The nucleolus of cooperative game defined by characteristic function $\hat{V}(S, z_2)$, is denoted by $\tilde{n}(z_2) = (\tilde{n}_1(z_2), \tilde{n}_2(z_2), \tilde{n}_3(z_2))$. It equals to the following one:

$$\tilde{n}(z_2) \approx (4.213, 6.894, 4.560) \neq \hat{n}(z_2).$$

The «new» nucleolus $\hat{n}(z_0)$ calculated for the game $\bar{G}(z_0)$, belongs to the imputation set of the cooperative game defined by characteristic function $\hat{V}(S, z_0)$

(nucleolus $\hat{n}(z_0)$ also belongs to the imputation set $I(z_0)$), but it does not coincide with the nucleolus of this cooperative game. The nucleolus of the cooperative game defined by characteristic function $\hat{V}(S, z_0)$, given above, is denoted by $\tilde{n}(z_0)$ and it equals

$$\tilde{n}(z_0) \approx (3.621, 2.720, 2.881) \neq \hat{n}(z_0).$$

§ 1.9 Strongly subgame consistency of the core in stochastic games with finite duration

In this section we consider the case when solution of the cooperative stochastic game is the set and contains more than one imputation. As an example of such a solution we examine the core. First, we describe the problem of subgame consistency and then find the sufficient conditions of strongly subgame consistency of the core. This problem was initially examined by Leon Petrosyan for differential games [36] and then for multicriteria problems of optimal control [37].

Suppose that the cores of stochastic game $\bar{G}(z_0)$ and any subgame $\bar{G}(z)$, $z \in CZ \setminus z_0$, are non-empty. When players cooperate, they come to an agreement about the realization of the cooperative strategy profile $\bar{\varphi}$ and expect to receive the components of the imputation belonging to the core $CO(z_0)$. Reaching the intermediate vertex $z \in CZ \setminus \{z_0\}$ of the cooperative subtree, player $i \in N$ chooses an action \bar{a}_i^z in accordance with the cooperative strategy $\bar{\varphi}_i$ and receives the payoff $K_i^z(\bar{a}^z)$. If the players recalculate the cooperative solution, i.e., find the solution of the cooperative subgame starting from vertex z , the recalculated solution will be the core $CO(z)$. It will be rational to require that the payoff received by the player in vertex z summarized with the expected sum of any imputations from solutions $CO(y)$, $y \in L(z)$, of the games of the cooperative subtrees following game $\Gamma(z)$, is equal to the imputation from solution $CO(z)$. If this property is satisfied for any vertex z of the cooperative subtree, the core of cooperative stochastic game $\bar{G}(z_0)$ is strongly subgame-consistent.

To introduce mathematically strict definition of strongly subgame-consistent core, it is necessary to define the so-called *expected core*. For any non-terminal vertex of the cooperative subtree we define the set of expected imputations belonging to the

cores which are the solutions of the subgames following the considered vertex. For any vertex $z \in CZ$, $L(z) \neq \emptyset$, define the *expected core*:

$$EC(L(z)) = \left\{ \alpha(L(z)) = \sum_{y \in L(z)} p(y|z, \bar{a}^z) \alpha(y) \mid \alpha(y) \in CO(y) \right\}. \quad (1.9.1)$$

The set $EC(L(z))$ consists of vectors $\alpha(L(z))$ which are the mathematical expectations of possible collection of the imputations from the cores of the subgames beginning from the vertices following vertex z with respect to the probability distribution $\{p(y|z, \bar{a}^z), y \in L(z)\}$.

We also define the distribution procedure of the players' payoffs in the vertices of a cooperative subtree. Refine Definition 1.9 of the imputation distribution procedure. The first condition in Definition 1.9 may be called the condition of "feasibility of the imputation distribution procedure" because it guarantees that in any vertex of the cooperative subtree the sum of the payments to the players equals the sum of the payoffs received by the players when they realize cooperative strategies. The second condition guarantees to the players that they receive the components of the initially chosen imputation from the core of cooperative game $\bar{G}(z_0)$ in the sense of mathematical expectation, if the payments to the players along the game are realized in accordance with imputation distribution procedure $\{\beta(z) : z \in CZ\}$.

Now we need to define the distribution procedure of the imputation $\alpha(z_0)$ from the core $CO(z_0)$ in a way that the core is strongly subgame-consistent.

Definition 1.15. *We call the core $CO(z_0)$ of cooperative stochastic game $\bar{G}(z_0)$ strongly subgame-consistent if for any the imputation from the core $CO(z_0)$ there exists the distribution procedure $\{\beta(z)\}_{z \in CZ}$ such that for each vertex $z \in CZ$ the inclusions take place:*

$$\beta(z) \oplus (1 - q_k)EC(L(z)) \subset CO(z), \quad (1.9.2)$$

$$B(z_0) \in CO(z_0), \quad (1.9.3)$$

where

$$\beta(z) \oplus (1 - q_k)EC(L(z)) = \left\{ \beta(z) + (1 - q_k)\alpha(L(z)) : \alpha(L(z)) \in EC(L(z)) \right\}.$$

And the imputation distribution procedure $\{\beta(z)\}_{z \in CZ}$ is called *strongly subgame-consistent*.³

Condition (1.9.2) means that the set of vectors which are equal to the sum of the imputation distribution procedure components at vertex z and the imputation from the expected core of the vertex z , belongs to the core of the subgame beginning from vertex z . This condition provides the restrictions on payments to the players in the games defined at vertices and often it is not satisfied for any game if the payments to the players are realised in accordance with initially defined payoff functions.

We impose additional restrictions on characteristic functions of subgames starting from the vertices of the cooperative subtree to obtain sufficient conditions of strongly subgame consistency of the core. Denote by $EV(S, L(z))$ the expected values of characteristic function calculated for coalition $S \subseteq N$ at the vertices following the vertex z :

$$EV(S, L(z)) = \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(S, y).$$

Denote by

$$\Delta V(S, z) = V(S, z) - (1 - q_k) EV(S, L(z))$$

the difference between the value of characteristic function at vertex z and expected value of characteristic function on condition that the game does not finish at vertex z . Denote by $\Delta CO(z)$ analogue of the core calculated using function $\Delta V(S, z)$. Now define sufficient condition of strongly subgame consistency of the imputation distribution procedure from the core $CO(z_0)$.

Theorem 1.6. *Let for each vertex $z \in CZ$ the core $CO(z)$ and the set $\Delta CO(z)$ be non-empty. For each vertex $z \in CZ$ distribution procedure $\{\beta(z) : z \in CZ\}$ of the imputation $\alpha(z_0)$ from the core $CO(z_0)$ satisfies the conditions:*

$$\beta(z) \in \Delta CO(z), \tag{1.9.4}$$

$$B(z_0) \in CO(z_0). \tag{1.9.5}$$

then imputation $\alpha(z_0) \in CO(z_0)$ and distribution procedure $\{\beta(z) : z \in CZ\}$ are strongly subgame-consistent.

³The sum denoted by sign \oplus is called Minkowski sum (see [207], in which some properties of this operator are proved).

Proof. We need to prove that any vector $\beta(z) \in \Delta CO(z)$ satisfying conditions (1.9.4) and (1.9.5) is strongly subgame-consistent distribution procedure of the imputation $\alpha(z_0) \in CO(z_0)$. So, the conditions (1.9.2) and (1.9.3) from Definition 1.15 hold. Condition (1.9.5) coincides with (1.9.3), therefore, it remains to show that the inclusion (1.9.2) holds for any vertex $z \in CZ$. Consider any vector $\alpha(L(z)) \in EC(L(z))$ for vertex z and calculate the sum $\beta(z) + (1 - q_k)\alpha(L(z))$. Verify if the latter vector belongs to the core $CO(z)$. Now calculate the sum of all components of the vector:

$$\begin{aligned} & \sum_{i \in N} \beta_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \sum_{i \in N} \alpha_i(y) \\ &= V(N, z) - (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(N, y) \\ &+ (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \sum_{i \in N} \alpha_i(y) = V(N, z), \end{aligned}$$

which carries out the property of collective rationality.

Now consider $S \subset N$:

$$\begin{aligned} & \sum_{i \in S} \beta_i(z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) \sum_{i \in S} \alpha_i(y) \\ & \geq V(S, z) + (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(S, y) \\ & - (1 - q_k) \sum_{y \in L(z)} p(y|z, \bar{a}^z) V(S, y) = V(S, z). \end{aligned}$$

By virtue of the arbitrariness of vertex $z \in CZ$, we make a conclusion that the imputation $\alpha(z_0)$ from the core of cooperative game $\bar{G}(z_0)$ and its procedure $\{\beta(z) : z \in CZ\}$ are strongly subgame-consistent. \square

When analogue of the core $\Delta CO(z)$ is non-empty for each vertex z of the cooperative subtree, Theorem 1.6 provides the method of construction of strongly subgame-consistent distribution procedure of the imputations from the core, equal $B_i(z_0)$ by condition (1.9.5). Notice that in a general case not all the imputations from the core can be realised using distribution procedure $\{\beta(z) : z \in CZ\}$ defined above.

Example 3. Consider stochastic game $G(z_0)$ defined on graph $\Psi(z_0)$ depicted on Fig. 1.3.

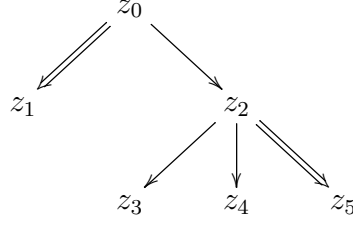


Figure 1.3: The tree $\Psi(z_0)$.

The set of the vertices of graph $\Psi(z_0)$ is $Z = \{z_0, \dots, z_5\}$. The set of the players is $N = \{1, 2, 3\}$. In each vertex of graph $G(z_0)$ the three-person normal-form game $\Gamma(z)$, $z \in Z$, is defined. The payoff matrices are

$$\begin{aligned} \Gamma(z_0) &: \left(\begin{pmatrix} (2, 2, 2) & (2, 2, 0) \\ (2, 2, 0) & (0, 0, 3) \end{pmatrix} \begin{pmatrix} (1, 1, 2) & (2, 2, 1) \\ (1, 3, 1) & (3, 0, 1) \end{pmatrix} \right), \\ \Gamma(z_2) &: \left(\begin{pmatrix} (1, 1, 1) & (2, 2, 0) \\ (2, 2, 0) & (0, 0, 3) \end{pmatrix} \begin{pmatrix} (1, 1, 2) & (2, 2, 1) \\ (1, 3, 1) & (3, 0, 1) \end{pmatrix} \right), \\ \Gamma(z_1), \Gamma(z_5) &: \left(\begin{pmatrix} (2, 0, 1) & (1, 0, 1) \\ (3, 1, 2) & (2, 2, 2) \end{pmatrix} \begin{pmatrix} (2, 2, 1) & (1, 1, 3) \\ (2, 1, 1) & (2, 1, 2) \end{pmatrix} \right), \\ \Gamma(z_3) &: \left(\begin{pmatrix} (1, 1, 1) & (2, 2, 2) \\ (3, 2, 0) & (1, 4, 1) \end{pmatrix} \begin{pmatrix} (2, 0, 1) & (2, 1, 1) \\ (4, 0, 1) & (0, 4, 1) \end{pmatrix} \right), \\ \Gamma(z_4) &: \left(\begin{pmatrix} (2, 1, 0) & (2, 1, 3) \\ (3, 1, 2) & (3, 6, 4) \end{pmatrix} \begin{pmatrix} (4, 5, 0) & (0, 5, 4) \\ (2, 8, 0) & (0, 8, 2) \end{pmatrix} \right). \end{aligned}$$

In each game the first player chooses a row, the second one chooses a column, the third one chooses a matrix. The strategy set of player $i \in N$ in game $\Gamma(z)$ is $A_i^z = \{1, 2\}$.

Define transition probabilities. If in game $\Gamma(z_0)$ the action profile $(1, 1, 1)$ is realised, the game process transits to vertex z_1 with a probability of $1/3$ and to vertex z_2 with a probability of $2/3$. If any action profile different from $(1, 1, 1)$ is realised (arrow \implies means the deterministic transition), the game process transits to vertex z_1 . If at vertex z_2 action profile $(2, 1, 2)$ is realised, transitions to vertices

z_3 and z_4 are made with probabilities of $1/3$, $2/3$ respectively. The game process transits to vertex z_5 with a probability of 1 from any other vertices.

The probabilities q_k that stochastic game $G(z_0)$ ends at stage k are given:

$$q_1 = 0.5, \quad q_2 = 0, \quad q_3 = 1.$$

To construct a cooperative version of stochastic game we find the cooperative strategy profile $\bar{\varphi}$. This profile $\bar{\varphi}$ prescribes to play action profile $(1, 1, 1)$ at vertex z_0 . The game ends at stage z_0 with probability 0.5 and transits to the next stage with a probability of 0.5. If the game does not end, it transits to stage z_1 with a probability of $1/3$, at which the players should realise any of action profiles $(2, 1, 1)$ or $(2, 2, 1)$, or with a probability of $2/3$ the game transits to vertex z_2 , at which the players should play action profile $(2, 1, 2)$. At vertex z_2 the game does not end because $q_2 = 0$ and transits to the vertices z_3 and z_4 with probabilities of $1/3$ and $2/3$ respectively. At vertices z_3 and z_4 the game terminates. Therefore, the set of the vertices of the cooperative subtree represented on Fig. 1.4 is $\bar{\Psi}(z_0) = \{z_0, z_1, z_2, z_3, z_4\}$.

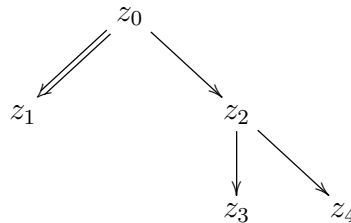


Figure 1.4: Cooperative subtree $\bar{\Psi}(z_0)$ of the game $\bar{G}(z_0)$.

Find the values of characteristic function using formulas (1.3.3) with boundary condition (1.3.4) for $S = N$, (1.3.5) with boundary condition (1.3.6) for $S \subset N$ and (1.3.7) for $S = \emptyset$. Results are given in Table 1.9. For further calculations we use package TUGlab of program Matlab [16].

Now we define the cores of subgames beginning from the vertices of cooperative subtree $\bar{\Psi}(z_0)$. We also assure that they all are non-empty to use the core as a cooperative solution of a stochastic game. The systems of linear inequalities and equities which determine the cores and their graphical representations are given in Tables 1.10 and 1.11. In figures, the imputation sets are depicted by light-gray color

Table 1.9: Characteristic functions $V(S, z)$ for $\bar{G}(z)$, $z \in \{z_0, z_1, z_2, z_3, z_4, z_5\}$.

$z \setminus S$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1, 2\}$	$\{2, 3\}$	$\{1, 3\}$	$\{1, 2, 3\}$
z_0	2	1	1.5	5.5	4.5	6	110/9
z_1	2	0	1	3	4	3	6
z_2	3	1	4/3	7	6	7	47/3
z_3	1	1	1	4	4	3	6
z_4	0	1	0	8	9	5	13
z_5	2	0	1	3	4	3	6

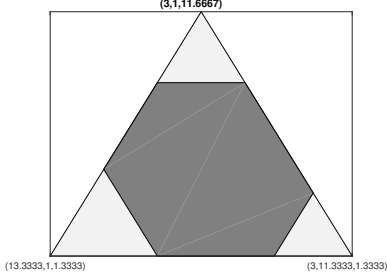
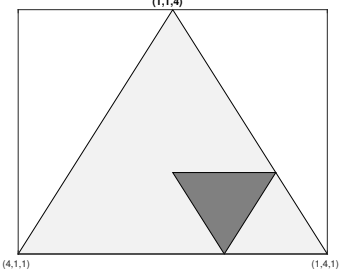
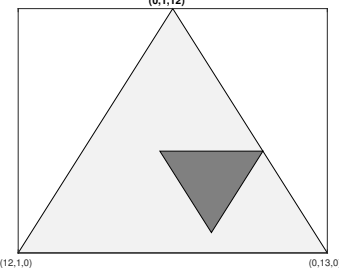
and the cores are dark-grey sets. Notice that at vertices z_1 and z_5 condition $\alpha_1 = 2$ holds for each element of the core. And the core is the segment connecting points $(2, 1, 3)$ and $(2, 3, 1)$.

Table 1.10: The core for vertices $z_0, z_1, z_5 \in CZ$.

z	Core	Graph of the core
z_0	$\begin{cases} \alpha_1 \geq 2 \\ \alpha_2 \geq 1 \\ \alpha_3 \geq 1.5 \\ \alpha_1 + \alpha_2 \geq 5.5 \\ \alpha_1 + \alpha_3 \geq 6 \\ \alpha_2 + \alpha_3 \geq 4.5 \\ \alpha_1 + \alpha_2 + \alpha_3 = 110/9 \end{cases}$	
z_1, z_5	$\begin{cases} \alpha_1 \geq 2 \\ \alpha_2 \geq 0 \\ \alpha_3 \geq 1 \\ \alpha_1 + \alpha_2 \geq 3 \\ \alpha_1 + \alpha_3 \geq 3 \\ \alpha_2 + \alpha_3 \geq 4 \\ \alpha_1 + \alpha_2 + \alpha_3 = 6 \end{cases}$	

For any vertex of cooperative subtree $\bar{\Psi}(z_0)$ we define the analogues of the cores denoted by $\Delta CO(z)$. Remind that for terminal vertices z_1, z_3, z_4 of set $\Delta CO(\cdot)$ coincide with the core $CO(\cdot)$. Systems of linear inequalities and equities determining sets $\Delta CO(z_0)$ and $\Delta CO(z_2)$ and also their graphs are presented in Tables 1.12.

Table 1.11: The core for the vertices $z_2, z_3, z_4 \in CZ$.

z_2	$\begin{cases} \alpha_1 \geq 3 \\ \alpha_2 \geq 1 \\ \alpha_3 \geq 4/3 \\ \alpha_1 + \alpha_2 \geq 7 \\ \alpha_1 + \alpha_3 \geq 7 \\ \alpha_2 + \alpha_3 \geq 6 \\ \alpha_1 + \alpha_2 + \alpha_3 = 47/3 \end{cases}$	
z_3	$\begin{cases} \alpha_1 \geq 1 \\ \alpha_2 \geq 1 \\ \alpha_3 \geq 1 \\ \alpha_1 + \alpha_2 \geq 4 \\ \alpha_1 + \alpha_3 \geq 3 \\ \alpha_2 + \alpha_3 \geq 4 \\ \alpha_1 + \alpha_2 + \alpha_3 = 6 \end{cases}$	
z_4	$\begin{cases} \alpha_1 \geq 0 \\ \alpha_2 \geq 1 \\ \alpha_3 \geq 0 \\ \alpha_1 + \alpha_2 \geq 8 \\ \alpha_1 + \alpha_3 \geq 5 \\ \alpha_2 + \alpha_3 \geq 9 \\ \alpha_1 + \alpha_2 + \alpha_3 = 13 \end{cases}$	

Notice that analogues of the cores $\Delta CO(\cdot)$ are non-empty for all vertices of the cooperative subtree. First, verify if the core is strongly subgame-consistent if the payments to the players are realised according to initially defined payoff functions, i. e., verify if payoff vectors in the vertices of the cooperative subtree belong to the

Table 1.12: Sets $\Delta CO(z)$ for vertices z_0 and z_2 .

z	$\Delta C(z)$	Graphs of $\Delta CO(z)$
z_0	$\begin{cases} \alpha_1 \geq 2/3 \\ \alpha_2 \geq 2/3 \\ \alpha_3 \geq 8/9 \\ \alpha_1 + \alpha_2 \geq 8/3 \\ \alpha_1 + \alpha_3 \geq 19/6 \\ \alpha_2 + \alpha_3 \geq 11/6 \\ \alpha_1 + \alpha_2 + \alpha_3 = 6 \end{cases}$	
z_2	$\begin{cases} \alpha_1 \geq 8/3 \\ \alpha_2 \geq 0 \\ \alpha_3 \geq 1 \\ \alpha_1 + \alpha_2 \geq 1/3 \\ \alpha_1 + \alpha_3 \geq 8/3 \\ \alpha_2 + \alpha_3 \geq -4/3 \\ \alpha_1 + \alpha_2 + \alpha_3 = 5 \end{cases}$	

corresponding sets $\Delta CO(\cdot)$ when the players realise cooperative strategy profile:

$$K^{z_0}(1, 1, 1) = (2, 2, 2) \in \Delta CO(z_0),$$

$$K^{z_1}(2, 2, 1) = (2, 2, 2) \in CO(z_1) = \Delta CO(z_1),$$

$$K^{z_2}(2, 1, 2) = (1, 3, 1) \notin \Delta CO(z_2),$$

$$K^{z_3}(1, 2, 1) = (2, 2, 2) \in CO(z_3) = \Delta CO(z_3),$$

$$K^{z_4}(2, 2, 1) = (3, 6, 4) \in CO(z_4) = \Delta CO(z_4).$$

We can easily see that at vertex z_2 the condition of inclusion is not satisfied and we can't guarantee strongly subgame consistency of an imputation from the core if the payments to the players are realised according to initially defined payoff functions.

We show that condition (1.9.2) does not hold at vertex z_2 . Following Definition 1.15 players may choose any imputation from the expected core of vertex z . Let them choose the imputations: $(1.5, 3, 1.5) \in CO(z_3)$ and $(0, 8, 5) \in CO(z_4)$, then the sum at the left-handed term of inclusion (1.9.2) takes form:

$$(1, 3, 1) + \frac{1}{3}(1.5, 3, 1.5) + \frac{2}{3}(0, 8, 5) = \left(\frac{3}{2}, \frac{28}{3}, \frac{29}{6} \right),$$

and this vector does not belong to the core $CO(z_2)$, which means that condition (1.9.2) does not hold and the core is not subgame-consistent.

Following Theorem 1.6, the set of vectors $\beta(z)$ belonging to $\Delta CO(z)$, $z \in CZ$, is the distribution procedure of an imputation from the core $CO(z_0)$ of initially defined game. By Theorem 1.6 we may also conclude that collection of vectors $(\beta(z) : z \in CZ)$ is not strongly subgame-consistent. For example, consider element from the set $\Delta C(z)$, $z \in CZ$: $\beta(z_0) = (4, 1, 1)$, $\beta(z_1) = (2, 2, 2)$, $\beta(z_2) = (3, 1, 1)$, $\beta(z_3) = (2, 2, 2)$, $\beta(z_4) = (3, 6, 4)$. Calculate the mathematical expectations of the players' payoffs if in the vertices of cooperative subtree they are paid in accordance with $\{\beta(\cdot)\}$:

$$\begin{aligned} B(z_0) &= (4, 1, 1) + 0.5 \left\{ \frac{1}{3}(2, 2, 2) + \frac{2}{3} \left((3, 1, 1) + \frac{1}{3}(2, 2, 2) + \frac{2}{3}(3, 6, 4) \right) \right\} = \\ &= \left(\frac{56}{9}, \frac{29}{9}, \frac{25}{9} \right). \end{aligned}$$

Obviously, $B(z_0) \in CO(z_0)$.

So, we have proposed a method of construction of strongly subgame-consistent imputation distribution procedure when the core is chosen as a cooperative solution.

Chapter 2

Cooperative Stochastic Games with Infinite Duration

§ 2.1 Noncooperative stochastic games with infinite duration

In this chapter we consider stochastic games with infinite duration defined by Shapley in the paper [210]. The main classical results on noncooperative stochastic games are presented in books [90, 149]. Similar to the previous chapter, the game is realised in a discrete time. The significant difference of this stochastic game from the game considered in Chapter 1 is that now the game has an infinite duration, the set of states which can be realised at any stage is finite and does not change over time. We define first a stochastic game and then describe the set of strategies and the payoff function of the player. Notice that the notations of this chapter which are widely used in modern literature on stochastic games are a bit different from the notations of Chapter 1.

Consider stochastic game G defined by

1. The finite set of players $N = \{1, \dots, n\}$.
2. The finite non-empty set of states $\Omega = \{1, \dots, \bar{\omega}\}$;
3. The finite, non-empty set of available actions A_i^ω of player $i \in N$ in state $\omega \in \Omega$. The action of player $i \in N$ in state $\omega \in \Omega$ is element $a_i^\omega \in A_i^\omega$. The action profile in state $\omega \in \Omega$ is a vector of players' actions $a^\omega = (a_i^\omega : i \in N)$. The set of action profiles in state ω is $A^\omega = A_1^\omega \times \dots \times A_n^\omega$.

4. The finite payoff function $K_i^\omega : \prod_{k \in N} A_k^\omega \rightarrow \mathbb{R}$, for every player $i \in N$ and every state $\omega \in \Omega$.
5. The transition function $p(\cdot|\omega, a^\omega) : \Omega \times A^\omega \rightarrow \Delta(\Omega)$ from state $\omega \in \Omega$ and action profile $a^\omega \in \prod_{i \in N} A_i^\omega$. Here $\Delta(\Omega)$ is probability distribution over set Ω .
6. The initial state determined by probability distribution

$$\pi_0 = (\pi_0^{\omega_1}, \dots, \pi_0^\omega, \dots, \pi_0^{\bar{\omega}}),$$

where π_0^ω is the probability that state ω is realised at the first stage of the game, $\sum_{\omega \in \Omega} \pi_0^\omega = 1$.

Time is discrete and game G lasts for an infinite number of stages denoted by t . Stochastic game G is realised in the following way:

1. Prior to the game, an initial state is chosen along the probability distribution π_0 , i. e., with probability π_0^ω stochastic game starts with state ω .
2. At the first stage, state ω is realised and players simultaneously choose their actions. Player i chooses action $a_i^\omega \in A_i^\omega$, $i \in N$. Thus the action profile $a^\omega = (a_i^\omega : i \in N) \in A_1^\omega \times \dots \times A_n^\omega$ is realised at the first stage. Player i receives payoff $K_i^\omega(a^\omega)$. Once a^ω is announced for all players, then the game transits to the next state $\omega' \in \Omega$ with probability $p(\omega'|\omega, a^\omega)$.
3. At the second stage, player $i \in N$ chooses action $a_i^{\omega'} \in A_i^{\omega'}$. Thus, at the second stage the action profile $a^{\omega'} = (a_i^{\omega'} : i \in N) \in A_1^{\omega'} \times \dots \times A_n^{\omega'}$ is played and player i receives payoff $K_i^{\omega'}(a^{\omega'})$.
4. The game further is played in the way described above.

Finally, let $\hat{a}_i^\omega \in \Delta(A_i^\omega)$ be a mixed action of player i in state ω , where $\Delta(A_i^\omega)$ is a probability measure over A_i^ω .

Definition 2.1. *A discounted stochastic game G is defined as*

$$G = \left\langle N, \Omega, \{A_i^\omega\}_{\substack{i \in N \\ \omega \in \Omega}}, \{K_i^\omega\}_{\substack{i \in N \\ \omega \in \Omega}}, \pi_0, \left\{ p(\omega''|\omega', a^{\omega'}) \right\}_{\substack{\omega', \omega'' \in \Omega \\ a^{\omega'} \in \prod_{i \in N} A_i^{\omega'}}}, \delta \right\rangle, \quad (2.1.1)$$

where $\delta \in (0, 1)$ is a discount factor, the same for all players.

Every state ω is determined by n -person normal-form game

$$\langle N, \{A_i^\omega\}_{i \in N}, \{K_i^\omega\}_{i \in N} \rangle.$$

A change of state may correspond to the presence of (positive or negative) shocks of different size. They will be reflected on the players' payoffs.

The subgame of noncooperative stochastic game G beginning from stage k is denoted by $G(k)$.

To solve a stochastic game, we need to define the class of players' strategies and the calculation method of players' payoffs in the whole game. First, define players' strategies and distinguish two classes of strategies:

- The behavior strategy of player $i \in N$ is a function $\varphi_i = \{\varphi_i(k)\}_{k=1}^\infty$ and $\varphi_i(k) : h(k) \times \Omega \mapsto \Delta(A_i^\omega)$, where $h(k)$ is a history of stage k , which is given by a collection of pairs consisting of states and action profiles which were realised at the previous stages until stage k : $((\omega(1), a(1)), (\omega(2), a(2)), \dots, (\omega(k-1), a(k-1)))$. Denote the set of behavior strategies of player i by Φ_i and behavior strategy profile in stochastic game by $\varphi = (\varphi_i : i \in N)$, $\varphi_i \in \Phi_i$.
- We also consider the subset of behavior strategies set, that is, the set of stationary strategies. A stationary strategy prescribes a player to choose the same strategy in the same state independently of the history of the stage. Denote a stationary strategy to distinguish behavior (not necessarily stationary) and stationary strategies. Denote a stationary strategy of player i by $\eta_i = \{\eta_i(k)\}_{k=1}^\infty$, $\eta_i(k) : \Omega \mapsto \Delta(A_i^\omega)$. Denote the profile of stationary strategies in a stochastic game by $\eta = (\eta_i : i \in N)$, and the set of stationary strategies of player i by H_i , while $H_i \subset \Phi_i$.

Now we determine players' payoffs in stochastic game (2.1.1):

- For the finite number of stages t a payoff of player i in a stochastic game is determined as a mathematical expectation:

$$E_i(\varphi) = E^{\omega(1), \varphi} \frac{1}{t} \sum_{k=1}^t K_i^{\omega(k)}(a(k)),$$

i. e., a mathematical expectation of a payoff with respect to the initial state $\omega(1)$ and strategy profile φ , while $K_i^{\omega(k)}(a(k))$ is a payoff of player i in state $\omega(k)$ realised at stage k , $a(k)$ is a strategy profile in state $\omega(k)$ realised at stage k in accordance with strategy profile φ .

- In case of infinite game G , a discounted payoff of player i is given by

$$E_i(\varphi) = E^{\omega(1),\varphi} \sum_{k=1}^{\infty} \delta^{k-1} K_i^{\omega(k)}(a(k)) \quad (2.1.2)$$

as a mathematical expectation of the payoff with respect to the initial state $\omega(1)$ and profile φ .

We formulate the main results on the existence of the value of a stochastic game and the Nash equilibrium in a stochastic game with two and more than two players which are used in the present work.

Theorem 2.1. [210] *A two-person zero-sum stochastic game with discount factor $\delta \in (0, 1)$ has a value for any initial state. Moreover, players' optimal strategies are stationary.*

This result was extended on the case of nonzero-sum games with more than two players by Fink and Takahashi in 1964:

Theorem 2.2. [91, 217] *A nonzero-sum stochastic game with many players with discount factor $\delta \in (0, 1)$ and finite set of states and strategies a Nash equilibrium exists for any initial state. Moreover, there exist equilibrium stationary strategies of the players.*

§ 2.2 Most functional equations for stochastic games with infinite duration and stationary strategies

In this section we provide formulas to calculate players' payoffs in a stochastic game when players use stationary strategies. Since the set of states Ω is finite, there are only $\bar{\omega}$ different subgames $G^{\omega_1}, \dots, G^{\bar{\omega}}$, each with initial states $\omega_1, \dots, \bar{\omega}$ respectively, because stationary strategies prescribe the same behavior in the same states

even with different histories of the current stage. We denote a non-cooperative stochastic subgame in stationary strategies with initial state $\omega \in \Omega$ by G^ω .

We now define the $\bar{\omega} \times \bar{\omega}$ -matrix of transition probabilities in G :

$$\Pi(\eta) = \begin{pmatrix} p(\omega_1|\omega_1, a^{\omega_1}) & \dots & p(\bar{\omega}|\omega_1, a^{\omega_1}) \\ p(\omega_1|\omega_2, a^{\omega_2}) & \dots & p(\bar{\omega}|\omega_2, a^{\omega_2}) \\ \dots & \dots & \dots \\ p(\omega_1|\bar{\omega}, a^{\bar{\omega}}) & \dots & p(\bar{\omega}|\bar{\omega}, a^{\bar{\omega}}) \end{pmatrix} \quad (2.2.1)$$

which is a function $p(\omega'|\omega, a^\omega)$ of a stationary strategy profile $\eta = (\eta_i : i \in N)$ such that $\eta_i(\omega) = a_i^\omega \in \Delta(A_i^\omega)$, $\omega \in \Omega$, $i \in N$, and $a^\omega = (a_1^\omega, \dots, a_n^\omega)$ for any state $\omega \in \Omega$. Matrix entry (2.2.1) which is the element of the j^{th} row and the j^{th} column is the probability to transit from state j^{th} to state j^{th} when players use strategy profile $\eta = (\eta_i : i \in N)$.

We simplify equation (2.1.2) for player i 's payoff, i.e., we calculate the mathematical expectation of his payoff in an explicit form. Let $E_i^\omega(\eta)$ be the expected payoff of player i in subgame G^ω when profile $\eta = (\eta_1, \dots, \eta_n)$ in stationary strategies is adopted. The vectorial form of the expected payoffs is $E_i(\eta) = (E_i^{\omega_1}(\eta), \dots, E_i^{\bar{\omega}}(\eta))^T$.

Hence a player i 's indirect utility function in subgame G^ω satisfies the following recurrent equation:

$$E_i^\omega(\eta) = K_i^\omega(a^\omega) + \delta \sum_{\omega' \in \Omega} p(\omega'|\omega, a^\omega) E_i^{\omega'}(\eta). \quad (2.2.2)$$

Given a matrix form of transition probabilities (2.2.1), rewrite equation (2.2.2) in a matrix form:

$$E_i(\eta) = K_i(a) + \delta \Pi(\eta) E_i(\eta), \quad (2.2.3)$$

where $K_i(a) = (K_i^{\omega_1}(a^1), \dots, K_i^{\bar{\omega}}(a^{\bar{\omega}}))^T$. Equation (2.2.3) is equivalent to the equation

$$E_i(\eta) = (\mathbb{I} - \delta \Pi(\eta))^{-1} K_i(a),$$

where \mathbb{I} is an identity matrix of size $\bar{\omega} \times \bar{\omega}$. Matrix $(\mathbb{I} - \delta \Pi(\eta))^{-1}$ always exists for discount factor $\delta \in (0, 1)$. The payoff of player i in game G in stationary strategies taking into account the initial state distributed with π_0 is

$$\bar{E}_i(\eta) = \pi_0 E_i(\eta) = \pi_0 (\mathbb{I} - \delta \Pi(\eta))^{-1} K_i(a). \quad (2.2.4)$$

§ 2.3 Cooperative stochastic game with infinite duration

We now develop the cooperative version of stochastic game G which is denoted by G_c . Suppose that players decide to cooperate by forming a grand coalition N with the aim to maximise total payoff. The existence of maximum of the discounted joint payoff follows from theorem proved in [210], according to which the cooperative strategy of the grand coalition that yields the maximal payoff is stationary. Denote the profile of pure stationary strategies of player i as $\eta_i \in H_i$, where $H_i \subset \Phi_i$.¹ The mixed stationary strategy is denoted as $\hat{\eta}_i \in \hat{H}_i$, with $H_i \subset \hat{H}_i$.

A *cooperative strategy profile* maximising the sum of the expected players' payoffs in G is denoted as $\eta^* = (\eta_1^*, \dots, \eta_n^*)$, where²

$$\max_{\eta \in \prod_{i \in N} H_i} \sum_{i \in N} \bar{E}_i(\eta) = \sum_{i \in N} \bar{E}_i(\eta^*). \quad (2.3.1)$$

We determine the values of a characteristic function for any coalition $S \subseteq N$. This function describes how much collective payoff players can gain by forming a coalition. We denote the characteristic function as $V(S) = (V^{\omega_1}(S), \dots, V^{\omega_{\bar{\omega}}}(S))$. Following [115], let $V(S)$ be the minmax value of two-person zero-sum game G_S between coalition S and coalition $N \setminus S$.³ Before introducing characteristic function, we first define the pure stationary strategies of coalition S and $N \setminus S$ as $\eta_S \in H_S = \prod_{i \in S} H_i$ and $\eta_{N \setminus S} \in H_{N \setminus S} = \prod_{i \in N \setminus S} H_i$, respectively.

Remark 2.1. *When we determine the characteristic function $V(S)$, $S \subseteq N$, we assume that players in S play in the interests of the coalition. Therefore, the actions and strategies of the players in S are correlated (Aumann, 1974).*

In state $\omega \in \Omega$, the correlated actions of the players from coalition S are $\hat{a}_S^\omega \in \Delta(A_S^\omega)$ where $A_S^\omega = \prod_{i \in S} A_i^\omega$. The correlated stationary strategy of players from coalition S and $N \setminus S$ are $\hat{\eta}_S(\omega) \in \Delta(A_S^\omega)$ and $\hat{\eta}_{N \setminus S}(\omega) \in \Delta(A_{N \setminus S}^\omega)$, respectively. Let the set of correlated stationary strategies of coalition S and $N \setminus S$ be \hat{H}_S and $\hat{H}_{N \setminus S}$, respectively.

¹From now on we use the notation η_i if player i uses the stationary strategy in the game. When a player i uses a behaviour strategy (not necessarily stationary), we use the notation φ_i .

²Without loss of generality we may find the maximum in equation (2.3.1) over the set of pure strategies of coalition N .

³The existence of the minmax value of two-player discounted stochastic game is proved by Shapley (1953a).

Begin the construction of the characteristic function by examining the grand coalition, $S = N$. The Bellman equation for the characteristic function $V(N)$ represents the discounted payoff of N :

$$V(N) = \max_{\eta \in \prod_{i \in N} H_i} \sum_{i \in N} \bar{E}_i(\eta) = \sum_{i \in N} K_i(a^*) + \delta \Pi(\eta^*) V(N), \quad (2.3.2)$$

where η^* is the cooperative strategy profile satisfying condition (2.3.1) and $\eta^*(\omega) = a^{\omega^*}$, $\omega \in \Omega$, and $K_i(a^*) = (K_i^{\omega_1}(a^{\omega_1^*}), \dots, K_i^{\bar{\omega}}(a^{\bar{\omega}^*}))^T$. From (2.3.2), we can infer the matrix form of $V(N)$:

$$V(N) = (\mathbb{I} - \delta \Pi(\eta^*))^{-1} \sum_{i \in N} K_i(a^*). \quad (2.3.3)$$

We define next the value of $V^\omega(S)$ of coalition S as the minmax payoff in the subgame G_S^ω starting from state ω :⁴

$$V^\omega(S) = \min_{\hat{\eta}_{N \setminus S}} \max_{\eta_S} \sum_{i \in S} E_i^\omega(\eta_S, \hat{\eta}_{N \setminus S}) = \max_{\hat{\eta}_S} \min_{\eta_{N \setminus S}} \sum_{i \in S} E_i^\omega(\hat{\eta}_S, \eta_{N \setminus S}). \quad (2.3.4)$$

The Bellman equation for the characteristic function $V^\omega(S)$ is

$$\begin{aligned} V^\omega(S) &= \min_{\hat{\eta}_{N \setminus S} \in \hat{H}_{N \setminus S}} \max_{\eta_S \in H_S} \sum_{i \in S} E_i^\omega(\eta_S, \hat{\eta}_{N \setminus S}) = \sum_{i \in S} E_i^\omega(\eta_S, \hat{\eta}_{N \setminus S}) \\ &= \sum_{i \in S} K_i^\omega(\mathbf{a}_S^\omega, \hat{\mathbf{a}}_{N \setminus S}^\omega) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, (\mathbf{a}_S^\omega, \hat{\mathbf{a}}_{N \setminus S}^\omega)) V^{\omega'}(S), \end{aligned} \quad (2.3.5)$$

where $(\mathbf{a}_S^\omega, \hat{\mathbf{a}}_{N \setminus S}^\omega)$ is a profile in correlated actions in state $\omega \in \Omega$ such that $\eta_S(\omega) = \mathbf{a}_S^\omega$, $\hat{\eta}_{N \setminus S}(\omega) = \hat{\mathbf{a}}_{N \setminus S}^\omega$, and $K_i(\mathbf{a}_S^\omega, \hat{\mathbf{a}}_{N \setminus S}^\omega) = (K_i^{\omega_1}(\mathbf{a}_S^{\omega_1}, \hat{\mathbf{a}}_{N \setminus S}^{\omega_1}), \dots, K_i^{\bar{\omega}}(\mathbf{a}_S^{\bar{\omega}}, \hat{\mathbf{a}}_{N \setminus S}^{\bar{\omega}}))$. We then rewrite equation (2.3.5) in a matrix form:

$$V(S) = (\mathbb{I} - \delta \Pi(\eta_S, \hat{\eta}_{N \setminus S}))^{-1} \sum_{i \in S} K_i(\mathbf{a}_S^\omega, \hat{\mathbf{a}}_{N \setminus S}^\omega). \quad (2.3.6)$$

Finally, we define the characteristic function $\bar{V}(S)$ for the whole stochastic game as:

$$\bar{V}(S) = \pi_0 V(S), \quad (2.3.7)$$

for any coalition $S \subseteq N$, where $V(S) = (V^{\omega_1}(S), \dots, V^{\bar{\omega}}(S))$, and $V^\omega(S)$ is the value of the characteristic function of subgame G^ω for S .

⁴In equation (2.3.4), the maximum in $\min_{\hat{\eta}_{N \setminus S}} \max_{\eta_S} \sum_{i \in S} E_i^\omega(\eta_S, \eta_{N \setminus S})$ is found over the set of pure strategies of coalition S , while the minimum in $\max_{\hat{\eta}_S} \min_{\eta_{N \setminus S}} \sum_{i \in S} E_i^\omega(\eta_S, \eta_{N \setminus S})$ is found over the set of pure strategies of coalition $N \setminus S$.

We assume that for any state $\omega \in \Omega$:

$$V^\omega(\emptyset) = 0. \quad (2.3.8)$$

The characteristic functions $\bar{V}(S)$ and $V^\omega(S)$ determined by (2.3.7) and (2.3.3)-(2.3.8), respectively, are superadditive [69]. In other words, for any disjoint coalitions $S, T \subset N$, and $S \cap T = \emptyset$, the inequality $V(S) + V(T) \leq V(S \cup T)$ holds. Superadditivity implies that the value of two disjoint coalitions is at least as great when they play together as when they act non-cooperatively. If superadditivity is not satisfied, then the coalition $S \cup T$ is not profitable, thus it will not be formed.⁵

We are now in a position to define the cooperative version of stochastic game 2.1 and its subgames.

Definition 2.2. *A cooperative stochastic game G_c , corresponding to a stochastic game G , is a set $\langle N, \bar{V} \rangle$, where N is the set of players and $\bar{V} : 2^N \rightarrow \mathbb{R}$ is the characteristic function calculated by (2.3.7). A cooperative stochastic subgame G_c^ω starting from state ω is a set $\langle N, V^\omega \rangle$, where $V^\omega : 2^N \rightarrow \mathbb{R}$ is the characteristic function calculated by (2.3.3), (2.3.4) and (2.3.8).*

When forming the grand coalition, players should decide not only what strategies to use to maximise the joint payoff but also how to allocate the total payoff. The next definitions display the allocation rule or solution (also called *imputation*) of G_c^ω and G_c , respectively. To determine an imputation of the joint payoff (2.3.1) we need to determine the values of the characteristic function for any coalition $S \subset N$.

Definition 2.3. *An imputation in the subgame G_c^ω , $\omega \in \Omega$, is a vector $\sigma^\omega = (\sigma_1^\omega, \dots, \sigma_n^\omega)$ satisfying: (i) $\sum_{i \in N} \sigma_i^\omega = V^\omega(N)$, and (ii) $\sigma_i^\omega \geq V^\omega(\{i\})$ for any $i \in N$. The set of imputations in G_c^ω is denoted as Σ^ω .*

Definition 2.4. *An imputation in the game G_c is a vector $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n)$, where $\bar{\sigma}_i = \pi_0 \sigma_i$, $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\omega_n})^T$, and $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$, $\omega \in \Omega$. The set of imputations in G_c is denoted as $\bar{\Sigma}$.*

⁵The property of superadditivity is not needed and it is often omitted in cooperative game theory, because in real life there are a lot of motivations to consider both profitable and non-profitable coalitions. As Aumann and Dreze (1974, p. 233) note, there are arguments for superadditivity that are quite persuasive, but, as they also note, superadditivity is quite problematic in some economic applications.

By Definition 2.3, an imputation satisfies the following conditions: (i) any player should obtain no less than she may get by non-cooperative play (*individual rationality condition*) and (ii) the sum of components of the imputation equals the value of the characteristic function corresponding to grand coalition (*group rationality condition*). The set of imputations is non-empty in any subgame G_c^ω , $\omega \in \Omega$ and in the whole cooperative stochastic game G_c , since the characteristic function determined by equations (2.3.2)-(2.3.8) is superadditive.

In the next paragraph we formulate and examine principles of stable cooperation for the class of stochastic games with infinite duration.

§ 2.4 Principles of stable cooperation in stochastic games with infinite duration

In cooperative games, the solution of a game is determined by an optimality principle. The optimality principle is assumed to be the subset of the imputation set. Therefore, the optimality principle contains one or more than one imputations but sometimes it may be empty. The Shapley value as an optimality principle always exists and contains a unique imputation. Therefore, the solution of a cooperative game always exists and it is unique according to this optimality principle. The solution of cooperative stochastic game means an imputation.⁶ Now we do not consider the problem of choosing a unique imputation from the set but assume that the optimality principle contains the only one imputation. The examples of one-point solutions are the Shapley value [211] and the nucleolus [206]. The realisation of an imputation in a cooperative stochastic game requires the satisfaction of some principles, which in turn ensure *stable cooperation* in a game. Following [44, 183], we formulate the main principles of stable cooperation including subgame consistency, strategic stability (or strategic support) and irrational-behaviour-proof of the solution of a cooperative stochastic game. Each principle of stable cooperation is defined and analysed separately.

⁶We further consider the case when the solution of a cooperative stochastic game is an imputation set.

§ 2.4.1 Subgame consistency of optimality principle in stochastic games with infinite duration

The principle of subgame consistency ensures that in any subgame cooperative solution is determined according to the initially chosen allocation rule. This concept deserves a detailed explanation. Players agree on cooperation before the game and adopt an imputation following the allocation mechanism. During the game, they play a cooperative strategy profile a_i^* , $i \in N$ which maximises their total payoff. In any subgame beginning in a certain state, a player is able to derive her expected payoff for the remainder of the game. If at some intermediate stage of the game players decide to calculate their expected payoffs in the subgame according to the initially defined payoff functions, then most often these expected discounted payoffs do not coincide with an imputation calculated in accordance with the initially chosen optimality principle. This means subgame inconsistency of a cooperative solution (or optimality principle). If for any subgame discounted players' payoffs coincide with the imputations calculated in accordance with initial optimality principle, cooperative solution (or optimality principle) is subgame consistent (see [35]). To make cooperative solution subgame consistent, we propose the transfer mechanism, called *imputation distribution procedure* (IDP).⁷ Originally, the idea of IDP was proposed by L. A. Petrosyan for differential games [43].

This mechanism leads to a modification of the players' payoffs in a dynamic game. We call the modified game as σ -regularisation, where σ is an initially chosen imputation in cooperative game G_c . This modified game ensures several advantages to the players. First, subgame consistency is ensured through the "new" payoff functions. Second, the expected payoffs in the regularised game will be equal to the components of the chosen imputation σ . Moreover, the sum of the stage payoffs in the regularised game is equal to the sum of the payoffs in the correspondent state of the initial game. For instance, suppose that players choose the Shapley value as an allocation rule in the game G_c . In this case, subgame consistency guarantees that, in each subgame, the vector of the players' payoff for the remaining stages is the

⁷Imputation distribution procedure was adapted for the class of discounted stochastic games in [72]. See Petrosjan and Danilov (1979), and Baranova and Petrosjan (2006).

Shapley value calculated for this subgame.

Let players adopt cooperative solution in stochastic game, i.e., they choose imputation $\sigma^\omega = (\sigma_1^\omega, \dots, \sigma_n^\omega)^T \in \Sigma^\omega$ for every subgame G_c^ω . The problem is to determine the transfers that ensure the expected payoff σ_i^ω for player i in every subgame G_c^ω .

When players implement cooperative stationary strategies in the stochastic game (2.1.1), the number of different subgames is equal to the number of possible states. Therefore, we need to determine a vector of payments $\beta_i = (\beta_i^{\omega_1}, \dots, \beta_i^{\bar{\omega}})^T$ for where β_i^ω is the payment of player $i \in N$ in state $\omega \in \Omega$.

Definition 2.5. *The set of payments $\{\beta_i\}_{i \in N}$ is IDP if the following conditions are satisfied:*

1. *In each state $\omega \in \Omega$, the sum of the payments is equal to the sum of players' payoffs in cooperative strategy profile η^* :*

$$\sum_{i \in N} \beta_i^\omega = \sum_{i \in N} K_i^\omega(a^{\omega*}). \quad (2.4.1)$$

2. *The expected sum of payments to player $i \in N$ in the game \bar{G} is equal to the i^{th} component of the initially chosen imputation $\bar{\sigma}$.*

We then define the conditions of subgame consistency for the imputation and IDP.

Definition 2.6. *Imputation $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n)$ and corresponding IDP $\{\beta_i\}_{i \in N}$ are called subgame consistent if the expected sum of payments to player i in each subgame G_c^ω is equal to the i^{th} component of the initially chosen imputation in subgame G_c^ω (in accordance with the principle imputation $\bar{\sigma}$ of the whole game is calculated).*

The following statement suggests the method of IDP construction for imputation $\bar{\sigma}$.

Lemma 2.1. *Let imputation $\bar{\sigma}$ be such that $(\bar{\sigma}_1, \dots, \bar{\sigma}_n) \in \bar{\Sigma}$ where $\bar{\sigma}_i = \pi_0 \sigma_i$, $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\bar{\omega}})^T$ and $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$, $\omega \in \Omega$. Then the collection $\{\beta_i\}_{i \in N}$ where β_i calculated by*

$$\beta_i = (\mathbb{I} - \delta \Pi(\eta^*)) \sigma_i, \quad (2.4.2)$$

is an imputation σ distribution procedure⁸ in game G .

Proof. Verify the first IDP condition:

$$\sum_{i \in N} \beta_i^\omega = \sum_{i \in N} K_i^\omega(a^{\omega*}),$$

where $a^{\omega*}$ is an action profile adopted under cooperative profile η^* in state ω . It is easy to show that β_i from (2.4.2) satisfies (2.4.1). Since $\sum_{i \in N} \beta_i^\omega$ is equal to $(\mathbb{I} - \delta\Pi(\eta^*)) \sum_{i \in N} \sigma_i = (\mathbb{I} - \delta\Pi(\eta^*))V(N)$, and $V(N)$ is determined by (2.3.3), then equation (2.4.1) holds.

The second IDP condition is satisfied since the expected total payoff of player i , denoted as B_i , with new payment β_i^ω in state $\omega \in \Omega$ satisfies the recurrent equation:

$$B_i^\omega = \beta_i^\omega + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) B_i^{\omega'},$$

or, in vectorial form:

$$B_i = \beta_i + \delta\Pi(\eta^*)B_i, \quad (2.4.3)$$

where $B_i = (B_i^{\omega_1}, \dots, B_i^{\omega_n})^T$. Equation (2.4.3) is equivalent to:

$$B_i = (\mathbb{I} - \delta\Pi(\eta^*))^{-1} \beta_i. \quad (2.4.4)$$

Given the second condition of IDP and equation (2.4.4) we obtain:

$$\sigma_i = (\mathbb{I} - \delta\Pi(\eta^*))^{-1} \beta_i, \quad (2.4.5)$$

where $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\omega_n})^T$, $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$. Equation (2.4.5) can be rewritten equivalently as:

$$\beta_i = (\mathbb{I} - \delta\Pi(\eta^*))\sigma_i. \quad (2.4.6)$$

Finally, equation (2.4.2) equals to:

$$\sigma_i = \beta_i + \delta\Pi(\eta^*)\sigma_i. \quad (2.4.7)$$

The second item in the right part of (2.4.7) is the expected value of the payments calculated for the subgame from the next stage onwards. Suppose that the

⁸Notice that IDP is uniquely defined by formula (2.4.2) if optimality principle provides unique cooperative solution $\bar{\sigma}$ (e.g., if the solution is nucleolus, the Shapley value or another single-valued solution). If the cooperative solution is the set of imputations containing more than one imputation, the method of IDP construction should be modified [171].

imputation for each subgame is chosen following the same allocation rule that has been chosen by the players at the beginning of the game. If players maintain cooperative strategy profile η^* , then the expected payoff of player i with new payments is equal to the correspondent component of imputation $\bar{\sigma}$ in cooperative stochastic game G_c . \square

Given Definition 2.5, for every imputation $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n) \in \bar{\Sigma}$, where $\bar{\sigma}_i = \pi_0 \sigma_i$, $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\bar{\omega}})^T$, $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$, we can define the regularization of stochastic game G as follows:

Definition 2.7. *A σ -regularisation of stochastic game G (subgame G^ω , $\omega \in \Omega$) is non-cooperative stochastic game G_σ (subgame G_σ^ω) if, for any player $i \in N$ in state ω , payoff function $K_i^{\sigma, \omega}(a^\omega)$ is defined as:*

$$K_i^{\sigma, \omega}(a^\omega) = \begin{cases} \beta_i^\omega, & \text{if } a^\omega = a^{\omega*}, \\ K_i^\omega(a^\omega), & \text{if } a^\omega \neq a^{\omega*}, \end{cases} \quad (2.4.8)$$

where β_i^ω is a component of PDP of player i defined by (2.4.2) and $a^{\omega*} = \eta^*(\omega)$.

Equation (2.4.8) determines the modified payoff function for game G .

Remark 2.2. *The σ -regularisation changes the payoff functions in any state $\omega \in \Omega$ only when action profiles $a^{\omega*} = \eta^*(\omega)$ are adopted. We may expect that players agree to modify the initial payoff functions to be sure that their cooperative solution satisfies the principle of subgame consistency.*

The following theorem shows that the players' payoffs in σ -regularization of initial stochastic game G satisfy the principle of subgame consistency.

Theorem 2.3. *Let $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n) \in \bar{\Sigma}$ be the initially chosen imputation in game G , where $\bar{\sigma}_i = \pi_0 \sigma_i$, $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\bar{\omega}})^T$, $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$, $\omega \in \Omega$, then σ -regularization of stochastic game G satisfies the principle of subgame consistency, i.e., the cooperative solution $\bar{\sigma}$ is subgame consistent in game G_σ .*

Proof. At the beginning of the game, players choose the following imputation: $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n) \in \bar{\Sigma}$, where $\bar{\sigma}_i = \pi_0 \sigma_i$, $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\bar{\omega}})^T$, $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in \Sigma^\omega$. A cooperative strategy profile is η^* . Consider the σ -regularization of game G

determined by Definition 2.7, thus the set of payments $\{\beta_i\}_{i \in N}$ defined by (2.4.2) is a IDP which follows from Lemma 2.1. To prove that the σ -regularisation of the game G satisfies the principle of subgame consistency, we need to calculate the discounted payoffs in every subgame of the game G_σ when a cooperative strategy profile η^* occurs. Consider any subgame G_σ^ω starting from state $\omega \in \Omega$. The discounted payoff of player i in this subgame is:

$$E_i^\omega(\eta^*) = \beta_i^\omega + \delta \sum_{\omega' \in \Omega} p(\omega'|\omega, a^{\omega^*}) E_i(\eta^*), \quad (2.4.9)$$

where $E_i(\eta^*) = (E_i^{\omega_1}(\eta^*), \dots, E_i^{\bar{\omega}}(\eta^*))^T$ and $E_i^\omega(\eta^*)$ is the discounted payoff of player i in subgame G_σ^ω starting from state ω when players adopt η^* . Equation (2.4.9) can be rewritten in a vector form:

$$E_i(\eta^*) = \beta_i + \delta \Pi(\eta^*) E_i(\eta^*),$$

or

$$E_i(\eta^*) = (\mathbb{I} - \delta \Pi(\eta^*))^{-1} \beta_i.$$

Since β_i satisfies (2.4.2), we obtain

$$E_i(\eta^*) = (\mathbb{I} - \delta \Pi(\eta^*))^{-1} (\mathbb{I} - \delta \Pi(\eta^*)) \sigma_i = \sigma_i.$$

This equation proves that σ -regularization of game G satisfies the principle of subgame consistency. \square

Definition 2.7 and Theorem 2.3 provide a method of constructing subgame consistent payments in every state of a stochastic game. The imputation distribution procedure $\beta_i^{\omega_1}, \dots, \beta_i^{\bar{\omega}}$ in states $\omega_1, \dots, \bar{\omega}$ ensures that a player i receives the same expected payoff in game G_σ (subgame G_σ^ω), as she planned to receive in cooperative stochastic game G_c (subgame G_c^ω). Moreover, the expected payoff from future payments is in line with the same allocation rule chosen by the players at the beginning of the game.

§ 2.4.2 Strategic stability of cooperative solution in stochastic games with infinite duration

The principle of strategic stability ensures that, along the game, an individual deviation from cooperative strategy profile in a regularized game does not yield a higher

payoff than cooperation. In other words, it guarantees the existence of the Nash equilibrium in a σ -regularized game with the same payoffs that players expect to receive with the cooperative solution (which was the basis of regularization). This principle was proposed in [38]. We reformulate the principle and then find conditions under which Nash equilibrium is subgame perfect (see [208] or [209]) in a σ -regularized game with the payoffs described above.

The subgame perfectness is important for dynamic games because it allows to guarantee the existence of the Nash equilibrium in any subgame with which the players' payoffs coincide with the cooperative ones. Comparing our approach with the standard analysis of deterministic (repeated) games, the condition of strategic stability for stochastic (or dynamic) games corresponds to the condition of the existence of subgame perfect Nash equilibrium in grim-trigger strategies. The main difference is that, in our setting, players first regularize the initial game by adapting the IDP to achieve subgame consistency.

Suppose players come to a cooperative agreement, i. e., find a cooperative strategy profile η^* that maximises the expected total payoff in the game. If a player deviates from the cooperative strategy profile, then the other players switch to trigger strategy from the next stage until forever to punish the deviating player. The strict definition of a behavior strategy used by players in Nash equilibrium is given below (see formula (2.4.12)). Here we assume that a stochastic game is the game with perfect monitoring, that is, all players know the state of a current stage and the history of the stage.

To begin with, we define the Nash equilibrium in a regularized stochastic game. Denote the expected payoff of player i in σ -regularisation of subgame G^ω starting from state ω as $E_i^{\omega,\sigma}$, $\omega \in \Omega$.

Definition 2.8. *A Nash equilibrium in the σ -regularised game G_σ is a behaviour strategy profile $\varphi^* = (\varphi_1^*, \dots, \varphi_n^*)$ such that, for any player $i \in N$, and for any state $\omega \in \Omega$, the condition*

$$E_i^{\omega,\sigma}(\varphi_i^*, \varphi_{N \setminus i}^*) \geq E_i^{\omega,\sigma}(\varphi_i, \varphi_{N \setminus i}^*) \quad (2.4.10)$$

holds for any behaviour strategy $\varphi_i \in \Phi_i$ of player i .

We assume that the behaviour strategy exhibits the following structure. If, in the

history of stage k , all players use their cooperative strategies, then they implement the cooperative correlated actions also in stage k . Conversely, if before stage k the individual deviation of a player $z \in N$ is observed, then the coalition $N \setminus z$ punishes player z . We assume that the punishment ensures that player z 's payoff is at most her minimax value in any subgame.⁹ Notice that, since we focus on a Nash equilibrium, we need to consider only individual deviations from this profile.¹⁰ If deviation occurs by more than one member of the coalition, the player may implement any strategy from the her set of strategies.

We now outline the condition under which the Nash equilibrium with players' payoffs equal the cooperative ones exists. For convenience, define

$$F(\{i\}) \equiv (F^{\omega_1}(\{i\}), \dots, F^{\bar{\omega}}(\{i\}))^T, \\ F^{\omega}(\{i\}) = \max_{\hat{a}_i^{\omega} \in \Delta(A_i^{\omega})} \left\{ K_i^{\omega}(\hat{a}_i^{\omega}, a_{N \setminus i}^{\omega*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, (\hat{a}_i^{\omega}, a_{N \setminus i}^{\omega*})) V^{\omega'}(\{i\}) \right\}.$$

The following inequality:

$$\sigma_i = (\mathbb{I} - \delta \Pi(\eta^*))^{-1} \beta_i \geq F(\{i\}), \quad (2.4.11)$$

compares two payoffs for each subgame: (i) the payoff when players adopt the cooperative strategy profile in the left hand side, and (ii) the payoff of deviation plus future punishment in the right hand side. If the first payoff is greater or equal to the second one, the player gets no benefit from deviation. If this is true for any player and any state, then the principle of strategic stability is satisfied. This result is summarised in the following proposition.

Theorem 2.4. *If in an σ -regularisation G_{σ} such that $\bar{\sigma} = \pi_0 \sigma$, inequality (2.4.11) holds for any player $i \in N$, then there exists behaviour strategy profile $\hat{\varphi}$ such that it is the Nash equilibrium with players' payoffs $(\bar{\sigma}_1, \dots, \bar{\sigma}_n)$.*

Proof. We determine the behaviour strategy profile $\hat{\varphi} = (\hat{\varphi}_1, \dots, \hat{\varphi}_n)$ where strate-

⁹The strict definition of the behaviour strategy is given in the proof of Proposition 2.4.

¹⁰Things change for subgame perfectness. In this case, we need to prove that eq. (2.4.10) holds for all possible histories and all stages. Therefore, we need to determine the strategy of a player even if more than one player deviates. Strategy (2.4.20) defines the behaviour of the player given any history.

gies $\widehat{\varphi}_i$, $i \in N$ are:

$$\widehat{\varphi}_i(h(k)) = \begin{cases} a_i^{\omega^*}, & \text{if } \omega(k) = \omega, h(k) \subset h^*; \\ \widehat{a}_i^\omega(z), & \text{if } \omega(k) = \omega, \text{ and } \exists l \in [1, k-1], \\ & z \in N, i \neq z: h(l) \subset h^*, \text{ and} \\ & (\omega(l), a(l)) \notin h^*, \text{ but} \\ & (\omega(l), (a_z^*(l), a_{N \setminus z}(l))) \in h^*; \\ \text{any} & \text{otherwise,} \end{cases} \quad (2.4.12)$$

where $a_i^{\omega^*}$ corresponds to the player i 's cooperative action, while $\widehat{a}_i^\omega(z) \in \Delta(A_i^\omega)$ is the player i 's punishment that, together with actions $\widehat{a}_{i'}^\omega(z) \in \Delta(A_{i'}^\omega)$, of players $i' \neq i$, $i' \in N \setminus z$, forms the action (either in pure or mixed strategies) of coalition $N \setminus z$ against player z .¹¹ The proof of the proposition follows from the folk theorem for stochastic games (Dutta, 1995) using the structure of the behaviour strategy (2.4.12). Notice that we do not define the reaction of players when they observe the deviations of more than one player. This because we focus here on the Nash equilibrium (not subgame perfect). When more than one player deviates, the player chooses any strategy from the player's set of strategies. We now prove that $\widehat{\varphi}(\cdot) = (\widehat{\varphi}_1(\cdot), \dots, \widehat{\varphi}_n(\cdot))$ determined in (2.4.12) is a NE in the stochastic game G_σ . Given strategy (2.4.12) and provided that all players do not deviate from a cooperative strategy profile η^* , the discounted payoff of player i in the subgame G_σ^ω , $\omega \in \Omega$, is:

$$E_i^\omega(\widehat{\varphi}) = E_i^\omega(\eta^*).$$

Let $E_i(\widehat{\varphi})$ be equal to the vector $(E_i^{\omega_1}(\widehat{\varphi}), \dots, E_i^{\omega_n}(\widehat{\varphi}))^T$. Then for any player $i \in N$ the next equation holds:

$$E_i(\widehat{\varphi}) = (\mathbb{I} - \delta\Pi(\eta^*))^{-1}\beta_i. \quad (2.4.13)$$

Consider next the profile of strategies $(\varphi_z, \widehat{\varphi}_{N \setminus z})$, when some player z deviates from strategy $\widehat{\varphi}_z$. For any k , there exists $l \in [1, k-1]$ such that $h(l) \subset h^*$ but $(\omega(k), a(k)) \notin h^*$ and $(\omega(k), (a_z^*(k), a_{N \setminus z}(k))) \in h^*$. Without loss of generality, we simplify $\omega(k) = \omega$. In words, the first individual deviation of player z occurs at

¹¹Notice that the actions of the players from coalition $N \setminus z$ are correlated.

stage k . We are now able to determine the total payoff of player z in the game G_σ with strategy profiles $(\varphi_z, \widehat{\varphi}_{N \setminus z})$ by

$$\bar{E}_z^\sigma(\varphi_z, \widehat{\varphi}_{N \setminus z}) = \pi_0 E_z^\sigma(\varphi_z, \widehat{\varphi}_{N \setminus z}),$$

where

$$E_z^\sigma(\varphi_z, \widehat{\varphi}_{N \setminus z}) = E_z^{\sigma, [1, k-1]}(\varphi_z, \widehat{\varphi}_{N \setminus z}) + \delta^{k-1} \Pi^{k-1}(\varphi_z, \widehat{\varphi}_{N \setminus z}) E_z^{\sigma, [k, \infty)}(\varphi_z, \widehat{\varphi}_{N \setminus z}). \quad (2.4.14)$$

The first term in the right hand side of (2.4.14) is the expected payoff of player z in the first $k - 1$ stages of the game G_σ , the second term is the expected payoff of player z in the subgame of G_σ beginning from stage k , where $E_z^{\sigma, [k, \infty)}(\varphi_z, \widehat{\varphi}_{N \setminus z})$ is the vector $(E_z^{\sigma, 1}(\varphi_z, \widehat{\varphi}_{N \setminus z}), \dots, E_z^{\sigma, \bar{\omega}}(\varphi_z, \widehat{\varphi}_{N \setminus z}))^T$, with $E_z^{\sigma, \omega}(\varphi_z, \widehat{\varphi}_{N \setminus z})$ being the player z 's expected payoff in the regularised subgame G_σ^ω beginning at state ω . Since there are no deviations from a cooperative strategy profile η^* up to stage $k - 1$, the following equalities hold:

$$\begin{aligned} E_z^{\sigma, [1, k-1]}(\varphi_z, \widehat{\varphi}_{N \setminus z}) &= E_z^{\sigma, [1, k-1]}(\eta^*), \\ \Pi^{k-1}(\varphi_z, \widehat{\varphi}_{N \setminus z}) &= \Pi^{k-1}(\eta^*). \end{aligned}$$

We now find the discounted payoff of player z in the subgame G_σ^ω beginning with stage k and when state $\omega(k)$ is equal to ω . The following formula takes place:

$$E_z^{\sigma, \omega}(\varphi_z, \widehat{\varphi}_{N \setminus z}) = K_z^\omega(\hat{a}_z^\omega, a_{N \setminus z}^{\omega*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, (\hat{a}_z^\omega, a_{N \setminus z}^{\omega*})) V^{\omega'}(\{z\}), \quad (2.4.15)$$

where $\hat{a}_z^\omega \in \Delta(A_z^\omega)$. Players from the coalition $N \setminus z$ punish player z by playing the strategies which allow player z to obtain her minmax payoff according to the definition of strategy profile $\widehat{\varphi}$. In (2.4.15), the value of the characteristic function $V^{\omega'}(\{z\})$ is determined by (2.3.4). Since the expected payoffs of player z in the strategy profiles $\widehat{\varphi}$ and $(\varphi_z, \widehat{\varphi}_{N \setminus z})$ do not change up to stage $k - 1$, then a deviation may increase player z 's payoff only at the expenses of the expected payoff in the subgame G_σ^ω , $\omega \in \Omega$. In particular, the strategy profile $(\varphi_z, \widehat{\varphi}_{N \setminus z})$ ensures the following expected payoff of player z from stage k :

$$F(\{z\}) = \max_{\hat{a}_z^\omega \in \Delta(A_z^\omega)} \left\{ K_z^\omega(\hat{a}_z^\omega, a_{N \setminus z}^{\omega*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, (\hat{a}_z^\omega, a_{N \setminus z}^{\omega*})) V^{\omega'}(\{z\}) \right\}. \quad (2.4.16)$$

According to the definition of PDP, the expected payoff of player z in the regularised subgame G_σ^ω with a profile of strategies $\widehat{\varphi}(\cdot)$ can be found from:

$$E_z^\sigma(\widehat{\varphi}) = (\mathbb{I} - \delta\Pi(\eta^*))^{-1}\beta_z = \sigma_z, \quad (2.4.17)$$

where $E_z^\sigma(\widehat{\varphi}) = (E_z^{\sigma,\omega_1}(\widehat{\varphi}), \dots, E_z^{\sigma,\bar{\omega}}(\widehat{\varphi}))^T$. Taking into account (2.4.11) from (2.4.16), (2.4.17) and the above discussion we get

$$E_z^\sigma(\widehat{\varphi}) \geq E_z^\sigma(\varphi_z, \widehat{\varphi}_{N\setminus z}),$$

which is satisfied when inequality

$$\sigma_z = (\mathbb{I} - \delta\Pi(\eta^*))^{-1}\beta_z \geq F(\{z\}) \quad (2.4.18)$$

is true. In inequality (2.4.18) is satisfied for any player $z \in N$, a player is not willing to deviate from the cooperative strategy profile in any subgame of the σ -regularisation of game G .

Thus the behaviour strategy profile (2.4.12) is a NE in the σ -regularisation of game G . The discounted payoff of i in the game G_σ with profile of strategies $\widehat{\varphi}$ is equal to $\bar{\sigma}_i$, where $\bar{\sigma}_i = \pi_0\sigma_i$, while $\sigma_i = (\sigma_i^{\omega_1}, \dots, \sigma_i^{\bar{\omega}})^T$ consists of i^{th} components of imputations $\sigma^{\omega_1}, \dots, \sigma^{\bar{\omega}}$ derived from the cooperative subgames $G^{\omega_1}, \dots, G^{\bar{\omega}}$ accordingly. \square

Notice that the players' strategies used in a punishing regime of the behaviour strategies (2.4.12) are not individually rational, i.e., player i punishing the deviated player z needs to implement the strategies minimizing the payoff of player z in a subgame which may be not profitable for player i and may motivate player i to deviate from strategy profile formed by (2.4.12). Therefore, the strategy profile determined by strategies (2.4.12) is not subgame perfect.

We investigate now the conditions to obtain a subgame perfect Nash equilibrium (SPNE) of the σ -regularisation of G . To do so, we need to determine the behaviour strategy profile such that, for any state occurring in any period with any history, individual deviation is not profitable.

We assume that, if the history of the stage differs from the cooperative history, then all players implement a Nash equilibrium of the game G denoted by $\eta^{ne} =$

$(\eta_1^{ne}, \dots, \eta_n^{ne})$ such that $\eta_i^{ne}(\omega) = \hat{a}_i^{\omega, ne} \in \Delta(A_i^\omega)$.¹² Again for convenience, define

$$Q(\{i\}) \equiv (Q^{\omega_1}(\{i\}), \dots, Q^{\bar{\omega}}(\{i\}))^T, \\ Q^\omega(\{i\}) = \max_{\hat{a}_i^\omega \in \Delta(A_i^\omega)} \left\{ K_i^\omega(\hat{a}_i^\omega, a_{N \setminus i}^{\omega*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, (\hat{a}_i^\omega, a_{N \setminus i}^{\omega*})) E_i^{\omega'}(\eta^{ne}) \right\},$$

and

$$\sigma_i = (\mathbb{I} - \delta \Pi(\eta^*))^{-1} \beta_i \geq Q(\{i\}). \quad (2.4.19)$$

The condition of existence of a SPNE are summarised in the following proposition. The validity of inequality (2.4.19) implies that the principle of strategic stability holds when the Nash equilibrium is subgame perfect.

Theorem 2.5. *If, in an σ -regularisation G_σ such that $\bar{\sigma} = \pi_0 \sigma$, inequality (2.4.19) holds for any player $i \in N$, then there exists behaviour strategy profile $\tilde{\varphi}$ which is a SPNE with players' payoffs $(\bar{\sigma}_1, \dots, \bar{\sigma}_n)$.*

Proof. The proof is similar to the proof of Theorem 2.4 using the structure of the “new” strategy profile. Determine this behaviour strategy profile as $\tilde{\varphi} = (\tilde{\varphi}_1, \dots, \tilde{\varphi}_n)$ where strategies $\tilde{\varphi}_i, i \in N$ are:

$$\tilde{\varphi}_i(h(k)) = \begin{cases} a_i^{\omega*}, & \text{if } \omega(k) = \omega, h(k) \subset h^*; \\ \hat{a}_i^{\omega, ne}, & \text{if } h(k) \not\subset h^*, \end{cases} \quad (2.4.20)$$

where $\hat{a}_i^{\omega, ne} \in \Delta(A_i^\omega)$ is player's i 's punishment, which can be either in pure or mixed strategies. Notice that, if a multi-player deviation is observed in the history, all players implement η^{ne} . \square

§ 2.4.3 Irrational-behaviour-proof in stochastic games with infinite duration

Subgame consistency and strategic support assume that the players are fully rational. However, in reality cooperation may be broken down by irrational reasons. For instance, a player may use irrational acts to extort additional gains if some circumstances allow it. Refusal of other players to yield to his extortion would result in the

¹²In the case of multiple Nash equilibria, one of them should be chosen for the realisation of the punishment. Notice that this can be implemented because players use correlated strategies.

dissolution of the cooperative scheme. Thus in this case, a deviation would imply an “irrational behaviour.”¹³

D.W.K. Yeung proposed a condition¹⁴ under which, even if an irrational behaviour emerges in the game, a player is certain to obtain at least her individual payoff [223]. This procedure can be explained as follows. Suppose two different scenarios. In the first scenario, a player cooperates until a certain period, and then the cooperation breaks up. In the second scenario, a player plays individually during the whole game. If the payoff in the first scenario is not less than the payoff in the second scenario, then the principle of irrational behaviour proof is satisfied. The following definition provides the condition to satisfy this principle.

Definition 2.9. *An imputation $\bar{\sigma} = \pi_0\sigma$ and the corresponding IDP satisfy the principle of irrational-behaviour-proof if*

$$E_i^{\sigma, [1, k]} + \delta^k \Pi^k(\eta^*) V(\{i\}) \geq V(\{i\}), \text{ for every } i \in N \text{ and any } k = 1, 2, \dots, \quad (2.4.21)$$

where $E_i^{\sigma, [1, k]}$ is the expected player i 's payoff at the first k stages in σ -regularisation G_σ .

The underlying assumption is that, before the beginning of each stage, players know if the cooperation has broken down or not, so that the information is not delayed. In the left hand side of inequality (2.4.21), the first term is equal to the expected value of player i 's payoff if, in the first k stages, players play cooperative strategy profile η^* and σ -regularization of game G is made. The second term is the expected payoff of player i from stage $k + 1$, when the cooperation breaks up. The right hand side of (2.4.21) is the payoff of player i if she plays individually from the start onwards.

Theorem 2.6. *If inequality*

$$(\mathbb{I} - \delta \Pi(\eta^*)) (\sigma_i - V(\{i\})) \geq 0 \quad (2.4.22)$$

holds for any $i \in N$, then the cooperative solution $\bar{\sigma} = \pi_0\sigma$ and the corresponding IDP $\{\beta_i\}_{i \in N}$ satisfy the principle of irrational-behaviour-proof.

¹³Note that it is possible to formulate an analogous condition for repeated games.

¹⁴The so-called Yeung's condition or principle of irrational-behaviour-proof was adopted for linear-quadratic games in [53, 133].

Proof. In what follows, we show that condition (2.4.22) is sufficient for inequality (2.4.21) to hold for any $k = 1, 2, \dots$. The proof is based on the mathematical induction method. First, we rewrite (2.4.21) for $k = 1$. Then we transform (2.4.22) by considering definition σ_i and using IDP (2.4.5). We get

$$V(\{i\}) \leq \beta_i + \delta \Pi(\eta^*) V(\{i\}). \quad (2.4.23)$$

The inequalities coincide and it proves Theorem for $k = 1$.

Suppose that (2.4.22) implies (2.4.21) for $k = l$. Rewriting (2.4.21) for $k = l$ we yield:

$$V(\{i\}) \leq \beta_i + \dots + \delta^{l-1} \Pi^{l-1}(\eta^*) \beta_i + \delta^l \Pi^l(\eta^*) V(\{i\}). \quad (2.4.24)$$

We adopt the same procedure for $k = l + 1$. Inequality (2.4.21) for $k = l + 1$ is:

$$V(\{i\}) \leq \beta_i + \dots + \delta^l \Pi^l(\eta^*) \beta_i + \delta^{l+1} \Pi^{l+1}(\eta^*) V(\{i\}). \quad (2.4.25)$$

Next we need to prove that, if (2.4.22) holds, then (2.4.21) holds for $k = l + 1$. After transformation the right hand side of (2.4.25) is:

$$\beta_i + \delta \Pi(\eta^*) \left\{ \beta_i + \delta \Pi(\eta^*) \beta_i + \dots + \delta^{l-1} \Pi^{l-1}(\eta^*) \beta_i + \delta^l \Pi^l(\eta^*) V(\{i\}) \right\}.$$

Taking into account (2.4.24), the expression in braces is not less than $V(\{i\})$. Therefore the right part of (2.4.25) is not less than $\beta_i + \delta \Pi(\eta^*) V(\{i\})$. From equation (2.4.2) and (2.4.22), we get (2.4.21) for $k = l + 1$, which proves the theorem. \square

Corollary 2.1. *For irrational-behaviour-proof principle it is sufficient that for each $i \in N$ the following inequality is true:*

$$K_i(\tilde{a}) - \beta_i \leq \delta \left(\sigma_i^{\min} - V^{\max}(\{i\}) \right), \quad (2.4.26)$$

where $K_i(\tilde{a}) = \left(\max_{a_i^{\omega_1} \in A_i^{\omega_1}} K_i^{\omega_1}(a_i^{\omega_1}, a_{N \setminus i}^{\omega_1*}), \dots, \max_{a_i^{\bar{\omega}} \in A_i^{\bar{\omega}}} K_i^{\bar{\omega}}(a_i^{\bar{\omega}}, a_{N \setminus i}^{\bar{\omega}}) \right)^T$, and $\max_{a_i^{\omega} \in A_i^{\omega}} K_i^{\omega}(a_i^{\omega}, a_{N \setminus i}^{\omega*})$ is the maximal payoff of player i which he obtains deviating from action profile $a^{\omega*}$ which is the part of cooperative strategy profile η^* satisfying condition (2.3.1), and $\tilde{a}_i = \arg \max_{a_i^{\omega} \in A_i^{\omega}} K_i^{\omega}(a_i^{\omega}, a_{N \setminus i}^{\omega*})$ for each state $\omega \in \Omega$ and each player $i \in N$:

$$\sigma_i^{\min} = \left(\min_{\omega \in \Omega} \sigma_i^{\omega}, \dots, \min_{\omega \in \Omega} \sigma_i^{\omega} \right)^T,$$

$$V^{\max}(\{i\}) = \left(\max_{\omega \in \Omega} V^{\omega}(\{i\}), \dots, \max_{\omega \in \Omega} V^{\omega}(\{i\}) \right)^T.$$

Proof. Let sufficient condition (2.4.26) be satisfied. It can be rewritten in the following way:

$$\beta_i + \delta\sigma_i^{\min} \geq K_i(\tilde{a}) + \delta V^{\max}(\{i\}). \quad (2.4.27)$$

Estimate the left- and right- hand parts of inequality (2.4.27). As matrix of transition probabilities $\Pi(\eta^*)$ is stochastic, we obtain:

$$\beta_i + \delta\sigma_i^{\min} = \beta_i + \delta\Pi(\eta^*)\sigma_i^{\min} \leq \beta_i + \delta\Pi(\eta^*)\sigma_i. \quad (2.4.28)$$

For the right-hand side of inequality (2.4.27), the equity is true:

$$K_i(\tilde{a}) + \delta V^{\max}(\{i\}) = K_i(\tilde{a}) + \delta\Pi(\tilde{\eta})V^{\max}(\{i\}), \quad (2.4.29)$$

where $\Pi(\tilde{a})$ is a stochastic matrix, and $\tilde{\eta} = (\tilde{\eta}_i : i \in N)$ is a profile in stationary strategies such that

$$\tilde{\eta}_j = \begin{cases} \arg \max_{\eta_i \in H_i} \Pi(\eta_i, \eta_{N \setminus i}^*) V(\{i\}), & \text{if } j = i \\ \eta_j^*, & \text{if } j \neq i \end{cases}$$

Therefore, we have the inequality:

$$\begin{aligned} K_i(\tilde{a}) + \delta\Pi(\tilde{\eta})V^{\max}(\{i\}) &= \max_{a_i \in A_i} K_i(a_i, a_{N \setminus i}^*) + \delta \max_{\eta_i \in \Xi_i} \left\{ \Pi(\eta_i, \eta_{N \setminus i}^*) V(\{i\}) \right\} \geq \\ &\max_{\eta_i \in H_i} \left\{ K_i(a_i, a_{N \setminus i}^*) + \delta\Pi(\eta_i, \eta_{N \setminus i}^*) V(\{i\}) \right\}. \end{aligned} \quad (2.4.30)$$

The inequalities (2.4.27)–(2.4.29) and (2.4.30) implies condition (2.4.22). Therefore, by Theorem 2.6 the principle of irrational-behaviour-proof is satisfied. \square

§ 2.5 Existence of stable cooperative solution in stochastic games with infinite duration

In this section we discuss the conditions guaranteeing the existence of a stable cooperative solution. First, we need to mention that the allocation rule adopted should give a non-empty subset of the imputation set. Cooperative solutions such as the Shapley value or the nucleolus always exist and we may calculate them for any sub-game using the values of the characteristic function given by (2.3.3), (2.3.4) and (2.3.8).

The existence of a subgame consistent cooperative solution follows from Theorem 2.3 and the method of construction of IDP for $\bar{\sigma}$. For a given cooperative solution $\bar{\sigma}$, the regularisation of a stochastic game determines new payoff functions to players in order to satisfy the principle of subgame consistency. Hence, the players' discounted payoffs in σ -regularisation of the initial game are equal to the components of cooperative solution $\bar{\sigma}$, which is subgame consistent.

Thus, if the payments to the players are modified through σ -regularisation, then subgame consistent cooperative solution $\bar{\sigma}$ exists in general.

To verify whether cooperative solution $\bar{\sigma}$ satisfies the principle of strategic stability and irrational-behaviour-proof, we need to check that the following system of inequalities holds:

$$\begin{cases} \sigma_i = (\mathbb{I} - \delta\Pi(\eta^*))^{-1}\beta_i \geq F(\{i\}), & i \in N, \\ (\mathbb{I} - \delta\Pi(\eta^*)) (\sigma_i - V(\{i\})) \geq 0, & i \in N. \end{cases} \quad (2.5.1)$$

These conditions on discount factor δ are similar to those necessary to prove that a cooperative strategy profile is SPNE in repeated games. This system is non-linear with respect to δ and the solution of the system cannot be obtained in an explicit form.

However, we may state the existence of a stable cooperative solution for the class of stochastic games in which the cooperative strategy profile coincides with the Nash equilibrium and the players are symmetric. In this case, the Shapley value satisfies the principles of stable cooperation. Further, we examine the solution of system (2.5.1) on a specific class of stochastic games with two states and two players.

Example. *Stochastic game of competition between asymmetric firms.* Consider Cournot duopoly with asymmetric firms. Describe it with a stochastic game setting like Prisoners' Dilemma. Let the set of states be $\Omega = \{\omega_1, \omega_2\}$, where in ω_j normal-form game $\langle N, A_1^{\omega_j}, A_2^{\omega_j}, K_1^{\omega_j}, K_2^{\omega_j} \rangle$ is defined, $j = 1, 2$, and $A_i^{\omega_j} = \{C_j, D_j\}$ is the set of actions of player $i = 1, 2$. Strategies C_j and D_j stands for "collude" and

“deviate”, respectively. For state ω_1 , players’ payoffs are:

$$\begin{array}{cc} & C_1 & D_1 \\ C_1 & (7, 7) & (1, 8) \\ D_1 & (8, 1) & (4, 5) \end{array}$$

whereas for state ω_2 players’ payoffs are:

$$\begin{array}{cc} & C_2 & D_2 \\ C_2 & (9, 9) & (1, 10) \\ D_2 & (16.5, 1) & (6, 5) \end{array}$$

State ω_1 can be interpreted as a market with a low demand, and state ω_2 as a market with a high demand. Both one-shot games have the unique Nash equilibrium when both firms deviate with outcomes $(4, 5)$ and $(6, 5)$ in states ω_1 and ω_2 respectively. Conversely, the cooperative action profile that maximizes the sum of the payoffs are “to collude” with outcomes $(7, 7)$ and $(9, 9)$ respectively. When playing the cooperative action profile, players get equal payoffs, but in the Nash equilibrium outcome they obtain asymmetric payoffs. In particular, with a low demand Firm 1 has a lower payoff than Firm 2, and with a high demand Firm 2 has lower payoff than Firm 1. This scenario could be interpreted as the result of technical features of firms’ production. For instance, Firm 2 can be endowed with a production technology being more efficient in producing low levels of output.

In state ω_2 , players also differ in the profiles when one firm “colludes” and the competitor “deviates”. In particular, Firm 1’s deviation payoff is larger than Firm 2’s one. Hence the asymmetry of the players influences the cooperative payoff imputation. Another feature of state ω_2 is that, when both firms collude, their summarized payoff is not much larger the one in action profile (D_2, C_2) (18 against 17.5). Therefore, if the probability of transiting from profile (D_2, C_2) to state ω_1 is larger than from profile (C_2, C_2) , then players may agree on playing profile (D_2, C_2) to avoid transition from high to low demand state.

Let transition probabilities from states ω_1 and ω_2 be

$$\left(\begin{array}{cc} (0.3, 0.7) & (0.9, 0.1) \\ (0.4, 0.6) & (0.3, 0.7) \end{array} \right), \quad \left(\begin{array}{cc} (0.9, 0.1) & (0.4, 0.6) \\ (0.1, 0.9) & (0.3, 0.7) \end{array} \right)$$

where the element (k, l) of the matrix consists of transition probabilities from state ω_j to states ω_1, ω_2 , on condition that player 1 chooses actions k^{th} and player 2 chooses l^{th} . We may mention that the probability of transiting to state ω_1 in action profile (C_2, C_2) is much higher than the probability to transit to this state in action profile (D_2, C_2) , that is 0.9 contrary to 0.1. Let the discount factor be $\delta = 0.99$ and the vector of initial distribution over the set of states be $\pi_0 = (0.5, 0.5)$.

Cooperative game. Determine cooperative game G_c based on stochastic game G . For it, we compute cooperative solution $\eta^* = (\eta_1^*, \eta_2^*)$ in stationary strategies using (2.2.4) and (2.3.1). We obtain a unique stationary strategy $\eta_1^* = (C_1, D_2)$ for player 1, and $\eta_2^* = (C_1, C_2)$ for player 2 which give maximal total players' payoff $\bar{V}(\{1, 2\}) = \pi_0 V(\{1, 2\}) = 1704.61$. Following this profile, in state ω_1 the profile of cooperative strategies (when both players collude) gives payoff 7 for each firm. In state ω_2 , with a cooperative strategy profile, Firm 1 deviates and Firm 2 colludes, and the payoff of firm 2 is less than its payoff in the Nash equilibrium. But this will be compensated by Firm 1 when they apply an imputation of their joint payoff. Therefore, the values of a characteristic function for a grand coalition are

$$V(\{1, 2\}) = \begin{pmatrix} E_1^{\omega_1}(\eta^*) + E_2^{\omega_1}(\eta^*) \\ E_1^{\omega_2}(\eta^*) + E_2^{\omega_2}(\eta^*) \end{pmatrix} = \begin{pmatrix} 1702.43 \\ 1706.80 \end{pmatrix}.$$

By definition (2.3.8) the values of characteristic function for the empty set are zero:

$$V(\emptyset) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Calculate the values of characteristic function $V(S) = (V^{\omega_1}(S), V^{\omega_2}(S))$ for coalitions $S = \{1\}$ and $S = \{2\}$ using (2.3.3):

$$V(\{1\}) = \begin{pmatrix} 538.60 \\ 540.60 \end{pmatrix}, \quad V(\{2\}) = \begin{pmatrix} 500.00 \\ 500.00 \end{pmatrix}.$$

These are Firms' payoffs in the Nash equilibrium when both firms deviate in all states, i.e., they adopt strategy profiles (D_1, D_1) and (D_2, D_2) .

Using (2.3.7), we may calculate $\bar{V}(S)$ for all coalitions:

$$\bar{V}(\emptyset) = 0.00, \quad \bar{V}(\{1\}) = 539.60, \quad \bar{V}(\{2\}) = 500.00, \quad \bar{V}(\{1, 2\}) = 1704.61.$$

Thus, we determine cooperative stochastic subgame $G_c^{\omega_j}$ as the set $\langle N, V^{\omega_j}(\cdot) \rangle$, $j = 1, 2$, and cooperative stochastic game G_c as the set $\langle N, \bar{V}(\cdot) \rangle$.

The Shapley value. We suppose that players choose the Shapley value as a cooperative solution of their total payoff in cooperative stochastic game G_c and in all subgames $G_c^{\omega_j}$, $j = 1, 2$. For two-player game the Shapley value is calculated by formula:

$$\sigma_i^{\omega_j} = V^{\omega_j}(\{i\}) + \frac{V^{\omega_j}(\{1, 2\}) - V^{\omega_j}(\{1\}) - V^{\omega_j}(\{2\})}{2},$$

where $i = 1, 2$ and $j \in \{1, 2\}$. The Shapley values in subgames are

$$\sigma_1 = \begin{pmatrix} 870.516 \\ 873.698 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 831.916 \\ 833.098 \end{pmatrix}.$$

Then taking into account the vector of initial distribution π_0 , we are able to determine the Shapley value $\bar{\sigma}$ in the whole game G_c by Definition 2.4:

$$\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2) = (872.107, 832.507).$$

Subgame consistency. Now we verify if the Shapley value satisfies the principles of stable cooperation and begin with subgame consistency. If firms receive stage payoffs according to their initially defined payoffs, then their discounted payoffs in the whole game will be equal $\bar{E}_1(\eta^*) = 1526.809$ and $\bar{E}_2(\eta^*) = 177.805$ which are different from the components of the Shapley value which are $\bar{\sigma}_1 = 872.107$ and $\bar{\sigma}_2 = 832.507$. Define the imputation distribution procedure or transfer payments to the players such that they finally receive the components of the Shapley value and the imputation distribution procedure is subgame-consistent. Using that $\bar{\sigma}$ equals $\pi_0 \sigma$ and equation (2.4.2), we obtain IDP:

$$\beta_1 = (\mathbb{I} - \delta\Pi(\eta^*))\sigma_1 = \begin{pmatrix} 6.500 \\ 9.052 \end{pmatrix}, \quad \beta_2 = (\mathbb{I} - \delta\Pi(\eta^*))\sigma_2 = \begin{pmatrix} 7.500 \\ 8.448 \end{pmatrix}.$$

Define σ -regularisation of initial stochastic game G using IDP and Definition 2.7. We redefine payoff functions of the players in the initial game in all states when players adopt cooperative strategy profile substituting the payoffs by corresponding components of the IDP. Therefore, players' payoffs in states ω_1 and ω_2 correspondingly equal:

$$\begin{pmatrix} (6.500, 7.500) & (1, 8) \\ (8, 1) & (4, 5) \end{pmatrix},$$

$$\begin{pmatrix} (9, 9) & (1, 10) \\ (9.052, 8.448) & (6, 5) \end{pmatrix}.$$

In a regularized game in the state with a low demand (state ω_1), both firms adopt action “to collude” and receive payoffs (6.5, 7.5). Notice that their payoffs in the initial game are (7, 7). Therefore, Firm 1 gives 0.5 to Firm 2 at each stage in this state. In the state with a high demand (state ω_2) Firm 1 plays action “to deviate” while Firm 2 “colludes”. This behavior is prescribed by the cooperative strategy profile. In state ω_2 players’ payoffs are (9.052, 8.448). Notice that the payoffs in the initial game are (16.5, 1). Therefore, Firm 2 gets $16.5 - 9.052 = 7.448$ from Firm 1 at each stage in this state. If the regularization of the initial game is made by the above described method, the Shapley value and the corresponding IDP are subgame-consistent.

Strategic stability. We now check for strategic stability of the Shapley value, i. e., we check if Firms have benefits from individual deviations from the cooperative strategy profile. First, consider state ω_1 . As the action profile played in cooperation is not the Nash equilibrium, then the players may have benefits from deviation. We verify if the inequality is true:

$$\sigma_i^{\omega_1} \geq F^{\omega_1}(\{i\}),$$

for each $i = 1, 2$, where

$$F^{\omega_1}(\{i\}) = \max_{\substack{a_i^{\omega_1} \in A_i^{\omega_1} \\ a_i^{\omega_1} \neq a_i^{\omega_1*}}} \left\{ K_i^{\omega_1}(a_i^{\omega_1}, a_{N \setminus i}^{\omega_1*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega_1, (a_i^{\omega_1}, a_{N \setminus i}^{\omega_1*})) V^{\omega'}(\{i\}) \right\}.$$

Inequality (2.4.11) for Firm 1 is written in this way:

$$870.516 \geq 8 + 0.99 \begin{pmatrix} 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 538.60 \\ 540.60 \end{pmatrix} = 542.402,$$

and for Firm 2:

$$831.916 \geq 8 + 0.99 \begin{pmatrix} 0.9 & 0.1 \end{pmatrix} \begin{pmatrix} 500.00 \\ 500.00 \end{pmatrix} = 503.$$

In state ω_2 , an action profile (D_2, C_2) is the Nash equilibrium. Therefore, players can’t increase their payoffs by deviations. Therefore, we may conclude that inequality (2.4.11) holds for state ω_2 . The condition of strategic stability is satisfied.

Irrational-behavior-proof. To verify the condition of irrational behavior proof, we need to compare players' payoffs in two cases:

- 1) A firm plays individually during the whole game,
- 2) A firm cooperates with the other firm until some step, and after this it starts playing individually.

Notice that in the second case, when the firms cooperate, they receive payoffs in accordance with IDP, constructed on the basis of the initially chosen cooperative solution.

If the player's payoff in case 1) is not greater than his payoff in case 2), then the principle of irrational behavior proof against irrational behavior is satisfied. This has been proved in Theorem 2.6, since

$$(\mathbb{I} - \delta\Pi(\eta^*))(\sigma_1 - V(\{1\})) = (\mathbb{I} - \delta\Pi(\eta^*))(\sigma_2 - V(\{2\})) = \begin{pmatrix} 2.500 \\ 3.448 \end{pmatrix} \geq 0.$$

Thus, we make a regularization of the initially defined game to ensure subgame consistency of the Shapley value and show that the principles of strategic stability and irrational behavior proof are satisfied in the game.

§ 2.6 Strong transferable equilibrium in stochastic games with infinite duration

Theorem 2.4 can be generalized to the case when several players deviate, i. e., we may prove that if the condition similar to inequality (2.4.11) is satisfied in σ -regularization G_σ of stochastic game G , there exists a strong transferable equilibrium with payoffs $(\bar{\sigma}_1, \dots, \bar{\sigma}_n)$. In this case, players can implement a specially constructed profile in trigger strategies, where as a punishment for deviated coalition, not deviated players will implement trigger strategies that allow a deviated coalition to obtain a minimax payoff in any subgame.

Define a strong transferable equilibrium¹⁵ and prove a theorem similar to Theorem 2.4.

¹⁵The definition of such a strong equilibrium can be found in [66].

Definition 2.10. (see [46]) We call profile $\tilde{\varphi} = (\tilde{\varphi}_1, \dots, \tilde{\varphi}_n)$ strong transferable equilibrium in regularized game G_σ if for any coalition $S \subseteq N$, $S \neq \emptyset$, inequality

$$\sum_{i \in S} \bar{E}_i^\sigma(\tilde{\varphi}) \geq \sum_{i \in S} \bar{E}_i^\sigma(\varphi_S, \tilde{\varphi}_{N \setminus S}) \quad (2.6.1)$$

holds for any behaviour strategy of coalition S : $\varphi_S = (\varphi_i : i \in S) \in \prod_{i \in S} \Phi_i$. Here $\bar{E}_i^\sigma(\cdot)$ is a discounted payoff of player i in σ -regularisation of game G .

We will prove a theorem allowing us to obtain a condition on the game parameters for which in regularized game G_σ there exists a transferable equilibrium with players' payoffs equal to the corresponding components of the cooperative solution according to which the initial stochastic game is regularized.

Theorem 2.7. If in regularized game G_σ such that cooperative solution satisfies condition $\bar{\sigma} = \pi_0 \sigma$, the inequality holds:

$$\sum_{i \in S} \beta_i \geq (\mathbb{I} - \delta \Pi(\eta^*)) \tilde{F}(S) \quad (2.6.2)$$

for any coalition $S \subseteq N$, $S \neq \emptyset$, where $\tilde{F}(S) = (\tilde{F}^{\omega_1}(S), \dots, \tilde{F}^{\omega_n}(S))^T$,

$$\tilde{F}^\omega(S) = \max_{\substack{a_S^\omega \in \prod_{i \in S} \Delta(A_i^\omega) \\ a_S^\omega \neq a_S^{\omega^*}}} \left\{ \sum_{i \in S} K_i^\omega(a_S^\omega, a_{N \setminus S}^{\omega^*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega^*}) V^{\omega'}(S) \right\},$$

then in regularized game G_σ there exists a strong transferable equilibrium with players' payoffs $(\bar{\sigma}_1, \dots, \bar{\sigma}_n)$.

Proof. The proof of the theorem is close the proof of Theorem 2.4 but instead of strategy (2.4.12) we use the following behaviour strategy φ'_i , $i \in N$:

$$\varphi'_i(h(k)) = \begin{cases} a_i^{\omega^*}, & \text{if } \omega(k) = \omega, h(k) \subset h^*; \\ a_i^{\omega'}(S), & \text{if } \omega(k) = \omega \text{ and } \exists l \in [1, k-1], \\ & S \subset N, i \notin S: h(l) \subset h^* \text{ and} \\ & (\omega(l), a(l)) \notin h^*, \text{ but} \\ & (\omega(l), (a_S^*(l), a_{N \setminus S}(l))) \in h^*; \\ \text{any} & \text{otherwise,} \end{cases} \quad (2.6.3)$$

where $a_i^{\omega^*}$ is an action of player i in a cooperative mode, while $a_i^{\omega'}(S) \in \Delta(A_i^\omega)$ is an action of player i in a trigger mode which jointly with actions $a_i^{\omega'}(S) \in \Delta(A_i^\omega)$

of players $i' \neq i$, $i' \in N \setminus S$ forms an action of coalition $N \setminus S$ against coalition S and allows coalition S to obtain minmax value $V^\omega(S)$ in subgame G^ω . \square

§ 2.7 Stochastic game with one absorbing state

In this section we consider a two-player game with two states. The set of players is $N = \{1, 2\}$. Let state ω_1 be given by:

$$\omega_1 : \begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \left(\begin{array}{cc} (a, a+1) & (c, b) \\ (b, c) & (d+1, d) \end{array} \right) \end{array} \quad (2.7.1)$$

Players have two pure actions, C (to cooperate) and D (to defect). The constants satisfy the inequalities:

$$b > a + 1, \quad a > d + 1, \quad d > c > 0.$$

We also assume

$$2a + 1 > b + c. \quad (2.7.2)$$

From inequality (2.7.2) it follows that players receive a larger total payoff by cooperating than defecting. The game represents Prisoners' Dilemma with asymmetric players: in action profile (C, C) the payoff of player 1 is less than the payoff of player 2, but in profile (D, D) the payoff of player 2 is less than the payoff of player 1. If action profiles (C, C) and (D, D) are chosen in state ω_1 , a game process remains in this state with probability 1. But if profiles (C, D) or (D, C) are chosen, the game process transits to state ω_2 which is "absorbing", i.e. this state will be realised in all following stages of the game with probability 1. In state ω_2 both players have a unique action D and their payoffs will be equal to d :

$$\omega_2 : \begin{array}{c} D \\ D \end{array} \left(d, d \right) \quad (2.7.3)$$

The matrices of transition probabilities from states ω_1 and ω_2 are

$$\begin{pmatrix} (1, 0) & (0, 1) \\ (0, 1) & (1, 0) \end{pmatrix}, \quad (0, 1).$$

The discount factor is $\delta \in (0, 1)$ and the vector of the initial distribution over the set of states is $\pi_0 = (1, 0)$, i.e., a game starts with state ω_1 .

Cooperative game. For this game we construct a cooperative game by determining the characteristic functions for all subgames and the whole game. We then show how we need to redistribute the stage payoffs adopting IDP to obtain the subgame consistency of the Shapley value. The condition of strategic stability gives the lower bound of the discount factor.

The first step is to determine cooperative form G_c of non-cooperative stochastic game G . In particular, we need to find a cooperative strategy profile and then calculate the values of characteristic functions for each subgame (starting from states ω_1 and ω_2) and for the whole game.

We compute cooperative strategy profile $\eta^* = (\eta_1^*, \eta_2^*)$ using (2.2.4) and (2.3.1). In a cooperative strategy profile both players choose C in state ω_1 and D in state ω_2 . The total players' payoff with profile η^* is equal to the value of the characteristic function of coalition N :

$$\bar{V}(\{1, 2\}) = \bar{E}_1(\eta^*) + \bar{E}_2(\eta^*) = 2a + 1 + \delta(2a + 1) + \dots = \frac{2a + 1}{1 - \delta}. \quad (2.7.4)$$

In particular, the values of characteristic function $V^\omega(\{1, 2\})$ for both subgames are

$$V(\{1, 2\}) = \begin{pmatrix} V^{\omega_1}(\{1, 2\}) \\ V^{\omega_2}(\{1, 2\}) \end{pmatrix} = \begin{pmatrix} \frac{2a + 1}{1 - \delta} \\ \frac{2d}{1 - \delta} \end{pmatrix}. \quad (2.7.5)$$

We can now calculate the values of characteristic functions of coalitions $\{1\}$ and $\{2\}$ for both states using (2.3.4):

$$\begin{aligned} V^{\omega_1}(\{1\}) &= \frac{d + 1}{1 - \delta}, \\ V^{\omega_1}(\{2\}) &= \frac{d}{1 - \delta}, \\ V^{\omega_2}(\{1\}) &= V^{\omega_2}(\{2\}) = \frac{d}{1 - \delta}. \end{aligned}$$

By equation (2.3.8), the values of the characteristic functions for the empty set are zero:

$$V(\emptyset) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Using (2.3.7), we then calculate the values of the characteristic function $\bar{V}(\cdot)$ for all possible coalitions taking into account the initial distribution of states $\pi_0 = (1, 0)$:

$$\bar{V}(\emptyset) = 0, \quad \bar{V}(\{1\}) = \frac{d+1}{1-\delta}, \quad \bar{V}(\{2\}) = \frac{d}{1-\delta}, \quad \bar{V}(\{1, 2\}) = \frac{2a+1}{1-\delta}.$$

In this way, we determine cooperative stochastic subgames $G_c^{\omega_j}$ as the set $\langle N, V^{\omega_j}(\cdot) \rangle$, $j = 1, 2$, and cooperative stochastic game G_c as the set $\langle N, \bar{V}(\cdot) \rangle$.

The Shapley value. We assume that players choose the Shapley value as an imputation of their total payoff in cooperative stochastic game G_c and in all subgames $G_c^{\omega_j}$, $j = 1, 2$. For a two-person game, this is given by:

$$\sigma_i^{\omega_j} = V^{\omega_j}(\{i\}) + \frac{V^{\omega_j}(\{1, 2\}) - V^{\omega_j}(\{1\}) - V^{\omega_j}(\{2\})}{2},$$

where $i = 1, 2$, $j = 1, 2$. The Shapley values for the subgames are:

$$\sigma_1 = \begin{pmatrix} \sigma_1^{\omega_1} \\ \sigma_1^{\omega_2} \end{pmatrix} = \begin{pmatrix} \frac{a+1}{1-\delta} \\ d \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} \sigma_2^{\omega_1} \\ \sigma_2^{\omega_2} \end{pmatrix} = \begin{pmatrix} \frac{a}{1-\delta} \\ d \end{pmatrix}.$$

Taking into account the vector of initial distribution π_0 , we are able to determine the Shapley value $\bar{\sigma}$ in game G_c by Definition 2.4:

$$\bar{\sigma} = (\bar{\sigma}_1, \bar{\sigma}_2) = \left(\frac{a+1}{1-\delta}, \frac{a}{1-\delta} \right).$$

Subgame consistency of the Shapley value. We are now in a position to verify the principles of stable cooperation. Begin with subgame consistency. If players get payoffs according to the initially defined payoff functions, their total payoffs will be $\frac{a}{1-\delta}$ and $\frac{a+1}{1-\delta}$ in contrast to the components of the Shapley value $\frac{a+1}{1-\delta}$ and $\frac{a}{1-\delta}$. In order to obtain subgame consistency, we compute IDP by equating $\bar{\sigma}$ to $\pi_0 \sigma$ by using (2.4.2):

$$\begin{aligned} \beta_1 &= (\mathbb{I} - \delta \Pi(\eta^*)) \sigma_1 = \begin{pmatrix} 1-\delta & 0 \\ 0 & 1-\delta \end{pmatrix} \begin{pmatrix} \frac{a+1}{1-\delta} \\ d \end{pmatrix} = \begin{pmatrix} a+1 \\ d \end{pmatrix}, \\ \beta_2 &= (\mathbb{I} - \delta \Pi(\eta^*)) \sigma_2 = \begin{pmatrix} 1-\delta & 0 \\ 0 & 1-\delta \end{pmatrix} \begin{pmatrix} \frac{a}{1-\delta} \\ d \end{pmatrix} = \begin{pmatrix} a \\ d \end{pmatrix}. \end{aligned}$$

We then determine σ -regularisation of the initial stochastic game G using the IDP and Definition 2.7. We re-establish the payoff functions of the initial game in state ω_1 when players adopt the cooperative action profiles. Therefore, the players' payoffs in state ω_1 are:

$$\begin{pmatrix} (a+1, a) & (c, b) \\ (b, c) & (d+1, d) \end{pmatrix}.$$

In state ω_1 , when both players adopt the cooperative strategy profile η^* (both players use action C in state ω_1), their payoffs are $(a+1, a)$. Since their payoffs in the initial game were $(a, a+1)$, player 2 transfers 1 to player 1. If the initial game is regularised by the method described above, the Shapley value and the corresponding PDP satisfy the principle of subgame consistency (see Theorem 2.3).

Strategic support of the Shapley value. We now evaluate the strategic stability of the Shapley value by checking if players may deviate from the cooperative strategy profile. We consider the possible deviations of players in state ω_1 (in state ω_2 players have the unique action). In this state the cooperative action profile is not the Nash equilibrium, thus players may benefit from deviation. We should check if the following inequality

$$\sigma_i^{\omega_1} \geq F^{\omega_1}(\{i\}), \quad (2.7.6)$$

is true for any $i = 1, 2$, where

$$F^{\omega_1}(\{i\}) = \max_{a_i^{\omega_1} \in \Delta(A_i^{\omega_1})} \left\{ K_i^{\omega_1}(a_i^{\omega_1}, a_{N \setminus i}^{\omega_1*}) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega_1, (a_i^{\omega_1}, a_{N \setminus i}^{\omega_1*})) V^{\omega'}(\{i\}) \right\}.$$

For player 1, inequality (2.7.6) yields:

$$\frac{a+1}{1-\delta} \geq b + \delta d + \delta^2 d + \dots = b + \frac{\delta d}{1-\delta},$$

for player 2:

$$\frac{a}{1-\delta} \geq b + \delta d + \delta^2 d + \dots = b + \frac{\delta d}{1-\delta}.$$

These two inequalities give the condition on δ when the principle of strategic stability is satisfied:

$$\delta \geq \frac{b-a}{b-d}.$$

Principle of Irrational-behaviour-proof. In order to verify irrational-behaviour proof, we need to compare the payoffs of each player when:

- 1) A player acts as an “individual player” during the whole game.
- 2) A player cooperates with a competitor until some stage and then plays individually.

If the payoff of 2) is not less than the payoff of 1), then this principle is satisfied. This is confirmed by Theorem 2.6, since:

$$\begin{aligned}
 (\mathbb{I} - \delta\Pi(\eta^*)) (\sigma_1 - V(\{1\})) &= \begin{pmatrix} 1 - \delta & 0 \\ 0 & 1 - \delta \end{pmatrix} \begin{pmatrix} \frac{a+1}{1-\delta} - \frac{d+1}{1-\delta} \\ \frac{d}{1-\delta} - \frac{d}{1-\delta} \end{pmatrix} = \begin{pmatrix} a - d \\ 0 \end{pmatrix} \geq 0, \\
 (\mathbb{I} - \delta\Pi(\eta^*)) (\sigma_2 - V(\{2\})) &= \begin{pmatrix} 1 - \delta & 0 \\ 0 & 1 - \delta \end{pmatrix} \begin{pmatrix} \frac{a}{1-\delta} - \frac{d}{1-\delta} \\ \frac{d}{1-\delta} - \frac{d}{1-\delta} \end{pmatrix} = \begin{pmatrix} a - d \\ 0 \end{pmatrix} \geq 0.
 \end{aligned}$$

Both players benefit from cooperation even if IDP is adopted initially at some stages and then the game is played as a non-cooperative one with initially defined payoff functions as compared with a game played individually by both players during the whole game.

Conclusions. To sum up, we can formulate the conditions under which the Shapley value in the described stochastic game satisfies the three principles of stable cooperation (subgame consistency, strategic support, irrational-behavior-proof):

1. A discount factor is to be $\delta \geq \frac{b-a}{b-d}$.
2. A stochastic game is σ -regularised, i. e., the players' payoffs in state ω_1 are:

$$\begin{array}{cc}
 & C & D \\
 C & \left((a+1, a) & (c, b) \right) \\
 D & \left((b, c) & (d+1, d) \right)
 \end{array}$$

and in state ω_2 they must not be changed.

§ 2.8 Strongly subgame consistency of the core in stochastic games with infinite duration

Now suppose that the solution of a cooperative stochastic game is the subset of the imputation set that contains more than one imputation. For definiteness, let such a

solution be the core. We formulate the problem of strongly subgame consistency of the core and propose sufficient conditions for strongly subgame consistency of the core in stochastic games with infinite duration defined by (2.1.1).

Suppose that the cores of stochastic game G_c and any subgame G_c^ω , $\omega \in \Omega$, are nonempty. In cooperation, players agree on the joint implementation of cooperative strategy profile η^* and expect to obtain the components of the imputation belonging to the core \overline{CO} . Reaching intermediate states $\omega \in \Omega$, player $i \in N$ chooses action $a_i^{\omega^*}$ in accordance with cooperative strategy η_i^* and gets payoff $K_i^\omega(a^{\omega^*})$. If the players recalculate the solution, i.e., they find a solution of cooperative subgame G_c^ω , then the current solution will be the core CO^ω . It would be reasonable to require that the payoff received by a player in state ω summarized with the expected sum of any imputations from the cores $CO^{\omega'}$, $\omega' \in \Omega$, following state ω , would be an imputation from the core CO^ω . If this property holds for any state $\omega \in \Omega$, then the core of cooperative stochastic game G_c is strongly subgame-consistent.

To determine a strongly subgame-consistent core, we define the so-called expected core in state ω , i.e., we define the set of expected imputations belonging to the cores which are the solutions of the following subgames. For each state $\omega \in \Omega$, we define the *expected core*:

$$EC(\omega) = \left\{ \sigma(\omega) = \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega^*}) \sigma^{\omega'} \mid \sigma^{\omega'} \in CO^{\omega'} \right\}. \quad (2.8.1)$$

Set $EC(\omega)$ contains vectors $\sigma(\omega)$ which are mathematical expectations of all possible sets of the imputations from the cores of subgames starting in the states which are realised after the current state with respect to probability distribution $\{p(\omega' | \omega, a^{\omega^*}), \omega' \in \Omega\}$.

Remind Definition 2.5 of the imputation distribution procedure. The first condition (2.4.1) in a definition can be called the condition of “attainability of the imputation distribution procedure” because it allows to ensure that in any realized state the sum of payments to the players is equal to the sum of their payoffs when they implement cooperative strategies. The second condition guarantees players to receive the components of the initially chosen imputation from the core of cooperative game G_c in the sense of mathematical expectation, if payments to the players throughout

the game will be made in accordance with distribution procedure $\{\beta^\omega : \omega \in \Omega\}$.

We now define the distribution procedure of imputation $\bar{\sigma} = (\bar{\sigma}_1, \dots, \bar{\sigma}_n)$, where $\bar{\sigma}_i = \pi_0 \sigma_i$, $(\sigma_1^\omega, \dots, \sigma_n^\omega) = \sigma^\omega \in CO^\omega$, such that the core is strongly subgame-consistent.

Definition 2.11. *We call the core \overline{CO} of cooperative stochastic game G_c strongly subgame-consistent if for any imputation from the core \overline{CO} there exists a distribution procedure $\{\beta^\omega : \omega \in \Omega\}$ such that for any state $\omega \in \Omega$ the following inclusions hold:*

$$\beta^\omega \oplus \delta EC(\omega) \subset CO^\omega, \quad (2.8.2)$$

$$B^\omega \in CO^\omega, \quad \omega \in \Omega \quad (2.8.3)$$

where

$$\beta^\omega \oplus \delta EC(\omega) = \left\{ \beta^\omega + \delta \sigma(\omega) : \sigma(\omega) \in EC(\omega) \right\}.$$

The imputation distribution procedure $\{\beta^\omega : \omega \in \Omega\}$ is called strongly subgame-consistent.

Condition (2.8.2) means that the set of vectors equal to the sum of the imputation distribution procedure of the player in state ω and an imputation from the expected core for this state, is contained in the core of subgame starting from state ω . This condition imposes restrictions on payments to the players in the realized states, and very often is not satisfied for an arbitrary game, if payments to the players are made in accordance with the initially defined payoff functions.

We impose additional restrictions on the characteristic functions of subgames starting from the states of set Ω in order to obtain sufficient conditions for strongly subgame consistency of the core. Denote by $EV^\omega(S)$ the expected value of the characteristic function calculated for coalition $S \subseteq N$ for subgames following state ω :

$$EV^\omega(S) = \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) V^{\omega'}(S).$$

Denote by

$$\Delta V^\omega(S) = V^\omega(S) - \delta EV^\omega(S)$$

the difference between the values of a characteristic function in state ω and the expected value of the characteristic function. We denote by ΔCO^ω an analog of

the core constructed with function $\Delta V^\omega(S)$. We formulate a sufficient condition for strongly subgame consistency of IDP and the core \overline{CO} .

Theorem 2.8. *Let for each state $\omega \in \Omega$ the core CO^ω and the set ΔCO^ω be nonempty. If for every state $\omega \in \Omega$ distribution procedure $\{\beta^\omega : \omega \in \Omega\}$ of the imputation from the core \overline{CO} satisfies conditions:*

$$\beta^\omega \in \Delta CO^\omega, \quad (2.8.4)$$

$$B^\omega \in CO^\omega, \quad \omega \in \Omega, \quad (2.8.5)$$

then this imputation from the core \overline{CO} and corresponding procedure $\{\beta^\omega : \omega \in \Omega\}$ are strongly subgame-consistent.

Proof. We prove that any vector $\beta^\omega \in \Delta CO^\omega$ satisfying conditions (2.8.4) and (2.8.5) is a strongly subgame-consistent distribution procedure of imputation $\bar{\sigma} \in \overline{CO}$, i. e., conditions (2.8.2) and (2.8.3) from Definition 2.11 hold. Condition (2.8.5) coincides with (2.8.3), so, we need to prove that inclusion (2.8.2) holds for each state $\omega \in \Omega$. In state ω consider any vector $\sigma(\omega) \in EC(\omega)$ and find sum $\beta^\omega + \delta\sigma(\omega)$. Now we verify if the latter vector belongs to the core \overline{CO} . First, calculate the sum of all components of the vector:

$$\begin{aligned} \sum_{i \in N} \beta_i^\omega + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) \sum_{i \in N} \sigma_i^{\omega'} &= \\ &= V^\omega(N) - \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) V^{\omega'}(N) + \\ &+ \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) \sum_{i \in N} \sigma_i^{\omega'} = V^\omega(N), \end{aligned}$$

which means that property of collective rationality holds.

Next, consider $S \subset N$:

$$\begin{aligned} \sum_{i \in S} \beta_i^\omega + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) \sum_{i \in S} \sigma_i^{\omega'} &\geq \\ &\geq V^\omega(S) + \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) V^{\omega'}(S) - \\ &- \delta \sum_{\omega' \in \Omega} p(\omega' | \omega, a^{\omega*}) V^{\omega'}(S) = V^\omega(S). \end{aligned}$$

Since the proof does not depend on the choice of state $\omega \in \Omega$, we conclude that the imputation from the core of game G_c and its procedure $\{\beta^\omega : \omega \in \Omega\}$ are strongly subgame-consistent. \square

When analogs of the cores ΔCO^ω are nonempty for any state ω , Theorem 2.8 provides a method of construction of a strongly subgame-consistent distribution procedure of imputations from the core. Notice that generally not all imputations from the core can be realised with distribution procedure $\{\beta^\omega : \omega \in \Omega\}$ described above.

Chapter 3

Dynamic Games Played over Event Trees

§ 3.1 Definitions of a dynamic game played over event tree

In this chapter we consider the class of dynamic games at which the game process is realized over so-called event tree. The event tree represents a finite tree graph or acyclic connected graph. The main property of event tree is that transition probabilities from nodes to the next nodes are initially given, do not change over time and they do not depend on strategies chosen by the players at this stage. The detailed description of this class of games can be found in [103, 226]. Notations used in this chapter are initially proposed in the mentioned works.

Let $\mathcal{T} = \{0, 1, \dots, T\}$ be the set of periods, $T \geq 1$. The exogenous stochastic process is represented by an event tree, which has a root node n^0 in period 0 and a set of nodes $\mathcal{N}^t = \{n_1^t, \dots, n_{N_t}^t\}$ in period $t = 1, \dots, T$. Each node $n_i^t \in \mathcal{N}^t$ represents a possible sample value of the history of the stochastic process up to time t . The tree graph structure represents the nesting of information as one time period succeeds the other. Denote by $a(n_i^t) \in \mathcal{N}^{t-1}$ the unique predecessor of node $n_i^t \in \mathcal{N}^t$ on the event-tree graph for $t = 1, \dots, T$, and by $\mathcal{S}(n_i^t) \subset \mathcal{N}^{t+1}$ the set of all possible direct successors of node $n_i^t \in \mathcal{N}^t$ for $t = 0, \dots, T-1$. A path from the root node n^0 to a terminal node n_i^T is called a *scenario*.

Each scenario has a probability and the probabilities of all scenarios sum up to 1. We denote by $\pi(n_i^t)$ the probability of passing through node n_i^t , which corresponds to the sum of the probabilities of all scenarios that contain this node. In particular, $\pi(n^0) = 1$ and $\pi(n_i^T)$ is equal to the probability of the single scenario that terminates in (leaf) node n_i^T . Observe that each node $n_i^t \in \mathcal{N}^t$ represents a possible sample

value of the history of the stochastic process up to time t . The tree graph structure represents the nesting of information as one time period succeeds the other. The given probabilities $\pi(n_i^t)$, $n_i^t \in \mathcal{N}^t$, satisfy the following condition:

$$\pi(n_i^t) = \sum_{n_k^{t+1} \in \mathcal{S}(n_i^t)} \pi(n_k^{t+1}). \quad (3.1.1)$$

Denote by $M = \{1, \dots, m\}$ the set of players. For each player $j \in M$, we define a set of controls indexed over the set of nodes. Let $U_j^{n_i^t} \subset \mathbb{R}^r$ be the control set of player j which is defined for any player $j \in M$ and any node $n_i^t \in \mathcal{N}^t$, $t = 0, 1, \dots, T-1$ with $\mu_j^{n_i^t}$ a given positive integer. Denote by $u_j(n_i^t) \in U_j^{n_i^t}$ the control of player j at node n_i^t , and let $u(n_i^t) = (u_1(n_i^t), \dots, u_m(n_i^t))$ be a control profile realised at node n_i^t . Denote by $U^{n_i^t} = U_1^{n_i^t} \times \dots \times U_j^{n_i^t} \times \dots \times U_m^{n_i^t}$ the product control sets or the set of control profiles at node n_i^t .

Let $X \subset \mathbb{R}^p$ be a state set with $p \in \mathbb{N}_+$. We assume that the state set X is the same for all nodes $n_i^t \in \mathcal{N}^t$, $t = 0, 1, \dots, T$. A transition function $f^{n_i^t}(\cdot, \cdot) : X \times U^{n_i^t} \mapsto X$ is associated with each node n_i^t . The state equations are given by

$$x(n_i^t) = f^{a(n_i^t)}(x(a(n_i^t)), u(a(n_i^t))), \quad (3.1.2)$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, t = 1, \dots, T. \quad (3.1.3)$$

At each non-terminal node n_i^t , $t = 0, \dots, T-1$, the reward to player j is a function of the state and the controls of all players, given by $\phi_j^{n_i^t}(x(n_i^t), u(n_i^t))$. At a terminal node n_i^T , the reward to player j is given by the function $\Phi_j^{n_i^T}(x(n_i^T))$ which does not directly depend on players' strategies.

We assume that player $j \in M$ maximizes his expected stream of payoffs discounted at rate λ_j ($0 < \lambda_j < 1$). The state equations and the reward functions define the following multistage game, where we let

$$\mathbf{x} = \{x(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T\},$$

$$\mathbf{u} = \{u(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T-1\},$$

and $J_j(\mathbf{x}, \mathbf{u})$ be the payoff to player j , that is,

$$\begin{aligned} J_j(\mathbf{x}, \mathbf{u}) &= \sum_{t=0}^{T-1} \lambda_j^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \phi_j^{n_i^t}(x(n_i^t), u(n_i^t)) \\ &\quad + \lambda_j^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) \Phi_j^{n_i^T}(x(n_i^T)), \quad j \in M, \end{aligned} \quad (3.1.4)$$

s.t.

$$x(n_i^t) = f^{a(n_i^t)}(x(a(n_i^t)), u(a(n_i^t))), \quad (3.1.5)$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, t = 1, \dots, T,$$

$$x(n^0) = x^0. \quad (3.1.6)$$

Definition 3.1. *An admissible S-adapted strategy of player j is a vector of admissible strategies of player j : $\mathbf{u}_j = \{u_j(n_i^t) : u_j(n_i^t) \in U_j^{n_i^t}, n_i^t \in \mathcal{N}^t, t = 0, \dots, T-1\}$ that is, a plan of actions adapted to the history of the random process represented by the event tree.*

Denote by $\mathbf{u} = (\mathbf{u}_j : j \in M)$ the S-adapted strategy vector of the players from M and call it strategy profile. We can thus define a game in normal form $\langle M, \{\mathbf{u}_j\}_{j \in M}, \{W_j(\mathbf{u}, x^0)\}_{j \in M} \rangle$, where $W_j(\mathbf{u}, x^0)$ is a payoff function of player j , $W_j(\mathbf{u}, x^0) = J_j(\mathbf{x}, \mathbf{u})$, \mathbf{x} is a collection of states obtained from \mathbf{u} as the unique solution of the state equations (3.1.2) that emanate from the initial state x^0 .

If the game is played noncooperatively, i. e., each player maximizes his payoff function (3.1.4) then the players will seek a Nash equilibrium in S-adapted strategies defined as follows:

Definition 3.2. *An S-adapted Nash equilibrium is an admissible S-adapted strategy profile $\mathbf{u}^N = (\mathbf{u}_j : j \in M)$ such that for every player $j \in M$ and any S-adapted strategy \mathbf{u}_j the following condition holds:*

$$W_j(\mathbf{u}^N, x^0) \geq W_j((\mathbf{u}_j, \mathbf{u}_{-j}^N), x^0),$$

where $(\mathbf{u}_j, \mathbf{u}_{-j}^N)$ is the S-adapted strategy profile when each player $i \in M$, $i \neq j$ uses his Nash equilibrium policy.

If the players agree to cooperate, then they will maximize the sum of their discounted payoffs throughout the entire horizon, that is,

$$\max_{\mathbf{u}_j: j \in M} \sum_{j \in M} W_j(\mathbf{u}, x^0).$$

Denote the resulting vector of cooperative controls by $\mathbf{u}^*(x^0)$, which solves the given cooperative problem:

$$\mathbf{u}^*(x^0) = \arg \max_{\mathbf{u}_j: j \in M} \sum_{j \in M} W_j(\mathbf{u}, x^0). \quad (3.1.7)$$

The cooperative state trajectory generated by the cooperative controls $\mathbf{u}^*(x^0)$ and equations (3.1.2) is denoted by $\mathbf{x}^* = \{x^*(n_l^t) : n_l^t \in \mathcal{N}^t, t = 0, 1, \dots, T\}$.

For later use, we also need to determine the subgame starting from state $x^*(n_l^t)$ at node $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T - 1$. This subgame takes place on a tree subgraph $\Gamma(n_l^t)$ of the initial graph. The root node of subtree $\Gamma(n_l^t)$ is node n_l^t . The payoff of player $j \in M$ in this subgame is given as follows:

$$W_j(\mathbf{u}(x^*(n_l^t)), x^*(n_l^t)) = \sum_{\theta=t}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_l^\theta} \pi(n_l^\theta | n_l^t) \phi_j^{n_l^\theta}(x(n_l^\theta), u(x(n_l^\theta))) \\ + \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_l^T} \pi(n_l^T | n_l^t) \Phi_j^{n_l^T}(x(n_l^T)), \quad (3.1.8)$$

where $\mathcal{N}_l^\theta = \mathcal{N}^\theta \cap \Gamma(n_l^t)$, $\mathbf{u}(x^*(n_l^t)) = (\mathbf{u}_j(x^*(n_l^t)) : j \in M)$ is an S -adapted strategy profile and $\mathbf{u}_j(x^*(n_l^t))$ is an admissible S -adapted strategy of player j in the subgame starting from node n_l^t , with initial state $x^*(n_l^t)$, $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T - 1$. The term $\pi(n_l^\theta | n_l^t)$ is the conditional probability that node n_l^θ will be realized if the subgame starts from node n_l^t . The conditional probability $\pi(n_l^\theta | n_l^t)$ can be calculated with the formula: $\pi(n_l^\theta | n_l^t) = \pi(n_l^\theta) / \pi(n_l^t)$ if $\pi(n_l^t) \neq 0$. Otherwise, the subgame starting from node n_l^t cannot materialize.

If the players play noncooperatively in the subgame beginning from node n_l^t and state $x^*(n_l^t)$ and find the Nash equilibrium according to Definition 3.2, we denote it as $\mathbf{u}^N(x^*(n_l^t)) = (\mathbf{u}_j^N(x^*(n_l^t)) : j \in M)$. The S -adapted equilibrium payoff of player j in the subgame beginning from node n_l^t and state $x^*(n_l^t)$ is equal to $W_j(\mathbf{u}^N(x^*(n_l^t)), x^*(n_l^t))$. Let $\mathbf{u}_j^N(x^*(n_l^t), [n_v^\tau, n_w^T])$ be the trajectory of equilibrium

strategies $\mathbf{u}_j^N(x^*(n_l^t))$ on the path emanating from node $n_v^\tau \in \mathcal{N}^\tau$, $\tau > t$, and terminating at node $n_w^T \in \mathcal{N}^T$.

If the players cooperate in the subgame starting from node n_l^t and state $x^*(n_l^t)$, then they maximize the sum of their total discounted payoffs, i.e.,

$$\max_{\mathbf{u}_j(x^*(n_l^t)):j \in M} \sum_{j \in M} W_j(\mathbf{u}(x^*(n_l^t)), x^*(n_l^t)).$$

The cooperative controls in the subgame starting from node n_l^t and state $x^*(n_l^t)$ are given as follows:

$$\mathbf{u}^*(x^*(n_l^t)) = \arg \max_{\mathbf{u}_j(x^*(n_l^t)):j \in M} \sum_{j \in M} W_j(\mathbf{u}(x^*(n_l^t)), x^*(n_l^t)). \quad (3.1.9)$$

Therefore, the payoff of player j in the cooperative subgame starting from node n_l^t , with initial state $x^*(n_l^t)$, $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T$, is equal to $W_j(\mathbf{u}^*(x^*(n_l^t)), x^*(n_l^t))$. Denote the trajectory of $\mathbf{u}_j^*(x^*(n_l^t))$ on the path emanating from node n_v^τ , $n_v^\tau \in \mathcal{N}^\tau$, $\tau > t$, and terminating at node $n_w^T \in \mathcal{N}^T$ as $\mathbf{u}_j^*(x^*(n_l^t), [n_v^\tau, n_w^T])$.

Remark 3.1. *The trajectories $\mathbf{u}_j^N(x^*(n_l^t), [n_v^\tau, n_w^T])$ and $\mathbf{u}_j^N(x^*(n_v^\tau), [n_v^\tau, n_w^T])$, $\tau > t$, do not, in general, coincide. One reason is that the trajectory $\mathbf{u}_j^N(x^*(n_l^t), [n_v^\tau, n_w^T])$ has been computed assuming that the players have cooperated only during the time interval $[0, t]$, whereas $\mathbf{u}_j^N(x^*(n_v^\tau), [n_v^\tau, n_w^T])$ is computed under the assumption of a cooperative mode of play on $[0, \tau]$, with $\tau > t$.*

Remark 3.2. *In this chapter, we suppose that the joint-optimization solution and the Nash equilibrium in the whole game and in any subgame are unique. The uniqueness for the joint-optimization solution requires, as usual, strict concavity of the objective function and the control set must be compact and convex. For uniqueness of S -adapted Nash equilibrium, we observe that the multistage game has a normal form representation, and therefore the conditions for uniqueness are the same as in classical games with continuous payoffs with constraints as established in [200].*

§ 3.2 Node consistency of the core in dynamic games played over event trees

Suppose that the players decide to cooperate during the whole duration of the game, and agree to use the core as an allocation mechanism. The players will then define

the cooperative outcomes $W_j(\mathbf{u}^*(x^0), x^0)$, $j \in M$. The Characteristic Function $V(G; x^*(n_l^t))$ in any subgame starting at node n_l^t in state $x^*(n_l^t)$, is defined by

$$V(G; x^*(n_l^t)) : 2^M \rightarrow \mathbb{R}, \quad (3.2.1)$$

where G is a coalition of players, $G \subset M$, and $V(\emptyset; x^*(n_l^t)) = 0$. In particular, $V(G; x^0)$ is the characteristic function in the whole game.

Let $Y(x^*(n_l^t))$ be the set of imputations, that is,

$$Y(x^*(n_l^t)) = \left\{ (y_1(x^*(n_l^t)), \dots, y_m(x^*(n_l^t))) \mid y_j(x^*(n_l^t)) \geq V(\{j\}; x^*(n_l^t)), \right. \\ \left. \forall j \in M \text{ and } \sum_{j \in M} y_j(x^*(n_l^t)) = V(M; x^*(n_l^t)) \right\}.$$

Both the characteristic function and the imputation set, and the imputation depend on the initial node of the subgame and the initial state of this subgame. We omit the node in the notation of argument for simplicity keeping only the initial state.

Definition 3.3. *An imputation $y(x^*(n_l^t)) = (y_1(x^*(n_l^t)), \dots, y_m(x^*(n_l^t)))$ is said to be unstable through a coalition G if $V(G; x^*(n_l^t)) > \sum_{j \in G} y_j(x^*(n_l^t))$. The imputation $y(x^*(n_l^t))$ is unstable if there is a coalition G such that $y(x^*(n_l^t))$ is unstable through G , otherwise $y(x^*(n_l^t))$ is stable.*

The set $C(x^0)$ of stable imputations is the core of the whole game, starting from node n^0 with state x^0 , and is given by

$$C(x^0) = \left\{ (y_1, \dots, y_m) \mid \sum_{j \in G} y_j \geq V(G; x^0) \text{ for } \forall G \subset M, \right. \\ \left. \text{and } \sum_{j \in M} y_j = V(M; x^0) \right\}. \quad (3.2.2)$$

Similarly, we can define the core of the subgame starting from any node n_l^t , with state value $x^*(n_l^t)$, that is,

$$C(x^*(n_l^t)) = \left\{ (y_1, \dots, y_m) \mid \sum_{j \in G} y_j \geq V(G; x^*(n_l^t)) \quad \forall G \subset M, \right. \\ \left. \text{and } \sum_{j \in M} y_j = V(M; x^*(n_l^t)) \right\}. \quad (3.2.3)$$

We assume from now on that the core $C(x^0)$ of the whole game and the core $C(x^*(n_l^t))$ of subgame starting from any node n_l^t with state $x^*(n_l^t)$ are non-empty. Otherwise, the issue of designing node-consistent payments would be irrelevant. The conditions of nonemptiness of the core are not considered in the thesis, since this problem has been thoroughly studied and described in the literature on non-dynamic cooperative games [51, 176].

To compute the characteristic function value $V(G; x^*(n_l^t))$ of coalition G , we need to make assumptions how to measure the “power” of the coalition, i. e., we need to make an assumption on the behavior of left-out-players, i.e., $N \setminus G$. There are several approaches to constructing a characteristic function for a cooperative game if a non-cooperative game is initially defined. These approaches are described in detail in Chapter 1. In this chapter we adopt the γ -characteristic function assumption [81], which states that when coalition G forms, the left-out-players would not take any particular joint action against it, but only use their individually best reply strategies. Consequently, the γ -characteristic-function value of coalition G is given by the S -adapted equilibrium outcome of G in the non-cooperative game between members of G maximizing their joint payoff, and non members playing individually, i.e., maximizing their individual payoffs. The core calculated using the γ -characteristic function is called the γ -core [81]. We now turn to the issue of finding the plan of payments to the players in the dynamic game such that the core is node-consistent.

Definition 3.4. *We call the collection of payments $(\{\beta_j(x^*(n_l^t))\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M)$, defined for each node n_l^t and each player $j \in M$, an imputation distribution procedure of the imputation $y(x^0) = (y_1(x^0), \dots, y_m(x^0))$, if for all $j \in M$ the following conditions hold:*

$$y_j(x^0) = \sum_{\theta=0}^T \lambda_j^\theta \sum_{n_l^\theta \in \mathcal{N}^\theta} \pi(n_l^\theta) \beta_j(x^*(n_l^\theta)), \quad (3.2.4)$$

$$\sum_{j \in M} \beta_j(x^*(n_l^t)) = \sum_{j \in M} \phi_j^{n_l^t}(x^*(n_l^t), u^*(x^*(n_l^t))), \quad (3.2.5)$$

$$\sum_{j \in M} \beta_j(x^*(n_l^T)) = \sum_{j \in M} \Phi_j^{n_l^T}(x^*(n_l^T)), \quad (3.2.6)$$

where condition (3.2.5) holds for any $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$ and condition (3.2.6)

holds for any $n_l^T \in \mathcal{N}^T$.

Equation (3.2.4) means that the expected sum of the discounted payments to player $j \in M$ is equal to the component of the imputation from the core that players have decided to choose as an allocation mechanism in the whole game. Equations (3.2.5), (3.2.6) are the conditions for “admissibility” of imputation distribution procedure, i.e., the sum of payments to the players in any node is equal to the sum of payoffs that they can obtain in this node using the cooperative controls $u_j^*(x^*(n_l^t))$, $j \in M$, at all nodes $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T - 1$.

Definition 3.5. *The imputation $y(x^0) \in C(x^0)$ and corresponding imputation distribution procedure*

$$\left(\left\{ \beta_j(x^*(n_l^t)) \right\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M \right)$$

are called *node-consistent (time-consistent) in the whole game* if for any node $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T$ with state $x^*(n_l^t)$, there exists $y(x^*(n_l^t)) = (y_1(x^*(n_l^t)), \dots, y_m(x^*(n_l^t))) \in C(x^*(n_l^t))$, satisfying the following condition:

$$\sum_{\theta=0}^{t-1} \lambda_j^\theta \sum_{n_k^\theta \in \mathcal{N}^\theta} \pi(n_k^\theta) \beta_j(x^*(n_k^\theta)) + \lambda_j^t \sum_{n_k^t \in \mathcal{N}^t} \pi(n_k^t) y_j(x^*(n_l^t)) = y_j(x^0). \quad (3.2.7)$$

Definition 3.6. *The core $C(x^0)$ in the whole game is a node-consistent allocation mechanism if any imputation $y(x^0)$ from the core $C(x^0)$ is node-consistent.*

Theorem 3.1. *If the core $C(x^0)$ of the whole game and the core $C(x^*(n_l^t))$ of the subgame starting from any node n_l^t with state $x^*(n_l^t)$ are non-empty, then the core $C(x^0)$ is node-consistent when the corresponding imputation distribution procedure for each imputation $y(x^0) \in C(x^0)$ satisfies the following conditions for $t = 0, \dots, T - 1$:*

$$\beta_j(x^*(n_l^t)) = y_j(x^*(n_l^t)) - \lambda_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \pi(n_k^{t+1} | n_l^t) y_j(x^*(n_k^{t+1})), \quad (3.2.8)$$

and for $t = T$:

$$\beta_j(x^*(n_l^T)) = y_j(x^*(n_l^T)), \quad (3.2.9)$$

where $y(x^*(n_l^t)) = (y_1(x^*(n_l^t)), \dots, y_m(x^*(n_l^t))) \in C(x^*(n_l^t))$ for any $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T$ and $\pi(n_k^{t+1} | n_l^t)$ is the conditional probability that node n_k^{t+1} is reached if node n_l^t has been already reached.

Proof. At first, we prove that $\beta_j(x^*(n_l^t))$ determined by (3.2.8) and (3.2.9) is an imputation distribution procedure for imputation $y(x^0)$ from the core $C(x^0)$. We show that conditions (3.2.4)–(3.2.6) are satisfied. The proof of (3.2.4) follows from imputaion distribution procedure construction by (3.2.8) and (3.2.9), and taking into account that $\pi(n_l^t)\pi(n_k^{t+1}|n_l^t) = \pi(n_k^{t+1})$ for any $n_k^{t+1} \in \mathcal{N}^{t+1}, t = 0, \dots, T$.

$$\begin{aligned}
& \sum_{\theta=0}^T \lambda_j^\theta \sum_{n_k^\theta \in \mathcal{N}^\theta} \pi(n_k^\theta) \beta_j(x^*(n_k^\theta)) = \sum_{\theta=0}^{T-1} \lambda_j^\theta \sum_{n_k^\theta \in \mathcal{N}^\theta} \pi(n_k^\theta) \left\{ y_j(x^*(n_l^t)) \right. \\
& \quad \left. - \lambda_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \pi(n_k^{t+1}|n_l^t) y_j(x^*(n_k^{t+1})) \right\} + \lambda_j^T \sum_{n_k^T \in \mathcal{N}^T} \pi(n_k^T) y_j(x^*(n_l^T)) \\
& = \pi(n^0) y_j(x^0) - \lambda_j \sum_{n_k^1 \in \mathcal{S}(n^0)} \pi(n^0) \pi(n_k^1|n^0) y_j(x^*(n_k^1)) + \\
& \quad + \lambda_j \sum_{n_k^1 \in \mathcal{S}(n^0)} \pi(n_k^1) y_j(x^*(n_k^1)) - \dots + \lambda_j^T \sum_{n_k^T \in \mathcal{S}(n_l^{T-1})} \pi(n_k^T) y_j(x^*(n_k^T)) - \\
& \quad - \lambda_j^T \sum_{n_k^T \in \mathcal{S}(n_l^{T-1})} \pi(n_l^{T-1}) \pi(n_k^T|n_l^{T-1}) y_j(x^*(n_k^T)) = y_j(x^0),
\end{aligned}$$

because $\pi(n^0) = 1$.

Now we show that conditions (3.2.5) and (3.2.6) are satisfied if we use the distribution procedure of imputation $y(x^0)$ defined by equations (3.2.8) and (3.2.9). For terminal nodes, condition (3.2.6) is obviously satisfied. Consider condition (3.2.5) for any non-terminal node n_l^t :

$$\begin{aligned}
& \sum_{j \in M} \beta_j(x^*(n_l^t)) = \sum_{j \in M} y_j(x^*(n_l^t)) - \sum_{j \in M} \lambda_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \pi(n_k^{t+1}|n_l^t) y_j(x^*(n_k^{t+1})) = \\
& = V(M; x^*(n_l^t)) - \left(\sum_{j \in M} W_j(\mathbf{u}^*(x^*), x^*(n_l^t)) - \sum_{j \in M} \phi_j^{n_l^t}(x^*(n_l^t), \mathbf{u}^*(x^*)) \right).
\end{aligned}$$

and given that $V(M; x^*(n_l^t)) = \sum_{j \in M} W_j(\mathbf{u}^*(x^*), x^*(n_l^t))$, and following the characteristic function construction we get equality (3.2.5).

If the core $C(x^0)$ of the whole game and the core $C(x^*(n_l^t))$ of subgame starting from any node $n_l^t \in \mathcal{N}^t, t = 0, \dots, T-1$ are non-empty, we can always find at least one imputation $y(x^*(n_l^t)) \in C(x^*(n_l^t))$ and using given imputations $y(x^*(n_l^t))$ for all nodes $n_l^t \in \mathcal{N}^t, t = 0, \dots, T$, construct the imputation distribution procedure

$(\{\beta_j(x^*(n_l^t))\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M)$, with formulas (3.2.8) and (3.2.9) for any imputation from the core $C(x^0)$. The proof that this imputation distribution procedure is node-consistent, i.e., it satisfies equation (3.2.7), can be obtained when we substitute $\beta_j(x^*(n_k^\theta))$ from (3.2.8) and (3.2.9) into (3.2.7) and taking into account that $\pi(n^0) = 1$ and $\pi(n_l^t)\pi(n_k^{t+1}|n_l^t) = \pi(n_k^{t+1})$ for any $n_k^{t+1} \in \mathcal{N}^t, t = 0, \dots, T$:

$$\begin{aligned} & \sum_{\theta=0}^{t-1} \lambda_j^\theta \sum_{n_k^\theta \in \mathcal{N}^\theta} \pi(n_k^\theta) \beta_j(x^*(n_k^\theta)) + \lambda_j^t \sum_{n_k^t \in \mathcal{N}^t} \pi(n_k^t) y_j(x^*(n_l^t)) = \pi(n^0) y_j(x^0) \\ & - \lambda_j \sum_{n_k^1 \in \mathcal{S}(n^0)} \pi(n^0) \pi(n_k^1 | n^0) y_j(x^*(n_k^1)) + \lambda_j \sum_{n_k^1 \in \mathcal{S}(n^0)} \pi(n_k^1) y_j(x^*(n_k^1)) \\ & - \dots - \lambda_j^t \sum_{n_k^t \in \mathcal{N}^t} \pi(n_k^{t-1}) \pi(n_k^t | n_k^{t-1}) y_j(x^*(n_l^t)) + \lambda_j^t \sum_{n_k^t \in \mathcal{N}^t} \pi(n_k^t) y_j(x^*(n_l^t)) \\ & = y_j(x^0). \end{aligned}$$

and we easily get (3.2.7) using that $\pi(n^0) = 1$ and $\pi(n_l^t)\pi(n_k^{t+1}|n_l^t) = \pi(n_k^{t+1})$ for any $n_k^{t+1} \in \mathcal{N}^t, t = 0, \dots, T$. \square

The implementation of the imputation distribution procedure $(\{\beta_j(x^*(n_l^t))\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M)$ requires the definition of side payments between the players according to the following rule for $\forall n_l^t \in \mathcal{N}^t, t = 0, \dots, T-1$:

$$\omega_j(n_l^t, x^*(n_l^t)) = \beta_j(x^*(n_l^t)) - \phi_j^{n_l^t}(x^*(n_l^t), u^*(x^*(n_l^t))), \quad (3.2.10)$$

and for $\forall n_l^T \in \mathcal{N}^T$:

$$\omega_j(n_l^T, x^*(n_l^T)) = \beta_j(x^*(n_l^T)) - \Phi_j^{n_l^T}(x^*(n_l^T)), \quad (3.2.11)$$

where $\omega_j(n_l^t, x^*(n_l^t))$ is the transfer payment that player j does in node n_l^t over the cooperative trajectory $x^*(n_l^t)$, such that

$$\sum_{j \in M} \omega_j(n_l^t, x^*(n_l^t)) = 0$$

for any node n_l^t over cooperative trajectory $x^*(n_l^t)$. Clearly, $\omega_j(n_l^t, x^*(n_l^t))$ can assume any sign depending on the sign of the difference in the right-hand sides of (3.2.10)-(3.2.11).

Finally, we make the following two comments:

1. We do not deal with the issue of selecting a specific imputation from the core when it contains more than one. This problem is far beyond the objective of this work. For illustrating purpose, we use in the next section the nucleolus¹ [206]. It is well-known that the nucleolus is unique and if the core is non-empty, then the nucleolus is in the core (see [176]).
2. We assumed that the players adopt the γ -characteristic function, and consequently computed the γ -core. There is absolutely no conceptual difficulty in extending our approach to α -, and β -characteristic functions.

§ 3.3 Node-consistent core in the game of pollution control

We illustrate our procedure for constructing a node-consistent core with an example of pollution control, which is a three-player stochastic version of the deterministic model in [98]. Denote by $M = \{1, 2, 3\}$ the set of players, and by $\mathcal{T} = \{0, 1, 2\}$ the set of periods. Let $u(n_i^t) = (u_1(n_i^t), u_2(n_i^t), u_3(n_i^t))$ be the vector of countries' emissions of some pollutant in node n_i^t in time period t , and denote by $x(n_i^t)$ the stock of pollution at this node. The evolution of this stock is governed by the following difference equation:

$$x(n_i^t) = (1 - \delta(n_i^t))x(n_i^{t-1}) + \sum_{j \in M} u_j(n_i^t), \quad (3.3.1)$$

and for the root node n^0 , we have

$$x(n^0) = (1 - \delta(n^0))x_{-0} + \sum_{j \in M} u_j(n^0), \quad (3.3.2)$$

where x_{-0} is given, $\delta(n_i^t)$ ($0 < \delta(n_i^t) < 1$) is the stochastic rate of pollution absorption by Mother Nature at node n_i^t . We suppose that $\delta(n_i^t)$ can take two possible values, that is, $\delta(n_i^t) \in \{\underline{d}, \bar{d}\}$, with $\underline{d} < \bar{d}$. The event tree is depicted in Figure 3.1. Let nodes n^0, n_1^1, n_1^2, n_3^2 correspond to the low level of pollution reduction \underline{d} , and nodes n_2^1, n_2^2, n_4^2 correspond to the high level of pollution reduction \bar{d} .

¹For the definition of the nucleolus the definition of excess of coalition S is needed. It represents the "attitude" of coalition S to the payoff vector, i.e., it is the difference between the sum of the payoffs and the characteristic function of the coalition. Then if we rank all the excesses from the largest to the smallest one and find the minimal possible vector of excesses over the set of imputations, we get the nucleolus (see [176]).

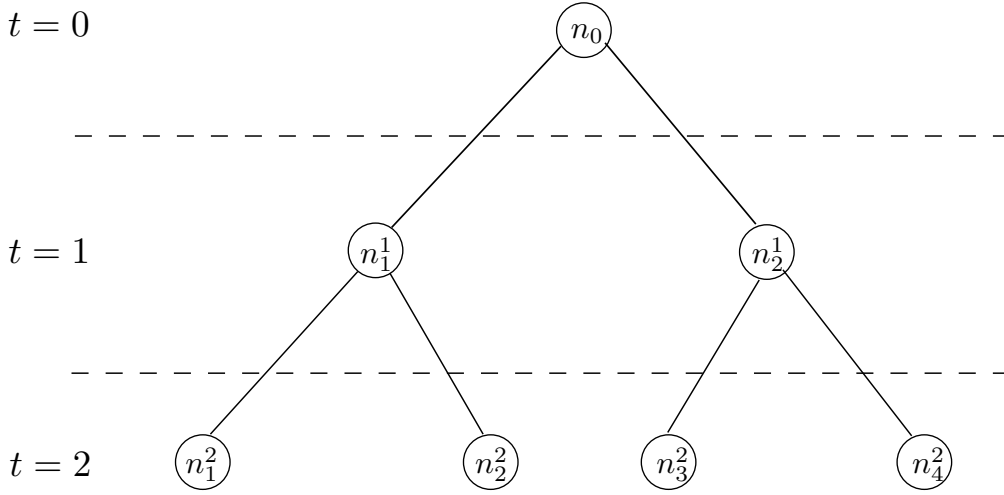


Figure 3.1: Event tree graph.

The damage cost is an increasing convex function in the pollution stock having the form $D_j(x(n_l^t)) = \alpha_j x^2(n_l^t)$, $j \in M$, where α_j is a strictly positive parameter. The cost of emissions is also given by a quadratic function $C_j(u_j(n_l^t)) = \frac{\gamma_j}{2} (u_j(n_l^t) - e)^2$, where e and γ_j are strictly positive constants.

The total discounted cost $J_j(\mathbf{x}, \mathbf{u})$ to be minimized by player $j \in M$ is given by

$$\sum_{t=0}^{T-1} \lambda_j^t \sum_{n_l^t \in \mathcal{N}^t} \pi(n_l^t) (C_j(u_j(n_l^t)) + D_j(x(n_l^t))) + \lambda_j^T \sum_{n_l^T \in \mathcal{N}^T} \pi(n_l^T) D_j(x(n_l^T)),$$

where $\mathbf{x} = \{x(n_l^t)\}$ and $\mathbf{u} = \{\mathbf{u}(n_l^t)\}$, $n_l^t \in \mathcal{N}^t$, $\lambda_j \in (0, 1)$ is a discount rate of player j , subject to (3.3.1), given initial stock $x_{-0} = 0$ before the game starts and constraints: $u_j(n_l^t) \in [0, e]$ for any player $j \in M$ and any node $n_l^t \in \mathcal{N}^T$.

We use the following parameters for the numerical simulation:

$$\begin{aligned} \alpha_1 &= 1, \alpha_2 = 1.5, \alpha_3 = 1.8, \\ \gamma_1 &= 2, \gamma_2 = 1.5, \gamma_3 = 1.4, \\ \underline{\delta} &= 0.15, \bar{\delta} = 0.45, e = 5, \lambda_1 = \lambda_2 = \lambda_3 = 0.9, \\ \pi(n_1^1) &= 0.3, \pi(n_2^1) = 0.7, \pi(n_1^2) = 0.2, \\ \pi(n_2^2) &= 0.1, \pi(n_3^2) = 0.3, \pi(n_4^2) = 0.4. \end{aligned}$$

Using (3.1.7) and (3.1.9), we compute the cooperative controls for each possible subgame and for the whole game. Consequently, we obtain the following cooperative

Time period	$t = 0$	$t = 1$		$t = 2$			
Node	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2
$V(\{1\})$	33.51298	23.60711	19.89234	4.43294	1.85601	4.11655	1.72354
$V(\{2\})$	50.39239	32.77819	31.00779	6.64941	2.78401	6.17483	2.58531
$V(\{3\})$	50.97086	34.33383	32.20934	7.97930	3.340813	7.40980	3.10237
$V(\{1, 2\})$	76.31863	43.69330	41.51454	2.45773	1.02901	2.50077	1.04703
$V(\{1, 3\})$	74.57715	42.49885	40.49344	2.24939	0.94178	2.30923	0.96684
$V(\{2, 3\})$	101.36325	67.11203	63.21713	14.62871	6.12482	13.58463	5.68768
$V(\{1, 2, 3\})$	112.16237	61.35616	59.92551	1.57532	0.65956	1.66402	0.69670

Table 3.1: Characteristic functions for the whole game and subgames.

trajectory:² $(x^*(n^0), x^*(n_1^1), x^*(n_2^1), x^*(n_1^2), x^*(n_2^2), x^*(n_3^2), x^*(n_4^2)) = (0.44621, 0.71208, 0.73186, 0.60527, 0.39165, 0.62208, 0.40252)$.

The γ -characteristic function values are given in Table 3.1. It is easy to verify that the γ characteristic function is subadditive at each node.

The nonempty cores of the whole game and subgames, which are computed using definitions (3.2.2) and (3.2.3), are provided in Table 3.2. (Note that inequalities in (3.2.2) and (3.2.3) are opposite because we consider a cost minimization problem in the example.) Assuming that the players in all subgames agree to adopt the nucleolus as a selection procedure of a single imputation from the core, then they will get the payoffs given in Table 3.3. Finally, based on this imputation (nucleolus), we determine the node-consistent imputation using equations (3.2.8) and (3.2.9) in Theorem 3.1. The results, that is, the $\beta_j(x^*(n_i^t)), j \in M$, are shown in Figure 3.2. There is vector of IDP $\beta(\cdot)$ near each node.

To interpret the results, let us consider the case of player 1. According to the nucleolus, player 1 is entitled to a payoff of 25.92624 in the whole game (see Table 3.3). Our imputation distribution procedure decomposes this amount over nodes, that is, it is equal to the expected sum of discounted payments received in the

²The calculations in this section were made in Maple with *Optimization* and *Simplex* packages.

n^0
$\{y = (y_1, y_2, y_3) : y_3 = 112.162 - y_1 - y_2,$ $61.192 \leq y_1 + y_2 \leq 76.319, 10.799 \leq y_1 \leq$ $\leq 33.513, 37.585 \leq y_2 \leq 50.392\}$
n_1^1
$\{y = (y_1, y_2, y_3) : y_3 = 61.356 - y_1 - y_2,$ $27.022 \leq y_1 + y_2 \leq 43.693, -5.756 \leq y_1 \leq$ $\leq 23.607, 18.857 \leq y_2 \leq 32.778\}$
n_2^1
$\{y = (y_1, y_2, y_3) : y_3 = 59.926 - y_1 - y_2,$ $27.716 \leq y_1 + y_2 \leq 41.515, -3.291 \leq y_1 \leq$ $\leq 19.892, 19.432 \leq y_2 \leq 31.008\}$
n_1^2
$\{y = (y_1, y_2, y_3) : y_3 = 1.575 - y_1 - y_2,$ $-6.404 \leq y_1 + y_2 \leq 2.458, -13.05 \leq y_1 \leq$ $\leq 4.433, -0.674 \leq y_2 \leq 6.649\}$
n_2^2
$\{y = (y_1, y_2, y_3) : y_3 = 0.660 - y_1 - y_2,$ $-2.681 \leq y_1 + y_2 \leq 1.029, -5.465 \leq y_1 \leq$ $\leq 1.856, -0.282 \leq y_2 \leq 2.784\}$
n_3^2
$\{y = (y_1, y_2, y_3) : y_3 = 1.664 - y_1 - y_2,$ $-5.746 \leq y_1 + y_2 \leq 2.501, -11.92 \leq y_1 \leq$ $\leq 4.117, -0.645 \leq y_2 \leq 6.175\}$
n_4^2
$\{y = (y_1, y_2, y_3) : y_3 = 0.697 - y_1 - y_2,$ $-2.406 \leq y_1 + y_2 \leq 1.047, -4.991 \leq y_1 \leq$ $\leq 1.724, -0.270 \leq y_2 \leq 2.585\}$

Table 3.2: Cores.

$t = 0$			$t = 1$		
n^0			n_1^1	n_2^1	
25.926			10.915	10.507	
43.989			25.818	25.220	
42.247			24.623	24.199	
$t = 2$					
n_1^2	n_2^2	n_3^2	n_4^2		
-4.192	-1.76	-3.67	-1.54		
2.988	1.251	2.765	1.158		
2.779	1.164	2.573	1.077		

Table 3.3: Nucleoli for the whole game and all subgames.

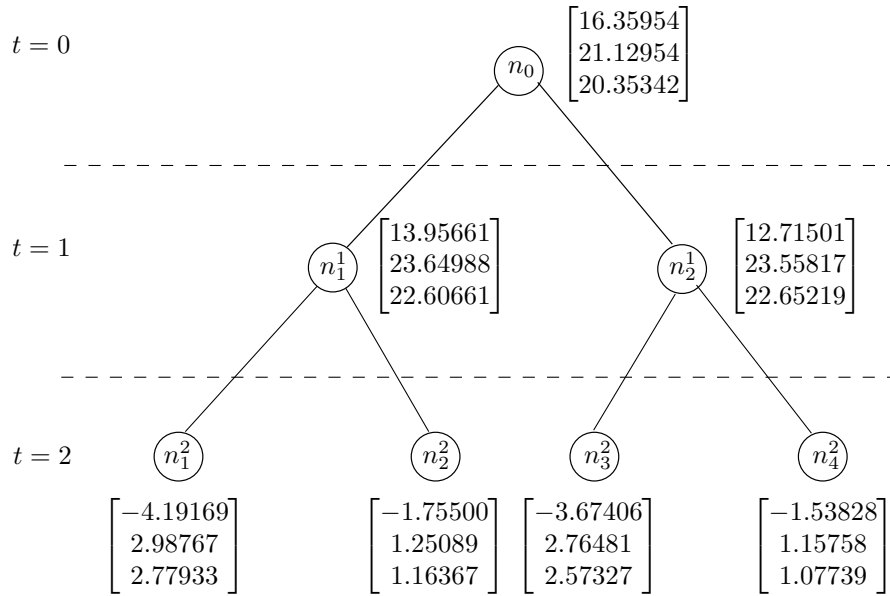


Figure 3.2: Imputation distribution procedure.

subgame emanating from this node (here the whole game). Indeed, we have

$$\begin{aligned} 25.92624 = & 16.35954 + 0.3 \cdot 0.9 \cdot 13.95661 + 0.7 \cdot 0.9 \cdot 12.71501 \\ & - 0.2 \cdot (0.9)^2 \cdot 4.19169 - 0.1 \cdot (0.9)^2 \cdot 1.75500 \\ & - 0.3 \cdot (0.9)^2 \cdot 3.67406 - 0.4 \cdot (0.9)^2 \cdot 1.53828. \end{aligned}$$

Similarly, if the players agree to implement the nucleolus solution in the subgame starting in node n_1^1 , then player 1 would collect 10.91510 in this subgame. It is easy to verify that this amount corresponds to the expected sum of discounted payments received in this subgame, that is,

$$10.91510 = 13.95661 - (2/3 \cdot 0.9 \cdot 4.19169) - (1/3 \cdot 0.9 \cdot 1.75500),$$

where $2/3$ and $1/3$ are the probabilities of reaching nodes n_1^2 and n_2^2 from node n_1^1 , respectively. Notice that the imputation procedure β_j is node-consistent and allows to implement the imputation (in our case, the nucleolus) in the dynamics. As an initial imputation from the core, any other imputation can be chosen. The imputation distribution procedure can be constructed in a similar way. Thus, we have demonstrated node consistency of the core under the following condition: the implementation of the imputations should be realized using the IDP, given by formulas (3.2.8) and (3.2.9).

§ 3.4 Subgame-perfect ε -equilibrium in dynamic games played over event trees

In this section, we consider the game played over event tree, described in the paragraph §3.1. Since the game has a finite duration, it is impossible to guarantee the existence of subgame perfect equilibrium with cooperative payoffs. Therefore, we verify the conditions for the existence of approximated or ε -equilibrium (see [128, 141]) for the considered class of dynamic games.

Now, we consider the game played over an event tree as a game in extensive form with closed-loop information structure. This means that each player knows not only the current node $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T$ and what she played on the (unique) path leading from the initial node n^0 to $a(n_i^t)$, but also what the other players did in

all previous periods. Let this path be $(n^0, n_{i_1}^1, \dots, n_{i_{t-1}}^{t-1}, n_i^t)$ and denote it by $P(n_i^t)$. The collection of nodes and corresponding strategy profiles realized on path $P(n_i^t)$, except node n_i^t , is called the *History of Node n_i^t* and is denoted by

$$H(n_i^t) = ((n^0, u(n^0)), (n_{i_1}^1, u(n_{i_1}^1)), \dots, (n_{i_{t-1}}^{t-1}, u(n_{i_{t-1}}^{t-1}))).$$

Definition 3.7. A behavior strategy $\sigma_j = \{\sigma_j(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T-1\}$ of player $j \in M$ in the game played over an event tree is a mapping that associates to each node n_i^t an action $u_j(n_i^t) \in U_j^{n_i^t}$ with each history $H(n_i^t)$, that is,

$$\sigma_j(n_i^t) : H(n_i^t) \longrightarrow U_j^{n_i^t}.$$

In other words, a behavior strategy tells a player what to do at each node n_i^t of the event tree $\Gamma(n^0)$. Denote by Σ_j the set of behavior strategies of player j , by $\sigma = (\sigma_1, \dots, \sigma_m)$ a behavior strategy profile, and by $\Sigma = \Sigma_1 \times \dots \times \Sigma_m$ the set of possible strategy profiles. For a given behavior strategy profile, we can compute the expected payoff in all subgames, including the whole game, for any given initial state. To avoid adding new notations, we denote the payoff of player j in the subgame starting at node n_i^t and in state $x(n_i^t)$ as a function of the behavior strategy profile, by

$$W_j(\sigma, x(n_i^t)) = W_j(\mathbf{u}(n_i^t), x(n_i^t)),$$

where $\mathbf{u}(n_i^t)$ is a trajectory of controls in the subgame starting at node n_i^t and determined by profile σ .

Definition 3.8. A behavior strategy profile $\hat{\sigma}$ is an ε -equilibrium if, for each player $j \in M$ and each strategy $\sigma_j \in \Sigma_j$, the following inequality holds:

$$W_j(\hat{\sigma}, x^0) \geq W_j((\hat{\sigma}_{-j}, \sigma_j), x^0) - \varepsilon. \quad (3.4.1)$$

Definition 3.9. A behavior strategy profile $\hat{\sigma}$ is a subgame-perfect ε -equilibrium if, for each player $j \in M$, each node $n_i^t \in \mathcal{N}^t$, each strategy $\sigma_j \in \Sigma_j$, and each history $H(n_i^t)$, the following inequality holds:

$$W_j(\hat{\sigma}, x(n_i^t)|H(n_i^t)) \geq W_j((\hat{\sigma}_{-j}, \sigma_j), x(n_i^t)|H(n_i^t)) - \varepsilon,$$

where $W_j(\hat{\sigma}, x(n_i^t)|H(n_i^t))$ is player j 's payoff in the subgame starting at node n_i^t in state $x(n_i^t)$ given by history $H(n_i^t)$ when the players use strategy profile $\hat{\sigma}$.

To strategically support cooperation in the finite-horizon dynamic game played over an event tree, we shall construct an approximated equilibrium in behavior strategies with a closed-loop information structure. The development is in line with what has been done in repeated games, modulo the fact that here we additionally have a vector of state variables evolving over time by equations (3.1.5) и (3.1.6). Before formally defining the trigger strategy that will be used in this construction, let us suppose

1. that the players want to realize the cooperative trajectory \mathbf{u}^* from (3.1.7),
2. that if a player j deviates from cooperation at node $a(n_i^t)$ realizing control $u_j(a(n_i^t)) \neq u_j^*(a(n_i^t))$, then cooperation breaks up and all players realize Nash equilibrium strategies in the subgame starting at n_i^t in state $x(n_i^t) = f^{a(n_i^t)}(x^*(a(n_i^t)), (u_{-j}^*(a(n_i^t)), u_j(a(n_i^t))))$.

If the cooperation breaks up the payoff of player $p \in M$ in the subgame starting at n_i^t in state $x(n_i^t)$ is $W_p(\mathbf{u}^N(x(n_i^t)), x(n_i^t))$ that is, the profit that player p achieves when all players implement Nash equilibrium strategies. Denote by $\hat{\sigma} = (\hat{\sigma}_p : p \in M)$ a behavior strategy profile such that prescribes player $p \neq j$ to realize cooperative control $u_p^*(n_i^t)$ in node n_i^t if in the history of this node no deviations from cooperative trajectory have been observed, and to realize $u_p^N(n_i^t)$ otherwise.

Denote by $\hat{\mathbf{u}}(x(n_i^t)) = \{\hat{u}(n_i^\theta) : n_i^\theta \in \Gamma(n_i^t)\}$ the collection of controls corresponding to strategy profile $\hat{\sigma}$ such that the trigger mode of behavior strategy is implemented in the subgame starting at node n_i^t in state $x(n_i^t)$. If in the history of n_i^θ the individual deviation of player j has been observed, we denote the collection of controls corresponding $\hat{\sigma}$ by $\hat{\mathbf{u}}^j(n_i^\theta)$.

The trigger behavior strategy of a player consists of two behavior types or two modes:

The nominal mode. If the history of node n_i^t coincides with

$$H^*(n_i^t) = ((n^0, u^*(n^0)), (n_{i_1}^1, u^*(n_{i_1}^1)), \dots, (n_{i_{t-1}}^{t-1}, u^*(n_{i_{t-1}}^{t-1}))), \quad (3.4.2)$$

i.e., all players used their cooperative controls on the path $P(n_{i_{t-1}}^{t-1})$, that is, from n^0 until $n_{i_{t-1}}^{t-1}$, then player $p, p \in M$ implements $u_p^*(n_i^t)$ in node n_i^t .

The trigger mode. If the history of node n_l^t is such that there exists a node n on the path $P\left(n_{i_{t-1}}^{t-1}\right)$ such that $u(n) \neq u^*(n)$, then player p 's strategy is Nash equilibrium strategy calculated for the subgame starting from the successor of n and corresponding state. Here the history of node n_l^t is such that there exists a node n and at least one deviating player $j \in M$, $j \neq p$, that is, the history $H(n)$ of node n is part of $H^*(n_l^t)$, and $(n, u(n))$ is not part of $H^*(n_l^t)$, but if we replace the control $u_j(n)$ of player j in node n by the cooperative control $u_j^*(n)$, then the pair $(n, (u_{-j}(n), u_j^*(n)))$ will be $(n, u^*(n))$ and part of history $H^*(n_l^t)$.

Formally speaking, the trigger behavior strategy of player $p \in M$ is defined as follows:

$$\hat{\sigma}_p(H(n_l^t)) = \begin{cases} u_p^*(n_l^t), & \text{if } H(n_l^t) = H^*(n_l^t), \\ \hat{u}_p(n_l^t), & \text{if there exists a node } n \text{ on path } P(n_l^t) \\ & \text{such that } u(n) \neq u^*(n) \end{cases} \quad (3.4.3)$$

where $\hat{u}_p(n_l^t)$ is player p 's control in node n_l^t . The control $\hat{u}_p(n_l^t)$ implements the punishing strategy in the subgame starting in the unique node belonging to the set $\mathcal{S}(n) \cap P(n_l^t)$. The control $\hat{u}_p(n_l^t)$ coincides with $u_p^N(n_l^t)$ calculated as a part of Nash equilibrium for subgame starting at node $n_1 = \mathcal{S}(n) \cap P(n_l^t)$ in state $x(n_1)$.

To avoid further complicating the notation, we omitted the state argument in the punishing control and the trigger strategy, but we stress that they depend on the state value. Let node n_1 be a direct successor of node n in which player j deviates. The collection of controls $(u_{-j}^*(n), u_j(n))$ is then realized, and the state value in node n_1 can be calculated using the state dynamics $x(n_1) = f^n(x^*(n), (u_{-j}^*(n), u_j(n)))$. The control $\hat{u}_p(n_1)$ is part of the control profile $\hat{u}(n_1) = (u_p^N(n_1) : p \in M)$ where $u_p^N(n_1)$ is a control of player p in node n_1 .

Now, in the subgame starting from node $n_l^t \in \mathcal{N}^t$ in state $x(n_l^t)$, the collection of controls punishing player j 's individual deviation is given by

$$\hat{\mathbf{u}}^j(n_l^t) = (\hat{\mathbf{u}}_p^j(n_l^t) : p \in M),$$

where $\hat{\mathbf{u}}_p^j(n_l^t) = \{\hat{u}_p^j(n_l^\theta) : n_l^\theta \in \Gamma(n_l^t)\}$. This collection of controls generates Nash

equilibrium trajectory of states in player j 's punishment in this subgame, that is,

$$\hat{\mathbf{x}}^j(n_l^t) = \{\hat{x}^j(n_l^\theta) : n_l^\theta \in \Gamma(n_l^t)\}.$$

To construct the trigger strategies, we need to find m punishing strategy profiles for each subgame.

Theorem 3.2. *For any $\varepsilon \geq \tilde{\varepsilon}$ in the game played over an event tree there exists a subgame-perfect ε -equilibrium in trigger strategies with players' payoffs $(W_1(\mathbf{u}^*, x^0), \dots, W_m(\mathbf{u}^*, x^0))$, and*

$$\tilde{\varepsilon} = \max_{j \in M} \max_{\substack{n_l^t \in \mathcal{N}^t \\ t=0, \dots, T-1}} \varepsilon_j(n_l^t), \quad (3.4.4)$$

where

$$\begin{aligned} \varepsilon_j(n_l^t) = & \max_{u_j(n_l^t) \in U_j^{n_l^t}} \left\{ \phi_j^{n_l^t}(x^*(n_l^t), (u_{-j}^*(n_l^t), u_j(n_l^t))) - \phi_j^{n_l^t}(x^*(n_l^t), u^*(n_l^t)) \right. \\ & + \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \left(\phi_j^{n_l^\theta}(\hat{x}^j(n_l^\theta), \hat{u}^j(n_l^\theta)) - \phi_j^{n_l^\theta}(x^*(n_l^\theta), u^*(n_l^\theta)) \right) \\ & \left. + \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \left(\Phi_j^{n_l^T}(\hat{x}(n_l^T)) - \Phi_j^{n_l^T}(x^*(n_l^T)) \right) \right\}, \quad (3.4.5) \end{aligned}$$

and $\hat{u}^j(n_l^\theta)$ is a control profile in node n_l^θ corresponding to a behavior strategy profile $\hat{\sigma}$ determined by (3.4.3) and when the trigger mode of the strategy begins in the subgame starting at the node belonging to the set $\mathcal{S}(n_l^t)$ and in state $f^{n_l^t}(x(n_l^t), (u_{-j}^*(n_l^t), u_j(n_l^t)))$. Therefore, the differences in the second and third lines of (3.4.5) also depend on the control $u_j(n_l^t)$. The state $\hat{x}^j(n_l^\theta)$, $n_l^\theta \in \Gamma(n_l^t)$ is a state trajectory corresponding to $\hat{\mathbf{u}}^j(n_l^t)$.

Proof. Consider the trigger behavior strategy $\hat{\sigma} = (\hat{\sigma}_p : p \in M)$ defined in (3.4.3), and the subgame starting from any node $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$. Consider possible histories of any node n_l^t , and compute the benefit of player j deviating at

node n_l^t . Her payoff in this subgame will be given by

$$\begin{aligned} W_j(\mathbf{u}^*(n_l^t), x^*(n_l^t)) &= \phi_j^{n_l^t}(x^*(n_l^t), u^*(n_l^t)) \\ &+ \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \phi_j^{n_l^\theta}(x^*(n_l^\theta), u^*(n_l^\theta)) \\ &+ \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \Phi_j^{n_l^T}(x^*(n_l^T)), \end{aligned} \quad (3.4.6)$$

where $\mathcal{N}_\Gamma^\theta = \mathcal{N}^\theta \cap \Gamma(n_l^t)$, $\mathbf{u}^*(n_l^t) = (u_j^*(n_l^t) : j \in M)$ is an S -adapted cooperative strategy profile.

First, let the history of node n_l^t be $H^*(n_l^t)$. Suppose player j deviates in node n_l^t from the cooperative trajectory. In this case, she may secure the following payoff in the subgame starting at node n_l^t , given the information that the behavior strategy profile $\hat{\sigma} = (\hat{\sigma}_p(\cdot) : p \in M)$ determined by (3.4.3) will materialize:

$$\begin{aligned} &\max_{u_j(n_l^t) \in U_j^{n_l^t}} \left\{ \phi_j^{n_l^t}(x^*(n_l^t), (u_{-j}^*(n_l^t), u_j(n_l^t))) \right. \\ &+ \left. \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \phi_j^{n_l^\theta}(\hat{x}^j(n_l^\theta), \hat{u}^j(n_l^\theta)) + \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \Phi_j^{n_l^T}(\hat{x}^j(n_l^T)) \right\}, \end{aligned} \quad (3.4.7)$$

where punishing Nash strategy starts to be implemented in nodes from $\mathcal{S}(n_l^t)$. Then, we may compute the benefit from deviation of player j in node n_l^t as a difference between (3.4.7) and (3.4.6), namely:

$$\begin{aligned} \varepsilon_j(n_l^t) &= \max_{u_j(n_l^t) \in U_j^{n_l^t}} \left\{ \phi_j^{n_l^t}(x^*(n_l^t), (u_{-j}^*(n_l^t), u_j(n_l^t))) - \phi_j^{n_l^t}(x^*(n_l^t), u^*(n_l^t)) \right. \\ &+ \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \left(\phi_j^{n_l^\theta}(\hat{x}^j(n_l^\theta), \hat{u}^j(n_l^\theta)) - \phi_j^{n_l^\theta}(x^*(n_l^\theta), u^*(n_l^\theta)) \right) \\ &\left. + \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \left(\Phi_j^{n_l^T}(\hat{x}^j(n_l^T)) - \Phi_j^{n_l^T}(x^*(n_l^T)) \right) \right\}, \end{aligned} \quad (3.4.8)$$

Second, suppose that the history of node n_l^t does not coincide with $H^*(n_l^t)$. This means that all players have switched from nominal mode to the cooperative one. Player j will have no benefit from the deviation in node n_l^t because the players

implement their Nash equilibrium strategies regardless of which player (or group of players) has deviated in the previous nodes.

Calculating the maximum benefit from deviation for any subgame and any player given by (3.4.8), we obtain the value of $\tilde{\varepsilon}$ from the theorem statement, that is,

$$\tilde{\varepsilon} = \max_{j \in M} \max_{\substack{n_i^t \in \mathcal{N}^t \\ t=1, \dots, T-1}} \varepsilon_j(n_i^t).$$

And for any $\varepsilon \geq \tilde{\varepsilon}$ the behavior strategy profile determined by (3.4.3) is a subgame-perfect ε -equilibrium by construction. \square

Remark 3.3. *The computed quantity $\tilde{\varepsilon}$ is the maximal benefit that a player may obtain in the whole game by deviating from cooperation. Clearly, the actual value of $\tilde{\varepsilon}$ depends on the variables' measurement units, as well as on other data of the game such as the number of players and the duration of the game. In this sense, there is no obvious way of deriving an upper bound for $\tilde{\varepsilon}$. However, one way of making sense of its value is in assessing it in relative terms, instead in absolute terms. One possibility is to compare it to the total cooperative payoff, and compute the following ratio that we call CoC for Cost of Cooperation:*

$$\text{CoC} = \frac{\tilde{\varepsilon}}{\sum_{j \in M} W_j(\mathbf{u}^*, x^0)}.$$

§ 3.5 Subgame perfect ε -equilibrium in the game of pollution control

To illustrate the results of the previous section, we consider a three-player stochastic version of the deterministic model of pollution control in [98]. Denote by $M = \{1, 2, 3\}$ the set of players, and by $\mathcal{T} = \{0, 1, 2, 3\}$ the set of periods. Let $u(n_i^t) = (u_1(n_i^t), u_2(n_i^t), u_3(n_i^t))$ be the vector of countries' emissions of some pollutant and denote by $x(n_i^t)$ the stock of pollution at node n_i^t in time period t . The evolution of this stock is governed by the following difference equation:

$$x(n_i^t) = (1 - \delta(a(n_i^t)))x(a(n_i^t)) + \sum_{j \in M} u_j(a(n_i^t)), \quad (3.5.1)$$

$$x(n^0) = x^0, \quad (3.5.2)$$

with the initial stock x^0 at root node n^0 being given, and $\delta(n_i^t)$ ($0 < \delta(n_i^t) < 1$) is the stochastic rate of pollution absorption by nature at node n_i^t . We suppose that $\delta(n_i^t)$ can take two possible values, that is, $\delta(n_i^t) \in \{\underline{\delta}, \bar{\delta}\}$, with $\underline{\delta} < \bar{\delta}$. The event tree is depicted in Figure 3.3. Let nodes n_2^1, n_2^2, n_4^2 correspond to the low level of pollution reduction $\underline{\delta}$, and nodes n^0, n_1^1, n_1^2, n_3^2 correspond to the high level of pollution reduction $\bar{\delta}$.

The damage cost is an increasing convex function in the pollution stock having the quadratic form $D_j(x(n_i^t)) = \alpha_j x^2(n_i^t)$, $j \in M$, where α_j is a strictly positive parameter. The cost of emissions is also given by a quadratic function $C_j(u_j(n_i^t)) = \frac{\gamma_j}{2} (u_j(n_i^t) - e)^2$, where e and γ_j are strictly positive constants.

At each node $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, 3$ the cost function of player $j \in M$ are defined by two functions: (i) $D_j(x(n_i^t)) = \alpha_j x^2(n_i^t)$, where $\alpha_j > 0$; (ii) $C_j(u_j(n_i^t)) = \frac{\gamma_j}{2} (u_j(n_i^t) - e)^2$, where e and γ_j are positive constants.

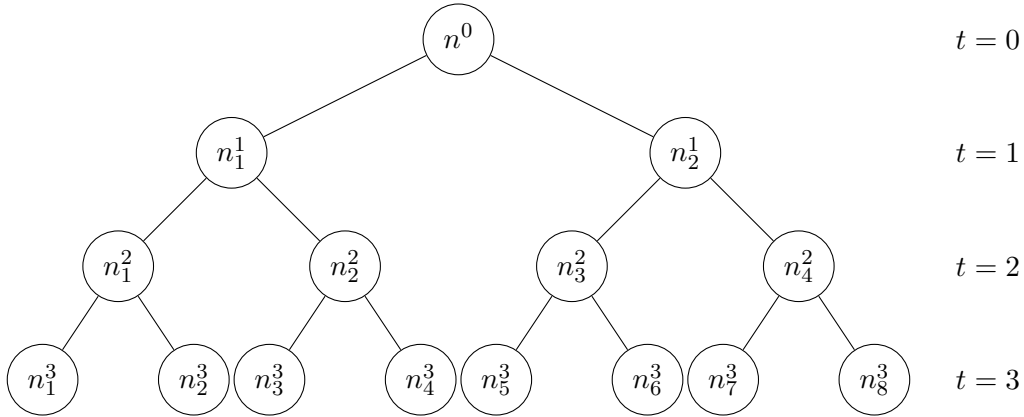


Figure 3.3: Event tree graph for $T = 3$.

The total discounted cost $J_j(\mathbf{x}, \mathbf{u})$ to be minimized by player $j \in M$ is given by

$$\sum_{t=0}^2 \lambda_j^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) (C_j(u_j(n_i^t)) + D_j(x(n_i^t))) + \lambda_j^3 \sum_{n_i^3 \in \mathcal{N}^3} \pi(n_i^3) D_j(x(n_i^3)),$$

where $\mathbf{x} = \{x(n_i^t)\}$ and $\mathbf{u} = \{u(n_i^t)\}$, $n_i^t \in \mathcal{N}^t$, $\lambda_j \in (0, 1)$ is a discount rate of player j , subject to (3.5.1), given initial stock $x^0 = 0$ before the game starts and constraints: $u_j(n_i^t) \in [0, e]$ for any player $j \in M$ and any node $n_i^t \in \mathcal{N}^t$, $t = 0, 1, 2$.

We use the following parameters for the numerical simulation:

$$\begin{aligned}\alpha_1 &= 0.1, \alpha_2 = 0.2, \alpha_3 = 0.3, \\ \gamma_1 &= 0.9, \gamma_2 = 0.8, \gamma_3 = 0.7, \\ \underline{\delta} &= 0.45, \bar{\delta} = 0.8, e = 30, \lambda_1 = \lambda_2 = \lambda_3 = 0.9, \\ \pi(n_1^1) &= 0.6, \pi(n_2^1) = 0.4, \\ \pi(n_1^2) &= 0.3, \pi(n_2^2) = 0.3, \pi(n_3^2) = 0.3, \pi(n_4^2) = 0.1, \\ \pi(n_1^3) &= 0.1, \pi(n_2^3) = 0.2, \pi(n_3^3) = 0.1, \pi(n_4^3) = 0.2, \\ \pi(n_5^3) &= 0.05, \pi(n_6^3) = 0.25, \pi(n_7^3) = 0.05, \pi(n_8^3) = 0.05.\end{aligned}$$

Using (3.1.7) and (3.1.9), we compute the cooperative controls for each possible subgame and for the whole game. The cooperative state trajectory is given by

$$\begin{array}{cccccccc} x^*(n^0) & x^*(n_1^1) & x^*(n_2^1) & x^*(n_1^2) & x^*(n_2^2) & x^*(n_3^2) & x^*(n_4^2) & x^*(n_1^3) \\ 0 & 55.177 & 55.177 & 63.4188 & 63.4188 & 79.0484 & 79.0484 & 72.8623 \\ x^*(n_2^3) & x^*(n_3^3) & x^*(n_4^3) & x^*(n_5^3) & x^*(n_6^3) & x^*(n_7^3) & x^*(n_8^3) & \\ 72.8623 & 88.6125 & 88.6125 & 75.0804 & 75.0804 & 94.7123 & 94.7123 & \end{array}$$

We use (3.1.8) to compute the players' costs in each of the subgames and report them in Table 3.4.

Table 3.4: Players' costs in the cooperative control profile.

Time period	$t = 0$	$t = 1$		$t = 2$			
Node	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2
$W_1(\mathbf{u}(\cdot), x^*(\cdot))$	249.371	213.06	242.807	122.401	161.771	149.748	201.349
$W_2(\mathbf{u}(\cdot), x^*(\cdot))$	381.031	344.644	394.864	214.701	279.02	267.534	351.835
$W_3(\mathbf{u}(\cdot), x^*(\cdot))$	517.495	479.554	550.625	308.23	398.087	386.625	504.397
	$t = 3$						
	n_1^3	n_2^3	n_3^3	n_4^3	n_5^3	n_6^3	n_7^3
	53.0891	53.0891	78.5218	78.5218	56.3706	56.3706	89.7041
	106.178	106.178	157.044	157.044	112.741	112.741	179.408
	159.267	159.267	235.565	235.565	169.112	169.112	269.112

For each player $j \in M$ and any node $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T - 1$ we need to solve the optimization problem defined in (3.4.8), with min instead of max. Once we obtain these costs, we compute the differences with the cooperative payoffs and

give them in Table 3.5. Based on these differences, we determine the values $\varepsilon_j(n_i^t)$ for $j \in M$ and $n_i^t \in \mathcal{N}^t$, $t = 0, 1, 2$ (see Table 3.6).

Table 3.5: Maximum benefits from deviation for any subgame.

Time period	$t = 0$	$t = 1$		$t = 2$			
Node	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2
Player 1	52.0734	45.6117	53.2126	23.4217	34.6419	24.8694	39.5753
Player 2	-24.5520	5.18347	11.0904	16.4602	24.3455	17.4776	27.8126
Player 3	-111.9650	-38.4121	-33.8928	10.2658	15.1836	10.9003	17.346

Table 3.6: Values of $\varepsilon_j(\cdot)$ for any subgame and any player.

Time period	$t = 0$	$t = 1$		$t = 2$			
Node	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2
ε_1	52.0734	45.6117	53.2126	23.4217	34.6419	24.8694	39.5753
ε_2	0	5.18347	11.0904	16.4602	24.3455	17.4776	27.8126
ε_3	0	0	0	10.2658	15.1836	10.9003	17.346

From Table 3.6 we see that at root node n^0 only player 1 benefits from deviating from cooperation. At time $t = 1$, players 1 and 2 can gain by deviating, whereas at terminal period 2, all players gain by deviating, which is expected. The largest benefit from deviation is realized for player 1 in node n_2^1 . Finally, we note that, in this example, $\tilde{\varepsilon}$ is equal to 53.2126. The cost of cooperation in this game is $53.2126 / (249.371 + 381.031 + 517.495)$, which is approximately 4.63%.

The game at hand belongs to the class of environmental games with negative externalities. Given the general result that cooperation is hard to achieve in such setting (see, e.g., the survey in [109]), the results in the above short-time horizon example come at no surprise. To get a hint of the impact of having a longer time horizon, we now let the set of periods be $\mathcal{T} = \{0, 1, \dots, 10\}$, with everything else being equal.

The event tree is depicted in Figure 3.4. It is a binary tree, i.e., each node in periods $t = 0, \dots, 9$ has two successors. The conditional probability of realization of the upward successor of any node is $\frac{1}{3}$ and is $\frac{2}{3}$ for a downward successor. So, for instance, we have probabilities $\pi(n_1^1) = \frac{1}{3}$ and $\pi(n_2^1) = \frac{2}{3}$ in period 1, and probabilities $\pi(n_1^2) = \frac{1}{9}$, $\pi(n_2^2) = \frac{2}{9}$, $\pi(n_3^2) = \frac{2}{9}$, $\pi(n_4^2) = \frac{4}{9}$ for $t = 2$. The root

node n^0 and all upward (or left-handed) nodes have the low rate $\underline{\delta}$ of pollution absorption by nature, and all downward (or right-handed) nodes have the high level $\bar{\delta}$ of pollution absorption.

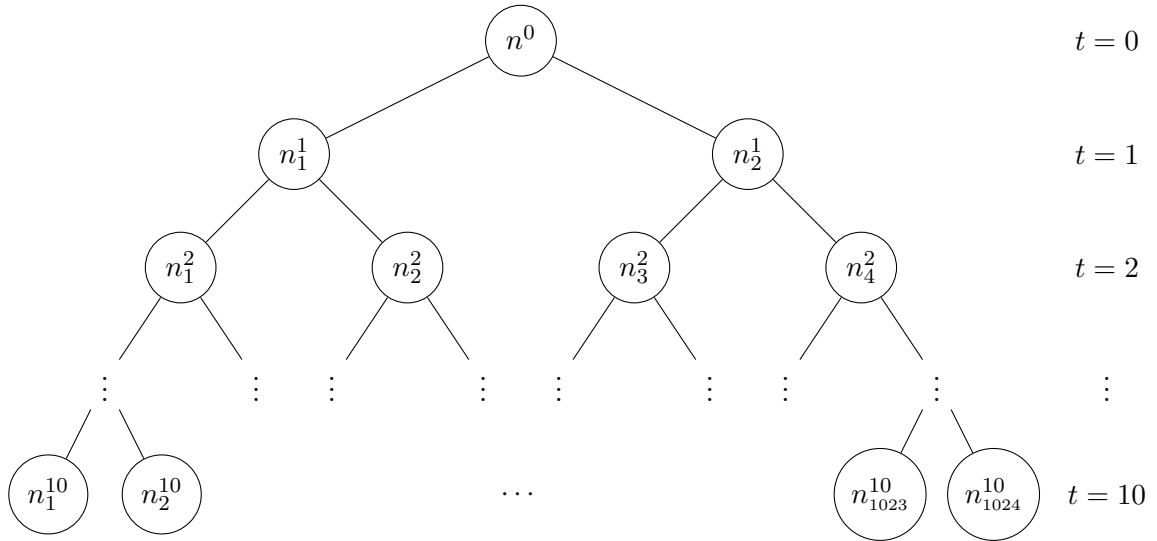


Figure 3.4: Event tree graph for $T = 10$.

As in the previous example, we compute the cooperative payoffs as well as the benefits from deviating from the cooperative solution in all subgames. To save on space, in Tables 3.7 and 3.8, we only print the results regarding the benefits of cheating on the agreement and the values of $\varepsilon_j(n_i^t)$. (More precisely, we only show the max values for each time period.) These tables show that the first time a player could benefit from deviating from cooperation is player 1 in period 0. Interestingly, player 2 would deviate in the period previous to the last one and player 3 would only deviate in the last period. The value of $\tilde{\varepsilon}$ is equal to 132.87, and the cost of cooperation is 4.37%. As under a cooperative regime, the total accumulated pollution is lower than under noncooperation, the results obtained here are encouraging, not only from an economic point of view, but also from environmental one.

Table 3.7: The maximal benefit from deviation in time period t calculated for the example with 11 periods.

Time period	0	1	2	3	4	5	6	7	8	9
Player 1	117.05	131.23	132.87	128.121	120.169	110.289	98.8774	85.8045	69.719	43.7459
Player 2	-162.213	-138.843	-118.765	-99.4618	-79.2663	-57.3184	-33.1637	-6.82434	19.3776	30.7436
Player 3	-497.143	-458.461	-414.163	-364.781	-309.905	-249.008	-181.732	-108.818	-33.8152	19.174

Table 3.8: The maximal ε_j in time period t calculated for the example with 11 periods.

Time period	0	1	2	3	4	5	6	7	8	9
$\max_{n_i^t \in \mathcal{N}^t} \varepsilon_1(n_i^t)$	117.05	131.23	132.87	128.121	120.169	110.289	98.8774	85.8045	69.719	45.1047
$\max_{n_i^t \in \mathcal{N}^t} \varepsilon_2(n_i^t)$	0	0	0	0	0	0	0	0	19.3776	30.7436
$\max_{n_i^t \in \mathcal{N}^t} \varepsilon_3(n_i^t)$	0	0	0	0	0	0	0	0	0	19.174

Figure 3.5 represents the relation between $\tilde{\varepsilon}$ and discount factor $\lambda_j = \lambda$, $j = 1, 2, 3$ for the game with 11 time periods. The larger the discount factor, the larger is $\tilde{\varepsilon}$.

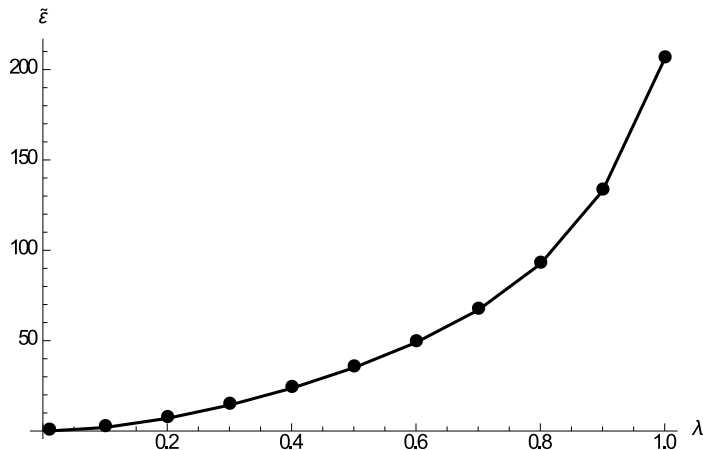
Figure 3.5: Value of $\tilde{\varepsilon}$ as a function of discount factor λ .

Table 3.9 gives the information on when each player has the first positive benefit from deviation. Player 3 will not deviate before the last period if $\lambda \geq 0.5$. But Player 1 has a positive benefit from deviation from period 0 onwards for any λ . With low discount factors, players have lower benefits from deviation, when compared with the benefits from deviation with high discount factors; however, with low discount factors, players may start to deviate earlier than with high discount factors.

Table 3.9: The first time period when the player has an incentive to deviate.

λ	0.01	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
Player 1	0	0	0	0	0	0	0	0	0	0	0
Player 2	0	0	0	0	1	1	5	7	7	8	8
Player 3	0	0	1	8	8	9	9	9	9	9	9

Figure 3.6 represents the cost of cooperation as a function of discount factor λ . We may notice that this value does not exceed 7% for any λ .

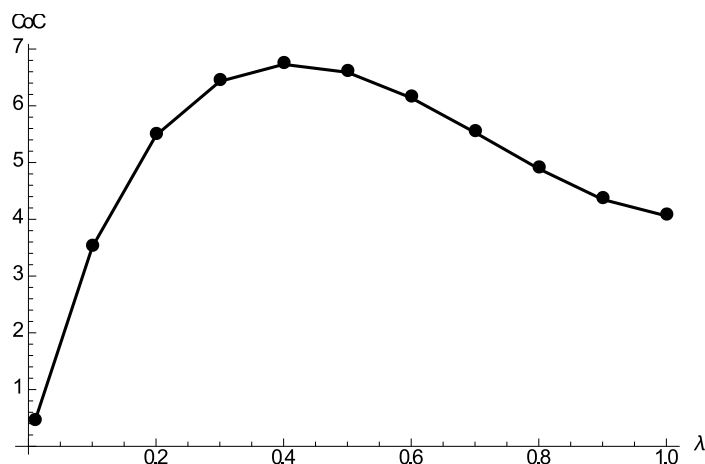


Figure 3.6: The cost of cooperation (in percents) as a function of discount factor λ .

§ 3.6 Strategic support of a cooperative solution in dynamic games played over event trees

We suppose that the players agreed to play cooperatively throughout the game, i.e., they agreed to form a grand coalition and maximize the summarized payoff of the grand coalition, and also decided to redistribute payments to players at each node of the event tree in accordance with the IDP to obtain node-consistent imputation. Thus, the initially defined payoff function of player $j \in M$ at node n_i^t , that is $\phi_j^{n_i^t}(x^*(n_i^t), u^*(n_i^t))$ will be replaced by the component of the IDP $\beta_j(n_i^t)$ for all $j \in M$ and $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$. Similarly, $\Phi_j^{n_i^T}(x^*(n_i^T))$ is replaced by $\beta_j(n_i^T)$, $j \in M$ for all $n_i^T \in \mathcal{N}^T$. Now the payoff function of player $j \in M$ in the subgame beginning at node $n_i^t \in \mathcal{N}^t$ with state $x^*(n_i^t)$ is defined by

$$\hat{W}_j(\mathbf{u}(x^*(n_i^t)), x^*(n_i^t)) = \sum_{\theta=t}^T \lambda_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_i^\theta | n_i^t) \beta_j(n_i^\theta). \quad (3.6.1)$$

When the redistribution of the players' payoffs has been made, we may examine the problem of strategic support of a cooperative solution, as in Chapter 2. In the case of a game with finite duration, we show that there exists a subgame perfect ε -equilibrium with players' payoffs equal to the payoffs in a non-cooperative game, in which the payments to the players are made in accordance with the imputation distribution procedure.

We represent the dynamic game played over event tree as a game in extensive form and assume that the players use closed-loop information structure when choosing their strategies. It means that each player knows not only the current node $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T$ and what she has played on the path leading from the initial node n^0 to $a(n_i^t)$, but also what the other players did in all previous periods. Remind that $P(n_i^t) = (n^0, n_{i_1}^1, \dots, n_{i_{t-1}}^{t-1}, n_i^t)$ is the unique path from initial node n^0 until n_i^t .

Now we consider the game in behavior strategies $\sigma_j = \left\{ \sigma_j^{n_i^t} \right\}_{n_i^t \in \mathcal{N}^t, t=0, \dots, T-1}$ where $\sigma_j^{n_i^t}: H(n_i^t) \rightarrow U_j^{n_i^t}$ and $H(n_i^t)$ is a history of node n_i^t . For a given behavior strategy profile $\sigma = (\sigma_j : j \in M)$ we may calculate the payoff of a player in the whole game and its any subgame. The payoff of player j in the subgame starting at node n_i^t with state $x(n_i^t)$ is a function of the behavior strategy profile, given by

$$\hat{W}_j(\sigma, x(n_i^t)) = W_j(\mathbf{u}(n_i^t), x(n_i^t)),$$

where $\mathbf{u}(n_i^t)$ is a trajectory of controls in the subgame starting at node n_i^t with the state $x(n_i^t)$ corresponding to the profile σ .

Below we describe the trigger strategy for the DGPET in detail. To strategically support cooperation in the finite-horizon dynamic game played over an event tree, we shall construct an approximated equilibrium (ε -equilibrium) in behavior strategies with a closed-loop information structure. We make the following assumptions about players' behavior:

1. The players want to realize the cooperative trajectory \mathbf{u}^* from (3.1.7).
2. They want to realize the imputation $y(x^0)$ along the cooperative trajectory \mathbf{u}^* .
3. Payments to the players at each node of the realized trajectory are made in accordance with the imputation distribution procedure defined by the recurrent formulas (3.2.8) and (3.2.9) based on the imputation $y(x^0)$.
4. If a player j deviates from cooperation at node $a(n_i^t)$, that is, implementing a control $u_j(a(n_i^t)) \neq u_j^*(a(n_i^t))$, then cooperation breaks down and all players switch to their Nash equilibrium strategies in the subgame starting at node n_i^t in state $x(n_i^t) = f^{a(n_i^t)}(x^*(a(n_i^t)), (u_{-j}^*(a(n_i^t)), u_j(a(n_i^t))))$.

5. If more than one player deviates at node $a(n_i^t)$, the grand coalition breaks and at node n_i^t players switch to the Nash equilibrium strategies, i. e., they play non-cooperatively in the remaining part of the game.

The trigger behavior strategy of a player consists of two behavior types or two modes:

The nominal mode. If the history of node n_i^t coincides with $H^*(n_i^t)$ i.e., all players used their cooperative controls on the path $P(n_{i_{t-1}}^{t-1})$, that is, from n^0 until $n_{i_{t-1}}^{t-1}$, then player $p \in M$ implements $u_p^*(n_i^t)$ in node n_i^t .

The trigger mode. If the history of node n_i^t is such that there exists a node n on the path $P(n_{i_{t-1}}^{t-1})$ such that $u(n) \neq u^*(n)$, then player p 's strategy is Nash equilibrium strategy calculated for the subgame starting from the successor of n and corresponding state. Here, the history of node n_i^t is such that there exists a node n and at least one deviating player $j \in M$, $j \neq p$, that is, the history $H(n)$ of node n is part of $H^*(n_i^t)$, and $(n, u(n))$ is not part of $H^*(n_i^t)$.

Formally speaking, the trigger behavior strategy of player $p \in M$ is defined as follows:

$$\hat{\sigma}_p(H(n_i^t)) = \begin{cases} u_p^*(n_i^t), & \text{if } H(n_i^t) = H^*(n_i^t), \\ \hat{u}_p(n_i^t), & \text{if there exists a node } n \text{ on path } P(n_i^t), \\ & \text{such that } u(n) \neq u^*(n), \end{cases} \quad (3.6.2)$$

where $\hat{u}_p(n_i^t)$ is player p 's control in node n_i^t . The control $\hat{u}_p(n_i^t)$ implements the punishing strategy in the subgame starting in the unique node belonging to the set $\mathcal{S}(n) \cap P(n_i^t)$. The control $\hat{u}_p(n_i^t)$ coincides with $u_p^N(n_i^t)$ calculated as a part of Nash equilibrium for the subgame starting at node $n_1 = \mathcal{S}(n) \cap P(n_i^t)$ in state $x(n_1)$.

To avoid further complicating the notation, we omitted the state argument in the punishing control and the trigger strategy, but we stress that they depend on the state value. Let node n_1 be a direct successor of node n in which player j deviates. The collection of controls $(u_{-j}^*(n), u_j(n))$ is then realized, and the state value in node n_1 can be calculated using the state dynamics $x(n_1) = f^n(x^*(n), (u_{-j}^*(n), u_j(n)))$.

The control $\hat{u}_p(n_1)$ is part of the control profile $\hat{u}(n_1) = (u_p^N(n_1) : p \in M)$ where $u_p^N(n_1)$ is a control of player p in node n_1 .

Now, in the subgame starting from node $n_l^t \in \mathcal{N}^t$ in state $x(n_l^t)$, the collection of controls punishing player j 's individual deviation is given by

$$\hat{\mathbf{u}}^j(n_l^t) = (\hat{\mathbf{u}}_p^j(n_l^t) : p \in M),$$

where $\hat{\mathbf{u}}_p^j(n_l^t) = \{\hat{u}_p^j(n_l^\theta) : n_l^\theta \in \Gamma(n_l^t)\}$. This collection of controls generates Nash equilibrium trajectory of states in player j 's punishment in this subgame, that is,

$$\hat{\mathbf{x}}^j(n_l^t) = \{\hat{x}^j(n_l^\theta) : n_l^\theta \in \Gamma(n_l^t)\}.$$

To construct the trigger strategies, we need to find m punishing strategy profiles (each profile corresponds to a particular player) for each subgame. Now we are ready to prove the main result of this paragraph.

Theorem 3.3. *Consider the game played over event tree when players' payoffs on the cooperative trajectory are determined by node-consistent IDP with equations (3.2.8) and (3.2.9) based on imputation $y(n^0)$. For any $\varepsilon \geq \hat{\varepsilon}$ in the game there exists subgame perfect ε -equilibrium in trigger strategies with players' payoffs $y_1(n^0), \dots, y_m(n^0)$, and*

$$\hat{\varepsilon} = \max_{j \in M} \max_{\substack{n_l^t \in \mathcal{N}^t \\ t=0, \dots, T-1}} \varepsilon_j(n_l^t), \quad (3.6.3)$$

where

$$\begin{aligned} \varepsilon_j(n_l^t) = & \max_{u_j(n_l^t) \in U_j^{n_l^t}} \left\{ \phi_j^{n_l^t}(x^*(n_l^t), (u_{-j}^*(n_l^t), u_j(n_l^t))) - \beta_j(n_l^t) \right. \\ & + \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_l^\theta | n_l^t) \left(\phi_j^{n_l^\theta}(\hat{x}^j(n_l^\theta), \hat{u}^j(n_l^\theta)) - \beta_j(n_l^\theta) \right) \\ & \left. + \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_\Gamma^T} \pi(n_l^T | n_l^t) \left(\Phi_j^{n_l^T}(\hat{x}(n_l^T)) - \beta_j(n_l^T) \right) \right\}, \quad (3.6.4) \end{aligned}$$

where $\hat{u}^j(n_l^\theta)$ is a control profile in node n_l^θ corresponding to a behavior strategy profile $\hat{\sigma}$ determined by (3.6.2) and when the trigger mode of the strategy begins in the subgame starting at the node belonging to the set $\mathcal{S}(n_l^t)$ and in state

$f^{n_i^t}(x(n_i^t), (u_{-j}^*(n_i^t), u_j(n_i^t)))$. Therefore, the differences in the second and third lines also depend on the control $u_j(n_i^t)$. The state $\hat{x}^j(n_i^\theta)$, $n_i^\theta \in \Gamma(n_i^t)$ is a state trajectory corresponding to $\hat{\mathbf{u}}^j(n_i^t)$.

Proof. Consider the trigger behavior strategy $\hat{\sigma} = (\hat{\sigma}_p : p \in M)$ defined in (3.6.2), and the subgame starting at any node $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$. Consider possible histories of any node n_i^t , and compute the benefit of player j deviating at node n_i^t . When she does not deviate at this node and proceeds realising the cooperative trajectory as the the players do, her cooperative payoff in this subgame will be given by

$$\begin{aligned} \hat{W}_j(\mathbf{u}^*(n_i^t), x^*(n_i^t)) &= \beta_j(n_i^t) + \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_i^\theta | n_i^t) \beta_j(n_i^\theta) \\ &\quad + \lambda_j^{T-t} \sum_{n_i^T \in \mathcal{N}_\Gamma^T} \pi(n_i^T | n_i^t) \beta_j(n_i^T), \end{aligned} \quad (3.6.5)$$

where $\mathcal{N}_\Gamma^\theta = \mathcal{N}^\theta \cap \Gamma(n_i^t)$, $\mathbf{u}^*(n_i^t) = (\mathbf{u}_j^*(n_i^t) : j \in M)$ is an S -adapted cooperative strategy profile. In the payoff we use the components of IDP instead of payoffs prescribed initially by payoff functions ϕ_j and Φ_j .

First, consider the case where the history of node n_i^t is $H^*(n_i^t)$. Suppose player j deviates in node n_i^t from the cooperative trajectory. In this case, she may secure the following payoff in the subgame starting at node n_i^t , given the information that the behavior strategy profile $\hat{\sigma} = (\hat{\sigma}_p(\cdot) : p \in M)$ determined by (3.6.2) will materialize:

$$\begin{aligned} &\max_{u_j(n_i^t) \in U_j^{n_i^t}} \left\{ \phi_j^{n_i^t}(x^*(n_i^t), (u_{-j}^*(n_i^t), u_j(n_i^t))) + \right. \\ &\quad + \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} \pi(n_i^\theta | n_i^t) \phi_j^{n_i^\theta}(\hat{x}^j(n_i^\theta), \hat{u}^j(n_i^\theta)) + \\ &\quad \left. + \lambda_j^{T-t} \sum_{n_i^T \in \mathcal{N}_\Gamma^T} \pi(n_i^T | n_i^t) \Phi_j^{n_i^T}(\hat{x}^j(n_i^T)) \right\}, \end{aligned} \quad (3.6.6)$$

where punishing Nash strategy starts to be implemented in nodes from $\mathcal{S}(n_i^t)$. If the deviation occurs, the players leave the cooperative trajectory and their payoffs are calculated by initially defined functions $\phi_j^{n_i^t}(\cdot)$ and $\Phi_j^{n_i^T}(\cdot)$.

Then, we may compute the benefit from deviation of player j at node n_l^t as a difference between payoffs without and with deviation at node n_l^t , that is,

$$\begin{aligned} \varepsilon_j(n_l^t) = & \max_{u_j(n_l^t) \in U_j^{n_l^t}} \left\{ \phi_j^{n_l^t}(x^*(n_l^t), (u_{-j}^*(n_l^t), u_j(n_l^t))) - \beta_j(n_l^t) + \right. \\ & + \sum_{\theta=t+1}^{T-1} \lambda_j^{\theta-t} \sum_{n_l^\theta \in \mathcal{N}_l^\theta} \pi(n_l^\theta | n_l^t) \left(\phi_j^{n_l^\theta}(\hat{x}^j(n_l^\theta), \hat{u}^j(n_l^\theta)) - \beta_j(n_l^\theta) \right) + \\ & \left. + \lambda_j^{T-t} \sum_{n_l^T \in \mathcal{N}_l^T} \pi(n_l^T | n_l^t) \left(\Phi_j^{n_l^T}(\hat{x}(n_l^T)) - \beta_j(n_l^T) \right) \right\}. \end{aligned} \quad (3.6.7)$$

Second, suppose that the history of node n_l^t does not coincide with $H^*(n_l^t)$. This means that all players have switched from nominal mode to the trigger one. Player j will have no benefit from deviating in node n_l^t because the players implement their Nash equilibrium strategies regardless of which player (or group of players) has deviated in the previous nodes.

Calculating the maximal benefit from deviation for any subgame and any player given by (3.6.7), we obtain the value of $\hat{\varepsilon}$ from the Theorem's statement, that is,

$$\hat{\varepsilon} = \max_{j \in M} \max_{\substack{n_l^t \in \mathcal{N}^t \\ t=0, \dots, T-1}} \varepsilon_j(n_l^t).$$

And for any $\varepsilon \geq \hat{\varepsilon}$ the behavior strategy profile determined by (3.6.2) is a subgame-perfect ε -equilibrium by construction. \square

§ 3.7 Subgame perfect ε -equilibrium using imputation distribution procedure in a game of pollution control

To illustrate the results of the previous section, we consider a three-player game considered in paragraphs §3.3 и §3.5 with some differences. Original version of the deterministic model of pollution control was proposed in [98]. Denote by $M = \{1, 2, 3\}$ the set of players, and by $\mathcal{T} = \{0, 1, \dots, 5\}$ the set of periods. Let $u(n_l^t) = (u_1(n_l^t), u_2(n_l^t), u_3(n_l^t))$ be the vector of countries' emissions of some pollutant and denote by $x(n_l^t)$ the stock of pollution at node n_l^t in time period t . The evolution of

this stock is governed by the following difference equation:

$$x(n_i^t) = (1 - \delta(a(n_i^t)))x(a(n_i^t)) + \sum_{j \in M} u_j(a(n_i^t)), \quad (3.7.1)$$

with the initial stock x^0 at root node n^0 being given, and $\delta(n_i^t)$ ($0 < \delta(n_i^t) < 1$) is the stochastic rate of pollution absorption by nature at node n_i^t . We suppose that $\delta(n_i^t)$ can take two possible values, that is, $\delta(n_i^t) \in \{\underline{\delta}, \bar{\delta}\}$, with $\underline{\delta} < \bar{\delta}$. The event tree is a binary tree, i.e., each node in periods $t = 0, \dots, 4$ has two successors (see Figure 3.7). The conditional probability of realization of the upward successor of any node is $\frac{1}{4}$ and is $\frac{3}{4}$ for a downward successor. So, for instance, we have probabilities $\pi(n_1^1) = \frac{1}{4}$ and $\pi(n_2^1) = \frac{3}{4}$ in period 1, and probabilities $\pi(n_1^2) = \frac{1}{16}$, $\pi(n_2^2) = \frac{3}{16}$, $\pi(n_3^2) = \frac{3}{16}$, $\pi(n_4^2) = \frac{9}{16}$ for $t = 2$. The root node n^0 and all upward (or left-handed) nodes have the low rate $\underline{\delta}$ of pollution absorption by nature, and all downward (or right-handed) nodes have the high level $\bar{\delta}$ of pollution absorption.

The damage cost is an increasing convex function in the pollution stock having the quadratic form $D_j(x(n_i^t)) = \alpha_j x^2(n_i^t)$, $j \in M$, where α_j is a strictly positive parameter. The cost of emissions is also given by a quadratic function $C_j(u_j(n_i^t)) = \frac{\gamma_j}{2} (u_j(n_i^t) - e)^2$, where e and γ_j are strictly positive constants.

At each node $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, 4$ cost function of player $j \in M$ is the sum of two functions: (i) $D_j(x(n_i^t)) = \alpha_j x^2(n_i^t)$, where $\alpha_j > 0$; (ii) $C_j(u_j(n_i^t)) = \frac{\gamma_j}{2} (u_j(n_i^t) - e)^2$, where e and γ_j are positive constants.

The total discounted cost $J_j(\mathbf{x}, \mathbf{u})$ to be minimized by player $j \in M$ is given by

$$\sum_{t=0}^4 \lambda_j^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) (C_j(u_j(n_i^t)) + D_j(x(n_i^t))) + \lambda_j^5 \sum_{n_i^5 \in \mathcal{N}^5} \pi(n_i^5) D_j(x(n_i^5)),$$

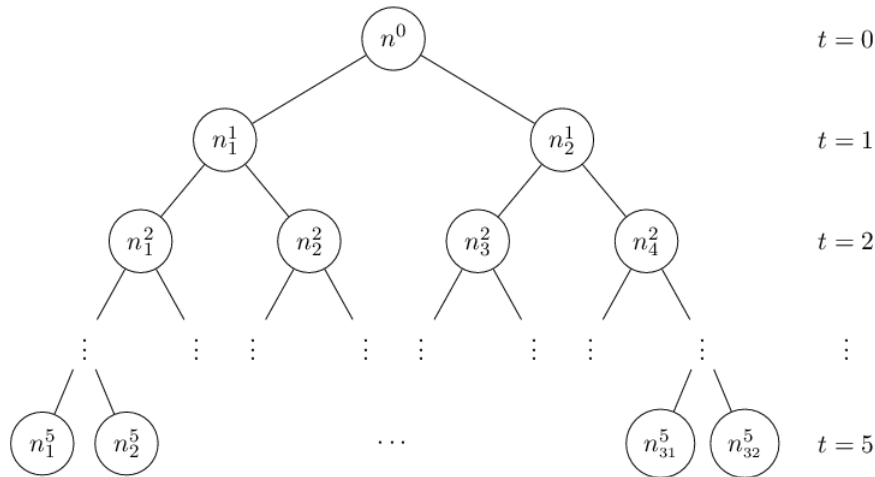
where $\mathbf{x} = \{x(n_i^t)\}$ and $\mathbf{u} = \{u(n_i^t)\}$, $n_i^t \in \mathcal{N}^t$, $\lambda_j \in (0, 1)$ is a discount rate of player j , subject to (3.7.1), given initial stock $x_0 = 0$ before the game starts and constraints: $u_j(n_i^t) \in [0, e]$ for any player $j \in M$ and any node $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, 4$.

We use the following parameters for the numerical simulation:

$$\alpha_1 = 0.1, \alpha_2 = 0.2, \alpha_3 = 0.3,$$

$$\gamma_1 = 0.9, \gamma_2 = 0.8, \gamma_3 = 0.7,$$

$$\underline{\delta} = 0.45, \bar{\delta} = 0.8, e = 30, \lambda_1 = \lambda_2 = \lambda_3 = 0.9.$$

Figure 3.7: Event tree graph for $T = 5$.

Tables 3.10 and 3.11 provide the results regarding the values of maximal benefits from deviation according to (3.6.4) and the corresponding values for the case when players do not adopt IDP. (More precisely, we only show the maximal values for each time period $t = 0, 1, 2, 3, 4$.)

Table 3.10: The maximal benefits from deviation in time period t calculated for the 6-period game when players adopt IDP.

Time period t	0	1	2	3	4
Player 1	-128.351	-96.4928	-59.492	-23.5305	-0.794881
Player 2	-118.967	-73.6691	-28.0224	12.5117	26.5563
Player 3	-105.177	-33.8832	25.4111	67.5529	61.8055

Table 3.11: The maximal benefits from deviation in time period t calculated for the 6-period game without IDP.

Time period t	0	1	2	3	4
Player 1	104.714	108.902	96.9909	75.0803	40.6216
Player 2	-103.318	-60.0593	-18.0474	18.2751	28.5478
Player 3	-353.892	-251.621	-141.231	-41.2657	17.8045

The total costs of the three players in the whole game are 2148.7. Tables 3.10 and 3.11 show the advantage of adopting the IDP in this game. When players adopt the IDP the maximal benefit of a deviating player or $\bar{\varepsilon}$ is equal to 67.5529 in comparison with 108.902 in the game without using IDP for payoff redistribution. Further, the

first time when a player can benefit from deviating is in period $t = 2$ while in the game without IDP the profitable deviation is observed in period $t = 0$. Players 1 and 2 have lower benefits from deviations in the game with IDP than in the game without IDP. But player 3 will have higher incentive to deviate from cooperative trajectory if players adopt the IDP.

§ 3.8 Dynamic games played over binary event trees with linear state equations

In this section we make several assumptions about event trees and the players' payoff functions. Then, we consider the non-cooperative and cooperative versions of the game, and also compare the payoffs of the players in these cases using the so-called "price of anarchy" [117].

The state equations are given by

$$x(n_i^t) = A(a(n_i^t))x(a(n_i^t)) + \sum_{j \in M} B_j(a(n_i^t))u_j(a(n_i^t)), \quad (3.8.1)$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, t = 1, \dots, T. \quad (3.8.2)$$

where $x(n_i^t) \in X$, $u_j(a(n_i^t)) \in U_j^{a(n_i^t)}$, for any node $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T$, matrices $A(a(n_i^t))$ and $B_j(a(n_i^t))$ are of sizes $q \times q$ and $q \times r$ respectively. The components of the state vector $x(n_i^t)$ of node n_i^t are linear combinations of the components of the state of its predecessor $a(n_i^t)$ and players' controls which players selected at this predecessor. The coefficients of the linear combinations are the row entries of the matrices $A(a(n_i^t))$ and $B_j(a(n_i^t))$, $j \in M$.

At each node n_i^t , $t = 0, \dots, T - 1$, the reward to player j is a function of the state and the controls of all players, given by

$$\begin{aligned} \phi_j^{n_i^t}(x(n_i^t), u(n_i^t)) &= \frac{1}{2}u'_j(n_i^t)C_{1j}(n_i^t)u_j(n_i^t) + \sum_{k \neq j} u'_k(n_i^t)C_{2j}(n_i^t)u_j(n_i^t) \\ &+ c'_j(n_i^t)u_j(n_i^t) + d'_j(n_i^t)x(n_i^t) \end{aligned}$$

where $C_{1j}(n_i^t)$, $C_{2j}(n_i^t)$ are matrices of size $r \times r$, $c_j(n_i^t) \in \mathbb{R}^r$ and $d_j(n_i^t) \in \mathbb{R}^q$. We assume the matrices C_{1j} , $j \in M$ are symmetric and negative definite.

In a terminal node n_l^T , the reward to player j is given by the function

$$\Phi_j^{n_l^T}(x(n_l^T)) = d'_j(n_l^T)x(n_l^T).$$

We assume that player $j \in M$ maximizes her expected stream of payoffs discounted³ at ϱ_j ($0 < \varrho_j < 1$). The state equations and the reward functions define the following multistage game, where we let

$$\begin{aligned} \mathbf{x} &= \{x(n_l^t) : n_l^t \in \mathcal{N}^t, t = 0, \dots, T\}, \\ \mathbf{u} &= \{u(n_l^t) : n_l^t \in \mathcal{N}^t, t = 0, \dots, T-1\}, \end{aligned}$$

and $J_j(\mathbf{x}, \mathbf{u})$ be the payoff to player j , that is,

$$J_j(\mathbf{x}, \mathbf{u}) = \sum_{t=0}^{T-1} \varrho_j^t \sum_{n_l^t \in \mathcal{N}^t} \pi(n_l^t) \phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) + \varrho_j^T \sum_{n_l^T \in \mathcal{N}^T} \pi(n_l^T) \Phi_j^{n_l^T}(x(n_l^T)), \quad (3.8.3)$$

s.t.

$$x(n_l^t) = A(a(n_l^t))x(a(n_l^t)) + \sum_{j \in M} B_j(a(n_l^t))u_j(a(n_l^t)), \quad (3.8.4)$$

$$u(a(n_l^t)) \in U^{a(n_l^t)}, \quad n_l^t \in \mathcal{N}^t, t = 1, \dots, T, \quad (3.8.5)$$

$$x(n^0) = x^0. \quad (3.8.6)$$

Consequently, the objective function (3.8.3) is then strictly concave in the control variables.

In a dynamic game, one needs to specify the piece of information that the players use when they make their decisions. When the strategies (as decision rules) are only function of time, we say that the information structure is in open loop, and when they depend on the state variables (and also possibly on time), the information structure is then in closed loop. In DGPET, the information structure is S -adapted, where the S stands for sample value of the stochastic process, that is, the players' actions are adapted to the history of this process. This property derives from the nature of the process considered here, namely, an event tree. If we consider any node in the tree, then we can reconstruct its history, that is, the unique path emanating from the root node to that particular node.

³In this section we denote the discount rate of player j by ϱ_j , not by λ_j , as in the previous paragraphs of Chapter 3, because in this paragraph notation λ_j will be used for Lagrange multiplier.

Similarly to the previous paragraphs we consider two versions of the game: cooperative and non-cooperative scenarios. If the game is played non-cooperatively, then the players will seek a Nash equilibrium in S -adapted strategies. In a cooperative scenario players maximize the discounted sum of the payoffs in the whole game, that is,

$$\max_{(\mathbf{u}_j: j \in M)} \sum_{j \in M} W_j(\mathbf{u}, x^0).$$

Denote the resulting vector of cooperative controls by \mathbf{u}^* , which is found for the game starting from node n^0 and state x^0 , that is,

$$\mathbf{u}^* = \arg \max_{(\mathbf{u}_j: j \in M)} \sum_{j \in M} W_j(\mathbf{u}, x^0). \quad (3.8.7)$$

Further, denote by $\mathbf{x}^* = \{x^*(n_l^t) : n_l^t \in \mathcal{N}^t, t = 0, 1, \dots, T\}$ the cooperative state trajectory generated by the cooperative controls \mathbf{u}^* .

Remark 3.4. *Additionally to Remark 3.2, we assume that given a profile of parameters of the linear-state game player over the event tree (the matrices $A(a(n))$, $B_j(a(n))$, $C_{1j}(n)$, $C_{2j}(n)$ and the vectors $c_j(n)$, $d_j(n)$ for any node $n \in \mathcal{N}$ and any player $j \in M$), the joint-optimization solution and the Nash equilibrium at each node n of the tree prescribe profiles of controls from the set U^n , and at each node n the state $x(n)$ driven by the state equation (3.8.1) for any S -adapted strategy profile remains in X .*

§ 3.8.1 Binary event tree with homogeneous players

To keep the model as parsimonious as possible, while still capturing the impact of uncertainty on the outcomes, we restrict our setting to considering a game on a binary tree meaning that if the game has not terminated, given an intermediate game period, we always have two possible realizations of a random variable for the next period. In section § 3.8 we also made an assumption about the linearity of the state equations, i.e., the dynamic game is of the linear-state variety, that is, the state dynamics and the players' rewards are linear in the state. We also suppose that the set of players is homogeneous, which is necessary to ensure that the PoA (or PoS) is due, and only due, to the difference in behavior of the players (cooperation or noncooperation) and not to their parameter conditions.

Let each node in any period $t = 1, \dots, T-1$ have two successors. The conditional probability of realization of the upward successor of any node is p and $1-p$ for a downward successor. For period $t = 1$ there are two nodes in the tree: n_1^1 and n_2^1 with corresponding probabilities $\pi(n_1^1) = p$ and $\pi(n_2^1) = 1-p$. And node n_1^1 has two successors n_1^2 and n_2^2 with the probabilities $\pi(n_1^2) = p^2$ and $\pi(n_2^2) = p(1-p)$. The binary event tree of a T -period game contains $2^{T+1} - 1$ nodes including the root node.

Let the state equations be given by

$$x(n_l^t) = \begin{cases} A_0 x(a(n_l^t)) + \sum_{j \in M} B_0 u_j(a(n_l^t)), & \text{if } a(n_l^t) = n^0, \\ A_1 x(a(n_l^t)) + \sum_{j \in M} B_1 u_j(a(n_l^t)), & \text{if } a(n_l^t) = n_k^{t-1} \text{ and} \\ & k = 2z + 1, z = 0, 1, \dots, \\ A_2 x(a(n_l^t)) + \sum_{j \in M} B_2 u_j(a(n_l^t)), & \text{if } a(n_l^t) = n_k^{t-1} \text{ and} \\ & k = 2z, z = 1, 2, \dots, \end{cases} \quad (3.8.8)$$

with $u_j(a(n_l^t)) \in U_j^{a(n_l^t)}$ for $n_l^t \in \mathcal{N}^t, t = 1, \dots, T$.

As the players are identical, they have the same reward functions in any non-terminal node n_l^t , which is given by

$$\phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) = \frac{1}{2} u_j'(n_l^t) C_1 u_j(n_l^t) + \sum_{k \neq j} u_k'(n_l^t) C_2 u_j(n_l^t) + c' u_j(n_l^t) + d' x(n_l^t), \quad (3.8.9)$$

where C_1, C_2, c and d do not depend on the node $n_l^t \in \mathcal{N}^t$ and the player.

In a terminal node n_l^T , the reward to player j is given by the function

$$\Phi_j^{n_l^T}(x(n_l^T)) = d' x(n_l^T). \quad (3.8.10)$$

The identical players have common discount factor $\varrho_j = \varrho, j \in M$.

§ 3.8.2 S -adapted Nash equilibrium in dynamic games played over event trees

We seek the S -adapted Nash equilibrium in the game played over a binary event tree with identical players described in Section § 3.8.1. Players want to maximize

function (3.8.3) with ϕ_j and Φ_j given by equations (3.8.9) and (3.8.10) subject to (3.8.5), (3.8.6), (3.8.8).

We formulate the necessary and sufficient conditions for an S -adapted equilibrium using Pontryagin maximum principle [5]. For any player j we define the pre-Hamiltonian function for the root node n^0 and any node $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T - 1$:

$$\begin{aligned} H_j^{n^0} &= \frac{1}{2} u'_j(n^0) C_1 u_j(n^0) + \sum_{k \neq j} u'_k(n^0) C_2 u_j(n^0) + c' u_j(n^0) + d' x^0 \\ &\quad + \varrho [p \lambda'_j(n_1^1) + (1 - p) \lambda'_j(n_2^1)] \left[A_0 x^0 + \sum_{j \in M} B_0 u_j(n^0) \right], \\ H_j^{n_l^t} &= \frac{1}{2} u'_j(n_l^t) C_1 u_j(n_l^t) + \sum_{k \neq j} u'_k(n_l^t) C_2 u_j(n_l^t) + c' u_j(n_l^t) + d' x(n_l^t) \\ &\quad + \varrho [p \lambda'_j(n_1^{t+1}) + (1 - p) \lambda'_j(n_2^{t+1})] \left[A_l x(n_l^t) + \sum_{j \in M} B_l u_j(n_l^t) \right], \quad l = 1, 2, \\ &\quad j \in M, \end{aligned}$$

where $n_l^t \in \{n_1^t, n_2^t\}$, index $l = 1$ identifies the upward node and $l = 2$ identifies the downward node and $\lambda_j(n_l^t)$ is the costate variable from \mathbb{R}^q .

Theorem 3.4. *If the matrix $C_1 + (m - 1)C_2$ is invertible, the controls*

$$u_j^{N}(n^0) = - \left[c' + \varrho d' \sum_{\tau=0}^{T-1} (\varrho A)^\tau B_0 \right] (C_1 + (m - 1)C_2)^{-1}, \quad (3.8.11)$$

$$\begin{aligned} u_j^{N}(n_l^t) &= - \left[c' + \varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_l \right] (C_1 + (m - 1)C_2)^{-1}, \quad (3.8.12) \\ &\quad t = 1, \dots, T - 1, \quad l = 1, 2, \quad j \in M, \end{aligned}$$

form the S -adapted Nash equilibrium in the game played over a binary event tree with identical players where $A = pA_1 + (1 - p)A_2$. The corresponding costate variables are

$$\lambda_j^{N}(n_l^t) = d' \left[I + \varrho \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau A_l \right], \quad t = 1, \dots, T - 1, \quad (3.8.13)$$

$$\lambda_j^{N}(n_l^T) = d', \quad (3.8.14)$$

where I is the identity matrix of size q . Moreover, the total payoff of any player

$j \in M$ in the S -adapted Nash equilibrium is

$$\begin{aligned}
J_j(\mathbf{x}^N, \mathbf{u}^N) = & d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 + u_j^N(n^0) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n^0) \\
& + \left[c' + m \varrho d' \sum_{t=0}^{T-1} (\varrho A)^t B_0 \right] u_j^N(n^0) \\
& + \sum_{t=1}^{T-1} \varrho^t \left(p u_j^N(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_1^t) \right. \\
& \quad \left. + (1-p) u_j^N(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_2^t) \right) \\
& + \sum_{t=1}^{T-1} \varrho^t \left(p \left[c' + m \varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_1 \right] u_j^N(n_1^t) \right. \\
& \quad \left. + (1-p) \left[c' + m \varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_2 \right] u_j^N(n_2^t) \right). \tag{3.8.15}
\end{aligned}$$

Proof. The following conditions hold for the existence of S -adapted Nash equilibrium for any $l = 1, 2$:

$$\begin{aligned}
0 &= \frac{\partial H_j^{n^0}}{\partial u_j(n^0)} = u_j'(n^0) C_1 + \sum_{k \neq j} u_k'(n^0) C_2 + c' \\
&\quad + \varrho [p \lambda_j'(n_1^1) + (1-p) \lambda_j'(n_2^1)] B_0, \\
0 &= \frac{\partial H_j^{n_t^t}}{\partial u_j(n_t^t)} = u_j'(n_t^t) C_1 + \sum_{k \neq j} u_k'(n_t^t) C_2 + c' \\
&\quad + \varrho [p \lambda_j'(n_1^{t+1}) + (1-p) \lambda_j'(n_2^{t+1})] B_l, \quad n_i^t \in \mathcal{N}^t, \quad t = 1, \dots, T-1, \\
\lambda_j'(n_t^t) &= \frac{\partial H_j^{n_t^t}}{\partial x(n_t^t)} = d' + \varrho [p \lambda_j'(n_1^{t+1}) + (1-p) \lambda_j'(n_2^{t+1})] A_l, \\
&\quad n_i^t \in \mathcal{N}^t, \quad t = 1, \dots, T-1, \\
\lambda_j'(n_t^T) &= \frac{\partial \Phi_j^{n_t^T}}{\partial x(n_t^T)} = d', \quad n_i^T \in \mathcal{N}^T
\end{aligned}$$

with the initial condition for the state variable $x(n^0) = x^0$.

From the last two equations and following the structure of the binary tree, we recurrently obtain the expressions for the costate variables in the forms given by equations (3.8.13) and (3.8.14). Substituting the expressions for the costate variables from (3.8.13) and (3.8.14), and solving the first two equations $\partial H_j^{n^0} / \partial u_j(n^0) = 0$

and $\partial H_j^{n_i^t} / \partial u_j(n_i^t) = 0$, we obtain the controls of the players in the S -adapted Nash equilibrium which are determined by (3.8.11) and (3.8.12). We may notice that the control of the player depends on the time period and the type of a node, i.e., it depends on if she is in the upward or downward successor at the current time period.

The total payoff of any player $j \in M$ in the S -adapted Nash equilibrium is calculated by equation (3.8.3) substituting the S -adapted Nash equilibrium controls and corresponding state trajectory:

$$\begin{aligned}
J_j(\mathbf{x}^N, \mathbf{u}^N) &= \sum_{t=0}^{T-1} \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \left[\frac{1}{2} u_j'^N(n_i^t) C_1 u_j^N(n_i^t) + \sum_{k \neq j} u_k'^N(n_i^t) C_2 u_j^N(n_i^t) \right. \\
&\quad \left. + c' u_j^N(n_i^t) + d' x^N(n_i^t) \right] + \varrho^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) d' x^N(n_i^T) \\
&= \sum_{t=0}^{T-1} \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \left[u_j'^N(n_i^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_i^t) + c' u_j^N(n_i^t) \right. \\
&\quad \left. + d' x^N(n_i^t) \right] + \varrho^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) d' x^N(n_i^T) \\
&= u_j'^N(n^0) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n^0) + c' u_j^N(n^0) \\
&\quad + \sum_{t=1}^{T-1} \varrho^t \left[p u_j'^N(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_1^t) \right. \\
&\quad \left. + (1-p) u_j'^N(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_2^t) \right] \\
&\quad + c' \sum_{t=1}^{T-1} \varrho^t [p u_j^N(n_1^t) + (1-p) u_j^N(n_2^t)] + \sum_{t=0}^T \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) d' x^N(n_i^t) \\
&= u_j'^N(n^0) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n^0) + c' u_j^N(n^0) \\
&\quad + \sum_{t=1}^{T-1} \varrho^t \left[p u_j'^N(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_1^t) \right. \\
&\quad \left. + (1-p) u_j'^N(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^N(n_2^t) \right]
\end{aligned}$$

$$\begin{aligned}
& +c' \sum_{t=1}^{T-1} \varrho^t [pu_j^N(n_1^t) + (1-p)u_j^N(n_2^t)] \\
& +d' \left[x^0 + \varrho \sum_{t=0}^{T-1} (\varrho A)^t [A_0 x^0 + mB_0 u_j^N(n^0)] \right. \\
& \quad \left. + m \sum_{t=2}^T \varrho^t \sum_{\tau=0}^{t-2} A^\tau [pB_1 u_j^N(n_1^{t-\tau-1}) + (1-p)B_2 u_j^N(n_2^{t-\tau-1})] \right] \\
= & d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 + u_j^N(n^0) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^N(n^0) \\
& + \left[c' + m\varrho d' \sum_{t=0}^{T-1} (\varrho A)^t B_0 \right] u_j^N(n^0) \\
& + \sum_{t=1}^{T-1} \varrho^t \left(pu_j^N(n_1^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^N(n_1^t) \right. \\
& \quad \left. + (1-p)u_j^N(n_2^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^N(n_2^t) \right) \\
& + \sum_{t=1}^{T-1} \varrho^t \left(p \left[c' + m\varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_1 \right] u_j^N(n_1^t) \right. \\
& \quad \left. + (1-p) \left[c' + m\varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_2 \right] u_j^N(n_2^t) \right).
\end{aligned}$$

Theorem is proved. □

Remark 3.5. *We make the following three observations:*

1. *The costate variables do not directly depend on the number of players.*
2. *If A_1 , A_2 and d are non-negative, then on the paths formed of either only upward nodes, or downward nodes, we have*

$$\lambda_j^N(n_l^1) \geq \dots \geq \lambda_j^N(n_l^T), l = 1, 2,$$

that is, the costate variable of player $j \in M$ as a function of the game period is non-increasing.

3. *If A_1 , A_2 , B_l , c , d are non-negative and C_1 , C_2 are negative-defined, then we have a similar property on the control variables, namely:*

$$u_j^N(n_l^1) \geq \dots \geq u_j^N(n_l^{T-1}), l = 1, 2.$$

Further, if $B_0 = B_l$, then $u_j^N(n^0) \geq u_j^N(n_l^1) \geq \dots \geq u_j^N(n_l^{T-1})$.

We observe that the suppositions spelled out in items 2 and 3 of the above remark are satisfied in a typical dynamic game in economics and management science, including in the pollution control problem considered in the next section.

§ 3.8.3 Joint optimization in dynamic games played over binary event trees

Suppose that the players maximize their joint payoffs, that is,

$$J_M(\mathbf{x}, \mathbf{u}) = \sum_{t=0}^{T-1} \varrho^t \sum_{n_l^t \in \mathcal{N}^t} \pi(n_l^t) \phi_M^{n_l^t}(x(n_l^t), u(n_l^t)) + \varrho^T \sum_{n_l^T \in \mathcal{N}^T} \pi(n_l^T) \Phi_M^{n_l^T}(x(n_l^T)), \quad (3.8.16)$$

subject to (3.8.5), (3.1.6), (3.8.8), where for any node $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T-1$:

$$\begin{aligned} \phi_M^{n_l^t}(x(n_l^t), u(n_l^t)) &= \sum_{j \in M} \left(\frac{1}{2} u'_j(n_l^t) C_1 u_j(n_l^t) + \sum_{k \neq j} u'_k(n_l^t) C_2 u_j(n_l^t) + c' u_j(n_l^t) \right) \\ &\quad + m d' x(n_l^t), \end{aligned}$$

and in any state $n_l^T \in \mathcal{N}^T$:

$$\Phi_M^{n_l^T}(x(n_l^T)) = m d' x(n_l^T).$$

For the grand coalition M we define the pre-Hamiltonian function for the root node n^0 and any node $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T-1$:

$$\begin{aligned} H_M^{n^0} &= \sum_{j \in M} \left(\frac{1}{2} u'_j(n^0) C_1 u_j(n^0) + \sum_{k \neq j} u'_k(n^0) C_2 u_j(n^0) + c' u_j(n^0) \right) + m d' x^0 \\ &\quad + \varrho [p \lambda'_M(n_1^1) + (1-p) \lambda'_M(n_2^1)] \left[A_0 x^0 + \sum_{j \in M} B_0 u_j(n^0) \right], \\ H_M^{n_l^t} &= \sum_{j \in M} \left(\frac{1}{2} u'_j(n_l^t) C_1 u_j(n_l^t) + \sum_{k \neq j} u'_k(n_l^t) C_2 u_j(n_l^t) + c' u_j(n_l^t) \right) + m d' x(n_l^t) \\ &\quad + \varrho [p \lambda'_M(n_1^{t+1}) + (1-p) \lambda'_M(n_2^{t+1})] \left[A_l x(n_l^t) + \sum_{j \in M} B_l u_j(n_l^t) \right], l = 1, 2. \end{aligned}$$

Theorem 3.5. *If the matrix $C_1 + (m - 1)(C_2 + C'_2)$ is invertible, the controls*

$$u_j^*(n^0) = - \left[c' + \varrho m d' \sum_{\tau=0}^{T-1} (\varrho A)^\tau B_0 \right] (C_1 + (m - 1)(C_2 + C'_2))^{-1}, \quad (3.8.17)$$

$$u_j^*(n_l^t) = - \left[c' + \varrho m d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_l \right] (C_1 + (m - 1)(C_2 + C'_2))^{-1}, \quad (3.8.18)$$

where $t = 1, \dots, T - 1$, $l = 1, 2$, form the cooperative solution in dynamic game played over a binary event tree with identical players. The corresponding costate variables are

$$\lambda_M^*(n_l^t) = m d' \left[I + \varrho \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau A_l \right], \quad t = 1, \dots, T - 1, \quad (3.8.19)$$

$$\lambda_M^*(n_l^T) = m d'. \quad (3.8.20)$$

Moreover, the total payoff of any player $j \in M$ in the cooperative solution is

$$\begin{aligned} J_j(\mathbf{x}^*, \mathbf{u}^*) &= d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 - u_j^*(n^0) \left[\frac{1}{2} C_1 + (m - 1) C'_2 \right] u_j^*(n^0) \\ &\quad - \sum_{t=1}^{T-1} \varrho^t \left(p u_j^*(n_1^t) \left[\frac{1}{2} C_1 + (m - 1) C'_2 \right] u_j^*(n_1^t) \right. \\ &\quad \left. + (1 - p) u_j^*(n_2^t) \left[\frac{1}{2} C_1 + (m - 1) C'_2 \right] u_j^*(n_2^t) \right). \end{aligned} \quad (3.8.21)$$

Proof. The following conditions hold for the existence of cooperative controls (3.8.7) for any $l = 1, 2$:

$$\begin{aligned} 0 &= \frac{\partial H_M^{n^0}}{\partial u_j(n^0)} = u_j'(n^0) C_1 + \sum_{k \neq j} u_k'(n^0) (C_2 + C'_2) + c' \\ &\quad + \varrho [p \lambda_M'(n_1^1) + (1 - p) \lambda_M'(n_2^1)] B_0, \\ 0 &= \frac{\partial H_M^{n_l^t}}{\partial u_j(n_l^t)} = u_j'(n_l^t) C_1 + \sum_{k \neq j} u_k'(n_l^t) (C_2 + C'_2) + c' \\ &\quad + \varrho [p \lambda_M'(n_1^{t+1}) + (1 - p) \lambda_M'(n_2^{t+1})] B_l, \\ n_l^t &\in \mathcal{N}^t, \quad t = 1, \dots, T - 1, \end{aligned}$$

$$\begin{aligned}\lambda'_M(n_i^t) &= \frac{\partial H_M^{n_i^t}}{\partial x(n_i^t)} = md' + \varrho [p\lambda'_M(n_1^{t+1}) + (1-p)\lambda'_M(n_2^{t+1})] A_l, \\ & n_i^t \in \mathcal{N}^t, \quad t = 1, \dots, T-1, \\ \lambda'_M(n_i^T) &= \frac{\partial \Phi_M^{n_i^T}}{\partial x(n_i^T)} = md', \quad n_i^T \in \mathcal{N}^T\end{aligned}$$

with the initial condition for the state variable $x(n^0) = x^0$.

From the last two equations and following the structure of the binary tree, we recurrently obtain the expressions for the costate variables in the forms given by equations (3.8.19) and (3.8.20). Substituting the expressions for the costate variables from (3.8.19) and (3.8.20), and solving the first two equations $\partial H_M^{n^0}/\partial u_j(n^0) = 0$ and $\partial H_M^{n_i^t}/\partial u_j(n_i^t) = 0$, we obtain the optimal controls of the players which are determined by (3.8.17) and (3.8.18). We may notice that the cooperative control of the player depends on the time period and the type of a node, i.e., it depends on if she is in the upward or downward successor at the current time period.

The total payoff of any player $j \in M$ in the cooperative equilibrium is calculated by equation (3.8.3) substituting the cooperative controls and corresponding state trajectory:

$$\begin{aligned}J_j(\mathbf{x}^*, \mathbf{u}^*) &= \sum_{t=0}^{T-1} \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \left[\frac{1}{2} u_j^*(n_i^t) C_1 u_j^*(n_i^t) + \sum_{k \neq j} u_k^*(n_i^t) C_2 u_j^*(n_i^t) \right. \\ &\quad \left. + c' u_j^*(n_i^t) + d' x^*(n_i^t) \right] + \varrho^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) d' x^*(n_i^T) \\ &= \sum_{t=0}^{T-1} \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) \left[u_j^*(n_i^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^*(n_i^t) \right. \\ &\quad \left. + c' u_j^*(n_i^t) + d' x^*(n_i^t) \right] + \varrho^T \sum_{n_i^T \in \mathcal{N}^T} \pi(n_i^T) d' x^*(n_i^T) \\ &= u_j^*(n^0) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^*(n^0) + c' u_j^*(n^0) \\ &\quad + \sum_{t=1}^{T-1} \varrho^t \left[p u_j^*(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^*(n_1^t) \right.\end{aligned}$$

$$\begin{aligned}
& + (1-p)u_j^*(n_2^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_2^t) \Big] \\
& + c' \sum_{t=1}^{T-1} \varrho^t [pu_j^*(n_1^t) + (1-p)u_j^*(n_2^t)] + \sum_{t=0}^T \varrho^t \sum_{n_i^t \in \mathcal{N}^t} \pi(n_i^t) d' x^*(n_i^t) \\
= & u_j^*(n^0) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n^0) + c' u_j^*(n^0) \\
& + \sum_{t=1}^{T-1} \varrho^t \left[pu_j^*(n_1^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_1^t) \right. \\
& \quad \left. + (1-p)u_j^*(n_2^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_2^t) \right] \\
& + c' \sum_{t=1}^{T-1} \varrho^t [pu_j^*(n_1^t) + (1-p)u_j^*(n_2^t)] \\
& + d' \left[x^0 + \varrho \sum_{t=0}^{T-1} (\varrho A)^t [A_0 x^0 + mB_0 u_j^*(n^0)] \right. \\
& \quad \left. + m \sum_{t=2}^T \varrho^t \sum_{\tau=0}^{t-2} A^\tau [pB_1 u_j^*(n_1^{t-\tau-1}) + (1-p)B_2 u_j^*(n_2^{t-\tau-1})] \right] \\
= & d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 + u_j^*(n^0) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n^0) \\
& + \left[c' + m\varrho d' \sum_{t=0}^{T-1} (\varrho A)^t B_0 \right] u_j^*(n^0) \\
& + \sum_{t=1}^{T-1} \varrho^t \left(pu_j^*(n_1^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_1^t) \right. \\
& \quad \left. + (1-p)u_j^*(n_2^t) \left[\frac{1}{2}C_1 + (m-1)C_2 \right] u_j^*(n_2^t) \right) \\
& + \sum_{t=1}^{T-1} \varrho^t \left(p \left[c' + m\varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_1 \right] u_j^*(n_1^t) \right. \\
& \quad \left. + (1-p) \left[c' + m\varrho d' \sum_{\tau=0}^{T-t-1} (\varrho A)^\tau B_2 \right] u_j^*(n_2^t) \right)
\end{aligned}$$

$$\begin{aligned}
&= d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 + u_j^*(n^0) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^*(n^0) \\
&\quad - u_j^*(n^0) [C_1 + (m-1)(C_2 + C'_2)] u_j^*(n^0) \\
&\quad + \sum_{t=1}^{T-1} \varrho^t \left(p u_j^*(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^*(n_1^t) \right. \\
&\quad \quad \left. + (1-p) u_j^*(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C_2 \right] u_j^*(n_2^t) \right) \\
&\quad - \sum_{t=1}^{T-1} \varrho^t \left(p u_j^*(n_1^t) [C_1 + (m-1)(C_2 + C'_2)] u_j^*(n_1^t) \right. \\
&\quad \quad \left. + (1-p) u_j^*(n_2^t) [C_1 + (m-1)(C_2 + C'_2)] u_j^*(n_2^t) \right) \\
&= d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 - u_j^*(n^0) \left[\frac{1}{2} C_1 + (m-1) C'_2 \right] u_j^*(n^0) \\
&\quad - \sum_{t=1}^{T-1} \varrho^t \left(p u_j^*(n_1^t) \left[\frac{1}{2} C_1 + (m-1) C'_2 \right] u_j^*(n_1^t) \right. \\
&\quad \quad \left. + (1-p) u_j^*(n_2^t) \left[\frac{1}{2} C_1 + (m-1) C'_2 \right] u_j^*(n_2^t) \right).
\end{aligned}$$

Theorem is proved. □

Remark 3.6. *We make the following three observations:*

1. *If A_1 , A_2 and d are non-negative, then on the paths formed of either only upward nodes, or downward nodes, we have*

$$\lambda_j^*(n_l^1) \geq \dots \geq \lambda_j^*(n_l^T), l = 1, 2,$$

that is, the costate variable of player $j \in M$ as a function of the game period is non-increasing.

2. *If A_1 , A_2 , B_l , c , d are non-negative and C_1 , C_2 are negative-defined, then we have a similar property on the control variables, namely:*

$$u_j^*(n_l^1) \geq \dots \geq u_j^*(n_l^{T-1}), l = 1, 2.$$

Further, if $B_0 = B_l$, then $u_j^(n^0) \geq u_j^*(n_l^1) \geq \dots \geq u_j^*(n_l^{T-1})$.*

3. *For any node $n \in \mathcal{N} \setminus \mathcal{N}^T$ and player $j \in M$, the cooperative controls $u_j^*(n)$ and controls in S -adapted Nash equilibrium $u_j^N(n)$ satisfy the following linear*

matrix equation:

$$u_j^*(n) = [mu_j'^N(n) (C_1 + (m - 1)C_2) + (m - 1)c'] (C_1 + (m - 1)(C_2 + C'_2))^{-1} \quad (3.8.22)$$

Using the expression of the cooperative equilibrium controls (3.8.22) and (3.8.3), the total payoff of any player $j \in M$ in the S -adapted Nash equilibrium can be expressed in an alternative form:

$$\begin{aligned} J_j(\mathbf{x}^N, \mathbf{u}^N) &= d' \left[I + \varrho \sum_{t=0}^{T-1} (\varrho A)^t A_0 \right] x^0 + u_j'^N(n^0) \left[\frac{1}{2} C_1 + (m - 1) C_2 \right] u_j^N(n^0) \\ &\quad - u_j^*(n^0) [C_1 + (m - 1)(C_2 + C'_2)] u_j^N(n^0) \\ &\quad + \sum_{t=1}^{T-1} \varrho^t \left(p u_j'^N(n_1^t) \left[\frac{1}{2} C_1 + (m - 1) C_2 \right] u_j^N(n_1^t) \right. \\ &\quad \left. + (1 - p) u_j'^N(n_2^t) \left[\frac{1}{2} C_1 + (m - 1) C_2 \right] u_j^N(n_2^t) \right) \\ &\quad - \sum_{t=1}^{T-1} \varrho^t (p u_j^*(n_1^t) [C_1 + (m - 1)(C_2 + C'_2)] u_j^N(n_1^t) \\ &\quad \left. + (1 - p) u_j^*(n_2^t) [C_1 + (m - 1)(C_2 + C'_2)] u_j^N(n_2^t) \right). \end{aligned}$$

that is, as function of cooperative and noncooperative controls.

§ 3.8.4 Price of anarchy in dynamic games played over binary event trees

In this subsection, we compare the payoffs of the players in the S -adapted Nash and cooperative equilibria. We consider the price of anarchy [117] as a measure of the difference between payoffs in two types of players' behavior—cooperative and non-cooperative.

Definition 3.10. *In the game played over event tree with payoff function $W_j(\mathbf{u}, x^0) = J_j(\mathbf{x}, \mathbf{u})$ determined by (3.8.3) subject to (3.8.4), (3.8.5), (3.8.6), the price of anarchy in the problem of profit maximization is given by*

$$PoA = \frac{\max_{\mathbf{u}} \sum_{j \in M} W_j(\mathbf{u}, x^0)}{\sum_{j \in M} W_j(\mathbf{u}^N, x^0)}.$$

In our case of binary event tree, homogeneous players and maximization problem, the price of anarchy becomes

$$PoA = \frac{J_j(\mathbf{x}^*, \mathbf{u}^*)}{J_j(\mathbf{x}^N, \mathbf{u}^N)}, \quad (3.8.23)$$

where $J_j(\mathbf{x}^*, \mathbf{u}^*)$ is the payoff of any player $j \in M$ in the cooperative equilibrium given by (3.8.21) and $J_j(\mathbf{x}^N, \mathbf{u}^N)$ is the payoff of any player $j \in M$ in the S -adapted Nash equilibrium given by (3.8.15).

Obviously, the price of anarchy is not less than one if the payoffs $J_j(\mathbf{x}^*, \mathbf{u}^*)$ and $J_j(\mathbf{x}^N, \mathbf{u}^N)$ are non-negative. Using (3.8.23), we may calculate the price of anarchy in explicit form.

Remark 3.7. *We introduce another formula for the price of anarchy. Having $u_j^N(n_l^t)$ and $u_j^*(n_l^t) \in \mathbb{R}^r$, $t = 0, \dots, T-1$ and $l = 1, 2$, we form vectors of cooperative controls:*

$$\mathbf{u}_j^*(l) = (u_j^*(n^0), u_j^*(n_l^1), \dots, u_j^*(n_l^{T-1})) \in \mathbb{R}^{rT}, \quad l = 1, 2,$$

and vectors formed the S -adapted equilibrium:

$$\mathbf{u}_j^N(l) = (u_j^N(n^0), u_j^N(n_l^1), \dots, u_j^N(n_l^{T-1})) \in \mathbb{R}^{rT}, \quad l = 1, 2.$$

Denoting $C = \frac{1}{2}C_1 + (m-1)C_2'$, we form the block diagonal matrix \mathbf{C} of size $rT \times rT$:

$$\mathbf{C} = \begin{pmatrix} C & 0 & \dots & 0 \\ 0 & C & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C \end{pmatrix}.$$

Therefore, we obtain the following formula for the price of anarchy:

$$PoA = \frac{p\mathbf{u}_j'^*(1)\mathbf{C}\mathbf{u}_j^*(1) + (1-p)\mathbf{u}_j'^*(2)\mathbf{C}\mathbf{u}_j^*(2)}{p\Psi(1) + (1-p)\Psi(2)},$$

where

$$\Psi(k) = \mathbf{u}_j'^*(k)\mathbf{C}\mathbf{u}_j^*(k) - (\mathbf{u}_j^*(k) - \mathbf{u}_j^N(k))' \mathbf{C} (\mathbf{u}_j^*(k) - \mathbf{u}_j^N(k)), \quad k = 1, 2.$$

§ 3.9 Price of Anarchy in an Example in Pollution Control

§ 3.9.1 Model

In this section, we illustrate our results in the context of a dynamic game of pollution control with m players representing countries or firms emitting pollutants in a geographical area, with emissions accumulating over time.

During the last three decades or so, a significant literature has looked at issues related to global pollution control problems, where the emissions by any player affect the global commons and the damage cost of all players. In such setting, it has been repeatedly shown that when players (firms, regions, countries) decide noncooperatively their pollution emissions levels, then each player pollutes more than when the game is played cooperatively. The main reason for this result is that under cooperation (i.e., joint optimization), each player internalizes the damage cost of all other players, whereas when acting selfishly a player only accounts for its own damage cost.

We retain a simple model, which, however, is very much in line with respect to what has been done in the literature; see the reviews by [109] и [125] on applications of dynamic games to pollution control. Indeed, our model captures the main features of a multi-player pollution control problem, namely: (i) emissions are needed to produce goods and services and therefore are the source of revenue; (ii) emissions accumulate over time and the stock of pollution harms the environment and inflicts a damage cost on all players; (iii) nature cleans itself by eliminating part of the pollution stock, but the rate at which oceans and forests absorb over time pollution is uncertain; (iv) each player aims at maximizing a stream of discounted rewards over a finite horizon, with the instantaneous payoff being the difference between revenues and the damage cost.

The game is played over a binary event tree with T periods starting from period $t = 0$ and with the conditional probability p of realization of the upward successor of any node and $1 - p$ for a downward successor.

Let $u_j(n_j^t) \in \mathbb{R}_+$ be player j 's emissions in node $n_j^t \in \mathcal{N}^t$, and denote by $x(n_j^t) \in$

\mathbb{R}_+ the stock of pollution at that node, with initial condition $x(n^0) = 0$.⁴ The state (stock) equation for any $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T$ has the form:

$$x(n_l^t) = \begin{cases} \sum_{j \in M} u_j(a(n_l^t)) & \text{if } a(n_l^t) = n^0, \\ \delta_1 x(a(n_l^t)) + \sum_{j \in M} u_j(a(n_l^t)) & \text{if } a(n_l^t) \text{ is an upward node,} \\ \delta_2 x(a(n_l^t)) + \sum_{j \in M} u_j(a(n_l^t)) & \text{if } a(n_l^t) \text{ is a downward node,} \end{cases} \quad (3.9.1)$$

where δ_1 and δ_2 are positive constants from the unit interval representing the natural decay rate of the stock in the corresponding node.

As alluded to it above, pollution is an inevitable by-product of production. If we assume a monotone increasing relationship between production and pollution, then the benefit from production for player j can be expressed as function of emissions, that is, $f_j(u)$. The stock of pollution causes an environmental damage, which is assumed to be well approximated by $\kappa x(n_l^t)$, where $\kappa > 0$. The linearity of the damage cost is for tractability and it is surely worth it to consider other functional forms in future investigations. We suppose the presence of an interaction term in players' emissions in the benefit function. This model corresponds to an oligopolistic competition à la Cournot.

Consider an industry formed of m oligopolistic firms producing a homogeneous product and competing à la Cournot, with a linear inverse demand function, that is,

$$\mathcal{P}(n_l^t) = \left(\alpha - \gamma \sum_{i \in M} u_i(n_l^t) \right). \quad (3.9.2)$$

The revenue function at node n_l^t , $t = 0, 1, \dots, T - 1$ is given by

$$f_j^{n_l^t}(u(n_l^t)) = u_j(n_l^t) \left(\alpha - \gamma \sum_{i \in M} u_i(n_l^t) \right), \quad (3.9.3)$$

and the reward (revenue minus cost) at that node by

$$\phi_j^{n_l^t}(x(n_l^t), u_j(n_l^t)) = u_j(n_l^t) \left(\alpha - \gamma \sum_{i \in M} u_i(n_l^t) \right) - \kappa x(n_l^t), \quad (3.9.4)$$

where α , γ and κ are positive constants. In a terminal node n_l^T , the reward to player j is given by the function

$$\Phi_j^{n_l^T}(x(n_l^T)) = -\kappa x(n_l^T). \quad (3.9.5)$$

⁴Having a positive initial stock does not change qualitatively the results.

Remark 3.8. *The usual assumption in the literature is that the revenue function only depends on own emissions, that is, it is given by $f_j(u_j)$. Here, we consider the more general case where the revenue function depends on all players' controls.*

§ 3.9.2 Results

Using Theorems 3.4 and 3.5, we obtain the S -adapted Nash and cooperative equilibrium controls:

$$u_j^N(n_l^t) = \frac{1}{(m+1)\gamma} (\alpha - \varrho\kappa\Delta(t)), \quad (3.9.6)$$

$$u_j^*(n_l^t) = \frac{1}{2m\gamma} (\alpha - m\varrho\kappa\Delta(t)), \quad (3.9.7)$$

for any player $j \in M$, $n_l^t \in \mathcal{N}^t$, $t = 0, 1, \dots, T-1$, where

$$\Delta(t) = \frac{1 - [\varrho(p\delta_1 + (1-p)\delta_2)]^{T-t}}{1 - \varrho(p\delta_1 + (1-p)\delta_2)}.$$

The difference between the S -adapted Nash equilibrium and cooperative controls is given by

$$u_j^N(n_l^t) - u_j^*(n_l^t) = \frac{m-1}{2m(m+1)\gamma} (\alpha + m\varrho\kappa\Delta(t)), \quad (3.9.8)$$

and it is positive for all nodes.

From Theorems 3.4 and 3.5 we also obtain the following payoffs to any player $j \in M$ in the S -adapted equilibrium and cooperative solution:

$$\begin{aligned} J_j(\mathbf{x}^N, \mathbf{u}^N) &= \sum_{t=0}^{T-1} \varrho^t [m^2\gamma u_j^N(n_l^t) - \alpha(m-1)] u_j^N(n_l^t) = \\ &= m\gamma \sum_{t=0}^{T-1} \varrho^t [2u_j^*(n_l^t) - u_j^N(n_l^t)] u_j^N(n_l^t), \end{aligned} \quad (3.9.9)$$

$$J_j(\mathbf{x}^*, \mathbf{u}^*) = m\gamma \sum_{t=0}^{T-1} \varrho^t [u_j^*(n_l^t)]^2. \quad (3.9.10)$$

Noticing that

$$2u_j^*(n_l^t) - u_j^N(n_l^t) = \frac{1}{m(m+1)\gamma} (\alpha - m^2\varrho\kappa\Delta(t)),$$

we may calculate the difference between payoffs in the S -adapted Nash equilibrium and the cooperative solution:

$$J_j(\mathbf{x}^N, \mathbf{u}^N) - J_j(\mathbf{x}^*, \mathbf{u}^*) = -m\gamma \sum_{t=0}^{T-1} \varrho^t [u_j^N(n_l^t) - u_j^*(n_l^t)]^2, \quad (3.9.11)$$

which is always negative.

Remark 3.9. As $\Delta(t)$ is decreasing in t , it suffices to assume that $\alpha > m\rho\kappa\Delta(0)$ to have interior solution in both solutions in all nodes. Further, for $J_j(\mathbf{x}^N, \mathbf{u}^N)$ to be positive, a sufficient condition is $\alpha > m^2\rho\kappa\Delta(0)$. We shall retain this assumption in the sequel, which means that all players will remain in the game for its whole duration.

Theorem 3.6. If $\alpha > m^2\rho\kappa\Delta(0)$, in the game with rewards (3.9.4) for non-terminal nodes and (3.9.5) for terminal nodes and state equation (3.9.1), then the price of anarchy PoA is given by

$$PoA = 1 + \frac{(m-1)^2 \sum_{t=0}^{T-1} \rho^t (\alpha + m\rho\kappa\Delta(t))^2}{4m \sum_{t=0}^{T-1} \rho^t (\alpha - m^2\rho\kappa\Delta(t))(\alpha - \rho\kappa\Delta(t))} \quad (3.9.12)$$

and is bounded as follows:

$$\begin{aligned} \underline{PoA} &= 1 + \frac{(m-1)^2(\alpha + m\rho\kappa)^2}{4m(\alpha - m^2\rho\kappa)(\alpha - \rho\kappa)} \leq PoA \\ &\leq 1 + \frac{(m-1)^2(\alpha + m\rho\kappa\Delta(0))^2}{4m(\alpha - m^2\rho\kappa\Delta(0))(\alpha - \rho\kappa\Delta(0))} = \overline{PoA}. \end{aligned} \quad (3.9.13)$$

Proof. We calculate PoA using equation (4.1.21):

$$\begin{aligned} PoA &= \frac{J_j(\mathbf{x}^*, \mathbf{u}^*)}{J_j(\mathbf{x}^N, \mathbf{u}^N)} = 1 + \frac{J_j(\mathbf{x}^*, \mathbf{u}^*) - J_j(\mathbf{x}^N, \mathbf{u}^N)}{J_j(\mathbf{x}^N, \mathbf{u}^N)} \\ &= 1 + \frac{m\gamma \sum_{t=0}^{T-1} \rho^t [u_j^N(n_i^t) - u_j^*(n_i^t)]^2}{m\gamma \sum_{t=0}^{T-1} \rho^t [2u_j^*(n_i^t) - u_j^N(n_i^t)] u_j^N(n_i^t)} \\ &= 1 + \frac{(m-1)^2 \sum_{t=0}^{T-1} \rho^t (\alpha + m\rho\kappa\Delta(t))^2}{4m \sum_{t=0}^{T-1} \rho^t (\alpha - m^2\rho\kappa\Delta(t))(\alpha - \rho\kappa\Delta(t))}. \end{aligned}$$

Here $u_j^N(n_i^t) - u_j^*(n_i^t)$ is a decreasing function of t and both $2u_j^*(n_i^t) - u_j^N(n_i^t)$ and

$u_j^N(n_t^t)$ are increasing functions of t , we may calculate the upper bound for the PoA:

$$\begin{aligned}
PoA &\leq 1 + \frac{m\gamma \sum_{t=0}^{T-1} \varrho^t [u_j^N(n^0) - u_j^*(n^0)]^2}{m\gamma \sum_{t=0}^{T-1} \varrho^t [2u_j^*(n^0) - u_j^N(n^0)] u_j^N(n^0)} \\
&= 1 + \frac{m\gamma \cdot \frac{1-\varrho^T}{1-\varrho} \cdot \frac{(m-1)^2}{4m^2(m+1)^2\gamma^2} \cdot (\alpha + m\rho\kappa\Delta(0))^2}{m\gamma \frac{1-\varrho^T}{1-\varrho} \cdot \frac{(\alpha - m^2\rho\kappa\Delta(0))(\alpha - \rho\kappa\Delta(0))}{m(m+1)^2\gamma^2}} \\
&= 1 + \frac{(m-1)^2(\alpha + m\rho\kappa\Delta(0))^2}{4m(\alpha - m^2\rho\kappa\Delta(0))(\alpha - \rho\kappa\Delta(0))}.
\end{aligned}$$

Determine the lower bound of the PoA:

$$\begin{aligned}
PoA &\geq 1 + \frac{m\gamma \sum_{t=0}^{T-1} \varrho^t [u_j^N(n^{T-1}) - u_j^*(n^{T-1})]^2}{m\gamma \sum_{t=0}^{T-1} \varrho^t [2u_j^*(n^{T-1}) - u_j^N(n^{T-1})] u_j^N(n^{T-1})} \\
&= 1 + \frac{m\gamma \cdot \frac{1-\varrho^T}{1-\varrho} \cdot \frac{(m-1)^2}{4m^2(m+1)^2\gamma^2} \cdot (\alpha + m\rho\kappa\Delta(T-1))^2}{m\gamma \frac{1-\varrho^T}{1-\varrho} \cdot \frac{(\alpha - m^2\rho\kappa\Delta(T-1))(\alpha - \rho\kappa\Delta(T-1))}{m(m+1)^2\gamma^2}} \\
&= 1 + \frac{(m-1)^2(\alpha + m\rho\kappa\Delta(T-1))^2}{4m(\alpha - m^2\rho\kappa\Delta(T-1))(\alpha - \rho\kappa\Delta(T-1))}.
\end{aligned}$$

Taking into account that $\Delta(T-1) = 1$ we obtain the lower bound of the PoA:

$$PoA \geq 1 + \frac{(m-1)^2(\alpha + m\rho\kappa)^2}{4m(\alpha - m^2\rho\kappa)(\alpha - \rho\kappa)}.$$

□

The price of anarchy and its bounds depend on all model's parameters, but γ . Recall that this parameter appears in the inverse demand function (3.9.2), where $\mathcal{P}(n_t^t)$ is interpreted as the price per unit of emissions (same as production here) at node n_t^t . Consequently, the higher γ the lower the price $\mathcal{P}(n_t^t)$ that the players can get. Now, from (3.9.9) and (3.9.10), it is easy to see that when we take the ratio of payoffs to compute the PoA, the parameter γ factors out. In economic terms, this is due to the assumption made here, as well as in the literature in general, that the product is homogeneous, that is, the impact of an additional emissions (production) unit by any player has the same impact on price. If the product was imperfectly substitutable, then clearly the result would have been different.

The following proposition sheds a light on the impact of varying the other parameters on the PoA and its bounds. The pollution dynamics parameters (δ_1, δ_2, p) do not have any particular interest as they simply allow to account for uncertainty.

Proposition 3.1. *Sensitivity analysis. For $\alpha > m^2 \varrho \kappa \Delta(0)$, the following holds:*

1. PoA and \overline{PoA} are strictly increasing in T , and

$$\lim_{T \rightarrow \infty} PoA = \lim_{T \rightarrow \infty} \overline{PoA} = 1 + \frac{(m-1)^2 (\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) + m\varrho\kappa)^2}{4m(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - m^2\varrho\kappa)(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - \varrho\kappa)}.$$

2. PoA , \underline{PoA} and \overline{PoA} are strictly increasing in the number of players m .

3. PoA , \underline{PoA} and \overline{PoA} are strictly decreasing in α .

4. PoA , \underline{PoA} and \overline{PoA} are strictly increasing in κ .

5. PoA , \underline{PoA} and \overline{PoA} are strictly increasing in discount factor ϱ .

Proof. 1. First, consider the upper bound from inequality (3.9.13) as a function of $\Delta(0)$. The first derivative of this bound is

$$\frac{\alpha\kappa\varrho(m-1)^2(m+1)^2\Delta(0)(\alpha - m\varrho\kappa\Delta(0))(\alpha + m\varrho\kappa\Delta(0))}{4(\alpha - \varrho\kappa\Delta(0))^2(\alpha - m^2\varrho\kappa\Delta(0))},$$

and it is positive because $\alpha > m^2\varrho\kappa\Delta(0)$. Therefore, the upper bound is an increasing function of $\Delta(0)$. Moreover, $\Delta(0)$ is an increasing function of T . Thus, the upper bound is an increasing function of T . The lower bound does not depend on T .

Consider the limit of the price of anarchy and its upper bound when T tends to infinity. The limit of the upper bound is

$$\lim_{T \rightarrow \infty} \left[1 + \frac{(m-1)^2(\alpha + m\varrho\kappa\Delta(0))^2}{4m(\alpha - m^2\varrho\kappa\Delta(0))(\alpha - \varrho\kappa\Delta(0))} \right] = 1 + \frac{(m-1)^2(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) + m\varrho\kappa)}{4m(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - m^2\varrho\kappa)(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - \varrho\kappa)},$$

taking into account that $\lim_{T \rightarrow \infty} \Delta(0) = 1/[1 - \varrho(p\delta_1 + (1-p)\delta_2)]$ and $\varrho(p\delta_1 + (1-p)\delta_2) \in (0, 1)$.

Now consider the price of anarchy. As $\lim_{T \rightarrow \infty} \Delta(t) = 1/(1 - \varrho(p\delta_1 + (1-p)\delta_2))$, we find the limit of the sum in the numerator:

$$\begin{aligned} \lim_{T \rightarrow \infty} \sum_{t=0}^{T-1} \varrho^t (\alpha + m\varrho\kappa\Delta(t))^2 &= \sum_{t=0}^{\infty} \varrho^t \left(\alpha + \frac{m\varrho\kappa}{(1 - \varrho(p\delta_1 + (1-p)\delta_2))} \right)^2 \\ &= \frac{(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) + m\varrho\kappa)^2}{(1 - \varrho)(1 - \varrho(p\delta_1 + (1-p)\delta_2))^2}. \end{aligned}$$

The limit of the sum in the denominator is

$$\begin{aligned} &\sum_{t=0}^{T-1} \varrho^t (\alpha - m^2\varrho\kappa\Delta(t))(\alpha - \varrho\kappa\Delta(t)) = \\ &= \frac{(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - m^2\varrho\kappa)(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - \varrho\kappa) \sum_{t=0}^{\infty} \varrho^t}{(1 - \varrho(p\delta_1 + (1-p)\delta_2))^2} \\ &= \frac{(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - m^2\varrho\kappa)(\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - \varrho\kappa)}{(1 - \varrho)(1 - \varrho(p\delta_1 + (1-p)\delta_2))^2}. \end{aligned}$$

Taking into account these limits we obtain

$$\begin{aligned} \lim_{T \rightarrow \infty} PoA &= 1 + \\ &+ \frac{(m-1)^2 (\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) + m\varrho\kappa)^2}{4m (\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - m^2\varrho\kappa) (\alpha(1 - \varrho(p\delta_1 + (1-p)\delta_2)) - \varrho\kappa)}. \end{aligned}$$

2. Consider the expression of the price of anarchy PoA , and the second summand is

$$\frac{(m-1)^2 \sum_{t=0}^{T-1} \varrho^t (\alpha + m\varrho\kappa\Delta(t))^2}{4m \sum_{t=0}^{T-1} \varrho^t (\alpha - m^2\varrho\kappa\Delta(t))(\alpha - \varrho\kappa\Delta(t))}. \quad (3.9.14)$$

The multiplier $\frac{(m-1)^2}{4m}$ increases with m for $m > 1$, the multiplier $\sum_{t=0}^{T-1} \varrho^t (\alpha + m\varrho\kappa\Delta(t))^2$ increases with m and the multiplier $\sum_{t=0}^{T-1} \varrho^t (\alpha - m^2\varrho\kappa\Delta(t))(\alpha - \varrho\kappa\Delta(t))$ decreases with m . Therefore, PoA increases with m . The same analysis can be applied for \underline{PoA} and \overline{PoA} and it proves that \underline{PoA} and \overline{PoA} increase with m .

3. First, consider the price of anarchy PoA . Its second summand (3.9.14) can be

represented in the following way:

$$\begin{aligned} & \frac{(m-1)^2 \sum_{t=0}^{T-1} \varrho^t (\alpha - m^2 \varrho \kappa \Delta(t)) (\alpha - \varrho \kappa \Delta(t)) + (m+1)^2 \alpha \varrho \kappa \sum_{t=0}^{T-1} \varrho^t \Delta(t)}{4m \sum_{t=0}^{T-1} \varrho^t (\alpha - m^2 \varrho \kappa \Delta(t)) (\alpha - \varrho \kappa \Delta(t))} \\ &= \frac{(m-1)^2}{4m} \left(1 + \frac{(m+1)^2 \alpha \varrho \kappa \sum_{t=0}^{T-1} \varrho^t \Delta(t)}{\sum_{t=0}^{T-1} \varrho^t (\alpha - m^2 \varrho \kappa \Delta(t)) (\alpha - \varrho \kappa \Delta(t))} \right). \end{aligned}$$

The last fraction decreases with α , therefore, the price of anarchy P_{oA} is strictly decreasing in α . The same result is true for $\underline{P_{oA}}$, $\overline{P_{oA}}$, they are both decreasing functions of α .

4. Consider the price of anarchy P_{oA} . In the expression of P_{oA} , the second summand takes form of (3.9.14). The numerator of this fraction increases with κ , and the denominator decreases with κ . Therefore, P_{oA} increases with κ . The same analysis is true for the lower $\underline{P_{oA}}$ and upper $\overline{P_{oA}}$ bounds. They both increase with κ .
5. The numerator in the second summand of the expression for $\underline{P_{oA}}$ in (3.9.13) is strictly increasing in ϱ while the denominator is strictly decreasing as its partial derivative w.r.t. ϱ is negative for $\alpha > m^2 \varrho \kappa \Delta(0)$. Therefore, $\underline{P_{oA}}$ is strictly increasing in ϱ .

The same result is true for $\overline{P_{oA}}$. The numerator in the second summand of the expression for $\overline{P_{oA}}$ (3.9.13) is strictly increasing in $\varrho \Delta(0)$ while the denominator is strictly decreasing in $\varrho \Delta(0)$. Therefore $\overline{P_{oA}}$ is strictly increasing in $\varrho \Delta(0)$ and ϱ .

Since $(\alpha - m^2 \varrho \kappa \Delta(t)) (\alpha - \varrho \kappa \Delta(t)) = (\alpha - m \varrho \kappa \Delta(t))^2 - (m-1)^2 \alpha \varrho \kappa \Delta(t)$ and $(\alpha + m \varrho \kappa \Delta(t))^2 = (\alpha - m \varrho \kappa \Delta(t))^2 + 4m \alpha \varrho \kappa \Delta(t)$, the expression for P_{oA} may be re-written in the following form:

$$P_{oA} = \frac{\frac{(m+1)^2}{4m} \cdot \sum_{t=0}^{T-1} \varrho^t (\alpha - m \varrho \kappa \Delta(t))^2}{\sum_{t=0}^{T-1} \varrho^t [(\alpha - m \varrho \kappa \Delta(t))^2 - (m-1)^2 \alpha \varrho \kappa \Delta(t)]}. \quad (3.9.15)$$

For $\alpha > m^2 \varrho \kappa \Delta(0)$, each summand in the denominator of PoA is decreasing in $\varrho \Delta(t)$ (and therefore in ϱ) faster than the summand in the numerator with the same index. Therefore the ratio, PoA , is strictly increasing in ϱ . □

§ 3.9.3 Interaction of Results

Now we provide the interaction of the results obtained in the previous paragraph.

Impact of planning horizon T : The longer the planning horizon, the larger is the collective loss that the players would incur if they fail to cooperate. Since for any given planning horizon, cooperation leads to at least as good collective payoff as noncooperation, extending the planning horizon, necessarily widens the gap. Interestingly, the result in item 1 shows that the PoA is upper bounded when we extend the analysis to an infinite-horizon game. Finally, we note that the lower bound \underline{PoA} corresponds to the price of anarchy for a deterministic game that is played once. Indeed, it is easy to verify that for $T = 1$ and $t = 0$, we have

$$\Delta(0) = \frac{1 - \varrho(p\delta_1 + (1-p)\delta_2)}{1 - \varrho(p\delta_1 + (1-p)\delta_2)} = 1,$$

and \underline{PoA} is then equal to PoA . In a one-shot game, there is no tree event involved, and the game is played at the root node. This implies that assessing the impact of varying the parameter values on the lower bound of the PoA , is equivalent to conducting a sensitivity analysis on the price of anarchy in a static game.

Impact of number of players m : The results show that the larger the number of players, the larger the cost of absence of cooperation. This is similar to what happens in network games, where selfishness leads to higher congestion and consequently to higher total cost. In our context, when the game is played non-cooperatively, each player ignores the impact of her emissions on the others, and pollutes more. Consequently, the larger the number of players, the larger the total pollution (which is qualitatively similar to an increase in congestion in routing games), and the larger the damage cost. In cooperation, the negative environmental externality is taken into account when choosing optimally the emissions, which leads to lower total pollution and damage cost than under non-cooperative mode of play.

Impact of α : Recalling that

$$\mathcal{P}(n_i^t) = \left(\alpha - \gamma \sum_{i \in M} u_i(n_i^t) \right),$$

is the price, then parameter α measures the maximum willingness-to-pay by consumer for a unit of emissions (or production). Our results show that increasing the value of α reduces the *PoA*, that is, cooperation delivers less, in relative terms, when consumers are willing to pay more. To explain this result, recall that the revenues at node n_i^t are given by

$$f_j^{n_i^t}(u(n_i^t)) = u_j(n_i^t) \left(\alpha - \gamma \sum_{i \in M} u_i(n_i^t) \right).$$

Substituting for non-cooperative and cooperative emissions, leads to the following revenues

$$\begin{aligned} f_j^{n_i^t}(u^N(n_i^t)) &= \frac{(\alpha - \rho\kappa\Delta(t)) (\alpha + m\rho\kappa\Delta(t))}{(m+1)^2\gamma}, \\ f_j^{n_i^t}(u^*(n_i^t)) &= \frac{(\alpha - m\rho\kappa\Delta(t)) (\alpha + m\rho\kappa\Delta(t))}{4m\gamma}. \end{aligned}$$

The marginal revenues are given by

$$\begin{aligned} \frac{\partial f_j^{n_i^t}(u^N(n_i^t))}{\partial u_j^N(n_i^t)} &= \rho\kappa\Delta(t), \\ \frac{\partial f_j^{n_i^t}(u^*(n_i^t))}{\partial u_j^*(n_i^t)} &= \frac{2m\alpha - (\alpha - m\rho\kappa\Delta(t)) (m+1)}{2m}. \end{aligned}$$

This shows that whereas the marginal revenues are independent of α in the non-cooperative case, they are increasing in this parameter in the cooperative case. This implies that there is a positive incentive to increase emissions in the cooperative case, which in turn leads to higher damage cost over time.

Impact of κ : Cooperation among players means that the determination of optimal emissions policies takes into account the damage cost of all players, and not only own damage as it is the case under noncooperation. This internalization of the damage cost leads in turn to lower emissions in the cooperative case than in the non-cooperative one. Consequently, the larger the marginal damage cost κ , the larger the gap in emissions between cooperation and noncooperation. This

translates into a larger cost for absence of cooperation among players when tackling the environmental problem.

Impact of ρ : Finally, the higher the discount factor ρ , the higher the price of anarchy. Increasing ρ means that we put a lower weight on future earnings, and the sound cooperative solution would deliver less benefit.

§ 3.10 Node-consistent Shapley Value in a game with random termination time

§ 3.10.1 Model

In this section we consider the model of a game played over event tree, in which the duration of the game process is random. This is the difference between the model considered here with the previous models considered above in this chapter. For such a game, a cooperative version of the game is considered and a method of constructing node-consistent Shapley value is proposed. It should be noticed that the problem of time consistency (subgame or node consistency — for the games played over graphs) of cooperative solutions for differential games with random duration was initially formulated in the paper [47]. Subsequently, this work was continued in the article [49]. The principles of stable cooperation for differential games with random duration have been formulated in [55], and the Hamilton-Jacobi- Bellman equations for the problems of this type have been obtained in [132].

The stochastic process is defined by two probabilities given for any node $n_i^t \in \mathcal{N}^t$, $t \in \mathcal{T}$, that is, $\tau(n_i^t)$ and $\pi(n_i^t)$. At any node $n_i^t \in \mathcal{N}^t$, the game may terminate with a probability $\tau(n_i^t) \in [0, 1]$.⁵ In particular, $\tau(n^0) = 0$. If the stochastic process does not terminate in node n_i^t , it transits to one of the successor nodes in $\mathcal{S}(n_i^t)$. We denote by $\pi(n_i^t)$ the probability of passing through node n_i^t , which corresponds to the sum of the probabilities of all scenarios that contain this node. In particular, $\pi(n^0) = 1$ and $\pi(n_i^T)$ is equal to the probability of the single scenario that terminates in (leaf) node n_i^T . Notice that $\tau(n_i^T) = \pi(n_i^T)$ for any $n_i^T \in \mathcal{N}^T$, therefore, the process will terminate in the nodes from \mathcal{N}^T for sure. The given probabilities satisfy

⁵By setting probabilities $\tau(n_i^t)$ for any $n_i^t \in \mathcal{N}^t$ for all $t = 0, \dots, T$ we determine the probability distribution of the process termination over the set of nodes. For the distribution it is that $\sum_{n \in \mathcal{N}} \tau(n) = 1$.

the following condition:

$$\pi(n_i^t) = \tau(n_i^t) + \sum_{n \in \mathcal{S}(n_i^t)} \pi(n), \quad (3.10.1)$$

that is, the probability of passing through node n_i^t is equal to the sum of the probability that the game terminates at this node and the sum of probabilities passing through the direct successors of this node.

As it was given above, $M = \{1, \dots, m\}$ is the set of players. For each player $j \in M$, we define a set of controls indexed over the set of nodes. Denote by $u_j(n_i^t) \in U_j^{n_i^t} \subset \mathbb{R}^{m_j}$ the control of player j at node n_i^t , and let $u(n_i^t) = (u_1(n_i^t), \dots, u_m(n_i^t))$ be a profile of controls of node n_i^t . For each node $n_i^t \in \mathcal{N}^t$, $t = 0, 1, \dots, T$, let $U_j^{n_i^t}$ be the control set of player j . Denote by $U^{n_i^t} = U_1^{n_i^t} \times \dots \times U_j^{n_i^t} \times \dots \times U_m^{n_i^t}$ the product control sets. Let $X \subset \mathbb{R}^p$ be a state set, the same for all nodes.

A transition function $f^{n_i^t}(\cdot, \cdot) : X \times U^{n_i^t} \mapsto X$ is associated with each node n_i^t . The state equations are given by

$$x(n_i^t) = f^{a(n_i^t)}(x(a(n_i^t)), u(a(n_i^t))), \quad (3.10.2)$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, t = 1, \dots, T, \quad (3.10.3)$$

where $x(n_i^t) \in X$ for any node n_i^t , $t = 0, 1, \dots, T$.

At a node n_i^t , the reward to player j is given by $\phi_j^{n_i^t}(x(n_i^t), u(n_i^t))$, if the game does not terminate at this node; otherwise, it is given by $\Phi_j^{n_i^t}(x(n_i^t))$. In particular, at $t = T$, the reward to player j is $\Phi_j^{n_i^T}(x(n_i^T))$. The information that a node n_i^t is terminal is revealed before the players choose their controls at that node, but not at $a(n_i^t)$. Observe that in any non-terminal node, the reward to a player depends on the state and the controls of all players.

Remind that ρ_j ($0 < \rho_j < 1$) is the discount factor of player $j \in M$ and suppose this players maximizes her expected discounted stream of payoffs over the planning horizon. The state equations and the reward functions define the following multistage game, where we let

$$\mathbf{x} = \{x(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T\},$$

$$\mathbf{u} = \{u(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, \dots, T - 1\},$$

and $J_j(\mathbf{x}, \mathbf{u})$ be the payoff to player j , that is,

$$\begin{aligned} J_j(\mathbf{x}, \mathbf{u}) &= \sum_{t=0}^{T-1} \rho_j^t \sum_{n_i^t \in \mathcal{N}^t} (\pi(n_i^t) - \tau(n_i^t)) \phi_j^{n_i^t}(x(n_i^t), u(n_i^t)) \\ &+ \sum_{t=1}^T \rho_j^t \sum_{n_i^t \in \mathcal{N}^t} \tau(n_i^t) \Phi_j^{n_i^t}(x(n_i^t)), \quad j \in M, \end{aligned} \quad (3.10.4)$$

s.t.

$$x(n_i^t) = f^{a(n_i^t)}(x(a(n_i^t)), u(a(n_i^t))), \quad (3.10.5)$$

$$u(a(n_i^t)) \in U^{a(n_i^t)}, \quad n_i^t \in \mathcal{N}^t, t = 1, \dots, T,$$

$$x(n^0) = x^0. \quad (3.10.6)$$

Let $\mathbf{u} = (\mathbf{u}_j : j \in M)$ be a profile of S -adapted strategies which satisfies Definition 3.1.

We may define a normal form game with payoff function of player j given by $W_j(\mathbf{u}, x^0) = J_j(\mathbf{x}, \mathbf{u})$, $j \in M$, where \mathbf{x} is a state trajectory begins with state x^0 and is uniquely defined from equations (3.10.5) given the strategy profile \mathbf{u} .

If the game is played noncooperatively, then the players will seek a Nash equilibrium in S -adapted strategies defined above (see Definition 3.2).

Observe that in an S -adapted information structure, the controls and the state equations are defined (indexed) over the set of nodes of the event tree, and not over t .

If the players cooperate, then they will maximize the sum of their discounted payoffs throughout the entire horizon, that is,

$$\max_{\mathbf{u}_j : j \in M} \sum_{j \in M} W_j(\mathbf{u}, x^0).$$

Denote the resulting vector of cooperative controls by \mathbf{u} found for the game starting from node n^0 and state x^0 , i.e.,

$$\mathbf{u}^* = \arg \max_{\mathbf{u}_j : j \in M} \sum_{j \in M} W_j(\mathbf{u}, x^0). \quad (3.10.7)$$

As before, $\mathbf{x}^* = \{x^*(n_i^t) : n_i^t \in \mathcal{N}^t, t = 0, 1, \dots, T\}$ is the cooperative state trajectory generated by the cooperative controls \mathbf{u}^* .

For later use, we also need to determine the subgame starting from state $x^*(n_i^t)$ at node $n_i^t \in \mathcal{N}^t$, $t = 1, \dots, T-1$. This subgame is defined as in the previous paragraphs. The payoff of player $j \in M$ in this subgame is given as follows:

$$\begin{aligned} W_j(\mathbf{u}(n_i^t), x^*(n_i^t)) &= \sum_{\theta=t}^{T-1} \rho_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} (\pi(n_i^\theta | n_i^t) - \tau(n_i^\theta | n_i^t)) \phi_j^{n_i^\theta}(x(n_i^\theta), u(n_i^\theta)) \\ &+ \sum_{\theta=t+1}^T \rho_j^{\theta-t} \sum_{n_i^\theta \in \mathcal{N}_\Gamma^\theta} \tau(n_i^\theta | n_i^t) \Phi_j^{n_i^\theta}(x(n_i^\theta)), \end{aligned} \quad (3.10.8)$$

where $\mathcal{N}_\Gamma^\theta = \mathcal{N}^\theta \cap \Gamma(n_i^t)$, $\mathbf{u}(n_i^t) = (\mathbf{u}_j(n_i^t) : j \in M)$ is an S -adapted strategy profile and $\mathbf{u}_j(n_i^t) = \{u_j(n_i^\theta) : n_i^\theta \in \Gamma(n_i^t)\}$ is an admissible S -adapted strategy of player j in the subgame starting from node n_i^t , with initial state $x^*(n_i^t)$. The term $\pi(n_k^\theta | n_i^t)$ is the conditional probability of passing through node n_k^θ if the subgame starts from node n_i^t . The term $\tau(n_k^\theta | n_i^t)$ is the conditional probability⁶ of terminating at node n_k^θ when the subgame starts from node n_i^t .

If the players play noncooperatively in the subgame beginning from state $x^*(n_i^t)$ and find the Nash equilibrium according to Definition 3.2, we denote it as $\mathbf{u}^N(x^*(n_i^t)) = (\mathbf{u}_j^N(x^*(n_i^t)) : j \in M)$. The S -adapted equilibrium payoff of player j is equal to $W_j(\mathbf{u}^N(x^*(n_i^t)), x^*(n_i^t))$.

If the players cooperate in the subgame starting from node n_i^t and in state $x^*(n_i^t)$, then they maximize the sum of their total discounted payoffs, i.e.,

$$\mathbf{u}^*(n_i^t) = \arg \max_{\mathbf{u}_j(n_i^t) : j \in M} \sum_{j \in M} W_j(\mathbf{u}(n_i^t), x^*(n_i^t)). \quad (3.10.9)$$

The payoff of player j in the cooperative subgame starting is $W_j(\mathbf{u}^*(n_i^t), x^*(n_i^t))$, $t = 1, \dots, T$.

If we think about the problem of uniqueness of the S -adaptive Nash equilibrium and the cooperative equilibrium, then Remark 3.2 is relevant.

⁶The conditional probabilities $\pi(n_k^\theta | n_i^t)$ and $\tau(n_k^\theta | n_i^t)$ can be calculated with the formulas: $\pi(n_k^\theta | n_i^t) = \pi(n_k^\theta) / (\pi(n_i^t) - \tau(n_i^t))$ and $\tau(n_k^\theta | n_i^t) = \tau(n_k^\theta) / (\pi(n_i^t) - \tau(n_i^t))$ respectively if $\pi(n_i^t) \neq 0$; otherwise, the subgame starting from node n_i^t cannot materialize.

§ 3.10.2 Node-consistent Shapley value in dynamic games played over event trees

Suppose that the players decide to cooperate during the game, and agree to use the Shapley Value⁷ as an allocation mechanism of their joint payoff $\sum_{j \in M} W_j(\mathbf{u}^*, x^0)$. Let $G \subset M$ be a coalition and denote by $v(G; x^*(n_l^t))$ the characteristic function⁸ value of this coalition in the subgame starting at node $n_l^t \in \mathcal{N}^t, t = 0, 1, \dots, T - 1$, in state $x^*(n_l^t)$. Recall that the characteristic function is defined by

$$v(G; x^*(n_l^t)) : 2^M \rightarrow \mathbb{R} \quad (3.10.10)$$

and $v(\emptyset; x^*(n_l^t)) = 0$.

To compute $v(G; x^*(n_l^t))$, we need to make an assumption on the behavior of left-out-players, i.e., $M \setminus G$. We adopt here the γ -characteristic function assumption, which states that when coalition G forms, the left-out-players would not take any particular joint action against it, but use their individually best reply strategies.⁹ Consequently, the γ -characteristic function value of coalition G is given by the S -adapted equilibrium outcome of G in the non-cooperative game between members of G maximizing their joint payoff (jointly play as one player), and non members playing individually, i.e., maximizing their individual payoffs.

Let $Y(x^*(n_l^t))$ be the set of imputations for the subgame starting at node n_l^t with state $x^*(n_l^t)$, that is,

$$Y(x^*(n_l^t)) = \left\{ (y_1(x^*(n_l^t)), \dots, y_m(x^*(n_l^t))) \text{ s.t. } y_j(x^*(n_l^t)) \geq v(\{j\}; x^*(n_l^t)) \right. \\ \left. \text{for any } j \in M \text{ and } \sum_{j \in M} y_j(x^*(n_l^t)) = v(M; x^*(n_l^t)) \right\}.$$

⁷Here, we choose the Shapley Value as a solution of the cooperative game, but the results of this section are also valid in the case of any other single-point optimality principle. If the solution of the cooperative game is a set of imputations, then it is necessary to make an assumption about its non-emptiness for any subgame and to use the results of § 3.2.

⁸We define the characteristic function $v(G; x^*(n_l^t))$ for any node $n_l^t \in \mathcal{N}^t$ based on the initial game assuming that the players implement the cooperative controls \mathbf{u}^* and moving along the cooperative trajectory \mathbf{x}^* . The values of characteristic function $v(G; x^*(n_l^t))$ also depend on the initial vertex n_l^t and on the cooperative controls, but we omit them to keep the notation simple.

⁹We retain here the γ -characteristic function, but our results would not change conceptually if we adopt a different definition, i.e., the α -, β - or δ -characteristic function. For a definition of these characteristic functions can be found in § 1.3.

The Shapley value $Sh(x^*(n_i^t)) = (Sh_1(x^*(n_i^t)), \dots, Sh_m(x^*(n_i^t)))$ is an imputation for the subgame starting at node n_i^t with state $x^*(n_i^t)$ and is given by

$$Sh_j(x^*(n_i^t)) = \sum_{G \subseteq M: G \ni j} \frac{(m-g)!(g-1)!}{m!} (v(G; x^*(n_i^t)) - v(G \setminus j; x^*(n_i^t))), \quad (3.10.11)$$

where $g = |G|$.

Definition 3.11. We call $(\{\beta_j(x^*(n_i^t))\}_{n_i^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M)$ an imputation distribution procedure (IDP) of the Shapley value $Sh(x^0) = (Sh_1(x^0), \dots, Sh_m(x^0))$ where

$$\beta_j(x^*(n_i^t)) = \begin{cases} \beta_j^c(x^0), & \text{if } n_i^t = n^0, \\ \beta_j^c(x^*(n_i^t)), & \text{if the process does not terminate at } n_i^t, \\ & t = 1, \dots, T-1, \\ \beta_j^r(x^*(n_i^t)), & \text{if the process terminates at } n_i^t, t = 1, \dots, T-1, \\ \beta_j^r(x^*(n_i^T)), & \text{if } n_i^t = n_i^T \end{cases} \quad (3.10.12)$$

for all $j \in M$ and the following conditions hold:¹⁰

$$Sh_j(x^0) = \sum_{t=0}^{T-1} \rho_j^t \sum_{n_i^t \in \mathcal{N}^t} (\pi(n_i^t) - \tau(n_i^t)) \beta_j^c(x^*(n_i^t)) + \quad (3.10.13)$$

$$+ \sum_{t=1}^T \rho_j^t \sum_{n_i^t \in \mathcal{N}^t} \tau(n_i^t) \beta_j^r(x^*(n_i^t)), \quad j \in M,$$

$$\sum_{j \in M} \beta_j^c(x^*(n_i^t)) = \sum_{j \in M} \phi_j^{n_i^t}(x^*(n_i^t), u^*(n_i^t)), \quad \text{if the process does not} \quad (3.10.14)$$

terminate at node $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$,

$$\sum_{j \in M} \beta_j^r(x^*(n_i^t)) = \sum_{j \in M} \Phi_j^{n_i^t}(x^*(n_i^t)), \quad \text{if the process terminates} \quad (3.10.15)$$

at node $n_i^t \in \mathcal{N}^t$, $t = 1, \dots, T$.

In the root node n^0 in which the stochastic process does not terminate, and nodes $n_i^T \in \mathcal{N}^T$ in which the process terminates, the components of the IDP are determined

¹⁰The indexes c and τ in the notation of β_j in equation (3.10.12) correspond to the words “continue” and “terminate”.

in a unique way because it is known if the game continues or terminates for sure. For an “intermediate” nodes $n_i^t \in \mathcal{N}^t$, $t = 1, \dots, T - 1$, we determine two possible IDP values for player j , corresponding to the two possibilities of the game continuing or terminating, namely, (i) $\beta_j^c(x^*(n_i^t))$ if the game does not terminate at node n_i^t , and (ii) $\beta_j^\tau(x^*(n_i^t))$ if it does. Equation (3.10.13) means that the component of the Shapley value corresponding to player $j \in M$ is equal to the discounted payments to player j in the whole game. Equations (3.10.14), (3.10.15) are the conditions for “admissibility” of an imputation distribution procedure, i.e., the sum of payments to the players in any node is equal to the sum of payoffs that they can obtain in this node using the cooperative controls $u^*(n_i^t)$.

Definition 3.12. *The Shapley value $Sh(x^0)$ and corresponding imputation distribution procedure $(\{\beta_j(x^*(n_i^t))\}_{n_i^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M)$ determined by (3.10.12) are called node consistent in the whole game if for any node $n_i^t \in \mathcal{N}^t$, $t = 0, \dots, T$, the following condition is satisfied:*

$$Sh_j(x^0) = \sum_{\theta=0}^{t-1} \rho_j^\theta \sum_{n_k^\theta \in \mathcal{N}^\theta} \{(\pi(n_k^\theta) - \tau(n_k^\theta)) \beta_j^c(x^*(n_k^\theta)) + \tau(n_k^\theta) \beta_j^\tau(x^*(n_k^\theta))\} \\ + \rho_j^t \sum_{n_k^t \in \mathcal{N}^t} \{(\pi(n_k^t) - \tau(n_k^t)) Sh_j(x^*(n_k^t)) + \tau(n_k^t) Sh_j^\tau(x^*(n_k^t))\}, \quad (3.10.16)$$

where $Sh_j(x^*(n_k^t))$ is the j th component of the Shapley value calculated for the subgame starting from node n_k^t in state $x^*(n_k^t)$ when the stochastic process does not terminate at node n_k^t , and $Sh_j^\tau(x^*(n_k^t))$ is the j th component of the Shapley value calculated for one-period game realizing at node n_k^t with state $x^*(n_k^t)$ when the stochastic process terminates at this node.¹¹

If the payoffs in the nodes are allocated according to the imputation distribution procedure, then node-consistency of the Shapley value $Sh(x^0)$ means that one can define a feasible distribution procedure under which the continuation values at every node are also the Shapley values of the continuation game.

¹¹We notice that if the game terminates at node n_k^t , the players does not choose any controls at this node and the components of the Shapley value $Sh_j^\tau(x^*(n_k^t))$ are equal to the corresponding values of their payoff functions $\Phi_j^{n_k^t}(x^*(n_k^t))$.

Theorem 3.7. *The Shapley value $Sh(x^0)$ is node consistent if the corresponding imputation distribution procedure determined by (3.10.12) satisfies the following conditions for $t = 0, \dots, T - 1$:*

$$\begin{aligned} \beta_j^c(x^*(n_l^t)) = Sh_j(x^*(n_l^t)) - \rho_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \{ & (\pi(n_k^{t+1}|n_l^t) - \tau(n_k^{t+1}|n_l^t)) Sh_j(x^*(n_k^{t+1})) \\ & + \tau(n_k^{t+1}|n_l^t) \Phi_j^{n_k^{t+1}}(x^*(n_k^{t+1})) \}, \end{aligned} \quad (3.10.17)$$

and for $t = 1, \dots, T$:

$$\beta_j^r(x^*(n_l^t)) = \Phi_j^{n_l^t}(x^*(n_l^t)). \quad (3.10.18)$$

Proof. At first, we prove that $\beta_j(x^*(n_l^t))$ determined by (3.10.17) and (3.10.18) is an imputation distribution procedure of the Shapley value $Sh(x^0)$, i.e., we show that conditions (3.10.13)–(3.10.15) are satisfied. Substituting the expressions of $\beta_j^c(x^*(n_l^t))$ and $\beta_j^r(x^*(n_l^t))$ given by (3.10.17) and (3.10.18) into the right-hand-side term of equation (3.10.13) and taking into account that

$$\pi(n_k^{t+1}|n_l^t) - \tau(n_k^{t+1}|n_l^t) = \frac{\pi(n_k^{t+1}) - \tau(n_k^{t+1})}{\pi(n_l^t) - \tau(n_l^t)}, \quad (3.10.19)$$

we obtain

$$\begin{aligned} & \sum_{t=0}^{T-1} \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} (\pi(n_l^t) - \tau(n_l^t)) \left[Sh_j(x^*(n_l^t)) - \rho_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \{ (\pi(n_k^{t+1}|n_l^t) - \tau(n_k^{t+1}|n_l^t)) \cdot \right. \\ & \cdot Sh_j(x^*(n_k^{t+1})) + \tau(n_k^{t+1}|n_l^t) \Phi_j^{n_k^{t+1}}(x^*(n_k^{t+1})) \} \left. \right] + \sum_{t=1}^T \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \tau(n_l^t) \Phi_j^{n_l^t}(x^*(n_l^t)) \\ & = (\pi(n^0) - \tau(n^0)) Sh_j(x^0) - \rho_j (\pi(n^0) - \tau(n^0)) \sum_{n_k^1 \in \mathcal{S}(n^0)} \{ (\pi(n_k^1|n^0) - \tau(n_k^1|n^0)) \cdot \\ & \cdot Sh_j(x^*(n_k^1)) + \tau(n_k^1|n^0) \Phi_j^{n_k^1}(x^*(n_k^1)) \} + \rho_j \sum_{n_k^1 \in \mathcal{N}^1} \tau(n_k^1) \Phi_j^{n_k^1}(x^*(n_k^1)) + \dots \\ & + \rho_j^{T-1} \sum_{n_l^{T-1} \in \mathcal{N}^{T-1}} (\pi(n_l^{T-1}) - \tau(n_l^{T-1})) \left[Sh_j(x^*(n_l^{T-1})) \right. \\ & - \rho_j \sum_{n_k^T \in \mathcal{S}(n_l^{T-1})} \{ (\pi(n_k^T|n_l^{T-1}) - \tau(n_k^T|n_l^{T-1})) Sh_j(x^*(n_k^T)) \\ & \left. + \tau(n_k^T|n_l^{T-1}) \Phi_j^{n_k^T}(x^*(n_k^T)) \} \right] + \rho_j^T \sum_{n_l^T \in \mathcal{N}^T} \tau(n_l^T) \Phi_j^{n_l^T}(x^*(n_l^T)) = Sh_j(x^0). \end{aligned}$$

We also use the condition that $Sh_j(x^*(n_k^T)) = \Phi_j^{n_k^T}(x^*(n_k^T))$ for any node $n_k^T \in \mathcal{N}^T$.

Now we show that conditions (3.10.14) and (3.10.15) are satisfied if we use the imputation distribution procedure defined by equations (3.10.17) and (3.10.18). For terminal nodes, condition (3.10.15) is satisfied because the Shapley value is an imputation and satisfies the property of efficiency. Further, we verify if condition (3.10.14) is true for the root node and any intermediate node $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T - 1$:

$$\begin{aligned} & \sum_{j \in M} \beta_j^c(x^*(n_l^t)) = \\ & = \sum_{j \in M} Sh_j(x^*(n_l^t)) - \sum_{j \in M} \rho_j \sum_{n_k^{t+1} \in \mathcal{S}(n_l^t)} \left\{ (\pi(n_k^{t+1}|n_l^t) - \tau(n_k^{t+1}|n_l^t)) Sh_j(x^*(n_k^{t+1})) \right. \\ & \left. + \tau(n_k^{t+1}|n_l^t) \Phi_j^{n_k^{t+1}}(x^*(n_k^{t+1})) \right\} \\ & = v(M; x^*(n_l^t)) - \left\{ \sum_{j \in M} W_j(\mathbf{u}^*(n_l^t), x^*(n_l^t)) - \sum_{j \in M} \phi_j^{n_l^t}(x^*(n_l^t), u^*(n_l^t)) \right\}, \end{aligned}$$

and given that

$$v(M; x^*(n_l^t)) = \sum_{j \in M} W_j(\mathbf{u}^*(n_l^t), x^*(n_l^t)),$$

and following the construction of the characteristic function, we obtain (3.10.14).

Finally, we prove that the Shapley value and the corresponding imputation distribution procedure are node consistent, i.e., they satisfy equation (3.10.16). It can be obtained when we substitute $\beta_j^c(x^*(n_k^\theta))$ from (3.10.17) and $\beta_j^r(x^*(n_k^\theta))$ from (3.10.18) into (3.10.16) and taking into account that $\pi(n^0) = 1$, $\tau(n^0) = 0$ and equation (3.10.19) for any $n_k^{t+1} \in \mathcal{N}^t$, $t = 0, \dots, T - 1$. Indeed, we have

$$\begin{aligned} & \sum_{\theta=0}^{t-1} \rho_j^\theta \sum_{n_l^\theta \in \mathcal{N}^\theta} \left\{ (\pi(n_l^\theta) - \tau(n_l^\theta)) Sh_j(x^*(n_l^\theta)) \right. \\ & - \rho_j \sum_{n_k^{\theta+1} \in \mathcal{S}(n_l^\theta)} [(\pi(n_k^{\theta+1}) - \tau(n_k^{\theta+1})) Sh_j(x^*(n_k^{\theta+1})) \\ & \left. + \tau(n_k^{\theta+1}) \Phi_j^{n_k^{\theta+1}}(x^*(n_k^{\theta+1})) \right\} + \rho_j^t \sum_{n_k^t \in \mathcal{N}^t} \left\{ (\pi(n_k^t) - \tau(n_k^t)) Sh_j(x^*(n_k^t)) \right. \\ & \left. + \tau(n_k^t) \Phi_j^{n_k^t}(x^*(n_k^t)) \right\} = Sh_j(x^0). \end{aligned}$$

This proves node consistency of the Shapley value when the corresponding IDP is defined by equations (3.10.17) and (3.10.18). \square

The implementation of the imputation distribution procedure

$$\left(\left\{ \beta_j \left(x^* \left(n_l^t \right) \right) \right\}_{n_l^t \in \mathcal{N}^t, t=0, \dots, T} : j \in M \right)$$

requires the definition of side payments between the players according to the following rule:

- if the game does not terminate at node $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T - 1$, then

$$\omega_j(n_l^t, x^*(n_l^t)) = \beta_j^c(x^*(n_l^t)) - \phi_j^{n_l^t}(x^*(n_l^t), u^*(n_l^t)), \quad (3.10.20)$$

- if the game terminates at node $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T$, then

$$\omega_j(n_l^t, x^*(n_l^t)) = 0, \quad (3.10.21)$$

where $\omega_j(n_l^t, x^*(n_l^t))$ is the transfer payment that player j does in node n_l^t over the cooperative state trajectory $x^*(n_l^t)$, such that $\sum_{j \in M} \omega_j(n_l^t, x^*(n_l^t)) = 0$ for any node n_l^t over $x^*(n_l^t)$. Clearly, $\omega_j(n_l^t, x^*(n_l^t))$ from (3.10.20) can assume any sign depending on the sign of the difference in the right-hand sides of (3.10.20).

§ 3.10.3 Necessary conditions for the S -adapted equilibrium

To find the values of the γ -characteristic function we need to find the S -adapted equilibrium in the game played over the event tree with unknown terminal time. In this section, we formulate the necessary conditions for the S -adapted equilibrium, in the form of Pontryagin's maximum principle.

First, rewrite the payoff of player j in the whole game given by (3.10.4) in another way, that is,

$$\begin{aligned} J_j(\mathbf{x}, \mathbf{u}) &= \phi_j^{n^0}(x^0, u(n^0)) \\ &+ \sum_{t=1}^{T-1} \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \left\{ (\pi(n_l^t) - \tau(n_l^t)) \phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) + \tau(n_l^t) \Phi_j^{n_l^t}(x(n_l^t)) \right\} \\ &+ \rho_j^T \sum_{n_l^T \in \mathcal{N}^T} \tau(n_l^T) \Phi_j^{n_l^T}(x(n_l^T)), \quad j \in M, \end{aligned}$$

where $\tau(n_l^T) = \pi(n_l^T)$ for any node $n_l^T \in \mathcal{N}^T$.

For each player $j \in M$, we form the Lagrangian

$$\begin{aligned}
\mathcal{L} &= \phi_j^{n_0}(x^0, u(x^0)) \tag{3.10.22} \\
&+ \sum_{t=1}^{T-1} \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \left\{ (\pi(n_l^t) - \tau(n_l^t)) \phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) + \tau(n_l^t) \Phi_j^{n_l^t}(x(n_l^t)) \right. \\
&+ \left. \pi(n_l^t) \lambda_j(n_l^t) (f^{a(n_l^t)}(x(a(n_l^t)), u(a(n_l^t))) - x(n_l^t)) \right\} \\
&+ \rho_j^T \sum_{n_l^T \in \mathcal{N}^T} \tau(n_l^T) \left\{ \Phi_j^{n_l^T}(x(n_l^T)) \right. \\
&+ \left. \lambda_j(n_l^T) (f^{a(n_l^T)}(x(a(n_l^T)), u(a(n_l^T))) - x(n_l^T)) \right\} = \\
&= \phi_j^{n_0}(x^0, u(x^0)) \\
&+ \sum_{t=1}^{T-1} \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \pi(n_l^t) \left\{ \frac{\pi(n_l^t) - \tau(n_l^t)}{\pi(n_l^t)} \phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) + \frac{\tau(n_l^t)}{\pi(n_l^t)} \Phi_j^{n_l^t}(x(n_l^t)) \right. \\
&+ \left. \lambda_j(n_l^t) (f^{a(n_l^t)}(x(a(n_l^t)), u(a(n_l^t))) - x(n_l^t)) \right\} \\
&+ \rho_j^T \sum_{n_l^T \in \mathcal{N}^T} \pi(n_l^T) \left\{ \Phi_j^{n_l^T}(x(n_l^T)) \right. \\
&+ \left. \lambda_j(n_l^T) (f^{a(n_l^T)}(x(a(n_l^T)), u(a(n_l^T))) - x(n_l^T)) \right\},
\end{aligned}$$

where $\pi(n_l^T) = \tau(n_l^T)$ for any $n_l^T \in \mathcal{N}^T$. For each player $j \in M$, we introduce a costate variable $\lambda_j(n_l^t)$ indexed over the set of nodes of periods $t = 1, \dots, T$. The costate variable $\lambda_j(n_l^t)$ has the same dimension as $x(n_l^t)$. Define the pre-Hamiltonian function for any player $j \in M$ and any node $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T - 1$ by

$$\begin{aligned}
H_j^{n_l^t}(x(n_l^t), u(n_l^t), \lambda_j(\mathcal{S}(n_l^t))) &= \frac{\pi(n_l^t) - \tau(n_l^t)}{\pi(n_l^t)} \phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) + \frac{\tau(n_l^t)}{\pi(n_l^t)} \Phi_j^{n_l^t}(x(n_l^t)) \\
&+ \rho_j \sum_{\nu \in \mathcal{S}(n_l^t)} \frac{\pi(\nu)}{\pi(n_l^t)} \lambda_j(\nu) f^{n_l^t}(x(n_l^t), u(n_l^t)), \tag{3.10.23}
\end{aligned}$$

where $\lambda_j(\mathcal{S}(n_l^t))$ is the collection of $\lambda_j(\nu)$, where $\nu \in \mathcal{S}(n_l^t)$.

Theorem 3.8. *Let \mathbf{u}^N be an S-adapted equilibrium in the game played over an event tree with initial state x^0 , generating the state trajectory $\mathbf{x}^N = \{x^N(n_l^t) :$*

$n_l^t \in \mathcal{N}^t, t = 1, \dots, T\}$. Then, for each player $j \in M$ there exist costate variables $\lambda_j(n_l^t), n_l^t \in \mathcal{N}^t, t = 1, \dots, T$ such that the following conditions hold for any $j \in M$, $u_j(n_l^t) = u_j^N(n_l^t), x(n_l^t) = x^N(n_l^t)$:

$$0 = \frac{\partial H_j^{n_l^t}}{\partial u_j(n_l^t)}, \quad n_l^t \in \mathcal{N}^t, \quad t = 0, \dots, T-1, \quad (3.10.24)$$

$$\lambda_j(n_l^t) = \frac{\partial H_j^{n_l^t}}{\partial x(n_l^t)}, \quad n_l^t \in \mathcal{N}^t, \quad t = 1, \dots, T-1 \quad (3.10.25)$$

$$\lambda_j(n_l^T) = \frac{\partial \Phi_j^{n_l^T}(x(n_l^T))}{\partial x(n_l^T)}, \quad n_l^T \in \mathcal{N}^T. \quad (3.10.26)$$

Proof. Consider the expression of Lagrangian (3.10.22) and rewrite it in the following form:

$$\begin{aligned} \mathcal{L} = & \sum_{t=0}^{T-1} \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \pi(n_l^t) \left\{ H_j^{n_l^t}(x(n_l^t), u(n_l^t), \lambda_j(\mathcal{S}(n_l^t))) - \lambda_j(n_l^t)x(n_l^t) \right\} \\ & + \rho_j^T \sum_{n_l^T \in \mathcal{N}^T} \pi(n_l^T) \left\{ \Phi_j^{n_l^T}(x(n_l^T)) - \lambda_j(n_l^T)x(n_l^T) \right\}. \end{aligned}$$

Taking into account (3.10.23), equations (3.10.24)-(3.10.26) are obtained by equating to zero the partial derivatives of the Lagrangian with respect to $x(n_l^t)$ and $u(n_l^t)$. \square

§ 3.10.4 Game-theoretical model of pollution control with random termination time

We illustrate the procedure for constructing a node-consistent Shapley value with an example of pollution control, which is a three-player stochastic version of the deterministic model in [98]. A description of the game with four periods $t = 0, \dots, 3$ can be found in the section § 3.5. State equations have the same form. The level of pollution absorption $\delta(n_l^t)$ at node n_l^t can take two possible values, that is, $\delta(n_l^t) \in \{\underline{\delta}, \bar{\delta}\}$, with $\underline{\delta} < \bar{\delta}$. The event tree is depicted in Figure 3.8. Let nodes $n_2^1, n_2^2, n_4^2, n_2^3, n_4^3, n_6^3, n_8^3$ correspond to the low level of pollution reduction $\underline{\delta}$, and nodes $n_1^1, n_1^2, n_3^2, n_1^3, n_3^3, n_5^3, n_7^3$ corresponds to the high level of pollution reduction $\bar{\delta}$.

The damage cost is an increasing convex function in the pollution stock having the quadratic form $D_j(x(n_l^t)) = \alpha_j x^2(n_l^t), j \in M$, where α_j is a strictly positive parameter. The cost of emissions is also given by a quadratic function $C_j(u_j(n_l^t)) =$

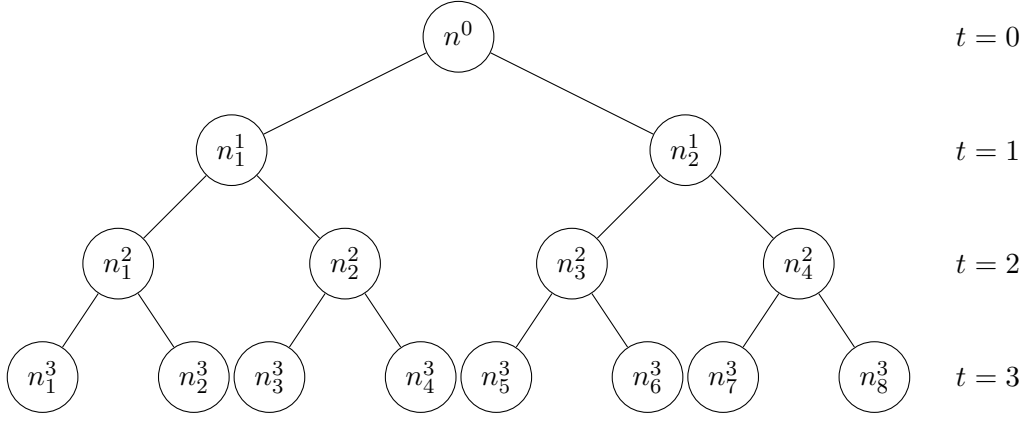


Figure 3.8: Event tree graph.

$\frac{\gamma_j}{2} (u_j(n_l^t) - e)^2$, where e and γ_j are strictly positive constants. The reward functions are given by $\phi_j^{n_l^t}(x(n_l^t), u(n_l^t)) = C_j(u_j(n_l^t)) + D_j(x(n_l^t))$ for any node $n_l^t \in \mathcal{N}^t$, $t = 0, \dots, T-1$, and by $\Phi_j^{n_l^t}(x(n_l^t)) = D_j(x(n_l^t))$ for any node $n_l^t \in \mathcal{N}^t$, $t = 1, \dots, T$.

The total discounted cost $J_j(\mathbf{x}, \mathbf{u})$ to be minimized by player $j \in M$ is given by

$$\sum_{t=0}^2 \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} (\pi(n_l^t) - \tau(n_l^t)) (C_j(u_j(n_l^t)) + D_j(x(n_l^t))) + \sum_{t=1}^3 \rho_j^t \sum_{n_l^t \in \mathcal{N}^t} \tau(n_l^t) D_j(x(n_l^t)),$$

where $\mathbf{x} = \{x(n_l^t)\}$ and $\mathbf{u} = \{u_j(n_l^t)\}$, $n_l^T \in \mathcal{N}^T$, subject to (3.3.1), given initial stock $x_0 = 0$ and constraints: $u_j(n_l^t) \in [0, e]$ for any player $j \in M$ and any node $n_l^t \in \mathcal{N}^t$, $t = 0, 1, 2$.

We use the following parameter values for the numerical simulation:

$$\alpha_1 = 0.1, \alpha_2 = 0.2, \alpha_3 = 0.3, \gamma_1 = 0.9, \gamma_2 = 0.8, \gamma_3 = 0.7,$$

$$\underline{\delta} = 0.45, \bar{\delta} = 0.8, e = 25, \lambda_1 = \lambda_2 = \lambda_3 = 0.9,$$

$$\pi(n_1^1) = 0.6, \pi(n_2^1) = 0.4, \pi(n_1^2) = 0.3, \pi(n_2^2) = \pi(n_3^2) = 0.25, \pi(n_4^2) = 0.1,$$

$$\pi(n_1^3) = 0.05, \pi(n_2^3) = 0.15, \pi(n_3^3) = 0.05, \pi(n_4^3) = 0.15,$$

$$\pi(n_5^3) = 0.05, \pi(n_6^3) = 0.15, \pi(n_7^3) = 0.025, \pi(n_8^3) = 0.025,$$

$$\tau(n_1^1) = \tau(n_2^1) = 0.05, \tau(n_1^2) = 0.1, \tau(n_2^2) = \tau(n_3^2) = \tau(n_4^2) = 0.05,$$

$$\tau(n_1^3) = 0.05, \tau(n_2^3) = 0.15, \tau(n_3^3) = 0.05, \tau(n_4^3) = 0.15,$$

$$\tau(n_5^3) = 0.05, \tau(n_6^3) = 0.15, \tau(n_7^3) = \tau(n_8^3) = 0.025.$$

Using (3.10.7) and (3.10.9), we compute the cooperative controls for the game

and any subgame. Consequently, we obtain the following cooperative trajectory:¹²

$x^*(n^0)$	0.000
$x^*(n_1^1), x^*(n_2^1)$	47.106
$x^*(n_1^2), x^*(n_2^2)$	54.843
$x^*(n_3^2), x^*(n_4^2)$	68.085
$x^*(n_1^3), x^*(n_2^3)$	61.002
$x^*(n_3^3), x^*(n_4^3)$	74.622
$x^*(n_5^3), x^*(n_6^3)$	62.881
$x^*(n_7^3), x^*(n_8^3)$	79.790

The γ -characteristic function values are given in Table 3.12. It is easy to verify that the γ -characteristic function is subadditive for any subgame.

Time period	$t = 0$		$t = 1$		$t = 2$			
Node	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2	
$V(\{1\})$	125.635	117.503	135.606	82.087	107.905	101.619	135.337	
$V(\{2\})$	260.689	241.361	278.095	166.723	219.625	205.948	275.036	
$V(\{3\})$	409.197	374.297	430.416	255.002	336.795	314.146	420.966	
$V(\{1, 2\})$	375.811	351.253	405.044	245.272	322.236	303.808	404.320	
$V(\{1, 3\})$	508.782	473.494	545.609	329.056	432.681	407.230	542.561	
$V(\{2, 3\})$	631.725	588.827	678.576	409.931	538.771	507.561	675.823	
$V(\{1, 2, 3\})$	666.538	642.349	742.643	463.653	604.231	579.041	762.632	
	$t = 3$							
Node	n_1^3	n_2^3	n_3^3	n_4^3	n_5^3	n_6^3	n_7^3	n_8^3
$V(\{1\})$	37.212	37.212	55.684	55.684	39.540	39.540	63.664	63.664
$V(\{2\})$	74.424	74.424	111.369	111.369	79.080	79.080	127.329	127.329
$V(\{3\})$	111.636	111.636	167.053	167.053	118.620	118.620	190.993	190.993
$V(\{1, 2\})$	111.636	111.636	167.053	167.053	118.620	118.620	190.993	190.993
$V(\{1, 3\})$	148.847	148.847	222.737	222.737	158.160	158.160	254.658	254.658
$V(\{2, 3\})$	186.059	186.059	278.422	278.422	197.700	197.700	318.322	318.322
$V(\{1, 2, 3\})$	223.271	223.271	334.106	334.106	237.240	237.240	381.987	381.987

Table 3.12: Characteristic functions for the game and all subgames.

The Shapley values of the whole game and all subgames, which are computed using formula (3.10.11), are provided in Table 3.13. Finally, based on the Shapley value, we determine the node-consistent imputation distribution procedure using equations (3.10.17) and (3.10.18) in Theorem 3.7. The results, that is, the $\beta_j^c(x^*(n_i^t))$

¹²Calculations of this section have been made in Mathematica 10.0.

and $\beta_j^\tau(x^*(n_t^j))$, $j \in M$, are shown in Figure 3.9. We should mention that in the root node n^0 only β^c is determined and in the nodes from \mathcal{N}^T only β^τ is determined, while in all “intermediate” nodes, both β^c and β^τ are determined by equations (3.10.17) and (3.10.18).

Time period	$t = 0$			$t = 1$			$t = 2$	
Node	n^0	n_1^1	n_2^1	n_1^2	n_2^2	n_3^2	n_4^2	
Sh_1	89.267	91.856	106.915	70.704	90.871	89.524	115.862	
Sh_2	218.265	211.452	244.643	153.459	199.776	191.853	252.342	
Sh_3	359.005	339.040	391.085	239.490	313.584	297.664	394.428	

$t = 3$								
Node	n_1^3	n_2^3	n_3^3	n_4^3	n_5^3	n_6^3	n_7^3	n_8^3
Sh_1	37.212	37.212	55.684	55.684	39.540	39.540	63.664	63.664
Sh_2	74.424	74.424	111.369	111.369	79.080	79.080	127.329	127.329
Sh_3	111.636	111.636	167.053	167.053	118.620	118.620	190.993	190.993

Table 3.13: The Shapley values for the game and all subgames.

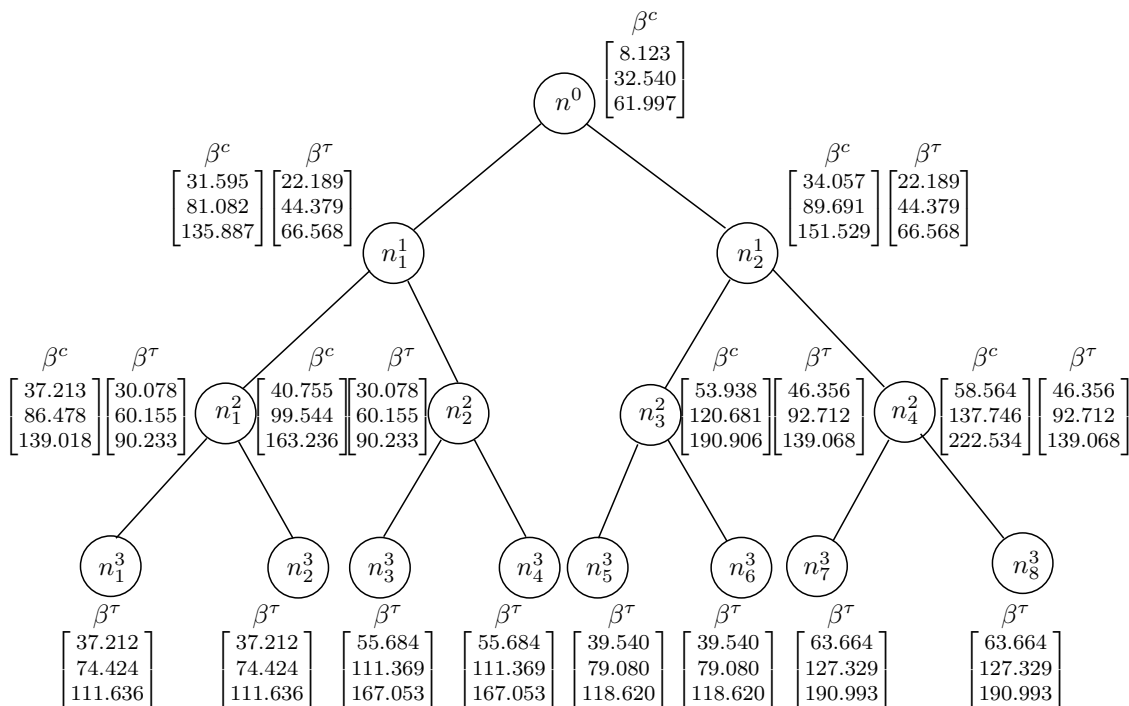


Figure 3.9: Imputation distribution procedure.

Let us show the decomposition of the component of the Shapley value of player 1 who is entitled to the costs of 89.267 in the whole game (see Table 3.13). Our imputation distribution procedure decomposes this amount over nodes, that is, it

is equal to the expected sum of discounted payments (the components of the imputation distribution procedure represented in Fig. 3.9) received in the subgame emanating from this node (here the whole game). Indeed, we have

$$\begin{aligned}
89.267 &= 8.123 + 0.9 \cdot \{(0.6 - 0.05) \cdot 31.595 + 0.05 \cdot 22.189 + \\
&\quad + (0.4 - 0.05) \cdot 34.057 + 0.05 \cdot 22.189\} \\
&+ 0.9^2 \cdot \{(0.3 - 0.1) \cdot 37.213 + 0.1 \cdot 30.078 + (0.25 - 0.05) \cdot 40.755 + \\
&\quad + 0.05 \cdot 30.078 + (0.25 - 0.05) \cdot 53.938 \\
&\quad + 0.05 \cdot 46.356 + (0.1 - 0.05) \cdot 58.564 + 0.05 \cdot 46.356\} \\
&+ 0.9^3 \cdot \{0.05 \cdot 37.212 + 0.15 \cdot 37.212 + 0.05 \cdot 55.684 + 0.15 \cdot 55.684 \\
&\quad + 0.05 \cdot 39.540 + 0.15 \cdot 39.540 + 0.025 \cdot 63.664 + 0.025 \cdot 63.664\}.
\end{aligned}$$

Similar calculations can be easily done for players 2 and 3. This decomposition of the Shapley value guarantees its node consistency. Consequently, players' payoffs in any subgame will be the corresponding components of the Shapley values calculated for these subgames.

Chapter 4

Applications of Stochastic Games

§ 4.1 Models of data transmission in wireless networks

In this section we present a method for modeling a complex telecommunication system by splitting it into small subsystems of simple topologies and solving the problem of optimizing the work of such small subsystems. The networks described have a simple structure and data transmission can be modeled by a stochastic game, in which data transfer devices (for example, routers) represent players. The state in the stochastic game is a vector which determines the presence of packages for transmission at each node of the network.

Game-theoretic modeling of data transmission process assumes that data transfer devices are players that performing some actions. Of course, this representation of technical devices is nominal and done in order to determine the rules of data transmission to maximize the network capacity.

We assume that in all games players are rational, i. e., they aim to maximize their expected payoffs.

§ 4.1.1 “Forwarder’s dilemma” game

We consider the first model of data transmission which is widely used in wireless network modeling, that is, “Forwarding dilemma” or “Forwarder’s dilemma”. Data transfer devices (hubs or routers) are presented in Fig. 4.1 by nodes 1 and 2. In game-theoretic modeling, it is assumed that players 1 and 2 are located at nodes 1 and 2, respectively. At the beginning of each time period, often called time slots in the literature, data packages of unit capacity (assumed for simplicity) appear at

these nodes with probabilities $a_1 \in [0, 1]$ and $a_2 \in [0, 1]$ respectively. Player 1's goal is to forward the package to node d_1 , which, as seen from the network representation, can be done only via player 2. Player 2's goal is to forward the package to node d_2 which can be done only via player 1. For successful data transfer, players must forward packages to the destination nodes via each other.

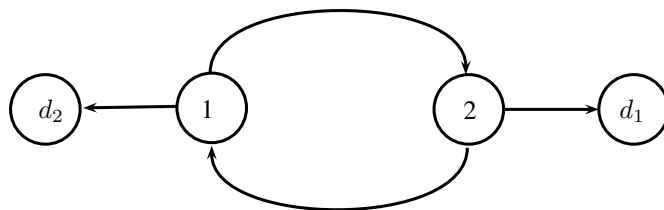


Figure 4.1: Data transmission scheme for Forwarder's dilemma.

We introduce the system of rewards for package transfer and costs for package transmission which is used to model the players' payoff functions:

- $c \in (0, 1)$ is the costs of the player for one data package transmission;
- 1 is the reward to the player for successful delivery of a package.

Note that when the player sends a package of the other player, he bears the costs of c . The reward for successful delivery of a package received by the player equals 1, and this reward is obtained by the player from whom this package originally appeared. If a player refuses to send a package to another player, this package remains with the player, from whom this package originally appeared, and at the next stage a new data package does not appear.

The state of a stochastic game determines whether there is a data package with players 1 and 2 or not. The set of states in the game is $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$, where the first entry of the vector denotes the number of packages belonging to player 1, the second entry is the number of packages belonging to player 2.

We consider all states of the game, define actions, payoff functions in these states, as well as transition probabilities.

- State $(0, 0)$.

The action of player 1 or 2 is W ("wait"). Players' payoffs are $(0, 0)$. Transition probabilities from state $(0, 0)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ are denoted

by $p_{1,1}$, $p_{1,2}$, $p_{1,3}$ and $p_{1,4}$ respectively. These probabilities form vector $p_1 = (p_{1,1}, p_{1,2}, p_{1,3}, p_{1,4})$ that is $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

- State $(0, 1)$.

The unique action of player 2 is T (“transmit”). Player 1 has two actions: F (“forward”) and D (“drop”). Players’ payoffs are represented in the matrix:

$$\begin{array}{c} T \\ F \left(\begin{array}{c} (-c, 1) \\ D \left(\begin{array}{c} (0, -c) \end{array} \right) \end{array} \right) \end{array} \quad (4.1.1)$$

Transition probabilities from state $(0, 1)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ are denoted by $p_{2,1}(F, T)$, $p_{2,2}(F, T)$, $p_{2,3}(F, T)$ and $p_{2,4}(F, T)$ respectively if action profile (F, T) has been played in the state. These probabilities form the vector of transition probabilities $p_2(F, T)$ that is $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

Transition probabilities from state $(0, 1)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ are denoted by $p_{2,1}(D, T)$, $p_{2,2}(D, T)$, $p_{2,3}(D, T)$ and $p_{2,4}(D, T)$ respectively if action profile (D, T) has been played in the state. These probabilities form the vector of transition probabilities $p_2(D, T)$ that is $(0, 1 - a_1, 0, a_1)$.

- State $(1, 0)$.

The unique action of player 1 is T (“transmit”). Player 2 has two actions: F (“forward”) and D (“drop”). Players’ payoffs are represented in the matrix:

$$\begin{array}{cc} F & D \\ T \left(\begin{array}{cc} (1, -c) & (-c, 0) \end{array} \right) \end{array} \quad (4.1.2)$$

The vector of transition probabilities from state $(1, 0)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ is denoted by $p_3(T, F) = (p_{3,1}(T, F), p_{3,2}(T, F), p_{3,3}(T, F), p_{3,4}(T, F))$ that is $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$ if action profile (T, F) has been played in the state.

Vector of transition probabilities from state $(1, 0)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ are denoted by $p_3(T, D)$ that is $(0, 0, 1 - a_2, a_2)$ if action profile (T, D) has been played in the state.

- State $(1, 1)$.

In the state, a package to transmit has appeared at any player at the beginning of the period. The strategies of player 1 or 2 are F (“forward”) and D (“drop”). So, if both players forward the packages of each other, then the packages are successfully delivered and each player obtains a payoff of 1 minus the costs for the other player’s package transmission. Thus, the players’ payoffs are

$$\begin{array}{cc} & \begin{array}{cc} F & D \end{array} \\ \begin{array}{c} F \\ D \end{array} & \begin{pmatrix} (1 - c, 1 - c) & (-c, 1) \\ (1, -c) & (0, 0) \end{pmatrix} \end{array} \quad (4.1.3)$$

In the state there are four pure action profiles. We write down transition probabilities from state $(1, 1)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ into Table 4.1.

Table 4.1: Transition probabilities from state $(1, 1)$ in “Forwarder’s dilemma”.

Action profile	$p_4(\cdot) = (p_{4,1}(\cdot), p_{4,2}(\cdot), p_{4,3}(\cdot), p_{4,4}(\cdot))$
(F, F)	$((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$
(F, D)	$(0, 0, 1 - a_2, a_2)$
(D, F)	$(0, 1 - a_1, 0, a_1)$
(D, D)	$(0, 0, 0, 1)$

We assume that players use stationary strategies which depend only on the current state and they are independent of time period and the history of the state.¹ We denote by $\eta_i \in H_i$ the strategy of player $i = 1, 2$ in game Γ . The strategy η_i is

¹The class of stationary strategies is a subclass of behavior strategies which depend on the history of the current period. We focus on the class of stationary strategies for modeling dynamic data transmission in wireless networks, since 1 ms is usually taken as the time period, and saving the history of the current period requires a lot of memory.

a function of a state which is a mapping from the set of states to the set of probability distributions over the set of pure actions of player i in this state. Stationary strategies (η_1, η_2) form a strategy profile in game Γ , which is denoted by η .

As a payoff of player i , we consider the expected discounted sum of step-by-step payoffs:

$$E_i(\eta) = \pi_0(\mathbb{I} - \delta\Pi(\eta))^{-1}K_i(\eta), \quad (4.1.4)$$

where π_0 is a vector of initial state distribution, \mathbb{I} is a unit matrix whose dimension is equal to the number of states of a game, δ is a discounting factor, $\Pi(\eta)$ is the matrix of transition probabilities whose rows are vectors p_1, \dots, p_4 , defined above, $K_i(\eta)$ is a payoff vector determining the players' payoffs in each state, provided that players realize the strategy profile η .

Definition 4.1. *Strategy profile $\eta^N = (\eta_1^N, \eta_2^N)$ is called the Nash equilibrium in stochastic game Γ if the inequalities hold:*

$$\begin{aligned} E_1(\eta_1^N, \eta_2^N) &\geq E_1(\eta_1, \eta_2^N) \text{ for } \forall \eta_1 \in H_1, \\ E_2(\eta_1^N, \eta_2^N) &\geq E_2(\eta_1^N, \eta_2) \text{ for } \forall \eta_2 \in H_2. \end{aligned}$$

Definition 4.2. *Strategy profile $\eta^* = (\eta_1^*, \eta_2^*)$ is called a cooperative solution in stochastic game Γ if the inequality*

$$\sum_{i=1}^2 E_i(\eta^*) \geq \sum_{i=1}^2 E_i(\eta)$$

is true for any $\eta \in H_1 \times H_2$.

We use the following theorem about the existence of the Nash equilibrium in a stochastic game with a finite set of strategies of each player.

Theorem 4.1. [91, 217] *Any non-zero sum stochastic game with many players with a discount factor $\delta \in (0, 1)$, finite sets of states and finite sets of the players' strategies in these states has the Nash equilibrium in stationary strategies.*

When we find the players' payoffs by formula (4.1.4), we estimate the expected number of data packages delivered by the player on condition of discounting. Calculation of the players' payoffs by formula (4.1.4) allows to evaluate the efficiency

of the data transfer protocol (rules for data transmission in the network) for a given network topology. The strategy profile in the game determines the protocol of data transfer, so the comparison of the protocol effectivenesses can be made by comparison of the players' payoffs in the profiles which determine these protocols.

You can use the following way to compare the cooperative and non-cooperative strategies of the players. To be specific, we find the strategy profiles that are optimal in case of cooperation and the absence of cooperation and make a conclusion about the effectiveness of cooperation by calculating the value of some coefficient if possible. As a measure of effectiveness of the cooperative players' behavior in comparison with the non-cooperative behavior in game Γ , we choose the ratio of the difference in the sum of the players' payoffs in cooperation and its absence to the sum of players' payoffs in the absence of cooperation, or the so-called *cost of cooperation rejection*:

$$\Delta(\Gamma) = \frac{\sum_{i=1}^2 E_i(\eta^*) - \sum_{i=1}^2 E_i(\eta^N)}{\sum_{i=1}^2 E_i(\eta^N)}, \quad (4.1.5)$$

where η^* is a stationary strategy profile maximizing sum $E_1(\eta) + E_2(\eta)$ over the set $\eta \in H_1 \times H_2$; η^N is the Nash equilibrium in stationary strategies². The coefficient (4.1.5) cannot be calculated if $E_1(\eta^N) + E_2(\eta^N) = 0$.

Remark 4.1. *The cost of cooperation rejection is a modified price of anarchy (see [117]) which is equal to the ratio of the total players' payoffs in the cooperative solution to their total payoffs in the worst Nash equilibrium. We have decided that the use of the cost of cooperation rejection of form (4.1.5), which is less than the price of anarchy per unit, is more convenient in the context of the problem under consideration.*

Remark 4.2. *In this paragraph, we do not determine a cooperative game, namely, we do not determine the characteristic function which allows to calculate the imputations or allocations of the jointly obtained payoffs. A detailed description of the construction of a cooperative stochastic game, including the construction of the characteristic function and imputations defining the allocations of the total payoff*

²If there exist more than one equilibrium, we arbitrarily choose the Nash equilibrium with the smallest total payoff

among players based on some axioms, as well as verification of their dynamic stability, can be found in the paper [163]. For the model of the data transfer in a wireless network, a cooperative game is constructed in [28], the Shapley value is calculated to redistribute the players' payoffs. The Shapley value distribution procedure is also found in [28] to achieve the subgame consistency of the Shapley value.

Because of complexity of the inverse matrix computing in a general form using formula (4.1.4) in the game "Forwarding dilemma", we find the Nash equilibrium and cooperative solution of the game for the given values of parameters: $a_1 = 0.8$, $a_2 = 0.2$, $c = 0.1$, $\delta = 0.99$.

We write down the players' payoffs calculated by the formula (4.1.4) into matrix (4.1.6), which rows (columns) correspond to the pure strategies of the first (second) player. The player's strategy consists of four entries, which are the player's actions in the states in the following order $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$.

$$\begin{array}{cccc}
 & WFFF & WTFD & WTDF & WTDD \\
 WFTF & (77.67, 12.33) & (64.23, 13.67) & (9.49, 18.39) & (-9.97, 20.30) \\
 WFTD & (78.58, 3.19) & (3.03, 0.12) & (43.20, 5.47) & (-0.34, 0.27) \\
 WDTF & (71.77, 10.79) & (62.16, 12.63) & (10.54, 17.84) & (-9.94, 20.20) \\
 WDTD & (79.70, -9.94) & (1.77, -0.22) & (77.80, -9.76) & (-0.21, 0.29)
 \end{array} \quad (4.1.6)$$

In the game "Forwarding dilemma", there exists the Nash equilibrium in pure stationary strategies $(WDTD, WTDD)$ which prescribes both players to "wait" in state $(0, 0)$; "drop" — to player 1 and "transmit" to player 2 in state $(0, 1)$; "transmit" — to player 1 and "drop" — to player 2 in state $(1, 0)$; "drop" to both players in state $(1, 1)$. Thus, players do not forward packets of each other in any state. The values of the players' payoffs let make a conclusion on how many their packages will be delivered. The payoffs of players 1 and 2 in the game when strategy profile $(WDTD, WTDD)$ is acted are -0.21 and 0.29 respectively, and the sum of their payoffs is 0.08 . If we take into account that the player's profit successful delivery of a data package is equal to one, then we can state that the implementation of the optimal behavior in the sense of Nash equilibrium will lead to system inactivity.

The cooperative solution in the game “Forwarder’s dilemma” is the strategy profile $(WFTF, WFFF)$ which prescribes both players to “wait” in state $(0, 0)$; “forward” — to player 1 and “transmit” to player 2 in state $(0, 1)$; “transmit” — to player 1 and “forward” — to player 2 in state $(1, 0)$; “forward” — to both players in state $(1, 1)$. With this strategy profile, players forward packages to each other in all states. Total payoffs of players 1 and 2 in the whole game when strategy profile $(WFTF, WFFF)$ are 77.67 and 12.33 respectively, and the total sum of payoffs is 90. Taking into account that the reward to the player for the successful package delivery is equal to one, then it becomes obvious that the system is active, i.e., when packages appear, they are always delivered to the final vertices.

The cost of cooperation rejection in the game “Forwarder’s dilemma” is calculated by the formula (4.1.5), it equals 1124. This value shows in how many times the players’ total payoff from transition to the cooperative solution exceeds their payoff in non-cooperative optimal behavior. The solution of the game shows that there is a need to coordinate the actions of transmitting devices in order to increase the capacity of the system with the topology shown in Fig. 4.1.

§ 4.1.2 “Joint package transmission” game

Consider the second data transfer scheme used in the modeling of wireless networks. Data transmission devices (players) are depicted in Fig. 4.2 by nodes 1 and 2. At vertex s , at the beginning of each time period a data package of unit capacity appears with probability $a_1 \in [0, 1]$. The goal of both players is to forward the package to node d . As one can recognize from the data transfer scheme, the package will be delivered to node d only if both players forward it. In case of successful delivery of a package, both players receive a reward of 1 minus the costs of package transmission. If player 1 forwards the package and player 2 does not forward it, only player 1 bears the costs of package transmission, and the undelivered package will be returned to node s . If player 1 does not forward the package, then none of the players bears the costs, and the package will be returned to node s .

The state of a stochastic game determines whether a data package exists or not at node s at the beginning of each period. The set of states in the game is $\{0, 1\}$,

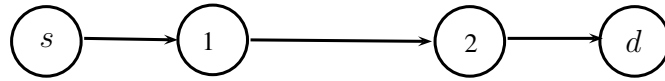


Figure 4.2: Data transmission scheme for “Joint package transmission game”.

equal to the number of packages at node s .

Consider each state of the game, determine players’ strategies, payoff functions in these states, as well as transition probabilities.

- State 0.

The action of any player 1 or 2 is W (“wait”). Players’ payoffs are $(0, 0)$. Transition probabilities from state 0 to states 0, 1 are denoted by $p'_{1,1}$, $p'_{1,2}$ respectively. These probabilities form the vector $p'_1 = (p'_{1,1}, p'_{1,2})$, that is $(1 - a_1, a_1)$.

- State 1.

In this state at node s at the beginning of the period the package has been appeared. The strategies of any player 1 or 2 are F (“forward”) and D (“drop”). Therefore, if both players forward the package, then it will be successfully delivered, and each player receives a payoff of 1 minus the costs of the other player’s package transmission. Players’ payoffs are

$$\begin{array}{cc}
 & \begin{array}{cc} F & D \end{array} \\
 \begin{array}{c} F \\ D \end{array} & \begin{pmatrix} (1 - c, 1 - c) & (-c, 0) \\ (0, 0) & (0, 0) \end{pmatrix}
 \end{array} \tag{4.1.7}$$

In this state there are four pure strategy profiles. Write down transition probabilities from state 1 to states 0 and 1 in Table 4.2.

Any player has two pure stationary strategies in the stochastic game. Let $\eta_{i,j}$ be the j th player’s strategy $i = 1, 2$, while $j = 1, 2$. The strategy $\eta_{i,1}$ is such that $\eta_{i,1}(1) = W$ and $\eta_{i,1}(2) = F$ for any i . The strategy $\eta_{i,2}$ is such that $\eta_{i,2}(1) = W$ and $\eta_{i,2}(2) = D$ for any i .

Table 4.2: Transition probabilities from state 1 in Joint package transmission game.

Profile	$p'_2(\cdot) = (p'_{2,1}(\cdot), p'_{2,2}(\cdot))$
(F, F)	$(1 - a_1, a_1)$
(F, D)	$(0, 1)$
(D, F)	$(0, 1)$
(D, D)	$(0, 1)$

Let π_0 be a vector $(\pi_{0,1}; 1 - \pi_{0,1})$. We calculate the players' payoffs in the profiles defined by these stationary strategies using the formula (4.1.4) and write them in the matrix:

$$\begin{matrix} & \eta_{2,1} & \eta_{2,2} \\ \eta_{1,1} & \left(\left(\frac{(1-c)((1-\delta)(1-\pi_{0,1})+a_1\delta)}{1-\delta}, \frac{(1-c)((1-\delta)(1-\pi_{0,1})+a_1\delta)}{1-\delta} \right) \right) & \left(-\frac{c((1-\delta)(1-\pi_{0,1})+a_1\delta)}{(1-\delta)(1-\delta(1-a_1))}, 0 \right) \\ \eta_{1,2} & (0, 0) & (0, 0) \end{matrix} \quad (4.1.8)$$

Theorem 4.2. *In Joint package transmission game, the profiles $(\eta_{1,1}, \eta_{2,1})$ and $(\eta_{1,2}, \eta_{2,2})$ are the Nash equilibria with players' payoffs $E_i(\eta_{1,1}, \eta_{2,1}) = (1 - c)((1 - \delta)(1 - \pi_{0,1}) + a_1\delta)/(1 - \delta)$, $i = 1, 2$ and $E_i(\eta_{1,1}, \eta_{2,1}) = 0$, $i = 1, 2$ respectively, for any vector of the initial distribution π_0 . Moreover, the profile $(\eta_{1,1}, \eta_{2,1})$ is a cooperative solution for any π_0 .*

Proof. Notice that the payoffs of both players in profile $(\eta_{1,1}, \eta_{2,1})$ are positive, and the payoff of player 1 in profile $(\eta_{1,1}, \eta_{2,2})$ is negative. In other profiles, the players' payoffs are null. First, we show that the profile $(\eta_{1,1}, \eta_{2,1})$ is the Nash equilibrium, and the sum of the players' payoffs with this profile is maximal. The players' payoffs are equal and positive, and the players can only reduce their payoffs if they individually deviate from this profile: player 1 will receive a zero payoff if he deviates to strategy $\eta_{1,2}$, and player 2 will also receive a zero payoff if he deviates to strategy $\eta_{2,2}$. Moreover, in this profile a sum of the players' payoffs is greater than a sum of their payoffs in any other profile. Thus, profile $(\eta_{1,1}, \eta_{2,1})$ is the Nash equilibrium and the cooperative solution.

Obviously, the profile $(\eta_{1,2}, \eta_{2,2})$ is also the Nash equilibrium with null players' payoffs. Deviating from strategy $\eta_{1,2}$ to strategy $\eta_{1,1}$ player 1 will also receive a

negative payoff of $-c((1 - \delta)(1 - \pi_{0,1}) + a_1\delta)/\{(1 - \delta)(1 - \delta(1 - a_1))\}$. Deviating to strategy $\eta_{2,1}$ player 2 will receive a null payoff. Therefore, the profile $(\eta_{1,2}, \eta_{2,2})$ is the Nash equilibrium. \square

Theorem 4.2 allows to assess the necessity of cooperative behavior of players in “Joint package transmission game”. As in the worst Nash equilibrium, the sum of the players’ payoffs is null, then the cost of cooperation rejection can not be calculated by formula (4.1.5). Certainly, if we may influence the players’ behaviour, the recommendation to implement the equilibrium $(\eta_{1,1}, \eta_{2,1})$ in the game will allow to realize the cooperative solution (socially optimal behavior). “Joint package transmission game” is a representative of the class of games in which the Nash equilibrium coincides with the cooperative solution, which is not frequent in real life.

Since the sum of players’ payoffs in the worst Nash equilibrium are null, the cost of cooperation rejection given by (4.1.5) is not defined.

§ 4.1.3 “Multiple access game”

Consider the third scheme of data transmission in wireless networks. Data transmission devices (players) are indicated in Fig. 4.3 by nodes 1 and 2. At the beginning of any time period a data package of unit capacity may appear in nodes 1 and 2 with probabilities $a_1 \in [0, 1]$ and $a_2 \in [0, 1]$ respectively. The goal of player $i = 1, 2$ is to deliver the package to node r_i . As one can see from the data transmission scheme, the package must pass the node of a unit capacity which is common for both players. In case when both players simultaneously forward the packages, these packages are returned to the initial nodes. For the successful delivery of the package, it is necessary that one of the players should transmit the package, and the other one should choose a strategy “wait”. If the package is successfully delivered, the player receives a reward of 1 minus costs of a package transmission (the costs equal $c \in (0, 1)$ as in the previous models).

The state of a stochastic game determines whether players 1 and 2 have got or not data packages at the beginning of each period. The set of states of the game is $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$, where the first (second) entry is a number of packages at node 1 (2).

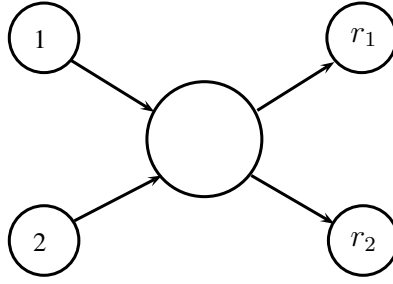


Figure 4.3: Data transmission scheme for Multiple access game.

Consider each state of the game, determine strategies, players' payoff functions in these states, as well as transition probabilities to the other states of the game.

- State $(0, 0)$.

An action of any player 1 or 2 is W ("wait"). Players' payoffs are $(0, 0)$. Transition probabilities from state 0 to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ are denoted by $p''_{1,1}$, $p''_{1,2}$, $p''_{1,3}$ and $p''_{1,4}$ respectively. These probabilities form the vector $p''_1 = (p''_{1,1}, p''_{1,2}, p''_{1,3}, p''_{1,4})$, that is $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

- State $(0, 1)$.

A unique action of player 1 is W ("wait"). A unique action of player 2 is T ("transmit"). In this state, player 2 transmits a package and successfully delivers it to the destination node r_2 . The players' payoffs are $(0, 1 - c)$.

Transition probabilities from state $(0, 1)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ are denoted by $p''_{2,1}$, $p''_{2,2}$, $p''_{2,3}$ and $p''_{2,4}$ respectively. These probabilities form the vector of transition probabilities $p''_2 = (p''_{2,1}, p''_{2,2}, p''_{2,3}, p''_{2,4})$, that is $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

- State $(1, 0)$.

A unique action of player 1 is T ("transmit"). A unique action of player 2 is W ("wait"). In this state, player 1 transmits a package and successfully delivers it to the destination node r_1 . The players' payoffs are $(1 - c, 0)$.

Transition probabilities from state $(1, 0)$ to the states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ form the vector of transition probabilities $p''_3 = (p''_{3,1}, p''_{3,2}, p''_{3,3}, p''_{3,4})$, that is $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.

- State $(1, 1)$.

In this state each player has got a package to transmit to the destination node. The actions of player 1 and 2 are T (“transmit”) and W (“wait”). The players’ payoffs are

$$\begin{array}{cc}
 & W & T \\
 \begin{array}{c} W \\ T \end{array} & \begin{pmatrix} (0, 0) & (0, 1 - c) \\ (1 - c, 0) & (-c, -c) \end{pmatrix} &
 \end{array} \tag{4.1.9}$$

Here we ignore the player’s costs for delay of data package transmission. By delaying we assume the fact that the package has not been transmitted in the current time period.

In this state, there are four pure stationary strategy profiles. Calculate the transition probabilities from state $(1, 1)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ and represent them in Table 4.3.

Table 4.3: Transition probabilities from state $(1, 1)$ in “Multiple access game”.

Profile	$p_4''(\cdot) = (p_{4,1}''(\cdot), p_{4,2}''(\cdot), p_{4,3}''(\cdot), p_{4,4}''(\cdot))$
(W, W)	$(0, 0, 0, 1)$
(W, T)	$(0, 0, 1 - a_2, a_2)$
(T, W)	$(0, 1 - a_1, 0, a_1)$
(T, T)	$((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$

In the stochastic game, player $i = 1, 2$ has got two pure stationary strategies: $\eta_{i,1}$ and $\eta_{i,2}$, represented in Table 4.4.

Table 4.4: Pure stationary strategies of players in “Multiple access game”.

State	Player 1		Player 2	
	$\eta_{1,1}$	$\eta_{1,2}$	$\eta_{2,1}$	$\eta_{2,2}$
$(0, 0)$	W	W	W	W
$(0, 1)$	W	W	T	T
$(1, 0)$	T	T	W	W
$(1, 1)$	W	T	W	T

Let π_0 be $(0.25, 0.25, 0.25, 0.25)$. We calculate the players' payoffs in the profiles which are formed by stationary strategies described in Table 4.4 using formula (4.1.4), and represent them by matrix:

$$\begin{array}{cc} & \eta_{2,1} & \eta_{2,2} \\ \eta_{1,1} & (E_1(\eta_{1,1}, \eta_{2,1}), E_2(\eta_{1,1}, \eta_{2,1})) & (E_1(\eta_{1,1}, \eta_{2,2}), E_2(\eta_{1,1}, \eta_{2,2})) \\ \eta_{1,2} & (E_1(\eta_{1,2}, \eta_{2,1}), E_2(\eta_{1,2}, \eta_{2,1})) & (E_1(\eta_{1,2}, \eta_{2,2}), E_2(\eta_{1,2}, \eta_{2,2})) \end{array}, \quad (4.1.10)$$

where

$$\begin{aligned} E_1(\eta_{1,1}, \eta_{2,1}) &= \frac{(1-c)(0.5(1-\delta) + a_1\delta(1.5-a_2))}{2(1-\delta(1-a_1a_2))}, \\ E_2(\eta_{1,1}, \eta_{2,1}) &= \frac{(1-c)(0.5(1-\delta) + a_2\delta(1.5-a_1))}{2(1-\delta(1-a_1a_2))}, \\ E_1(\eta_{1,1}, \eta_{2,2}) &= \frac{0.5(1-c)(0.5 + (1-a_1)(a_2-0.5)\delta^2 + (1.5a_1 - a_1a_2 - a_2)\delta)}{(1-\delta)(1-(1-a_1)a_2\delta)}, \\ E_2(\eta_{1,1}, \eta_{2,2}) &= \frac{(1-c)(0.5(1-\delta) + a_2\delta)}{1-\delta}, \\ E_1(\eta_{1,2}, \eta_{2,1}) &= \frac{(1-c)(0.5(1-\delta) + a_1\delta)}{1-\delta}, \\ E_2(\eta_{1,2}, \eta_{2,1}) &= \frac{0.5(1-c)(0.5 + (1-a_1)(a_2-0.5)\delta^2 + (1.5a_1 - a_1a_2 - a_2)\delta)}{(1-\delta)(1-(1-a_1)a_2\delta)}, \\ E_1(\eta_{1,2}, \eta_{2,2}) &= \frac{(1-\delta)(0.25 - 0.5c) + a_1\delta(1-c-a_2)}{1-\delta}, \\ E_2(\eta_{1,2}, \eta_{2,2}) &= \frac{(1-\delta)(0.25 - 0.5c) + a_2\delta(1-c-a_1)}{1-\delta}. \end{aligned}$$

Theorem 4.3. *Let π_0 be $(0.25, 0.25, 0.25, 0.25)$. In Multiple access game, both profiles $(\eta_{1,1}, \eta_{2,2})$ and $(\eta_{1,2}, \eta_{2,1})$ are the Nash equilibria with players' payoffs*

$$\begin{aligned} E_1(\eta_{1,1}, \eta_{2,2}) &= \frac{0.5(1-c)(0.5 + (1-a_1)(a_2-0.5)\delta^2 + (1.5a_1 - a_1a_2 - a_2)\delta)}{(1-\delta)(1-(1-a_1)a_2\delta)}, \\ E_2(\eta_{1,1}, \eta_{2,2}) &= \frac{(1-c)(a_2\delta + 0.5(1-\delta))}{1-\delta} \end{aligned}$$

and

$$\begin{aligned} E_1(\eta_{1,2}, \eta_{2,1}) &= \frac{(1-c)(a_1\delta + 0.5(1-\delta))}{1-\delta}, \\ E_2(\eta_{1,2}, \eta_{2,1}) &= \frac{0.5(1-c)(0.5 + (1-a_1)(a_2-0.5)\delta^2 + (1.5a_1 - a_1a_2 - a_2)\delta)}{(1-\delta)(1-(1-a_1)a_2\delta)}. \end{aligned}$$

The cooperative solution is profile $(\eta_{1,1}, \eta_{2,2})$ if $a_2 > a_1$, and profile $(\eta_{1,2}, \eta_{2,1})$ if $a_2 < a_1$. In case $a_1 = a_2$, both profiles $(\eta_{1,1}, \eta_{2,2})$, $(\eta_{1,2}, \eta_{2,1})$ are cooperative solutions.

Proof. Inequality $E_1(\eta_{1,1}, \eta_{2,1}) < E_1(\eta_{1,2}, \eta_{2,1})$ is true because difference $E_1(\eta_{1,1}, \eta_{2,1}) - E_1(\eta_{1,2}, \eta_{2,1})$ equals

$$\frac{(1-c) \{-0.5(1-\delta)^2 - a_1\delta(1-\delta)(0.5 + 2a_2) - 2a_1^2a_2\delta^2\}}{2(1-\delta(1-a_1a_2))(1-\delta)} < 0.$$

Similarly, we have $E_2(\eta_{1,1}, \eta_{2,1}) < E_2(\eta_{1,1}, \eta_{2,2})$ because the difference $E_2(\eta_{1,1}, \eta_{2,1}) - E_2(\eta_{1,1}, \eta_{2,2})$ is

$$\frac{(1-c) \{-0.5(1-\delta)^2 - a_2\delta(1-\delta)(0.5 + 2a_1) - 2a_2^2a_1\delta^2\}}{2(1-\delta(1-a_1a_2))(1-\delta)} < 0.$$

The expression $E_1(\eta_{1,1}, \eta_{2,2}) - E_1(\eta_{1,2}, \eta_{2,2})$ is positive if and only if

$$c > -\frac{\delta(1-a_1)(1-a_2)}{1-\delta+a_1\delta},$$

which is always true, since the expression on the right-hand side of the inequality is negative, and $c \in (0, 1)$.

Similarly, the difference $E_2(\eta_{1,2}, \eta_{2,1}) - E_2(\eta_{1,2}, \eta_{2,2})$ is positive if and only if

$$c > -\frac{\delta(1-a_2)(1-a_1)}{1-\delta+a_2\delta},$$

which is always true, since the expression on the right-hand side of the inequality is negative, and $c \in (0, 1)$.

Therefore, the profiles $(\eta_{1,2}, \eta_{2,1})$ and $(\eta_{1,1}, \eta_{2,2})$ are the Nash equilibria. The players' payoffs can be calculated using formula (4.1.4), the result is given in the theorem.

Now we verify which strategy profile is a cooperative solution. Obviously, the sum of the players' payoffs in the whole game with profiles $(\eta_{1,1}, \eta_{2,2})$ and $(\eta_{1,2}, \eta_{2,1})$ is greater than the sum of the payoffs with profiles $(\eta_{1,1}, \eta_{2,1})$ and $(\eta_{1,2}, \eta_{2,2})$. Now we find the conditions under which the cooperative solution is the profile $(\eta_{1,1}, \eta_{2,2})$, or the profile $(\eta_{1,2}, \eta_{2,1})$. For this, we calculate and transform the difference of the sums of the players' payoffs in these profiles:

$$\begin{aligned} E_1(\eta_{1,1}, \eta_{2,2}) + E_2(\eta_{1,1}, \eta_{2,2}) - (E_1(\eta_{1,2}, \eta_{2,1}) + E_2(\eta_{1,2}, \eta_{2,1})) &= \\ &= \frac{(a_2 - a_1)(1 - a_1)(1 - a_2)(\delta a_1 a_2 + 0.25(1 - \delta))\delta^2}{(1 - \delta)(1 - (1 - a_1)a_2\delta)(1 - (1 - a_2)a_1\delta)}. \end{aligned} \quad (4.1.11)$$

This difference is positive if and only if $a_2 > a_1$. In this case, the profile $(\eta_{1,1}, \eta_{2,2})$ is a cooperative solution. The difference (4.1.11) is negative if and only if $a_2 < a_1$, and in this case the profile $(\eta_{1,2}, \eta_{2,1})$ is a cooperative solution. If the difference (4.1.11) equals null ($a_1 = a_2$), both profiles $(\eta_{1,1}, \eta_{2,2})$, $(\eta_{1,2}, \eta_{2,1})$ are cooperative solutions. \square

From Theorem 4.3 it follows that the cooperative solution is also the Nash equilibrium. Thus, the players maximizing the total payoff act individually rational, which means that they implement the Nash equilibrium strategies.

But since in Multiple access game the cooperative solution is the Nash equilibrium with the greatest total payoff, then the cost of cooperation rejection (4.1.5) is null only in case when $a_1 = a_2$. This indicates that there is no need to coordinate the actions of transmitting devices at nodes 1 and 2, when $a_1 = a_2$. If $a_1 > a_2$ or $a_1 < a_2$, then the cost of cooperation rejection is greater than zero. In these cases, the coordination of the actions of transmitting devices increases the total payoff.

Remark 4.3. *We have considered three models of data transmission. For two of three games discussed above, it is possible to obtain theoretical results in a general form of a game, while for one game the solution of the game has been found for a given set of parameters. In two of three games, coordinating the actions of the transmitting devices will increase the players' payoffs, thereby increase the network capacity. In one game, the cooperative strategies of the players are individually rational (with the equal probability of the players' package appearance), so there is no need to implement control of such a network. Therefore, we conclude that before one starts managing the network participant's strategies, it is useful to build a mathematical model of data transmission to understand whether the management is necessary. But it may happen that even if the coordination of actions of network participants may increase the total payoff, this increase of the payoffs may be small in comparison with the costs of coordinating the players' actions. In order to understand whether there is an economic advantage in network management, we propose to calculate the cost of cooperation rejection, which allows to make a cost assessment of coordination of network participant's strategies.*

§ 4.1.4 “Multiple access game” with the lack of information

Unlike the previous paragraph, we assume that players do not have information about the presence of a package at the other player in the current time period, which means that player 1 does not distinguish in any time period in which state of the set $\{(0, 0), (0, 1)\}$ is realised, similarly for the states from the set $\{(1, 0), (1, 1)\}$. Player 2 does not distinguish the states of the set $\{(0, 0), (1, 0)\}$, as well as the states of the set $\{(0, 1), (1, 1)\}$.

Let players know the vector of initial distribution $\pi = (\pi_{(0,0)}, \pi_{(0,1)}, \pi_{(1,0)}, \pi_{(1,1)})$, where π_ω is the probability that state ω is realised in the first time period, and $\sum_{\omega \in \Omega} \pi_\omega = 1$.

Remark 4.4. *Given the problem statement, we may consider vector $\pi = ((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$ as the vector of initial distribution over the set of states. Taking the vector in a particular form, we assume that there were no data packages at the nodes of the network before the chance move.*

We describe the process of data transmission in such a network with the theory of stochastic games. To determine a stochastic game, it is necessary to specify the players' actions and payoff functions in each state and determine the class of strategies in which solution of the game will be found.

We will consider the class of stationary strategies which do not depend on time, but depend only on the state of the system (hereinafter, game). In case when player $i = 1, 2$ does not have a package for transmission, her unique action is to wait. Denote action “wait” by w_i . Since players do not have full information about the presence of a package at the other player, the actions of player 1 in the states of set $\{(1, 0), (1, 1)\}$ should be the same: w_1 (“wait”) and s_1 (“transmit a package”). Then the mixed action of player 1 in the states of the set $\{(1, 0), (1, 1)\}$ is the vector $(1 - \xi_1, \xi_1)$, where $\xi_1 \in [0, 1]$ is the probability of choosing action s_1 , or the probability of package transmission by player 1 in states $(1, 0), (1, 1)$. Thus, the stationary strategy of player 1 in the stochastic game is determined by probability $\xi_1 \in [0, 1]$. Similarly, we define the strategy of the second player. Player 2's actions in the states of the set $\{(0, 1), (1, 1)\}$ are w_2 (“wait”) and s_2 (“transmit a package”).

Then the mixed action of player 2 in the states of set $\{(0, 1), (1, 1)\}$ is the vector $(1 - \xi_2, \xi_2)$, where $\xi_2 \in [0, 1]$ is the probability of package transmission by player 2 in states $(0, 1), (1, 1)$. Thus, the stationary strategy of player 2 in stochastic game is determined by probability $\xi_2 \in [0, 1]$.

Now we consider all states of the game, determine the players' payoff functions in these states as functions of the strategies ξ_1 and ξ_2 of players 1 and 2 respectively, and transition probabilities in one time period from the current state to all possible states, which are generally the functions of the players' strategies:

- State $(0, 0)$.

The unique action of player $i = 1, 2$ is action w_i . Players' payoff functions are $K_1^{(0,0)}(\xi_1, \xi_2) = K_2^{(0,0)}(\xi_1, \xi_2) = 0$. Transition probabilities from state $(0, 0)$ to states $(0, 0), (0, 1), (1, 0), (1, 1)$ form vector $p_{(0,0)}$, that is

$$p_{(0,0)}(\xi_1, \xi_2) = ((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2).$$

- State $(0, 1)$.

The unique action of player 1 is w_1 . Player 2 has got two pure actions: s_2 and w_2 . In this state, player 2 chooses action s_2 with probability $\xi_2 \in [0, 1]$. Players' payoff functions are $K_1^{(0,1)}(\xi_1, \xi_2) = 0$, $K_2^{(0,1)}(\xi_1, \xi_2) = \xi_2(1 - c) + (1 - \xi_2)(-d)$. Transition probabilities from state $(0, 1)$ to states $(0, 0), (0, 1), (1, 0), (1, 1)$ form vector $p_{(0,1)}$, that is

$$p_{(0,1)}(\xi_1, \xi_2) = (\xi_2(1 - a_1)(1 - a_2), \xi_2(1 - a_1)a_2 + (1 - \xi_2)(1 - a_1), \\ \xi_2a_1(1 - a_2), \xi_2a_1a_2 + (1 - \xi_2)a_1).$$

- State $(1, 0)$.

Player 1 has got two pure actions: s_1 and w_1 . Player 1 chooses action s_1 with probability $\xi_1 \in [0, 1]$. The unique action of player 2 is w_2 . Players' payoff functions are $K_1^{(1,0)}(\xi_1, \xi_2) = \xi_1(1 - c) + (1 - \xi_1)(-d)$, $K_2^{(1,0)}(\xi_1, \xi_2) = 0$. Transition probabilities from state $(1, 0)$ to states $(0, 0), (0, 1), (1, 0), (1, 1)$ form vector $p_{(1,0)}$, that is

$$p_{(1,0)}(\xi_1, \xi_2) = (\xi_1(1 - a_1)(1 - a_2), \xi_1a_1(1 - a_2) + (1 - \xi_1)(1 - a_2), \\ \xi_1(1 - a_1)a_2, \xi_1a_1a_2 + (1 - \xi_1)a_2).$$

- State $(1, 1)$.

In this state, each player has got a package for transmission at the beginning of a time period. Player 1 has got two pure actions: s_1 and w_1 , player 2 also has got two pure actions: s_2 and w_2 . Player $i = 1, 2$ chooses action s_i with probability $\xi_i \in [0, 1]$. Players' payoff functions are $K_1^{(1,1)}(\xi_1, \xi_2) = \xi_1\xi_2(-c-d) + \xi_1(1-\xi_2)(1-c) + (1-\xi_1)(-d)$, $K_2^{(1,1)}(\xi_1, \xi_2) = \xi_1\xi_2(-c-d) + \xi_2(1-\xi_1)(1-c) + (1-\xi_2)(-d)$.

Transition probabilities from state $(1, 1)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ form vector $p_{(1,1)}$, that is

$$p_{(1,1)}(\xi_1, \xi_2) = (0, \xi_1(1-\xi_2)(1-a_1), (1-\xi_1)\xi_2(1-a_2), \\ (1-\xi_1)(1-\xi_2) + \xi_1\xi_2 + (1-\xi_1)\xi_2a_2 + \xi_1(1-\xi_2)a_1).$$

Stochastic game Γ of data transmission in the network with topology represented in Fig. 4.3, is determined by the set of states Ω , the set of players $N = \{1, 2\}$, the sets of strategies of any player in all states from set Ω , the players' payoff functions K_i^ω , $i \in N$ in all states $\omega \in \Omega$, and also by transition functions p_ω , which are probability distributions on the set of states Ω and are determined for any $\omega \in \Omega$.

As mentioned above, we will find the solutions of game Γ in stationary strategies. Stationary strategy η_i of player i in game Γ is determined by probability $\xi_i \in [0, 1]$; therefore, in order to avoid additional notation, we denote the player i 's stationary strategy in game Γ by ξ_i .

We calculate the mathematical expectation of player i 's payoff, $i = 1, 2$, in game Γ assuming that the vector of initial distribution π is known to players, using the formula:

$$E_i(\xi_1, \xi_2) = \pi(\mathbb{I} - \delta\Pi(\xi_1, \xi_2))^{-1}K_i(\xi_1, \xi_2),$$

where \mathbb{I} is an identity matrix of size 4×4 , $\delta \in (0, 1)$ is a discount factor, $\Pi(\xi_1, \xi_2)$ is a matrix of of size 4×4 whose entries are transition probabilities, i.e.,

$$\Pi(\xi_1, \xi_2) = \begin{pmatrix} p_{(0,0)}(\xi_1, \xi_2) \\ p_{(0,1)}(\xi_1, \xi_2) \\ p_{(1,0)}(\xi_1, \xi_2) \\ p_{(1,1)}(\xi_1, \xi_2) \end{pmatrix},$$

and vector $K_i(\xi_1, \xi_2)$ of player i 's payoff function in all states is

$$K_i(\xi_1, \xi_2) = \begin{pmatrix} K_i^{(0,0)}(\xi_1, \xi_2) \\ K_i^{(0,1)}(\xi_1, \xi_2) \\ K_i^{(1,0)}(\xi_1, \xi_2) \\ K_i^{(1,1)}(\xi_1, \xi_2) \end{pmatrix}.$$

Solutions of the game. Price of anarchy. We use noncooperative and cooperative approaches to find a solution of the considered game. By a noncooperative solution of a game, we mean the Nash equilibrium.

When we find a cooperative solution of the game, we make the assumption that the players cooperate in one coalition and maximize the total expected payoff of the coalition in game Γ . Moreover, like in the noncooperative game, players do not have perfect information about the presence of the package at the other player. We define a solution of the game under the cooperative approach as follows:

Definition 4.3. *Strategy profile $\tilde{\xi} = (\tilde{\xi}_1, \tilde{\xi}_2)$ is a cooperative solution of stochastic game Γ if the inequality is true:*

$$\sum_{i=1}^2 E_i(\tilde{\xi}) \geq \sum_{i=1}^2 E_i(\xi) \text{ for } \forall \xi \in [0, 1] \times [0, 1].$$

A cooperative solution determines the ‘‘social optimum’’ of a network with a given topology when the goal is to maximize the total system payoff or maximize the network capacity in case of its centralized control.

As a measure of benefit from the cooperative behavior of the players in comparison with the noncooperative behavior in game Γ , we use the *price of anarchy*, which is equal to the ratio of the total players' payoff with the cooperative solution to their total payoffs in the worst Nash equilibrium (the Nash equilibrium with the smallest total payoff):

$$PoA(\Gamma) = \frac{\sum_{i=1}^2 E_i(\tilde{\xi})}{\min_{\xi^* \in NE(\Gamma)} \sum_{i=1}^2 E_i(\xi^*)}, \quad (4.1.12)$$

where $NE(\Gamma)$ is the set of the Nash equilibria in game Γ . The price of anarchy (4.1.12) can be undefined if the sum in the denominator is null.

Theorem 4.4. *In stochastic game Γ of data transmission in a network with the topology presented in Fig. 4.3, there always exists the Nash equilibrium and a cooperative solution.*

Proof. Since the set of players is finite, the sets of actions of all players in all states of the game are finite, and the game is solved in the class of stationary strategies, then the existence of the cooperative solution follows from the Shapley's theorem [210], and the existence of the Nash equilibrium follows from the theorems proved by Fink and Takahashi [91, 217]. \square

Due to the fact that there are six parameters in the model $(a_1, a_2, c, d, \pi_0, \delta)$, and one parameter is a vector, then it is impossible to represent the explicit form of the Nash equilibrium strategies or the strategies forming a cooperative solution for a general case. We consider an example of finding the Nash equilibrium, the cooperative solution, and the price of anarchy for one network.

Let the probability of a data packet appearance at player 1 be greater than at player 2, namely, $a_1 = 0.8$, $a_2 = 0.2$. Suppose that the costs of any player for the package transmission are $c = 0.1$. For a package delay of one time period, the player bears costs of $d = 0.05$. The discount factor for the payoffs of both players is $\delta = 0.99$. Suppose that the game starts from the state where both players have a package to transit $\pi = (0, 0, 0, 1)$.

All calculations were made in Wolfram Mathematica 11.1 [126]. The payoff function of player 1 is:

$$\begin{aligned}
& 181.152 / (- 36.2304 + (-36.9142 - 10.734\xi_2)\xi_2 + (\xi_1)^2(1.42088 + \\
& + \xi_2(-2.42088 + \xi_2)) + \xi_1(-3.1894 + \xi_2(53.5078 + 19.1515\xi_2))) + \\
& + (\xi_2(42.8875 + 27.0034\xi_2) + \xi_1(-3470.67 + (2899.78 - 2686.02\xi_2)\xi_2) + \\
& + (\xi_1)^2(546.75 + \xi_2(1097.15 + 1578.1\xi_2))) / (- 36.2304 + \\
& + (-36.9142 - 10.734\xi_2)\xi_2 + (\xi_1)^2(1.42088 + \xi_2(-2.42088 + \xi_2)) + \\
& + \xi_1(-3.1894 + \xi_2(53.5078 + 19.1515\xi_2))).
\end{aligned}$$

The payoff function of player 2 is:

$$\begin{aligned}
& 181.152 / \left(-36.2304 + (-36.9142 - 10.734\xi_2)\xi_2 + (\xi_1)^2(1.42088 + \right. \\
& \left. + \xi_2(-2.42088 + \xi_2)) + \xi_1(-3.1894 + \xi_2(53.5078 + 19.1515\xi_2)) \right) + \\
& \left((-3400.79 - 780.872\xi_2)\xi_2 + (\xi_1)^2(-7.10438 + \right. \\
& \left. + (-435.212 - 222.744\xi_2)\xi_2) + \xi_1(15.947 + \xi_2(3308.04 + 1557.71\xi_2)) \right) / \\
& \left(-36.2304 + (-36.9142 - 10.734\xi_2)\xi_2 + (\xi_1)^2(1.42088 + \right. \\
& \left. + \xi_2(-2.42088 + \xi_2)) + \xi_1(-3.1894 + \xi_2(53.5078 + 19.1515\xi_2)) \right).
\end{aligned}$$

Players' payoffs as functions of probabilities ξ_1 , ξ_2 are presented in Fig. 4.4. The total players' payoff function for game Γ as a function of strategies ξ_1 and ξ_2 is presented in Fig. 4.5.

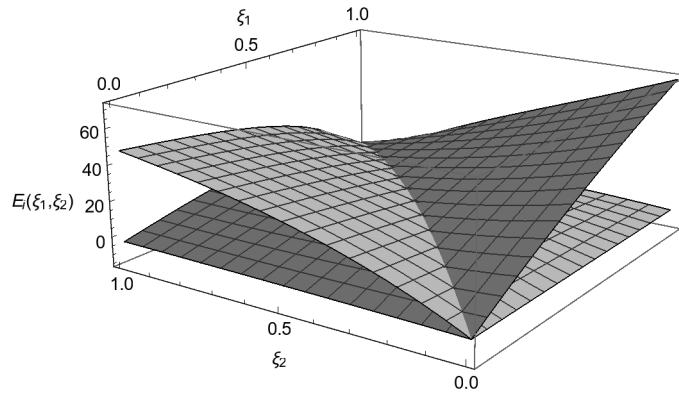


Figure 4.4: Players' payoff functions (player 1: dark-gray, player 2: light-gray).

The Nash equilibrium is determined by the probabilities $\xi_1^* = 0.955$, $\xi_2^* = 0.647$, which can be interpreted in the following way. Player 1, having a package for transmission, transmits it with probability $\xi_1^* = 0.955$ and waits with probability $1 - \xi_1^* = 0.045$. Player 2, having a package for transmission, transmits it with probability $\xi_2^* = 0.647$ and waits with probability $1 - \xi_2^* = 0.353$. Player payoffs in the Nash equilibrium are equal to $E_1(\xi_1^*, \xi_2^*) = 23.538$, $E_2(\xi_1^*, \xi_2^*) = 0.496$.

A cooperative solution in the game is profile $(\tilde{\xi}_1, \tilde{\xi}_2) = (1, 0)$, at which the players' payoffs are equal to $E_1(\tilde{\xi}_1, \tilde{\xi}_2) = 72.18$, $E_2(\tilde{\xi}_1, \tilde{\xi}_2) = -5$ with a total players' payoff of 67.18.

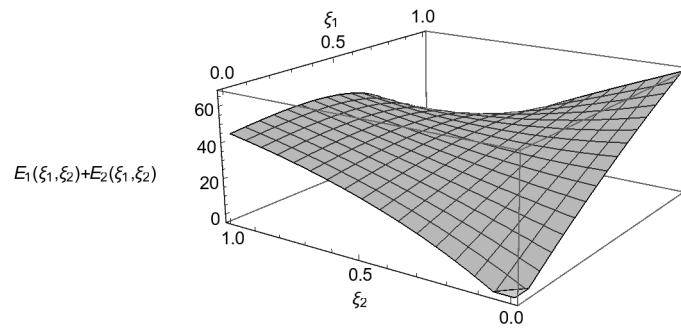


Figure 4.5: Total players' payoff function.

We use formula (4.1.12) and calculate the price of anarchy in this game:

$$PoA(\Gamma) = \frac{67.18}{23.538 + 0.496} = 2.79521,$$

which indicates the need to coordinate players' strategies in this network.

§ 4.1.5 “Forward and transmit” game

In this section, we consider a data transmission model with a larger number of parameters and demonstrate the construction of a cooperative game “Forward and transmit”, then we verify if the principles of stable cooperation are satisfied. Nodes 1 and 2 independently generate data packages of unit capacity at the beginning of each time period with probabilities a_1 and a_2 , respectively. The package may be generated only if the queue of the node is empty, i. e., it does not contain any data packages at the end of the previous time period. The scheme of data transmission is depicted in Fig. 4.6.

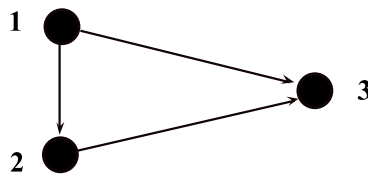


Figure 4.6: Data transmission scheme for “Forward and transmit” game.

We make some assumptions about the system organization:

1. Nodes 1 and 2 (Players 1 and 2, respectively) tend to send unit-capacity data packages to the common destination node — node 3.
2. The maximal capacity of any node is equal to one. The destination node can only accept one package in one time period. It is assumed that multi-package transmission and simultaneous acceptance and transmission of data by any node in any time period is impossible.
3. If players simultaneously transmit packages to the destination node, then the destination node rejects both packages and they are returned to the corresponding initial nodes. Therefore, in the next time period the new packages can not appear at nodes 1 and 2.
4. All transmitted packages have the same capacity, and it takes one time period to transfer the package from one node to another if these nodes are connected by an arc in the given graph (see Fig. 4.6).
5. Player 1 has a possibility to transmit a package directly to node 3 or forward a package to player 2 in order to transmit this package to the destination node 3 in the next time period.
6. If player 1 (node 1) transmits a package to node 2 which already has a package in its queue, then player 2 rejects player 1's package. Otherwise, player 2 decides to accept or reject the package transmitted by player 1.

We introduce a system of rewards and costs:

- $f \geq 0$ is a reward to player 1 or 2 for each successful delivery of the package to destination node 3.
- Player 1 receives a reward of $c \geq 0$ from player 2 for package delivery to player 2, who subsequently receives a reward of f only if this package is successfully delivered to the destination node in the next time period.
- The delay in package transmission in a time period brings the costs of $d \geq 0$ to the player in whose queue the package is located (regardless of which node it originally appeared in).

- D_{ij} is the cost of sending a package from node i to node j connected by an arc.

Suppose that the players' payoffs are discounted with the same factor $0 < \delta < 1$. As in the previous cases, we model the problem of data transmission in dynamics using a stochastic game. The state of the game is a pair (Q_1, Q_2) , where $Q_i \in \{0, 1\}$ is the number of packages at node i , $i = 1, 2$. Thus, there are four possible states in the game: $Q = \{(0, 0); (0, 1); (1, 0); (1, 1)\}$.

Consider all states and for each state we determine the sets of players' actions, the players' payoff functions and transition probabilities from the state to any state from set Q under the condition of realizing any possible action profile in the game:

1. State $(0, 0)$. Players 1 and 2 have a unique action W ("wait"). The players' payoffs in this state are $(0, 0)$. The transition probabilities form the vector $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.
2. The state $(0, 1)$. Player 1 has a unique action W ("wait"), player 2 has a unique action $T3$ ("send to node 3"). The players' payoffs in this state are $(0, f - D_{23})$. The transition probabilities form the vector $((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$.
3. State $(1, 0)$. Player 1 has two actions: i) $T3$ ("transmit to node 3"), ii) $T2$ ("transmit to node 2"). Player 2 also has two actions: i) Ac ("accept a package from node 1"), ii) Rej ("reject a package from node 1").

Players' payoff functions are represented in the matrix:

$$\begin{array}{cc} & \begin{array}{c} Ac \\ Rej \end{array} \\ \begin{array}{c} T3 \\ T2 \end{array} & \left(\begin{array}{cc} (f - D_{13}, 0) & (f - D_{13}, 0) \\ (c - D_{12}, -c) & (-d - D_{12}, 0) \end{array} \right) \end{array}$$

In this state there are four pure action profiles, and we write the transition probabilities from state $(1, 0)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ in Tab. 4.5.

4. State $(1, 1)$. Player 1 has two actions: i) $T3$ ("transmit to node 3"), ii) W ("wait"). Player 2 also has two actions: i) $T3$ ("transmit to node 3"), ii) W ("wait").

Table 4.5: Transition probabilities from state $(1, 0)$ in “Forward and transmit” game.

Profile	$p_3(\cdot) = (p_{3,1}(\cdot), p_{3,2}(\cdot), p_{3,3}(\cdot), p_{3,4}(\cdot))$
$(T3, Ac)$	$((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$
$(T3, Rej)$	$((1 - a_1)(1 - a_2), (1 - a_1)a_2, a_1(1 - a_2), a_1a_2)$
$(T2, Ac)$	$(0, 1 - a_1, 0, a_1)$
$(T2, Rej)$	$(0, 0, 1 - a_2, a_2)$

Players’ payoff functions are represented in the matrix:

$$\begin{array}{cc} & \begin{array}{c} T3 \\ W \end{array} & \begin{array}{c} W \\ W \end{array} \\ \begin{array}{c} T3 \\ W \end{array} & \begin{pmatrix} (-d - D_{13}, -d - D_{23}) & (f - D_{13}, -d) \\ (-d, f - D_{23}) & (-d, -d) \end{pmatrix} & \end{array}$$

In state $(1, 1)$ there are four pure action profiles, and we write the transition probabilities from state $(1, 1)$ to states $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ in Tab. 4.6.

Table 4.6: Transition probabilities from state $(1, 1)$ in “Forward and transmit” game.

Profile	$p_4(\cdot) = (p_{4,1}(\cdot), p_{4,2}(\cdot), p_{4,3}(\cdot), p_{4,4}(\cdot))$
$(T3, T3)$	$(0, 0, 0, 1)$
$(T3, W)$	$(0, 1 - a_1, 0, a_1)$
$(W, T3)$	$(0, 0, 1 - a_2, a_2)$
(W, W)	$(0, 0, 0, 1)$

Denote the pure strategy of player i in stochastic game as $\eta_i = (\eta_{ij} : j \in Q)$, where η_{ij} is the player i ’s strategy in state j . The strategy profile in the game is $\eta = (\eta_1, \eta_2)$. Denote the set of pure stationary strategies of player i by Ξ_i , $i = 1, 2$. The sets of strategies are

$$\Xi_1 = \{(W, W, T3, W), (W, W, T3, T3), (W, W, T2, W), (W, W, T2, T3)\},$$

$$\Xi_2 = \{(W, T3, Ac, T3), (W, T3, Ac, W), (W, T3, Rej, T3), (W, T3, Rej, W)\}.$$

For example, the pure strategy $\eta_1^1 = (W, W, T3, W)$ of player 1 prescribes her to choose action $T3$ in state $(1, 0)$ and action W in states $(0, 0)$, $(0, 1)$ and $(1, 1)$. Each player has four pure stationary strategies in the stochastic game. Therefore, in the stochastic game there are 16 pure stationary strategy profiles.

For each stationary strategy profile, transition probabilities can be written in matrix form: $\Pi(\eta)$.

For example, if we consider profile $\eta = (\eta_1, \eta_2) = ((W, W, T3, T3), (W, T3, Ac, T3))$, the matrix of transition probabilities becomes

$$\Pi(\eta) = \begin{pmatrix} (1-a_1)(1-a_2) & (1-a_1)a_2 & a_1(1-a_2) & a_1a_2 \\ (1-a_1)(1-a_2) & (1-a_1)a_2 & a_1(1-a_2) & a_1a_2 \\ (1-a_1)(1-a_2) & (1-a_1)a_2 & a_1(1-a_2) & a_1a_2 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

For each profile $\eta \in \Xi = \prod_{i=1}^2 \Xi_i$ we can calculate the expected payoffs of the players in the subgames, which are denoted by $E_i(\eta) = (E_i^{(0,0)}(\eta), E_i^{(0,1)}(\eta), E_i^{(1,0)}(\eta), E_i^{(1,1)}(\eta))'$, where

$$E_i(\eta) = (\mathbb{I}_4 - \delta\Pi(\eta))^{-1}K_i(\eta), \quad (4.1.13)$$

and $K_i(\eta)$, $\Pi(\eta)$ are defined above.

The expected payoff of player i in the whole game, including the chance move which chooses the initial state of the game, is

$$\bar{E}_i(\eta) = \pi^0 E_i(\eta), \quad (4.1.14)$$

where $\pi^0 = (\pi_{(0,0)}^0, \pi_{(0,1)}^0, \pi_{(1,0)}^0, \pi_{(1,1)}^0)$ is the vector of the initial distribution over the set of states, where π_k^0 is the probability that state $k \in Q$ is realized at the first time period of the stochastic game. Vector π^0 is given.

We consider a game in a cooperative setting, which means that the players' strategies are coordinated by one center (the strategies are correlated [67]) in order to improve the network productivity. Coordination of actions of transmitting devices is useful for increasing the rate of data transfer or network capacity. To solve a cooperative version of a stochastic game, we suppose that a player has information not only about the state of its queue, but also the state of the queue of the other player.

Algorithm of finding a solution of a cooperative stochastic game. Here we describe the steps of finding a solution of a stochastic "Forward and transmit" game for the network represented in Fig. 4.6.

1. For any state $k \in Q$ and every pure stationary strategy profile $\eta = (\eta_1, \eta_2)$, $\eta_i \in \Xi_i$, $i = 1, 2$, we calculate the expected payoff of the player $E_i^k(\eta)$ in subgame G^k by formula (4.1.13) and the expected payoffs of the player $\bar{E}_i(\eta)$ in the whole game by formula (4.1.14).
2. Find the cooperative solution $\bar{\eta}$ using condition (2.3.1).
3. Calculate the values of characteristic functions $v^k(S)$ for any state $k \in Q$ and any coalition $S \subset N$, using the equations (2.3.3), (2.3.4), (2.3.8). Then we compute the values of $\bar{v}(S)$ for all coalitions $S \subset N$ by formula (2.3.7).
4. We find the Shapley values $\alpha^k = (\alpha_1^k, \dots, \alpha_n^k)$ for any subgame G^k starting from state $k \in Q$ by formula [211]:

$$\alpha_i^k = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} (v^k(S \cup \{i\}) - v^k(S)). \quad (4.1.15)$$

Then we find the Shapley value for the whole game $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_2)$ by formula $\bar{\alpha}_i = \pi^0 \alpha_i$.

5. Compute the components of IDP β_i^k , $i = 1, 2$, $k \in Q$ by formula (2.4.2).
6. In order to make the Shapley value subgame-consistent, we define α -regularization G_α by redefining the payoff functions using formula (2.4.8).
7. Verify if there is the Nash equilibrium in G_α in behavior strategies with payoffs $(\bar{\alpha}_1, \bar{\alpha}_2)$. For this we use the inequality (2.4.11).

Consider a numerical example of “Forward and transmit” game of a data transmission in a wireless network. Define the parameters for modeling. Let the probability of package appearance at node 1 be greater than at node 2: $a_1 = 0.4$, $a_2 = 0.1$. The discount factor is $\delta = 0.99$, which can be interpreted as follows: the probability that the game ends at any stage is 0.01. The rewards and costs of players are as follows: $f = 1$, $d = 0.1$, $c = 0.3$, $D_{12} = 0.1$, $D_{13} = 0.6$, $D_{23} = 0.2$. One can notice that the cost of a package transmission from node 1 to node 3 is three times larger than the cost of a package transmission from node 2 to node 3. Thus, the cooperation of nodes 1 and 2 can be beneficial for both players. Let the game may start from any state with equal probability, i. e., $\pi^0 = (0.25, 0.25, 0.25, 0.25)$.

Tables 4.7 and 4.8 contain the values of players' expected payoffs

$$E_i(\eta) = \left(E_i^{(0,0)}(\eta), E_i^{(0,1)}(\eta), E_i^{(1,0)}(\eta), E_i^{(1,1)}(\eta) \right)'$$

for any pure stationary strategy profile η for all players $i = 1, 2$, and the sum of the payoffs (last column) equal to $\bar{E}_1 + \bar{E}_2$ taking into account the vector of the initial distribution over the set of states π^0 .

The cooperative solution maximizing the total players' payoff is strategy profile

$$\bar{\eta} = \eta^{11} = (0, 0, 1, 1),$$

in which player 1's strategy $\eta_1^{11} = (0, 0)$ prescribes him to "transmit a package to node 2" in state $(1, 0)$, if at node 1 there is a package, and there is no package at node 2. The player 2's strategy is $\eta_2^{11} = (1, 1)$ prescribes him to "accept a package" from player 1 in state $(1, 0)$. If the game is in state $(1, 1)$, player 1 "waits", and player 2 "transmits a package" to node 3.

The maximal total payoff of the players in the whole game is

$$\max_{\eta \in \Xi} \sum_{i \in N} \bar{E}_i(\eta) = \sum_{i \in N} \bar{E}_i(\bar{\eta}) = 26.9472.$$

We calculate the values of characteristic functions for subgames using formulas (2.3.3), (2.3.4), (2.3.8):

$$\begin{aligned} v(\{1\}) &= (-5.10968, -5.10968, -4.70968, -10.0)^T, \\ v(\{2\}) &= (-8.735960, -7.93596, -8.97396, -10.0)^T, \\ v(\{1, 2\}) &= (26.3784, 27.1784, 26.8766, 27.3553)^T, \end{aligned}$$

The characteristic function for the whole game is given by formula (2.3.7):

$$\begin{aligned} \bar{v}(\{1\}) &= -6.23226, \\ \bar{v}(\{2\}) &= -8.91147, \\ \bar{v}(\{1, 2\}) &= 26.9472. \end{aligned}$$

Then we calculate the Shapley values for all subgames and the whole game using formula (4.1.15):

Table 4.7: Players' payoffs in "Forward and transmit" game with profiles η^1, \dots, η^8 .

η	$E_1(\eta)$	$E_2(\eta)$	$\bar{E}_1 + \bar{E}_2$
$\eta^1 = (1, 1, 1, 1)$	-53.0129	-22.9935	-81.7048
	-53.0129	-22.1935	
	-52.6129	-22.9935	
	-70.1000	-30.0000	
$\eta^2 = (1, 1, 1, 0)$	15.8400	6.76818	23.1355
	15.8400	7.56818	
	16.2400	6.76818	
	16.2400	7.27732	
$\eta^3 = (1, 0, 1, 1)$	14.7353	7.9200	23.1855
	14.7353	8.7200	
	15.1353	7.9200	
	14.8563	8.7200	
$\eta^4 = (1, 0, 1, 0)$	-5.10968	-7.02581	-13.8016
	-5.10968	-6.22581	
	-4.70968	-7.02581	
	-10.0000	-10.0000	
$\eta^5 = (1, 1, 0, 1)$	-53.0129	-22.9935	-81.7048
	-53.0129	-22.1935	
	-52.6129	-22.9935	
	-70.0000	-30.0000	
$\eta^6 = (1, 1, 0, 0)$	15.8400	6.76818	23.1355
	15.8400	7.56818	
	16.2400	6.76818	
	16.2400	7.27732	
$\eta^7 = (1, 0, 0, 1)$	14.7353	7.9200	23.1855
	14.7353	8.7200	
	15.1353	7.9200	
	14.8563	8.7200	
$\eta^8 = (1, 0, 0, 0)$	-5.10968	-7.02581	-13.8016
	-5.10968	-6.22581	
	-4.70968	-7.02581	
	-10.0000	-10.0000	

- for subgames:

$$\alpha_1 = (15.0023, 15.0023, 15.5705, 13.6776)^T,$$

$$\alpha_2 = (11.376, 12.176, 11.3062, 13.6776)^T,$$

Table 4.8: Players' payoffs in "Forward and transmit" game with profiles η^9, \dots, η^{16} .

η	$E_1(\eta)$	$E_2(\eta)$	$\bar{E}_1 + \bar{E}_2$
$\eta^9 = (0, 1, 1, 1)$	-64.7464	-27.3398	-94.3241
	-64.7464	-26.5398	
	-65.9794	-27.9446	
	-70.0000	-30.0000	
$\eta^{10} = (0, 1, 1, 0)$	11.5347	13.4228	25.4926
	11.5347	14.2228	
	11.8060	13.6218	
	12.0060	13.8218	
$\eta^{11} = (0, 0, 1, 1)$	4.90248	21.4759	26.9472
	4.90248	22.2759	
	5.04298	21.8337	
	4.87602	22.4792	
$\eta^{12} = (0, 0, 1, 0)$	-8.93504	-8.73596	-18.1458
	-8.93504	-7.93596	
	-9.06741	-8.97396	
	-10.0000	-10.0000	
$\eta^{13} = (0, 1, 0, 1)$	-64.2491	-26.728	-93.4537
	-64.2491	-25.928	
	-65.4128	-27.248	
	-70.0000	-30.000	
$\eta^{14} = (0, 1, 0, 0)$	-8.4855	5.60847	-2.43475
	-8.4855	6.40847	
	-8.8128	5.57380	
	-7.6828	6.13680	
$\eta^{15} = (0, 0, 0, 1)$	-18.532	7.9200	-10.426
	-18.532	8.7200	
	-19.010	7.9200	
	-18.910	8.7200	
$\eta^{16} = (0, 0, 0, 0)$	-10.559	-8.8313	-19.49526
	-10.559	-8.0313	
	-10.917	-9.0826	
	-10.000	-10.000	

- for the whole game:

$$\bar{\alpha}_1 = 14.8132,$$

$$\bar{\alpha}_2 = 12.134.$$

The imputation distribution procedure β_1 for player 1 and β_2 for player 2 is found by equation (2.4.2) based on the Shapley values:

- $\beta_1 = (0, 0, 1.24274, -1.54974)^T$,
- $\beta_2 = (0, 0.8, -1.34274, 2.24974)^T$,

where β_i^k is the payment to player i in state k . Recall that the players' payoffs in the states initially determined by the payoff functions are as follows:

- $K_1 = (0, 0, 0.2, -0.1)^T$,
- $K_2 = (0, 0.8, -0.3, 0.8)^T$,

It should be noted that in states $(0, 0)$ and $(0, 1)$, the components of imputation distribution procedure coincide with the payoffs given by functions K_1 and K_2 . In states $(1, 0)$ and $(1, 1)$, there is a redistribution of players' payoffs. In state $(1, 0)$, players receive a total payoff of -0.1 and according to the imputation distribution procedure player 1 gets 1.24274 instead of 0.2, and player 2 gets -1.34274 contrary to -0.3 . Thus, player 2 pays 1.04274 to player 1 in order for the cooperative solution (the Shapley value) to become subgame consistent. In state $(1, 1)$, players get a total payoff of 0.7 and according to the imputation distribution procedure, player 1 gets -1.54974 instead of -0.1 , and player 2 gets 2.24974 instead of 0.8. Thus, player 1 pays 1.44974 to player 2 in order for the cooperative solution to become dynamically stable.

Therefore, the Shapley value $\bar{\alpha} = (14.8132, 12.134)^T$ is subgame consistent if payments to players in states are made in accordance with the IDP $\beta_1 = (0, 0, 1.24274, -1.54974)^T$, and $\beta_2 = (0, 0.8, -1.34274, 2.24974)^T$.

Now we examine the problem of strategic support of a cooperative solution. First, we compute values $F^k(\{i\})$ for $i = 1, 2$ and $k \in Q$, defined in Theorem 2.4:

1. $F(\{1\}) = (-5.10968, -5.10968, -4.70968, -5.28632)^T$,
2. $F(\{2\}) = (-8.73596, -7.93596, -8.97396, -8.1858)^T$.

Now we verify if inequality (2.4.11) is satisfied for any player and any state. For player 1, inequality (2.4.11) is

$$\begin{pmatrix} 0 \\ 0 \\ 1.24274 \\ -1.54974 \end{pmatrix} \geq \begin{pmatrix} -0.186662 \\ -0.186662 \\ 0.418853 \\ -0.566649 \end{pmatrix}$$

The inequality is not true. In state (1, 1), player 1 has a possibility to increase her payoff by deviating, since her payoff in accordance with imputation distribution procedure in this state is -1.54974 , which is less than his payoff of -0.566649 in case of deviation. This means that cooperation can not be strategically supported by the Nash equilibrium of form (2.4.12).

For player 2, the inequality (2.4.11) takes the form:

$$\begin{pmatrix} 0 \\ 0.8 \\ -1.34274 \\ 2.24974 \end{pmatrix} \geq \begin{pmatrix} -0.0718427 \\ 0.728157 \\ -1.01842 \\ 0.620393 \end{pmatrix}$$

Again, the inequality is not satisfied. In state (1, 0), player 2 can increase her payoff by deviating, since his payoff in this state in accordance with the imputation distribution procedure is -1.34274 , which is less than his payoff of -1.01842 in case of deviation. We can state that profile (2.4.12) is not the Nash equilibrium, so a cooperative solution can not be supported by profile (2.4.12).

§ 4.1.6 Stochastic game of data transmission in the presence of buffers of finite capacity

Consider the data transmission ALOHA-like scheme represented in Fig. 4.7. Players (Nodes) 1 and 2 tend to send data packages to the nodes r_1 and r_2 , respectively. As one can see from the data transmission scheme, the package should go through the node of a unit capacity which is common for both players. Player $i = 1, 2$ has a buffer of capacity k_i which means that it can possess from 0 to k_i data packages of a unit capacity at each time period. At the beginning of each time period, Players

1 and 2 may receive a data package of a unit capacity with probability $\nu_1 \in (0, 1)$ and $\nu_2 \in (0, 1)$ respectively if at the beginning of the period he possesses less than k_i packages.

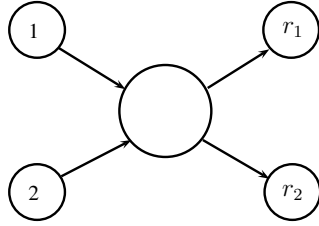


Figure 4.7: Data transmission scheme.

In each time period, Player $i = 1, 2$ can transmit only one package to the destination node r_i . In case both players simultaneously transmit the packages, the packages are back to the Nodes. For successful package delivery, it is necessary that only one Player transmits the package. If the package is delivered to the destination node, the player receives a payoff of 1 minus the costs of a package transmission which is equal to $c \in (0, 1)$. The player bears the costs of $d \in [0, 1)$ for one time period delay per each unit package, $d \ll 1$.

By the state of the system in time period t , we mean the pair $(\omega_1(t), \omega_2(t))$, where $\omega_i(t) \in \{0, 1, \dots, k_i\}$ is the number of data packages at Player i 's buffer, $i = 1, 2$. The set of the system states at any time period t is denoted by Ω , $|\Omega| = (k_1 + 1)(k_2 + 1)$. Let the states be $\omega^{(1)}, \dots, \omega^{(m)}$, where $m = (k_1 + 1)(k_2 + 1)$.

On the basis of the given assumptions we define a stochastic game with a finite state set and finite action sets.

The time is assumed to be discrete. The set of states at each time period is $\Omega = \{\omega = (\omega_1, \omega_2) : \omega_i \in [0, k_i], i = 1, 2\}$. At state ω the set of player i 's actions is A_i^ω that is

$$A_i^\omega = \begin{cases} \{t, w\}, & \text{if } \omega_i > 0, \\ \{w\}, & \text{if } \omega_i = 0, \end{cases} \quad (4.1.16)$$

where action t means “to transmit” a package, w — “to wait”.

In state $\omega = (\omega_1, \omega_2)$, where $\omega_1 = \omega_2 = 0$, the payoff function of player i is equal $u_i^\omega(a_1^\omega, a_2^\omega) = 0$ for any $i = 1, 2$, and both actions a_1^ω and a_2^ω are w . If $\omega = (\omega_1, \omega_2)$,

$\omega_1 > 0$ and $\omega_2 = 0$, then the payoff functions are defined by matrix

$$\begin{pmatrix} (1 - c - d(\omega_1 - 1); 0) \\ (-d\omega_1; 0) \end{pmatrix}, \quad (4.1.17)$$

where Player 1 chooses rows (row 1 correspond to action t , row 2 — to action w), Player 2 chooses a column. In case when $\omega = (\omega_1, \omega_2)$, $\omega_1 = 0$ and $\omega_2 > 0$, then the players' payoff functions are defined by matrix

$$\left((0; 1 - c - d(\omega_2 - 1)) \quad (0; -d\omega_2) \right). \quad (4.1.18)$$

If $\omega = (\omega_1, \omega_2)$, $\omega_1 > 0$ and $\omega_2 > 0$, then each player has two actions t and w and players' payoff functions in this state are defined by

$$\begin{pmatrix} (-c - d\omega_1; -c - d\omega_2) & (1 - c - d(\omega_1 - 1); -d\omega_2) \\ (-d\omega_1; 1 - c - d(\omega_2 - 1)) & (-d\omega_1; -d\omega_2) \end{pmatrix}. \quad (4.1.19)$$

Let assume that the players' strategies in the game be stationary which seems to be a natural assumption in data transmission modeling. We suppose that the player does not know the number of packages at the other player's buffer. And we also assume that Player i uses the same strategy in any state (ω_i, ω_j) , where $\omega_i \in \{1, \dots, k_i - 1\}$, and maybe the other strategy in state (ω_i, ω_j) , where $\omega_i = k_i$. Therefore, Player i 's mixed stationary strategy η_i is (p_i^f, p_i^{nf}) , $i = 1, 2$, where $p_i^f \in [0, 1]$ is a probability of choosing action t in any state (ω_i, ω_j) , $j \neq i$, $j = 1, 2$, where $\omega_i = k_i$, $\omega_j \in [0, k_j]$ (probability of transmitting a package when the buffer of player i is full); $p_i^{nf} \in [0, 1]$ is a probability of choosing action t in any state (ω_i, ω_j) , $j \neq i$, $j = 1, 2$, where $\omega_i \in [1, k_i - 1]$, $\omega_j \in [0, k_j]$ (probability of transmitting a package when the buffer of Player i is not full). The stationary strategy profile is $(\eta_1, \eta_2) = ((p_1^f, p_1^{nf}), (p_2^f, p_2^{nf}))$. Denote by Ξ_i the set of stationary strategies of Player i . The set of pure stationary strategies of Player $i = 1, 2$ is $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

Now we define transition probabilities $\pi(\omega''/\omega', \eta)$ to state $\omega'' \in \Omega$ from state $\omega' \in \Omega$ if strategy profile η is realised. They are $\pi(\omega''/\omega', \eta)$ which is represented in Tables 4.9–4.10, where $\omega = (\omega_i, \omega_j)$, $i = 1, 2$, $j = 1, 2$, $i \neq j$ and $\eta = (\eta_i, \eta_j) = ((p_i^f, p_i^{nf}), (p_j^f, p_j^{nf}))$.

We have defined two-person stochastic game G by a tuple

$$\langle \Omega, \{A_i^\omega\}_{i=1,2;\omega \in \Omega}, \{u_i^\omega\}_{i=1,2;\omega \in \Omega}, \{\Xi_i\}_{i=1,2}, \{\pi(\omega''/\omega', \eta)\}_{\omega'' \in \Omega, \omega' \in \Omega, \eta \in \Xi_1 \times \Xi_2}, \delta \rangle,$$

Table 4.9: Transition probabilities

$\pi(\omega''/\omega', \eta)$	$\omega' = (\omega'_i, \omega'_j)$	$\omega'' = (\omega''_i, \omega''_j)$
$\nu_i \nu_j$	$\omega'_i = \omega'_j = 0$	$\omega''_i = \omega''_j = 1$
$\nu_i(1 - \nu_j)$	$\omega'_i = \omega'_j = 0$	$\omega''_i = 1, \omega''_j = 0$
$(1 - \nu_i)(1 - \nu_j)$	$\omega'_i = \omega'_j = 0$	$\omega''_i = \omega''_j = 0$
$(1 - \nu_i)(1 - \nu_j)p_j^{nf}$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 0,$ $\omega''_j = \omega'_j - 1$
$(1 - \nu_i)[p_j^{nf} \nu_j$ $+ (1 - p_j^{nf})(1 - \nu_j)]$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 0,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)(1 - p_j^{nf})\nu_j$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 0,$ $\omega''_j = \omega'_j + 1$
$\nu_i(1 - \nu_j)p_j^{nf}$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 1,$ $\omega''_j = \omega'_j - 1$
$\nu_i[p_j^{nf} \nu_j$ $+ (1 - p_j^{nf})(1 - \nu_j)]$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 1,$ $\omega''_j = \omega'_j$
$\nu_i(1 - p_j^{nf})\nu_j$	$\omega'_i = 0,$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = 1,$ $\omega''_j = \omega'_j + 1$
$(1 - \nu_i)(1 - \nu_j)p_i^{nf}(1 - p_j^{nf})$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = \omega'_i - 1,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)\nu_j p_i^{nf}(1 - p_j^{nf})$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = \omega'_i - 1,$ $\omega''_j = \omega'_j + 1$
$(1 - \nu_i)(1 - \nu_j)p_i^{nf}p_j^{nf}$ $+ \nu_i(1 - \nu_j)p_i^{nf}(1 - p_j^{nf})$ $+ (1 - \nu_i)\nu_j(1 - p_i^{nf})p_j^{nf}$ $+ (1 - \nu_i)(1 - \nu_j) \cdot$ $\cdot (1 - p_i^{nf})(1 - p_j^{nf})$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = \omega'_i,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)\nu_j(1 - p_i^{nf})(1 - p_j^{nf})$ $+ (1 - \nu_i)\nu_j p_i^{nf} p_j^{nf}$ $+ \nu_i \nu_j p_i^{nf}(1 - p_j^{nf})$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = \omega'_i,$ $\omega''_j = \omega'_j + 1$
$\nu_i \nu_j(1 - p_i^{nf})(1 - p_j^{nf})$ $+ \nu_i \nu_j p_i^{nf} p_j^{nf}$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j \in [1, k_j - 1]$	$\omega''_i = \omega'_i + 1,$ $\omega''_j = \omega'_j + 1$
$(1 - \nu_i)(1 - \nu_j)p_j^f$	$\omega'_i = 0,$ $\omega'_j = k_j$	$\omega''_i = 0,$ $\omega''_j = \omega'_j - 1$
$(1 - \nu_i)[1 - p_j^f(1 - \nu_j)]$	$\omega'_i = 0,$ $\omega'_j = k_j$	$\omega''_i = 0,$ $\omega''_j = \omega'_j$
$\nu_i(1 - \nu_j)p_j^f$	$\omega'_i = 0,$ $\omega'_j = k_j$	$\omega''_i = 1,$ $\omega''_j = \omega'_j - 1$

Table 4.10: Transition probabilities (cntd)

$\nu_i[1 - p_j^f(1 - \nu_j)]$	$\omega'_i = 0,$ $\omega'_j = k_j$	$\omega''_i = 1,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)p_i^f(1 - p_j^f)$	$\omega'_i = k_i,$ $\omega'_j = k_j$	$\omega''_i = \omega'_i - 1,$ $\omega''_j = \omega'_j$
$(1 - p_i^f)(1 - p_j^f)$ $+ \nu_i p_i^f(1 - p_j^f)$ $+ \nu_j(1 - p_i^f)p_j^f + p_i^f p_j^f$	$\omega'_i = k_i,$ $\omega'_j = k_j$	$\omega''_i = \omega'_i,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)p_i^{nf}(1 - p_j^f)$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j = k_j$	$\omega''_i = \omega'_i - 1,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)(1 - \nu_j)(1 - p_i^{nf})p_j^f$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j = k_j$	$\omega''_i = \omega'_i,$ $\omega''_j = \omega'_j - 1$
$\nu_i p_i^{nf} p_j^f$ $+ \nu_i(1 - p_i^{nf})(1 - p_j^f)$ $+ \nu_i \nu_j(1 - p_i^{nf})p_j^f$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j = k_j$	$\omega''_i = \omega'_i + 1,$ $\omega''_j = \omega'_j$
$(1 - \nu_i)p_i^{nf} p_j^f$ $+ \nu_i p_i^{nf}(1 - p_j^f)$ $+ (1 - \nu_i)(1 - p_i^{nf})(1 - p_j^f)$ $+ (1 - \nu_i)\nu_j(1 - p_i^{nf})p_j^f$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j = k_j$	$\omega''_i = \omega'_i,$ $\omega''_j = \omega'_j$
$\nu_i(1 - \nu_j)(1 - p_i^{nf})p_j^f$	$\omega'_i \in [1, k_i - 1],$ $\omega'_j = k_j$	$\omega''_i = \omega'_i + 1,$ $\omega''_j = \omega'_j - 1$

where $\delta \in (0, 1)$ is a common discount rate.

We consider the discounted expected payoff as a payoff in stochastic game G given by

$$E_i(\eta) = \pi_0(\mathbb{I} - \delta\Pi(\eta))^{-1}u_i(\eta), \quad (4.1.20)$$

where π_0 is an initial probability distribution over the set of states, \mathbb{I} is an identity matrix of size m , $\Pi(\eta)$ is a $m \times m$ matrix of transition probabilities $\pi(\cdot/\cdot, \eta)$ whose (l, n) th entry is a probability of transition from state $\omega^{(l)}$ to state $\omega^{(n)}$ when strategy profile η is realised.

We consider two approaches (cooperative and non-cooperative) to find a solution in game G . We consider the Nash equilibrium as an optimality principle within a non-cooperative approach. Following a cooperative approach we find the cooperative solution maximizing the total players' expected payoff in game G .

As proved before, there exists the Nash equilibrium and cooperative solution in game G , because the set of states in game G is finite and the set of actions of any player is also finite for any state ω . The cooperative solution η^* is an argument of a well-known dynamic program problem and it always exists. Moreover, it is a pure stationary strategy profile.

To estimate the selfishness in the network we calculate the price of anarchy given by

$$PoA(G) = \frac{\sum_{i=1}^2 E_i(\eta^*)}{\min_{\eta \in NE(G)} \sum_{i=1}^2 E_i(\eta)}, \quad (4.1.21)$$

where $NE(G)$ is the set of the Nash equilibria in game G .

Simulation study. As an example we consider game G with the following parameters: $c = 0.2$, $d = 0.03$. Let Player 1 have a smaller buffer of capacity $k_1 = 2$ contrary to Player 2's buffer of capacity $k_2 = 4$. Both Players have the same discount factor 0.99. The probabilities of package appearance at Players 1 and 2 are $\nu_1 = 0.6$ and $\nu_2 = 0.2$ respectively. There are 15 states in the game. Let the game start with the state $(0, 0)$ when there are no packages at the nodes. The set of pure stationary strategies for each player is $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Player i 's mixed stationary strategy η_i is (p_i^f, p_i^{nf}) , $i = 1, 2$. Transition matrix is of size 15×15 and it is a function of stationary strategy profile $\eta = (\eta_1, \eta_2)$. To find a cooperative solution we need to calculate the total players' payoff in the whole game for each pure stationary strategy profile (there are 16 such profiles) and find the maximal one. To find the Nash equilibria in stochastic game G we use Lemke-Howson algorithm [105, 123, 198]. For this we calculate the matrices of Player 1 and 2's expected payoffs by formula (4.1.20) for each pure stationary strategy profile. Therefore we obtain two payoff matrix \mathcal{A} and \mathcal{B} for Player 1 and Player 2 respectively:

$$\mathcal{A} = \begin{pmatrix} -5.8518 & -5.8518 & -5.8518 & -5.8518 \\ 47.5200 & -2.7638 & 5.0169 & -5.8518 \\ 43.7829 & 34.8935 & -12.9591 & -20.9107 \\ 47.5200 & 38.7075 & -7.5023 & -20.9767 \end{pmatrix}$$

$$\mathcal{B} = \begin{pmatrix} -10.6254 & 15.8400 & 5.5014 & 15.8400 \\ -10.6254 & 15.3433 & 3.7837 & 15.8400 \\ -10.6254 & -11.8717 & -26.0360 & -27.5207 \\ -10.6254 & -12.3383 & -21.0293 & -28.2416 \end{pmatrix}$$

Then we use these matrices in Lemke-Howson algorithm.

There are four Nash equilibria in bimatrix game with payoff matrices $(\mathcal{A}, \mathcal{B})$: (1) $\xi_1 = (0, 0.0619, 0, 0.93819)$, $\xi_2 = (1, 0, 0, 0)$ with Players' 1 and 2 payoffs 47.5200 and -10.6254 respectively; (2) $\xi_1 = (0, 0.9697, 0, 0.0303)$, $\xi_2 = (0, 0.2672, 0, 0.7328)$ with Players' 1 and 2 payoffs -5.0266 and 14.5050 respectively; (3) $\xi_1 = (0, 1, 0, 0)$, $\xi_2 = (0, 0, 0, 1)$ with Players' 1 and 2 payoffs -5.8518 and 15.8400 respectively; (4) $\xi_1 = (1, 0, 0, 0)$, $\xi_2 = (0, 0, 0, 1)$ with Players' 1 and 2 payoffs -5.8518 and 15.8400 respectively. The "worst" Nash equilibria, i. e., the Nash equilibria with the smallest total Players' payoff, is the second one $\xi_1 = (0, 0.9697, 0, 0.0303)$, $\xi_2 = (0, 0.2672, 0, 0.7328)$. The equilibria in bimatrix game corresponds to the Nash equilibria in stochastic game G in stationary strategies (η_1, η_2) . Mixed strategy η_1 prescribes to choose pure strategy $(0, 1)$ with probability 0.9697 and $(1, 1)$ with probability 0.0303. Mixed strategy η_2 prescribes to choose pure strategy $(0, 1)$ with probability 0.2672 and $(1, 1)$ with probability 0.7328. Both Players randomize between two strategies $(0, 1)$ and $(1, 1)$, but Player 1 with much higher probability 0.9697 chooses strategy $(0, 1)$ contrary to Player 2 who chooses strategy $(0, 1)$ with probability 0.2672. The total Players' payoff in the Nash equilibrium is 9.4784.

The cooperative solution is (η_1, η_2) , where $\eta_1 = (1, 1)$, $\eta_2 = (0, 0)$, with total Players' payoff 36.8946. In cooperation Player 1 transmits a package in any state regardless his buffer is full or not, and Player 2 never transmits packages. The price of anarchy in game G is 3.8925.

§ 4.2 Dynamically stable coalition structures as solutions of stochastic game

§ 4.2.1 Basic definitions. An auxiliary stochastic game

A triple (N, v, π) is called an n -person cooperative game with coalition structure π where N is a finite set of players, $|N| = n \geq 2$, a real-valued function $v : 2^N \mapsto \mathbb{R}$ is the characteristic function, and π is a coalition structure, i.e. a partition $\pi = \{B_1, \dots, B_m\}$ of set N such that $B_1 \cup \dots \cup B_m = N$, $m \leq n$ and $B_i \cap B_\ell = \emptyset$ for all $i, \ell = 1, \dots, m$, $i \neq \ell$. Denote the set of all coalition structures for set N by Π and the coalition containing player i by $B(i) \in \pi$.

A profile $x^\pi = (x_1^\pi, \dots, x_n^\pi) \in \mathbb{R}^n$ is a payment distribution (or a payoff vector) in the game (N, v, π) if it is efficient, i.e. equality $\sum_{i \in B_\ell} x_i^\pi = v(B_\ell)$ holds for any coalition $B_\ell \in \pi$. A payment distribution x^π is an allocation in a game (N, v, π) if the individual rationality condition holds, i.e. $x_i^\pi \geq v(\{i\})$ for any player $i \in N$. The solution (or a solution concept) of the game (N, v, π) is a rule that assigns a payment distribution to the game.³

Now we proceed to the construction of a stochastic game of a special form defining all its elements such as game states, actions, payoff functions, a transition function, strategies and a transition matrix. Let time be discrete, $t = 1, 2, \dots$. We call each time moment by a stage.

The set of states. Given the set of players N , we can determine the number of different coalition structures s which the players can form, i.e. $\Pi = \{\pi_1, \dots, \pi_s\}$. Number s is the n th Bell number.⁴ We consider a coalition structure as a game state, thus the set of states is Π .

The set of actions. Let at the first stage cooperation of players be restricted by coalition structure $\pi_j = \{B_1, \dots, B_m\} \in \Pi$, i.e. the process starts from state π_j . In this state each player i chooses pure action $a_i^{\pi_j}$ from her finite action set $A_i^{\pi_j}$. Set $A_i^{\pi_j} \subset \Pi$ is the subset of coalition structures which are realised if player i either

³In classical cooperative games the solution is considered as an allocation due to the superadditive characteristic function. In the present study we consider the solution as a payment distribution since we do not claim superadditivity for the characteristic function.

⁴The n th Bell number $\mathcal{B}(n)$ is determined by recurrence formula $\mathcal{B}(n) = \sum_{k=0}^{n-1} \binom{n-1}{k} \mathcal{B}(k)$ with $\mathcal{B}(0) = 1$.

leaves coalition $B(i)$ or stays in $B(i)$, i.e. $A_i^{\pi_j} = \{\pi_k \in \Pi : \pi_k = \{B(i) \setminus \{i\}, B_\ell \cup \{i\}, \pi_{-B(i) \cup B_\ell}\}\}$ where $B_\ell \in \pi_j \cup \emptyset$ and $\pi_{-B(i) \cup B_\ell} = \pi_j \setminus \{B(i) \cup B_\ell\}$ denotes the set of coalitions from π_j without coalitions $B(i)$ and B_ℓ . It is obvious that π_j belongs to $A_i^{\pi_j}$ for any player $i \in N$ and any coalition structure $\pi_j \in \Pi$.

A mixed action of player i in state π_j is a profile $\xi_i^{\pi_j} = \{\xi_i(\pi|\pi_j)\}$ such that

$$\xi_i(\pi|\pi_j) \geq 0 \quad \text{and} \quad \sum_{\pi \in A_i^{\pi_j}} \xi_i(\pi|\pi_j) = 1 \quad \text{for all } \pi \in A_i^{\pi_j}.$$

The entry $\xi_i(\pi|\pi_j)$ is the conditional probability of choosing coalition structure π by player i being in π_j . The set of mixed actions of player i in state π_j is denoted by $\Xi_i^{\pi_j}$. Let $\xi^{\pi_j} = (\xi_1^{\pi_j}, \dots, \xi_n^{\pi_j})$ denote a mixed action profile, and $\Xi^{\pi_j} = \prod_{i \in N} \Xi_i^{\pi_j}$.

Example 1. For a three-person game, the third Bell number equals five, thus we have five coalition structures $\Pi = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\}$, $\pi_1 = \{\{1\}, \{2\}, \{3\}\}$, $\pi_2 = \{\{1, 2\}, \{3\}\}$, $\pi_3 = \{\{1, 3\}, \{2\}\}$, $\pi_4 = \{\{1\}, \{2, 3\}\}$, $\pi_5 = \{\{1, 2, 3\}\}$, which are the game states. Let the current state be π_2 . Then $A_1^{\pi_2} = \{\pi_1, \pi_2, \pi_3\}$, $A_2^{\pi_2} = \{\pi_1, \pi_2, \pi_4\}$ and $A_3^{\pi_2} = \{\pi_2, \pi_5\}$. For instance, Player 1 has three pure actions in state π_2 : “to deviate to state π_1 ”, “to stay in π_2 ”, and “to deviate to state π_3 ”. A mixed strategy of Player 1 in state π_2 is a non-negative profile

$$\xi_1^{\pi_2} = (\xi_1(\pi_1|\pi_2), \xi_1(\pi_2|\pi_2), \xi_1(\pi_3|\pi_2)).$$

Payoff function. For every player $i \in N$ we define stage payoff function u_i^π as the i 's component of the solution of the game (N, v, π) for a given coalition structure (a state) π , i.e. $u_i^\pi = x_i^\pi$, $\pi \in \Pi$. Below to specify a particular solution of the game (N, v, π) , we consider the Shapley value for the game with coalition structure (or the Aumann–Drèze value) $\phi^\pi = (\phi_1^\pi, \dots, \phi_n^\pi)$ as the solution⁵ (see [68]), whose entries are defined by formula

$$\phi_i^\pi = \sum_{S \subseteq B(i); S \ni i} \frac{(|B(i)| - |S|)! (|S| - 1)!}{|B(i)|!} [v(S) - v(S \setminus \{i\})], \quad i \in N. \quad (4.2.1)$$

⁵One may consider another solution of a cooperative game with a given coalition structure, the coalition value. According to [155], the elements of a coalition structure are called a priori unions, and the solution is defined as a two-step procedure using the Shapley value at both steps. At the first step the grand coalition worth is allocated among the unions, whereas the obtained values are allocated among the members of all unions at the second step. In our case, the Aumann–Drèze value allocates the worth of any union, given by the characteristic function, using the Shapley value.

Thus, $u_i^\pi = \phi_i^\pi$ for all $i \in N$.

Transition function. According to the definition of a stochastic game, we have to determine a transition rule from one state to any other. We determine this rule by means of transition probabilities. The probability to transit from state π_j to state π_k is supposed to be positive only if there exists at least one player $i \in N$ such that $\pi_k \in A_i^{\pi_j}$. Otherwise, the transition probability is zero. This means that the only states which can be realised with a positive probability are the states from set $A_1^{\pi_j} \cup \dots \cup A_n^{\pi_j}$. These states can be reached from state π_j in one stage by players' actions. Let in state π_j mixed action profile ξ^{π_j} be realised. Given the profile ξ^{π_j} , the non-zero probability $\lambda(\pi_k|\pi_j, \xi^{\pi_j})$ to transit from state π_j to state π_k is defined as follows. Assuming that players choose their mixed action profiles independently of each other, $\lambda(\pi_k|\pi_j, \xi^{\pi_j})$ is a probability of “deviation from coalition structure π_j to π_k ” by one player, while all other players “stay in π_j ”. More formally,

$$\lambda(\pi_k|\pi_j, \xi^{\pi_j}) = \begin{cases} \sum_{\substack{i \in N \\ \pi_k \in A_i^{\pi_j}}} \xi_i(\pi_k|\pi_j) \prod_{\ell \neq i} \xi_\ell(\pi_j|\pi_j), & \text{if } k \neq j, \quad \pi_k \in \bigcup_{i \in N} A_i^{\pi_j}, \\ 1 - \sum_{\ell \neq k} \lambda(\pi_\ell|\pi_j, \xi^{\pi_j}), & \text{if } k = j, \\ 0, & \text{if } \pi_k \notin \bigcup_{i \in N} A_i^{\pi_j}. \end{cases} \quad (4.2.2)$$

At the next stage, the process moves to a new state only if one player “deviates”. Otherwise, the next state remains the same. The key point is that only one player should deviate to transit to another state.

Note, that if we consider only pure actions of players, the transition rule (4.2.2) specifies the deterministic transition from one state to any other.

Remark 4.5. *It is worth noting that there is no specific rule of transition from one state to any other. The other way of determining a transition rule from state π_j is that a new state $\pi_k \in A_1^{\pi_j} \cup \dots \cup A_n^{\pi_j}$ can be realised with positive probability if one of two options is satisfied: (i) only one player i chooses action $\pi_k \in A_j^{\pi_j}$ in state π_j while other players choose to stay in π_j , or (ii) all players $m \in N$ with $A_m^{\pi_j} \ni \pi_k$ choose “deviation from state π_j to state π_k ” while other players do not deviate from π_j . According to this transition rule, the probability $\lambda(\pi_k|\pi_j, \xi^{\pi_j})$ to transit from*

state π_j to state π_k is of the form:

$$\lambda(\pi_k|\pi_j, \xi^{\pi_j}) = \begin{cases} \sum_{\substack{i \in N \\ \pi_k \in A_i^{\pi_j}}} \xi_i(\pi_k|\pi_j) \prod_{\ell \neq i} \xi_\ell(\pi_j|\pi_j) \\ + \prod_{\substack{i \in N \\ \pi_k \in A_i^{\pi_j}}} \xi_i(\pi_k|\pi_j) \cdot \prod_{\substack{\ell \in N \\ \pi_k \notin A_\ell^{\pi_j}}} \xi_\ell(\pi_j|\pi_j), & \text{if } k \neq j, \pi_k \in \bigcup_{i \in N} A_i^{\pi_j}, \\ 1 - \sum_{\ell \neq k} \lambda(\pi_\ell|\pi_j, \xi^{\pi_j}), & \text{if } k = j, \\ 0, & \text{if } \pi_k \notin \bigcup_{i \in N} A_i^{\pi_j}. \end{cases}$$

A more general transition rule is considered in [116], but the rule does not take into account neither pure nor mixed players' actions.

Stationary strategies. In the remaining part of the paper we suppose that players use only stationary strategies. Denote a stationary strategy of player $i \in N$ by $\eta_i = (\eta_i(\pi_1), \dots, \eta_i(\pi_s))$ where $\eta_i(\pi) = \xi_i^\pi \in \Xi_i^\pi$ for all $\pi \in \Pi$. The stationary strategy η_i of player i prescribes an action to the player in any game state, i.e. it prescribes the probability distribution $\xi_i^\pi \in \Xi_i^\pi$ over the set of his pure actions A_i^π for any state $\pi \in \Pi$. The set of all stationary strategies of player i is denoted by H_i . Let $\eta = (\eta_1, \dots, \eta_n)$ denote a stationary strategy profile, and $H = \prod_{i \in N} H_i$.

Transition matrix. Given stationary strategy profile $\eta \in H$, the transition function can be rewritten as transition matrix $\Lambda(\eta)$ which is a function of stationary strategy profile η . The transition matrix is of the following form:

$$\Lambda(\eta) = \begin{matrix} & \pi_1 & \dots & \pi_j & \dots & \pi_s \\ \begin{matrix} \pi_1 \\ \vdots \\ \pi_j \\ \vdots \\ \pi_s \end{matrix} & \left(\begin{array}{cccccc} \lambda(\pi_1|\pi_1, \xi^{\pi_1}) & \dots & \lambda(\pi_j|\pi_1, \xi^{\pi_1}) & \dots & \lambda(\pi_s|\pi_1, \xi^{\pi_1}) \\ \vdots & & \vdots & & \vdots \\ \lambda(\pi_1|\pi_j, \xi^{\pi_j}) & \dots & \lambda(\pi_j|\pi_j, \xi^{\pi_j}) & \dots & \lambda(\pi_s|\pi_j, \xi^{\pi_j}) \\ \vdots & & \vdots & & \vdots \\ \lambda(\pi_1|\pi_s, \xi^{\pi_s}) & \dots & \lambda(\pi_j|\pi_s, \xi^{\pi_s}) & \dots & \lambda(\pi_s|\pi_s, \xi^{\pi_s}) \end{array} \right) \end{matrix}$$

Here (j, k) -entry $\lambda(\pi_k|\pi_j, \xi^{\pi_j})$ of the matrix is the transition probability from state π_j to state π_k , provided that mixed action profile ξ^{π_j} is realised.

Now we may determine a stochastic game.

Definition 4.4. A stochastic game is a system $\Gamma = \langle N, \Pi, \{A_i, u_i\}_{i \in N}, \Lambda, \delta \rangle$, where $\delta \in (0, 1)$ is a common discount factor.

In stochastic game Γ in stationary strategies $\eta_i \in H_i$, $i \in N$, the expected discounted payoff to player i is defined as the solution of the following recurrent equation (in a vector form):

$$E_i(\eta) = u_i + \delta \Lambda(\eta) E_i(\eta) = \phi_i + \delta \Lambda(\eta) E_i(\eta),$$

or in an equivalent form

$$E_i(\eta) = (I_s - \delta \Lambda(\eta))^{-1} \phi_i \quad (4.2.3)$$

where $E_i(\eta) = (E_i^{\pi_1}(\eta), \dots, E_i^{\pi_s}(\eta))^T$, $E_i^{\pi_j}(\eta)$ is a mathematical expectation of the payoff to player i in game Γ starting from state π_j , $\phi_i = (\phi_i^{\pi_1}, \dots, \phi_i^{\pi_s})^T$ is a vector whose entries are the components of the solutions in all states, and I_s is the identity matrix of size s .

Definition 4.5. A stationary strategy profile $\eta^* = (\eta_1^*, \dots, \eta_n^*) \in H$ is the Nash equilibrium if for any player $i \in N$ and any stationary strategy $\eta_i \in H_i$ inequality (in a vector form) holds:

$$E_i(\eta_1^*, \dots, \eta_{i-1}^*, \eta_i^*, \eta_{i+1}^*, \dots, \eta_n^*) \geq E_i(\eta_1^*, \dots, \eta_{i-1}^*, \eta_i, \eta_{i+1}^*, \dots, \eta_n^*). \quad (4.2.4)$$

The existence of the Nash equilibrium in a stochastic game in stationary strategies determined above follows from [91, 217].

Theorem 4.5. In stochastic game Γ there always exists a Nash equilibrium $\eta^* = (\eta_1^*, \dots, \eta_n^*)$, $\eta_i^* \in H_i$ in stationary strategies.

§ 4.2.2 D-stable coalition structures

Consider stochastic game Γ in stationary strategies. Let $\eta^* = (\eta_1^*, \dots, \eta_n^*) \in H$ such that $\eta_i^*(\pi_j) = \xi_i^{\pi_j^*}$ for all $\pi_j \in \Pi$ and any player $i \in N$.

Definition 4.6. Coalition structure π_k is called d -stable if in game Γ there exists the Nash equilibrium in stationary strategies η^* such that $\lambda(\pi_k | \pi_k, \xi^{\pi_k^*}) = 1$, and for any other coalition structure $\pi_j \in \Pi$, $\pi_j \neq \pi_k$, the ℓ -step transition probability from π_j to π_k is positive, and ℓ is a finite number.

Whenever we refer to the d-stable coalition structure, we imply that this structure is stable with respect to the a priori prescribed solution. The solution of a game (N, v, π) , namely, the Shapley value, determines the stage payoff function u_i^π , $\pi \in \Pi$ of player $i \in N$ in stochastic game Γ of a special form. To shorten, we omit the “with respect to the Shapley value” part.

The d-stable coalition structure can be interpreted as a “limit” state of the stochastic process generated by the Nash equilibrium η^* (and transition matrix $\Lambda(\eta^*)$ as a result). Therefore, for an arbitrary initial coalition structure, the Nash equilibrium strategy profile may lead to one of the d-stable coalition structures. Note that we do not state the uniqueness of the d-stable coalition structure.

The important characteristic of the d-stable coalition structure is the expected number of steps before the game comes into this structure from any initial coalition structure π_j . This characteristic is the expected number of steps before being absorbed in a Markov chain when starting from state π_j . The d-stable coalition structure is an absorbing state in a Markov chain generated by transition matrix $\Lambda(\eta^*)$. A Markov chain is called the absorbing chain if there is at least one absorbing state and it is possible to reach at least one absorbing state from any state in a finite number of steps. Following Definition 4.6, we may say that if η^* is the Nash equilibrium and there exists more than one absorbing state in a Markov chain generated by transition matrix $\Lambda(\eta^*)$, the d-stable coalition structure exists.

To calculate the expected number of stages before the process comes into the d-stable coalition structure, we use the Markov chain theory. Let an absorbing Markov chain with transition matrix $\Lambda(\eta^*)$ have $s-p$ transient states and m absorbing states. Then $\Lambda(\eta^*)$ can be represented as

$$\Lambda(\eta^*) = \begin{pmatrix} \Theta(\eta^*) & R(\eta^*) \\ \mathbf{0} & I_p \end{pmatrix}$$

where $\Theta(\eta^*)$ is a $(s-p) \times (s-p)$ matrix, $R(\eta^*)$ is a non-zero $(s-p) \times p$ matrix, $\mathbf{0}$ is a $p \times (s-p)$ zero matrix, and I_p is the identity matrix of size p . When players realise the Nash equilibrium η^* , the expected number of steps before the game process comes into the d-stable coalition structure from all transient states is defined by entries of vector $\zeta'(\eta^*) = (I_{s-p} - \Theta(\eta^*))^{-1} \mathbf{1}$, where matrix $(I_{s-p} - \Theta)^{-1}$ is called

the fundamental matrix of transition matrix $\Lambda(\eta^*)$, and $\mathbf{1}$ is a length- $(s-p)$ column vector whose all entries are equal to 1. Following Definition 4.6, one may state that any d-stable coalition structure can be generated by the Nash equilibrium η^* such that $p = 1$ for matrix $\Lambda(\eta^*)$. Adding zero entry to vector $\zeta'(\eta^*)$ on the appropriate place (on the entry which corresponds to the d-stable coalition structure), we obtain vector $\zeta(\eta^*)$ of size s of the expected number of stages before the game comes into the d-stable coalition structure.

§ 4.2.3 D-stable coalition structures in two-person games

Consider a two-person game with characteristic function $v(\{1, 2\}) = v_{12}$, $v(\{1\}) = v_1$, $v(\{2\}) = v_2$. In stochastic game Γ the set of states is $\Pi = \{\pi_1, \pi_2\}$ where $\pi_1 = \{\{1, 2\}\}$, $\pi_2 = \{\{1\}, \{2\}\}$. Let $\phi_1 = (\phi_1^{\pi_1}, \phi_1^{\pi_2})$, $\phi_2 = (\phi_2^{\pi_1}, \phi_2^{\pi_2})$, where $\phi_1^{\pi_1} = (v_{12} + v_1 - v_2)/2$, $\phi_1^{\pi_2} = v_1$, $\phi_2^{\pi_1} = (v_{12} - v_1 + v_2)/2$, $\phi_2^{\pi_2} = v_2$.

In this section we show that in any two-person game there exists the Nash equilibrium in pure stationary strategies which generates a d-stable coalition structure. Let $\eta_1 = (\eta_1(\pi_1), \eta_1(\pi_2))$ be a stationary strategy of Player 1, $\eta_2 = (\eta_2(\pi_1), \eta_2(\pi_2))$ be a stationary strategy of Player 2 where

$$\begin{aligned}\eta_i(\pi_1) &= (\xi_i(\pi_1|\pi_1), \xi_i(\pi_2|\pi_1)), \\ \eta_i(\pi_2) &= (\xi_i(\pi_1|\pi_2), \xi_i(\pi_2|\pi_2)), \quad i = 1, 2.\end{aligned}$$

Given stationary strategy profile η , transition matrix $\Lambda(\eta)$ is as follows:

$$\Lambda(\eta) = \begin{array}{cc} & \begin{array}{c} \pi_1 \\ \pi_2 \end{array} \\ \begin{array}{c} \pi_1 \\ \pi_2 \end{array} & \begin{pmatrix} 1 - \lambda(\pi_2|\pi_1, \xi^{\pi_1}) & \lambda(\pi_2|\pi_1, \xi^{\pi_1}) \\ \lambda(\pi_1|\pi_2, \xi^{\pi_2}) & 1 - \lambda(\pi_1|\pi_2, \xi^{\pi_2}) \end{pmatrix} \end{array}$$

where

$$\begin{aligned}\lambda(\pi_2|\pi_1, \xi^{\pi_1}) &= \xi_1(\pi_2|\pi_1)\xi_2(\pi_1|\pi_1) + \xi_2(\pi_2|\pi_1)\xi_1(\pi_1|\pi_1), \\ \lambda(\pi_1|\pi_2, \xi^{\pi_2}) &= \xi_1(\pi_1|\pi_2)\xi_2(\pi_2|\pi_2) + \xi_2(\pi_1|\pi_2)\xi_1(\pi_2|\pi_2).\end{aligned}$$

In pure stationary strategies, transition matrix $\Lambda(\eta)$ can be one of four types:

$$\Lambda_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad \Lambda_2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad \Lambda_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Lambda_4 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

1. If $\Lambda(\eta) = \Lambda_1$, coalition structure π_1 is d-stable. The expected number of stages for reaching a d-stable coalition structure is 0 and 1 for states π_1 and π_2 respectively. Four pure stationary strategy profiles lead to transition matrix Λ_1 , namely, $\{\eta_1 = ((1, 0), (0, 1)), \eta_2 = ((1, 0), (1, 0))\}$, $\{\eta_1 = ((1, 0), (1, 0)), \eta_2 = ((1, 0), (0, 1))\}$, $\{\eta_1 = ((0, 1), (0, 1)), \eta_2 = ((0, 1), (1, 0))\}$, $\{\eta_1 = ((0, 1), (1, 0)), \eta_2 = ((0, 1), (0, 1))\}$. The players' payoffs in Γ are

$$E_i(\eta) = (I_2 - \delta\Lambda_1)^{-1}\phi_i = \begin{pmatrix} \phi_i^{\pi_1}/(1 - \delta) \\ v_i + \delta\phi_i^{\pi_1}/(1 - \delta) \end{pmatrix}, \quad i = 1, 2.$$

2. If $\Lambda(\eta) = \Lambda_2$, coalition structure π_2 is d-stable. The expected number of stages for reaching the d-stable coalition structure is 1 and 0 for states π_1 and π_2 respectively. Four pure stationary strategy profiles lead to transition matrix Λ_2 , namely, $\{\eta_1 = ((0, 1), (0, 1)), \eta_2 = ((1, 0), (0, 1))\}$, $\{\eta_1 = ((1, 0), (0, 1)), \eta_2 = ((0, 1), (0, 1))\}$, $\{\eta_1 = ((0, 1), (1, 0)), \eta_2 = ((1, 0), (1, 0))\}$, $\{\eta_1 = ((1, 0), (1, 0)), \eta_2 = ((0, 1), (1, 0))\}$. The players' payoffs in Γ are

$$E_i(\eta) = (I_2 - \delta\Lambda_2)^{-1}\phi_i = \begin{pmatrix} \phi_i^{\pi_1} + \delta v_i/(1 - \delta) \\ v_i/(1 - \delta) \end{pmatrix}, \quad i = 1, 2.$$

3. If $\Lambda(\eta) = \Lambda_3$, there are no d-stable coalition structures, yet four pure stationary strategy profiles lead to transition matrix Λ_3 , namely, $\{\eta_1 = ((1, 0), (0, 1)), \eta_2 = ((1, 0), (0, 1))\}$, $\{\eta_1 = ((1, 0), (1, 0)), \eta_2 = ((1, 0), (1, 0))\}$, $\{\eta_1 = ((0, 1), (0, 1)), \eta_2 = ((0, 1), (0, 1))\}$, $\{\eta_1 = ((0, 1), (1, 0)), \eta_2 = ((0, 1), (1, 0))\}$. The players' payoffs in Γ are

$$E_i(\eta) = (I_2 - \delta\Lambda_3)^{-1}\phi_i = \begin{pmatrix} \phi_i^{\pi_1}/(1 - \delta) \\ v_i/(1 - \delta) \end{pmatrix}, \quad i = 1, 2.$$

4. If $\Lambda(\eta) = \Lambda_4$, there are no d-stable coalition structures, yet four pure stationary strategy profiles lead to transition matrix Λ_4 , namely, $\{\eta_1 = ((0, 1), (1, 0)), \eta_2 = ((1, 0), (0, 1))\}$, $\{\eta_1 = ((1, 0), (1, 0)), \eta_2 = ((0, 1), (0, 1))\}$, $\{\eta_1 = ((0, 1), (0, 1)), \eta_2 = ((1, 0), (1, 0))\}$, $\{\eta_1 = ((1, 0), (0, 1)), \eta_2 = ((0, 1), (1, 0))\}$. The players' payoffs in Γ are

$$E_i(\eta) = (I_2 - \delta\Lambda_4)^{-1}\phi_i = \begin{pmatrix} (\phi_i^{\pi_1} + \delta v_i)/(1 - \delta^2) \\ (\delta\phi_i^{\pi_1} + v_i)/(1 - \delta^2) \end{pmatrix}, \quad i = 1, 2.$$

Theorem 4.6. *In two-person stochastic game Γ , in which the transition function is determined by rule (4.2.2), the following statements are true:*

1. *If $v_{12} > v_1 + v_2$, coalition structure π_1 is d-stable, and any stationary strategy profile η such that $\Lambda(\eta) = \Lambda_1$, is the Nash equilibrium.*
2. *If $v_{12} < v_1 + v_2$, coalition structure π_2 is d-stable, and any stationary strategy profile η such that $\Lambda(\eta) = \Lambda_2$, is the Nash equilibrium.*
3. *If $v_{12} = v_1 + v_2$, i.e. the characteristic function is additive, both coalition structures π_1 and π_2 are d-stable.*

Proof. If the following inequalities

$$\begin{aligned} \phi_i^{\pi_1}/(1-\delta) &\geq \phi_i^{\pi_1} + \delta v_i/(1-\delta), & v_i + \delta \phi_i^{\pi_1}/(1-\delta) &\geq v_i/(1-\delta), \\ \phi_i^{\pi_1}/(1-\delta) &\geq \phi_i^{\pi_1}/(1-\delta), & v_i + \delta \phi_i^{\pi_1}/(1-\delta) &\geq v_i/(1-\delta), \\ \phi_i^{\pi_1}/(1-\delta) &\geq (\phi_i^{\pi_1} + \delta v_i)/(1-\delta^2), & v_i + \delta \phi_i^{\pi_1}/(1-\delta) &\geq (\delta \phi_i^{\pi_1} + v_i)/(1-\delta^2) \end{aligned}$$

are satisfied for all $i = 1, 2$, then stationary strategy profile η such that $\Lambda(\eta) = \Lambda_1$, is the Nash equilibrium. These inequalities are equivalent to $\phi_i^{\pi_1} \geq v_i$, $i = 1, 2$. Taking into account inequality $v_{12} > v_1 + v_2$, we get $\phi_i^{\pi_1} > v_i$, $i = 1, 2$. From the definition of a d-stable coalition structure, we conclude that coalition structure π_1 is d-stable with respect to the Shapley value.

If the following inequalities

$$\begin{aligned} \phi_i^{\pi_1} + \delta v_i/(1-\delta) &\geq \phi_i^{\pi_1}/(1-\delta), & v_i/(1-\delta) &\geq v_i + \delta \phi_i^{\pi_1}/(1-\delta), \\ \phi_i^{\pi_1} + \delta v_i/(1-\delta) &\geq \phi_i^{\pi_1}/(1-\delta), & v_i/(1-\delta) &\geq v_i/(1-\delta), \\ \phi_i^{\pi_1} + \delta v_i/(1-\delta) &\geq (\phi_i^{\pi_1} + \delta v_i)/(1-\delta^2), & v_i/(1-\delta) &\geq (\delta \phi_i^{\pi_1} + v_i)/(1-\delta^2) \end{aligned}$$

are satisfied for all $i = 1, 2$, then stationary strategy profile η such that $\Lambda(\eta) = \Lambda_2$, is the Nash equilibrium. These inequalities are equivalent to $\phi_i^{\pi_1} \leq v_i$, $i = 1, 2$. Since $v_{12} < v_1 + v_2$, we have $\phi_i < v_i$, $i = 1, 2$. From the definition of a d-stable coalition structure, we conclude that coalition structure π_2 is d-stable with respect to the Shapley value.

The third statement of the theorem is obvious. □

§ 4.2.4 D-stable coalition structures in three-person games. Simulation study

Here we consider four numerical examples for superadditive and non-superadditive characteristic functions as well as the case of antimatroid as the collection of feasible coalitions.

Example 2. Consider a three-person game with non-superadditive characteristic function: $v(\{1, 2, 3\}) = v(\{1, 3\}) = 5$, $v(\{1, 2\}) = v(\{2, 3\}) = 6$, $v(\{1\}) = 2$, $v(\{2\}) = 3$, $v(\{3\}) = 1$. In the game there are five coalition structures $\Pi = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\}$. The coalition structures and the corresponding solutions calculated by formula (4.2.1) are represented in Table 4.11.

Let discount factor δ be 0.95. A stationary strategy of player $i \in N$ in game Γ is profile $\eta_i = (\xi_i^{\pi_1}, \xi_i^{\pi_2}, \xi_i^{\pi_3}, \xi_i^{\pi_4}, \xi_i^{\pi_5})$ where

$$\begin{aligned} \eta_1 &= ((\xi_1(\pi_1|\pi_1), \xi_1(\pi_2|\pi_1), \xi_1(\pi_3|\pi_1)), (\xi_1(\pi_1|\pi_2), \xi_1(\pi_2|\pi_2), \xi_1(\pi_3|\pi_2)), \\ &\quad (\xi_1(\pi_1|\pi_3), \xi_1(\pi_2|\pi_3), \xi_1(\pi_3|\pi_3)), (\xi_1(\pi_4|\pi_4), \xi_1(\pi_5|\pi_4)), (\xi_1(\pi_4|\pi_5), \xi_1(\pi_5|\pi_5))), \\ \eta_2 &= ((\xi_2(\pi_1|\pi_1), \xi_2(\pi_2|\pi_1), \xi_2(\pi_4|\pi_1)), (\xi_2(\pi_1|\pi_2), \xi_2(\pi_2|\pi_2), \xi_2(\pi_4|\pi_2)), \\ &\quad (\xi_2(\pi_3|\pi_3), \xi_2(\pi_5|\pi_3)), (\xi_2(\pi_1|\pi_4), \xi_2(\pi_2|\pi_4), \xi_2(\pi_4|\pi_4)), (\xi_2(\pi_3|\pi_5), \xi_2(\pi_5|\pi_5))), \\ \eta_3 &= ((\xi_3(\pi_1|\pi_1), \xi_3(\pi_3|\pi_1), \xi_3(\pi_4|\pi_1)), (\xi_3(\pi_2|\pi_2), \xi_3(\pi_5|\pi_2)), \\ &\quad (\xi_3(\pi_1|\pi_3), \xi_3(\pi_3|\pi_3), \xi_3(\pi_4|\pi_3)), (\xi_3(\pi_1|\pi_4), \xi_3(\pi_3|\pi_4), \xi_3(\pi_4|\pi_4)), \\ &\quad (\xi_3(\pi_2|\pi_5), \xi_3(\pi_5|\pi_5))). \end{aligned}$$

Each player has 108 pure stationary strategies, therefore, there are 1 259 712 pure stationary strategy profiles.

Entries $\lambda(\pi_k|\pi_j, \xi^{\pi_j})$ of a transition matrix $\Lambda(\eta) = \{\lambda(\pi_k|\pi_j, \xi^{\pi_j})\}_{j,k=1,\dots,5}$ are defined by (4.2.2). For instance, the first row of matrix $\Lambda(\eta)$ consists of the following entries:

$$\begin{aligned} \lambda(\pi_1|\pi_1, \xi^{\pi_1}) &= 1 - \lambda(\pi_2|\pi_1, \xi^{\pi_1}) - \lambda(\pi_3|\pi_1, \xi^{\pi_1}) - \lambda(\pi_4|\pi_1, \xi^{\pi_1}), \\ \lambda(\pi_2|\pi_1, \xi^{\pi_1}) &= [\xi_1(\pi_2|\pi_1)\xi_2(\pi_1|\pi_1) + \xi_2(\pi_2|\pi_1)\xi_1(\pi_1|\pi_1)] \cdot \xi_3(\pi_1|\pi_1), \\ \lambda(\pi_3|\pi_1, \xi^{\pi_1}) &= [\xi_1(\pi_3|\pi_1)\xi_3(\pi_1|\pi_1) + \xi_3(\pi_3|\pi_1)\xi_1(\pi_1|\pi_1)] \cdot \xi_2(\pi_1|\pi_1), \\ \lambda(\pi_4|\pi_1, \xi^{\pi_1}) &= [\xi_2(\pi_4|\pi_1)\xi_3(\pi_1|\pi_1) + \xi_3(\pi_4|\pi_1)\xi_2(\pi_1|\pi_1)] \cdot \xi_1(\pi_1|\pi_1), \\ \lambda(\pi_5|\pi_1, \xi^{\pi_1}) &= 0. \end{aligned}$$

There are 3 125 different possible transition matrices $\Lambda(\eta)$ in game Γ . Calculations show that there are 3 001 Nash equilibria in pure stationary strategies, and all these equilibria lead to 11 different matrices $\Lambda(\eta^*)$. The number of matrices which generate d -stable coalition structures is five (the results of calculations are represented in Table 4.12). Analyzing matrices $\Lambda(\eta^*)$, we obtain that coalition structures π_2, π_3, π_4 are d -stable.

Table 4.11: The Shapley values for Example 2.

π	ϕ_1^π	ϕ_2^π	ϕ_3^π
$\pi_1 = \{\{1\}, \{2\}, \{3\}\}$	2	3	1
$\pi_2 = \{\{1, 2\}, \{3\}\}$	2.5	3.5	1
$\pi_3 = \{\{1, 3\}, \{2\}\}$	3	3	2
$\pi_4 = \{\{1\}, \{2, 3\}\}$	2	4	2
$\pi_5 = \{\{1, 2, 3\}\}$	1.5	2.5	1

Table 4.12 presents the results of calculations for Example 2, namely: d -stable coalition structures (the first column); the number of corresponding Nash equilibria in pure stationary strategies that generate the particular transition probability matrices, and the matrices themselves (the second column); the expected payoffs of the players in game Γ and the vectors whose elements are equal to the expected number of steps before the game process enters the d -stable coalition structure (the last column).

Example 3. Now consider a three-person game with a superadditive characteristic function: $v(\{1, 2, 3\}) = 10$, $v(\{1, 3\}) = 5$, $v(\{1, 2\}) = v(\{2, 3\}) = 6$, $v(\{1\}) = 2$, $v(\{2\}) = 3$, $v(\{3\}) = 1$. The coalition structures and the corresponding solutions calculated by formula (4.2.1) are represented in Table 4.13.

Let δ be 0.95. Stationary strategy profile η and transition matrix $\Lambda(\eta)$ are defined in the previous example. In game Γ there are 3 125 different transition matrices $\Lambda(\eta)$. Calculations show that there are 3 043 Nash equilibria in pure stationary strategies, and all these equilibria lead to 26 different matrices $\Lambda(\eta^*)$. The number of matrices which generate d -stable coalition structures is 19 (the results of calculations are represented in Tables 4.14–4.16). Analyzing matrices $\Lambda(\eta^*)$, we state that there

Table 4.12: Calculations for Example 2.

π	Number of equilibria	$\Lambda_1(\eta^*)$	$E_1(\eta^*)$	$E_2(\eta^*)$	$E_3(\eta^*)$	$\zeta(\eta^*)$
π_2	2	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 48.122 \\ 50.000 \\ 49.550 \\ 48.550 \\ 49.000 \end{pmatrix}$	$\begin{pmatrix} 69.072 \\ 70.000 \\ 68.550 \\ 69.550 \\ 69.000 \end{pmatrix}$	$\begin{pmatrix} 20.950 \\ 20.000 \\ 21.000 \\ 21.000 \\ 20.000 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 0 \\ 2 \\ 2 \\ 1 \end{pmatrix}$
		$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 49.072 \\ 50.000 \\ 49.550 \\ 48.550 \\ 49.000 \end{pmatrix}$	$\begin{pmatrix} 68.122 \\ 70.000 \\ 68.550 \\ 69.550 \\ 69.000 \end{pmatrix}$	$\begin{pmatrix} 20.950 \\ 20.000 \\ 21.000 \\ 21.000 \\ 20.000 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 0 \\ 2 \\ 2 \\ 1 \end{pmatrix}$
		$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 48.122 \\ 50.000 \\ 49.550 \\ 48.550 \\ 49.000 \end{pmatrix}$	$\begin{pmatrix} 60.000 \\ 60.500 \\ 60.000 \\ 60.976 \\ 59.975 \end{pmatrix}$	$\begin{pmatrix} 39.000 \\ 39.000 \\ 40.000 \\ 38.147 \\ 38.050 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 3 \\ 2 \end{pmatrix}$
		$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 56.268 \\ 59.500 \\ 60.000 \\ 57.124 \\ 58.025 \end{pmatrix}$	$\begin{pmatrix} 60.927 \\ 60.500 \\ 60.000 \\ 60.976 \\ 59.975 \end{pmatrix}$	$\begin{pmatrix} 37.240 \\ 39.000 \\ 40.000 \\ 38.147 \\ 38.050 \end{pmatrix}$	$\begin{pmatrix} 4 \\ 1 \\ 0 \\ 3 \\ 2 \end{pmatrix}$
		$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 40.000 \\ 40.500 \\ 40.976 \\ 40.000 \\ 39.975 \end{pmatrix}$	$\begin{pmatrix} 79.000 \\ 79.500 \\ 77.124 \\ 80.000 \\ 78.025 \end{pmatrix}$	$\begin{pmatrix} 39.000 \\ 39.000 \\ 38.147 \\ 40.000 \\ 38.050 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \\ 3 \\ 0 \\ 2 \end{pmatrix}$

is a unique d -stable coalition structure π_5 .

Table 4.13: The Shapley values for Example 3.

π	ϕ_1^π	ϕ_2^π	ϕ_3^π
$\pi_1 = \{\{1\}, \{2\}, \{3\}\}$	2	3	1
$\pi_2 = \{\{1, 2\}, \{3\}\}$	2.5	3.5	1
$\pi_3 = \{\{1, 3\}, \{2\}\}$	3	3	2
$\pi_4 = \{\{1\}, \{2, 3\}\}$	2	4	2
$\pi_5 = \{\{1, 2, 3\}\}$	$3\frac{1}{6}$	$4\frac{1}{6}$	$2\frac{2}{3}$

Tables 4.14–4.16 present the results of calculations for Example 3, namely: d -stable coalition structures (the first column); the number of corresponding Nash equilibria in pure stationary strategies that generate particular transition probability matrices, and matrices themselves; the expected payoffs of the players in game Γ and vectors whose elements are equal to the expected number of steps before the game process enters the d -stable coalition structure π_5 (the last column).

Table 4.14: Results for Example 3.

Number of equilibria	$\Lambda_1(\eta^*)$	$E_1(\eta^*)$	$E_2(\eta^*)$	$E_3(\eta^*)$	$\zeta(\eta^*)$			
1	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 61.558 \\ 61.480 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 82.508 \\ 81.383 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 50.482 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \\ 3 \\ 1 \\ 0 \end{pmatrix}$			
	2	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.457 \\ 62.667 \\ 61.457 \\ 61.533 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 81.407 \\ 82.667 \\ 81.407 \\ 82.533 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 49.529 \\ 51.667 \\ 50.529 \\ 51.083 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 1 \\ 3 \\ 2 \\ 0 \end{pmatrix}$		
		1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 62.008 \\ 60.363 \\ 63.167 \\ 60.908 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 81.058 \\ 80.455 \\ 82.167 \\ 81.005 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 48.958 \\ 52.667 \\ 50.482 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 3 \\ 1 \\ 3 \\ 0 \end{pmatrix}$	
			2	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.908 \\ 62.667 \\ 63.167 \\ 62.008 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.955 \\ 82.667 \\ 82.167 \\ 82.058 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 50.432 \\ 51.667 \\ 52.667 \\ 52.033 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 0 \end{pmatrix}$
				1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.955 \\ 62.667 \\ 62.058 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.908 \\ 82.667 \\ 82.008 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 50.432 \\ 51.667 \\ 52.033 \\ 52.667 \\ 53.333 \end{pmatrix}$

Table 4.15: Results for Example 3 (cntd).

Number of equilibria	$\Lambda_1(\eta^*)$	$E_1(\eta^*)$	$E_2(\eta^*)$	$E_3(\eta^*)$	$\zeta(\eta^*)$			
1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.955 \\ 61.455 \\ 62.058 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.908 \\ 81.408 \\ 82.008 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 50.432 \\ 50.432 \\ 52.033 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 0 \end{pmatrix}$			
	1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 62.008 \\ 62.508 \\ 63.167 \\ 61.383 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 81.058 \\ 81.558 \\ 82.167 \\ 81.480 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 52.667 \\ 50.482 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \\ 1 \\ 3 \\ 0 \end{pmatrix}$		
		1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.407 \\ 62.667 \\ 62.533 \\ 61.407 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.457 \\ 82.667 \\ 81.533 \\ 81.457 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 49.529 \\ 51.667 \\ 51.083 \\ 50.529 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 1 \\ 2 \\ 3 \\ 0 \end{pmatrix}$	
			1	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 62.508 \\ 63.167 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 81.558 \\ 82.167 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 52.667 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \\ 1 \\ 1 \\ 0 \end{pmatrix}$
				2	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 60.455 \\ 61.005 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 80.363 \\ 80.908 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 48.958 \\ 50.482 \\ 52.667 \\ 53.333 \end{pmatrix}$
2					$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 62.008 \\ 61.558 \\ 63.167 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 81.058 \\ 82.508 \\ 82.167 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 52.667 \\ 52.667 \\ 53.333 \end{pmatrix}$
	1				$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 61.558 \\ 63.167 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 82.508 \\ 82.167 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 52.667 \\ 52.667 \\ 53.333 \end{pmatrix}$

Table 4.16: Results for Example 3 (cntd).

Number of equilibria	$\Lambda_1(\eta^*)$	$E_1(\eta^*)$	$E_2(\eta^*)$	$E_3(\eta^*)$	$\zeta(\eta^*)$			
1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.406 \\ 61.558 \\ 61.480 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.314 \\ 82.508 \\ 81.383 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 48.958 \\ 51.033 \\ 50.482 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 4 \\ 2 \\ 3 \\ 1 \\ 0 \end{pmatrix}$			
	2	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.908 \\ 61.408 \\ 63.167 \\ 62.008 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.955 \\ 81.455 \\ 82.167 \\ 82.058 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 50.432 \\ 50.432 \\ 52.667 \\ 52.033 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 3 \\ 1 \\ 2 \\ 0 \end{pmatrix}$		
		1	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 62.667 \\ 62.058 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 82.667 \\ 82.008 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.667 \\ 52.033 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}$	
			1	$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 62.008 \\ 62.508 \\ 63.167 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 81.058 \\ 81.558 \\ 82.167 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 51.033 \\ 52.667 \\ 52.667 \\ 53.333 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \\ 1 \\ 1 \\ 0 \end{pmatrix}$
				1	$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.336 \\ 62.667 \\ 62.533 \\ 61.407 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.384 \\ 82.667 \\ 81.533 \\ 81.457 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 49.003 \\ 51.667 \\ 51.083 \\ 50.529 \\ 53.333 \end{pmatrix}$
1					$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 60.314 \\ 62.508 \\ 63.167 \\ 61.383 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 80.406 \\ 81.558 \\ 82.167 \\ 81.480 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 48.958 \\ 51.033 \\ 52.667 \\ 50.482 \\ 53.333 \end{pmatrix}$
	1				$\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 61.058 \\ 61.455 \\ 62.058 \\ 62.167 \\ 63.333 \end{pmatrix}$	$\begin{pmatrix} 82.008 \\ 81.408 \\ 82.008 \\ 83.167 \\ 83.333 \end{pmatrix}$	$\begin{pmatrix} 51.033 \\ 50.432 \\ 52.033 \\ 52.667 \\ 53.333 \end{pmatrix}$

§ 4.3 A stochastic Prisoners' Dilemma with incomplete information on the discount factors

§ 4.3.1 Model

We consider a stochastic, infinitely-repeated prisoners' dilemma with two states $\{s_1, s_2\}$ and two players $\{i, j\}$. Time is discrete and, in any time period $t = 1, 2, \dots$, one of two states of nature can be realized. Each state is represented by a one-shot Prisoners' Dilemma. The action set of any player ν in any state s is $\mathbf{X} = \{C, D\}$. Action C stands for "to cooperate", while action D is "to deviate". The payoff function of any player ν in state s is $u^s : \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$, given by the payoff matrix:

$$\begin{array}{cc}
 & \begin{array}{c} j \\ C \quad D \end{array} \\
 \begin{array}{c} i \\ C \\ D \end{array} & \begin{array}{|cc|} \hline a^s; a^s & c^s; b^s \\ \hline b^s; c^s & d^s; d^s \\ \hline \end{array}
 \end{array} \tag{4.3.1}$$

Payoffs are symmetric among players, but change according to the state of the world.

We imply the standard assumptions about Prisoner's dilemma:

$$b^s > a^s > d^s > c^s,$$

and

$$2a^s > b^s + c^s$$

for any $s \in \{s_1, s_2\}$. The latter condition entails that, if players maximize the sum of their payoffs, they prefer the action profile (C, C) to profiles (D, C) or (C, D) in any state. This condition is not strictly necessary but aims at focusing on players who try to sustain (C, C) in each time period [101]. We assume perfect monitoring: all players observe the occurring state and the history of any time period.

The following assumption ensures that, in the one-shot game, individual deviation from action profile (C, C) is more profitable in state $s = s_2$ than in state $s = s_1$.

Assumption 4.1. *Let $b^{s_2} - a^{s_2} \geq b^{s_1} - a^{s_1}$.*

The stochastic game may start with any initial state from $\{s_1, s_2\}$. The transition from a particular state does not depend on time period and the action profile realized

in the state. The probability that the game transits from state s to state s_1 (s_2) is equal to π^s ($1 - \pi^s$). A transition from a state to another one may be interpreted as a shock.

§ 4.3.2 Publicly known discount factors

First, we consider the game where players' discount factor is public information. This analysis is convenient to later define the players' classes according to their intrinsic degree of patience. Moreover, the strategy profiles for the game with complete information are also used in the second phase of the Bayesian game, when players act as if they know the discount factor of their competitors (see Section § 4.3.3 for details).

Strategy profiles.

In this section we consider some behaviour strategy profiles which will be examined below and the conditions for which these strategy profiles are subgame perfect. We focus the analysis on pure strategies. Denote the behaviour strategy profile by $\sigma = (\sigma_\nu : \nu \in \{i, j\})$, where strategy σ_ν determines player ν 's action for any time period and any state depending on the history of the stage.

We restrict our analysis considering three behaviour strategy profiles in which players (i) play action D in any state forever (non-cooperative strategy profile $\sigma_{\mathbf{n}}$), (ii) cooperate only in state s_1 and deviate in state s_2 and transit to playing action D forever if they observe a deviation from the described behaviour in the history (semi-cooperative strategy profile $\sigma_{\mathbf{sc}}$), (iii) cooperate by playing action C in any state and transit to playing action D forever if they observe deviation in the history (cooperative strategy profile $\sigma_{\mathbf{c}}$).

There are strict mathematical definitions of the profiles used for the model. The behaviour strategy profile is given by

$$\sigma = (\sigma_\nu : \nu \in \{i, j\}). \quad (4.3.2)$$

In (4.3.2), $\sigma_\nu = \{\sigma_{\nu,t}^s\}_{t=1}^\infty$, where $\sigma_{\nu,t+1}^s : H(t) \rightarrow \mathbf{X}$ is an action of player ν in time period $t + 1$ and state $s \in \{s_1, s_2\}$. $H(t) = ((s(1), x(1)), \dots, (s(t), x(t)))$ is a history of time period t , where $s(t)$ is the state in time period t and $x(t)$ is the action profile played in state $s(t)$.

Definition 4.7. A non-cooperative strategy of player ν is denoted as $\sigma_{\mathbf{n},\nu} = \{\sigma_{\nu,t}^s\}_{t=1,\dots,\infty}^{s=s_1,s_2}$ such $\sigma_{\nu,t+1}^s(H(t)) = D$ for any $s = s_1, s_2, t = 1, \dots, \infty$ and any history $H(t)$.

We call the profile $\sigma_{\mathbf{n}} = (\sigma_{\mathbf{n},\nu} : \nu \in \{i, j\})$ as *non-cooperative strategy profile*.

Definition 4.8. A semi-cooperative strategy of player ν is denoted as $\sigma_{\mathbf{sc},\nu} = \{\sigma_{\nu,t}^s\}_{t=1,\dots,\infty}^{s=s_1,s_2}$ such that

$$\sigma_{\nu,t+1}^s(H(t)) = \begin{cases} C, & \text{if } s = s_1 \text{ and } H(t) = H_{\mathbf{sc}}(t), \\ D, & \text{otherwise,} \end{cases}$$

while $H_{\mathbf{sc}}(t)$ is a history of time period t containing only the elements $(s_1, (C, C))$ and $(s_2, (D, D))$.

We call the profile $\sigma_{\mathbf{sc}} = (\sigma_{\mathbf{sc},\nu} : \nu \in \{i, j\})$ as *semi-cooperative strategy profile*, according to which players choose action C in state s_1 and action D in state s_2 if the deviation from history $H_{\mathbf{sc}}(t)$ is not observed. Otherwise, the players switch to playing action D in any state forever.

Definition 4.9. A cooperative strategy of player ν is denoted as $\sigma_{\mathbf{c},\nu} = \{\sigma_{\nu,t}^s\}_{t=1,\dots,\infty}^{s=s_1,s_2}$ such that

$$\sigma_{\nu,t+1}^s(H(t)) = \begin{cases} C, & \text{if } H(t) = H_{\mathbf{c}}(t), \\ D, & \text{otherwise,} \end{cases}$$

and $H_{\mathbf{c}}(t) = ((s(1), (C, C)), \dots, (s(t), (C, C)))$ is a history at time period t according to which both players choose action C in all time periods before $t + 1$.

We call the profile $\sigma_{\mathbf{c}} = (\sigma_{\mathbf{c},\nu} : \nu \in \{i, j\})$ as *cooperative strategy profile*, which prescribes players to choose action C in time period $t + 1$ if the history shows past cooperation (i.e., no deviations are observed in the previous time periods). If a player observes deviation from action profile (C, C) , then she chooses action D forever.

Expected payoffs.

Let δ_ν denote the discount factor of a player ν 's payoff. We are interested in finding conditions for which the pure strategy profiles described above are subgame perfect. The strategy profile is subgame perfect if, for any time period and any state, a vector of restricted strategies form the Nash equilibrium in the subgame.

A player's discounted payoff when a strategy profile σ is implemented is

$$V(\sigma, \delta_\nu) = \sum_{t=1}^{\infty} \delta_\nu^{t-1} \Pi^{t-1} U_{\nu,t}^s(\sigma),$$

where $V(\sigma, \delta_\nu) = (V^{s_1}(\sigma, \delta_\nu), V^{s_2}(\sigma, \delta_\nu))'$, $U_{\nu,t}^s(\sigma) = (u_{\nu,t}^{s_1}(\sigma), u_{\nu,t}^{s_2}(\sigma))'(\sigma)$, with $()'$ representing the transpose matrix, $u_{\nu,t}^s(\sigma)$ is the payoff of the player ν in time period t and state s , corresponding to the strategy profile σ , and

$$\Pi = \begin{pmatrix} \pi^{s_1} & 1 - \pi^{s_1} \\ \pi^{s_2} & 1 - \pi^{s_2} \end{pmatrix}.$$

We denote the discounted payoffs of player ν in equilibria $\sigma_{\mathbf{n}}$, $\sigma_{\mathbf{sc}}$ and $\sigma_{\mathbf{c}}$ as $V_{\mathbf{n}}^s(\delta_\nu)$, $V_{\mathbf{sc}}^s(\delta_\nu)$ and $V_{\mathbf{c}}^s(\delta_\nu)$, respectively, where subscripts \mathbf{n} , \mathbf{sc} and \mathbf{c} stand for “non-cooperative”, “semi-cooperative” and “cooperative” equilibrium, while superscript $s \in \{s_1, s_2\}$ indicates the state of the game in the first period.

Lemma 4.1. *For any $\delta_\nu \in (0, 1)$, $V_{\mathbf{c}}(\delta_\nu) > V_{\mathbf{sc}}(\delta_\nu) > V_{\mathbf{n}}(\delta_\nu)$.*

Proof. We derive the value of the expected payoffs. For convenience, define

$$\tilde{\Pi}(\delta_\nu) \equiv \frac{1}{(1 - \delta_\nu)(1 - \delta_\nu(\pi^{s_1} - \pi^{s_2}))} \begin{pmatrix} 1 - \delta_\nu(1 - \pi^{s_2}) & \delta_\nu(1 - \pi^{s_1}) \\ \delta_\nu\pi^{s_2} & 1 - \delta_\nu\pi^{s_1} \end{pmatrix}.$$

We can easily calculate the player's payoff in any equilibria $\sigma_{\mathbf{n}}$, $\sigma_{\mathbf{c}}$ or $\sigma_{\mathbf{sc}}$:

1. The discounted payoff of player ν in equilibrium $\sigma_{\mathbf{n}}$ is

$$V_{\mathbf{n}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{n}}^{s_1}(\delta_\nu) \\ V_{\mathbf{n}}^{s_2}(\delta_\nu) \end{pmatrix} = \tilde{\Pi}(\delta_\nu) \begin{pmatrix} d^{s_1} \\ d^{s_2} \end{pmatrix}. \quad (4.3.3)$$

2. The discounted payoff of player ν in equilibrium $\sigma_{\mathbf{c}}$

$$V_{\mathbf{c}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{c}}^{s_1}(\delta_\nu) \\ V_{\mathbf{c}}^{s_2}(\delta_\nu) \end{pmatrix} = \tilde{\Pi}(\delta_\nu) \begin{pmatrix} a^{s_1} \\ a^{s_2} \end{pmatrix}. \quad (4.3.4)$$

3. The discounted payoff of player ν in equilibrium $\sigma_{\mathbf{sc}}$

$$V_{\mathbf{sc}}(\delta_\nu) = \begin{pmatrix} V_{\mathbf{sc}}^{s_1}(\delta_\nu) \\ V_{\mathbf{sc}}^{s_2}(\delta_\nu) \end{pmatrix} = \tilde{\Pi}(\delta_\nu) \begin{pmatrix} a^{s_1} \\ d^{s_2} \end{pmatrix}. \quad (4.3.5)$$

We obtain these formulas by calculating the payoff of player ν according to the profile definitions. The discounted payoff of player ν in equilibrium $\sigma_{\mathbf{n}}$ is

$$V_{\mathbf{n}}(\delta_{\nu}) = \begin{pmatrix} V_{\mathbf{n}}^{s_1}(\delta_{\nu}) \\ V_{\mathbf{n}}^{s_2}(\delta_{\nu}) \end{pmatrix} = \begin{pmatrix} d_{s_1} + \delta_{\nu} \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_{\nu}) \\ d_{s_2} + \delta_{\nu} \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_{\nu}) \end{pmatrix},$$

or in vectorial form:

$$V_{\mathbf{n}}(\delta_{\nu}) = \begin{pmatrix} d_{s_1} \\ d_{s_2} \end{pmatrix} + \delta_{\nu} \Pi V_{\mathbf{n}}(\delta_{\nu}).$$

This equation gives:

$$V_{\mathbf{n}}(\delta_{\nu}) = (\mathbb{I} - \delta_{\nu} \Pi)^{-1} \begin{pmatrix} d_{s_1} \\ d_{s_2} \end{pmatrix},$$

where \mathbb{I} is an identity matrix of size 2×2 . We denote $(\mathbb{I} - \delta_{\nu} \Pi)^{-1}$ by $\tilde{\Pi}(\delta_{\nu})$ and obtain the result. The inverse matrix $(\mathbb{I} - \delta_{\nu} \Pi)^{-1}$ always exists for any $\delta_{\nu} \in (0, 1)$.

Second, we calculate the discounted payoff of player ν in equilibrium $\sigma_{\mathbf{c}}$ that is:

$$V_{\mathbf{c}}(\delta_{\nu}) = \begin{pmatrix} V_{\mathbf{c}}^{s_1}(\delta_{\nu}) \\ V_{\mathbf{c}}^{s_2}(\delta_{\nu}) \end{pmatrix} = \begin{pmatrix} a^{s_1} \\ a^{s_2} \end{pmatrix} + \delta_{\nu} \begin{pmatrix} \pi^{s_1} & 1 - \pi^{s_1} \\ \pi^{s_2} & 1 - \pi^{s_2} \end{pmatrix} \begin{pmatrix} V_{\mathbf{c}}^{s_1}(\delta_{\nu}) \\ V_{\mathbf{c}}^{s_2}(\delta_{\nu}) \end{pmatrix}.$$

Rewriting this equation in vectorial form, we obtain equation (4.3.4).

Third, we calculate the discounted payoff of player player ν in equilibrium $\sigma_{\mathbf{sc}}$ that is:

$$V_{\mathbf{sc}}(\delta_{\nu}) = \begin{pmatrix} V_{\mathbf{sc}}^{s_1}(\delta_{\nu}) \\ V_{\mathbf{sc}}^{s_2}(\delta_{\nu}) \end{pmatrix} = \begin{pmatrix} a^{s_1} \\ d^{s_2} \end{pmatrix} + \delta_{\nu} \begin{pmatrix} \pi^{s_1} & 1 - \pi^{s_1} \\ \pi^{s_2} & 1 - \pi^{s_2} \end{pmatrix} \begin{pmatrix} V_{\mathbf{sc}}^{s_1}(\delta_{\nu}) \\ V_{\mathbf{sc}}^{s_2}(\delta_{\nu}) \end{pmatrix}.$$

Rewriting this equation in a vectorial form, we obtain equation (4.3.5).

Notice that, in expressions (4.3.3), (4.3.4) and (4.3.5), matrix $\tilde{\Pi}(\delta_{\nu})$ is the same.

The proof of lemma is easily obtained by expected players' payoffs $V_{\mathbf{n}}(\delta_{\nu})$, $V_{\mathbf{c}}(\delta_{\nu})$ and $V_{\mathbf{sc}}(\delta_{\nu})$ given above, together with that $a^{s_2} > d^{s_2}$ and $a^{s_1} > d^{s_1}$. \square

We are now in a position to examine the critical value of δ for which each strategy profile is a subgame perfect equilibrium. We also define

$$\delta_1^* = \frac{\Delta_1 - \pi^{s_1}(b^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_1} - a^{s_1})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1})}, \quad (4.3.6)$$

$$\delta_2^* = \frac{\Delta_2 - \pi^{s_2}(a^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) - \pi^{s_1}(b^{s_2} - a^{s_2})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2})} \quad (4.3.7)$$

where $\delta_1^* \in (0, 1)$, $\delta_2^* \in (0, 1)$, $\mathbf{p}^s = (\pi^s, 1 - \pi^s)$ for every $s \in \{s_1, s_2\}$ and

$$\begin{aligned}\Delta_1 &= ((\pi^{s_1}(b^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_1} - a^{s_1}))^2 \\ &\quad + 4(b^{s_1} - a^{s_1})(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1}))^{1/2}, \\ \Delta_2 &= ((\pi^{s_2}(a^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) + \pi^{s_1}(b^{s_2} - a^{s_2}))^2 \\ &\quad + 4(b^{s_2} - a^{s_2})(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2}))^{1/2}.\end{aligned}$$

The next theorem summarises the conditions on the discount factors for which each particular strategy profile is a subgame perfect Nash equilibrium (SPNE).

Theorem 4.7. *Let Assumption 4.1 hold and player $\nu \in \{i, j\}$ have a discount factor δ_ν which is public information. The cooperative strategy profile is SPNE iff $\delta_\nu \geq \delta_2^*$ for any $\nu \in \{i, j\}$. A semi-cooperative strategy profile is SPNE iff $\delta_\nu \geq \delta_1^*$ for any $\nu \in \{i, j\}$. A non-cooperative strategy profile is SPNE for any $\delta_\nu \in (0, 1)$, $\nu \in \{i, j\}$.*

Proof. We prove that the strategy profiles considered are subgame perfect using principle of non-profitability of one-shot deviations. Consider first the non-cooperative strategy profile. The fact that the strategy profile $\sigma_{\mathbf{n}}$ is a SPNE for every $\delta_\nu \in (0, 1)$, $\nu \in \{i, j\}$, discount factor δ is true because this profile prescribes players to choose the Nash equilibrium strategies in any state and one-shot deviations are non-profitable for any player ν with discount factor $\delta_\nu \in (0, 1)$.

Consider next the semi-cooperative strategy profile. The player ν 's payoff in this profile is given by Lemma 4.1. We need to prove that, if

$$\delta_\nu \geq \frac{\Delta_1 - \pi^{s_1}(b^{s_1} - d^{s_1}) - (1 - \pi^{s_2})(b^{s_1} - a^{s_1})}{2(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1})},$$

where

$$\begin{aligned}\Delta_1 &= ((\pi^{s_1}(b^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_1} - a^{s_1}))^2 \\ &\quad + 4(b^{s_1} - a^{s_1})(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1}))^{0.5},\end{aligned}$$

then $\sigma_{\mathbf{sc}}$ is SPNE. If a deviation is observed, in the next stage of the game the deviating player is punished by getting the Nash equilibrium payoff in any state (d^{s_1} in state s_1 and d^{s_2} in state s_2). Clearly, deviation in state s_2 is not profitable for any

$\delta_\nu \in (0, 1)$. Consider state s_1 . The strategy profile σ_{sc} is a subgame perfect Nash equilibrium if for any player $\nu \in \{i, j\}$ there is no gain from deviation in state s_1 :

$$a^{s_1} + \delta_\nu \mathbf{P}^{s_1} V_{\text{sc}}(\delta_\nu) \geq b^{s_1} + \delta_\nu \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_\nu).$$

Substituting the expressions of $V_{\text{sc}}(\delta_\nu)$ and $V_{\mathbf{n}}(\delta_\nu)$ from Lemma 4.1 and rearranging, we obtain:

$$\begin{aligned} f_1(\delta_\nu) &= \delta_\nu^2 (\pi^{s_2} - \pi^{s_1}) (b^{s_1} - d^{s_1}) + \delta_\nu \{ \pi^{s_1} (b^{s_1} - d^{s_1}) + (1 - \pi^{s_2}) (b^{s_1} - a^{s_1}) \} \\ &\quad - (b^{s_1} - a^{s_1}) \geq 0. \end{aligned}$$

Function $f_1(\delta_\nu)$ is a quadratic function satisfying the following conditions: i) $f_1(0) \leq 0$, and ii) $\lim_{\delta \rightarrow 1} f_1(\delta_\nu) > 0$. Therefore, there is a unique solution $\delta_\nu = \delta_1^*$ of the equation $f_1(\delta_\nu) = 0$ for $\delta_\nu \in (0, 1)$, where δ_1^* is given by (4.3.6). The solution of the inequality $f_1(\delta_\nu) \geq 0$ is $\delta_\nu \in [\delta_1^*, 1)$. Therefore, if for any player $\nu \in \{i, j\}$ discount factor $\delta_\nu \in [\delta_1^*, 1)$, the semi-cooperative strategy profile σ_{sc} is SPNE.

Finally, consider the cooperative strategy profile. The player ν 's payoff is given by Lemma 4.1. We need to prove that, for any

$$\delta_\nu \geq \frac{\Delta_2 - \pi^{s_2} (a^{s_1} - d^{s_1}) - (1 - \pi^{s_2}) (b^{s_2} - d^{s_2}) - \pi^{s_1} (b^{s_2} - a^{s_2})}{2(\pi^{s_2} - \pi^{s_1}) (b^{s_2} - d^{s_2})},$$

where

$$\begin{aligned} \Delta_2 &= \left((\pi^{s_2} (a^{s_1} - d^{s_1}) + (1 - \pi^{s_2}) (b^{s_2} - d^{s_2}) + \pi^{s_1} (b^{s_2} - a^{s_2}))^2 \right. \\ &\quad \left. + 4(b^{s_2} - a^{s_2}) (\pi^{s_2} - \pi^{s_1}) (b^{s_2} - d^{s_2}) \right)^{0.5}, \end{aligned}$$

the strategy profile σ_c is SPNE. If a deviation is observed, in the next stage of the game the deviating player is punished by getting the Nash payoff in any state (d^{s_1} in state s_1 and d^{s_2} in state s_2). The strategy profile σ_c is SPNE if, for any player $\nu \in \{i, j\}$, there is no gain from deviation in any state. Player ν does not deviate in state s_1 if

$$a^{s_1} + \delta \mathbf{P}^{s_1} V_c(\delta_\nu) \geq b^{s_1} + \delta_\nu \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_\nu),$$

and in state s_2 if

$$a^{s_2} + \delta_\nu \mathbf{P}^{s_2} V_c(\delta_\nu) \geq b^{s_2} + \delta_\nu \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_\nu).$$

Substituting the expressions of $V_{\mathbf{c}}(\delta_\nu)$ and $V_{\mathbf{n}}(\delta_\nu)$ from Lemma 4.1 and rearranging, we obtain:

$$f_3(\delta_\nu) = \delta_\nu^2(\pi^{s_2} - \pi^{s_1})(b^{s_1} - d^{s_1}) \\ + \delta_\nu \{ \pi^{s_1}(b^{s_1} - d^{s_1}) + (1 - \pi^{s_1})(a^{s_2} - d^{s_2}) + (1 - \pi^{s_2})(b^{s_1} - a^{s_1}) \} - (b^{s_1} - a^{s_1}) \geq 0,$$

$$f_2(\delta_\nu) = \delta_\nu^2(\pi^{s_2} - \pi^{s_1})(b^{s_2} - d^{s_2}) \\ + \delta_\nu \{ \pi^{s_2}(a^{s_1} - d^{s_1}) + (1 - \pi^{s_2})(b^{s_2} - d^{s_2}) + \pi^{s_1}(b^{s_2} - a^{s_2}) \} - (b^{s_2} - a^{s_2}) \geq 0.$$

Here $f_2(\delta_\nu)$ and $f_3(\delta_\nu)$ are quadratic functions satisfying the following conditions. Given Assumption 4.1, $0 \geq f_3(0) \geq f_2(0)$ and $\lim_{\delta \rightarrow 1} f_3(\delta_\nu) = \lim_{\delta \rightarrow 1} f_2(\delta_\nu) > 0$. Therefore, there is a unique solution $\delta_\nu = \delta_3^*$ of the equation $f_3(\delta_\nu) = 0$ for $\delta_\nu \in (0, 1)$ and a unique solution $\delta_\nu = \delta_2^*$ of the equation $f_2(\delta_\nu) = 0$ for $\delta_\nu \in (0, 1)$, where δ_2^* is given by (4.3.7). Taking into account that both $f_2(\delta_\nu)$ and $f_3(\delta_\nu)$ are non-decreasing functions, we may easily prove that the solution of the system of inequalities

$$\begin{cases} f_3(\delta_\nu) \geq 0, \\ f_2(\delta_\nu) \geq 0, \end{cases}$$

is $\delta_\nu \in [\delta_2^*, 1)$. Therefore, if for any player $\nu \in \{i, j\}$ discount factor $\delta_\nu \in [\delta_2^*, 1)$ the cooperative strategy profile $\sigma_{\mathbf{c}}$ is SPNE. \square

Depending on the parameters of the model, there are either $\delta_1^* < \delta_2^*$ or $\delta_1^* \geq \delta_2^*$. If $\delta_1^* < \delta_2^*$, there are three possible combinations of the players' i and j discount factors:

- $\delta_i \in (0, \delta_1^*)$, $\delta_j \in (0, 1)$: only $\sigma_{\mathbf{n}}$ is SPNE;
- $\delta_i \in [\delta_1^*, \delta_2^*)$, $\delta_j \in [\delta_1^*, 1)$: both $\sigma_{\mathbf{n}}$ and $\sigma_{\mathbf{sc}}$ are SPNE, but $V_{\mathbf{sc}}(\delta_\nu) > V_{\mathbf{n}}(\delta_\nu)$ by Lemma 4.1;
- $\delta_i \in [\delta_1^*, 1)$, $\delta_j \in [\delta_1^*, 1)$: strategy profiles $\sigma_{\mathbf{n}}$, $\sigma_{\mathbf{sc}}$, $\sigma_{\mathbf{c}}$ are SPNE. Again by Lemma 4.1, the payoffs in cooperative strategy profile are the largest ones.

If $\delta_1^* \geq \delta_2^*$, the interval $(\delta_1^*, 1]$, where the semi-cooperative strategy profile is SPNE, is contained into $(\delta_2^*, 1]$, where also the cooperative strategy profile is SPNE.

Hence, by Lemma 4.1, a semi-cooperative strategy profile is never played because the players' payoffs in $\sigma_{\mathbf{sc}}$ are less than in $\sigma_{\mathbf{c}}$. In this case, the analysis will be the same as in [101] where, in the interval $(0, \delta_2^*)$, only $\sigma_{\mathbf{n}}$ is SPNE (unique in the set $\{\sigma_{\mathbf{n}}, \sigma_{\mathbf{sc}}, \sigma_{\mathbf{c}}\}$), while in interval $(\delta_2^*, 1)$ the equilibrium with the largest players' payoffs is $\sigma_{\mathbf{c}}$.⁶ We then focus on the first case, and we assume the following:

Assumption 4.2. *Let $\delta_2^* > \delta_1^*$.*

Player's classes.

Based on Theorem 4.7 and Assumption 4.2, we define the players' classes according to their discount factors, as follows.

Definition 4.10. *A player ν belongs to class ℓ_ν and there are three classes $\ell_\nu \in \{I, M, P\}$:*

- i. I (impatient), whose discount factor is denoted as $\delta_\nu = \delta_I$ and satisfies $\delta_I \in (0, \delta_1^*)$;*
- ii. M (mildly patient), whose discount factor is denoted as $\delta_\nu = \delta_M$ and satisfies $\delta_M \in [\delta_1^*, \delta_2^*)$;*
- iii. P (patient), whose discount factor is denoted as $\delta_\nu = \delta_P$ and satisfies $\delta_P \in [\delta_2^*, 1)$.*

The cooperative strategy profile is SPNE if and only if both players are of class P . Otherwise, a player who is not of class P will deviate from the cooperative strategy profile because the deviation is profitable. A semi-cooperative strategy profile is SPNE if (i) both players are of class P , (ii) both players are of class M or (iii) if one is of class P and the other is of class M . If at least one of two players is of class I , then neither a cooperative nor a semi-cooperative strategy profile is SPNE.

⁶To see this, consider three options of combination of the players i and j 's discount factors:

- $\delta_i \in (0, \delta_2^*)$, $\delta_j \in (0, 1)$: only $\sigma_{\mathbf{n}}$ is SPNE.
- $\delta_i \in [\delta_2^*, \delta_1^*)$, $\delta_j \in [\delta_2^*, 1)$: both $\sigma_{\mathbf{n}}$ and $\sigma_{\mathbf{c}}$ are SPNE. But in the cooperative equilibrium the players' payoffs are larger than in the non-cooperative equilibrium.
- $\delta_i \in [\delta_2^*, 1)$, $\delta_j \in [\delta_1^*, 1)$: strategy profiles $\sigma_{\mathbf{n}}$, $\sigma_{\mathbf{sc}}$, $\sigma_{\mathbf{c}}$ are SPNE. The equilibrium with the largest players' payoffs is $\sigma_{\mathbf{c}}$.

§ 4.3.3 Two-phase game with unknown discount factors

We now turn the analysis on the case where a player's discount factor is private information. Possibly, this game may exhibit several classes of equilibria. We focus our analysis on some equilibria for the game which consist on two phases [101].

The first phase is *learning*, where the players' discount factors are private information and players try to recognize the competitor's class. In this phase players' strategies are Markovian: they are based on beliefs on the competitor's class, and not on the game history. The second phase is *cooperation*, where any player uses a behaviour strategy from the set $\{\sigma_{\mathbf{c},\nu}, \sigma_{\mathbf{sc},\nu}, \sigma_{\mathbf{n},\nu}\}$. The strategy chosen during the cooperation phase is determined by the beliefs on the competitor's class at the last period of the learning phase.

Learning phase: strategies and rules for updating beliefs.

In this section, we describe how the process of learning the competitor's class takes place. In time period t , a player believes the other player to be of class P with probability α_t , to be of class M with probability β_t , and to be of class I with probability $\gamma_t = 1 - \alpha_t - \beta_t$. We define the strategies for every $t = 1, \dots, T$: the player chooses her strategy in period t based on beliefs α_t and β_t . We assume that the initial beliefs about the other player's class in period 1 are given and known, and $\alpha_1 \in (0, 1)$, $\beta_1 \in (0, 1)$ are known and $\alpha_1 + \beta_1 \in (0, 1)$.

We denote player ν 's strategy in the learning phase as ψ_ν and the set of player ν 's strategies in the learning phase is $\Psi(\ell_\nu)$ which depends on the player's class $\ell_\nu \in \{I, M, P\}$. The set of Markovian strategies of a player of class P is $\Psi(P) = \{q_t^s, t = 1, \dots, T, s = s_1, s_2\}$, where $q_t^s : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is a function of α_t and β_t and corresponds to the probability of choosing action C in state s in time period t . Conversely, the set of Markovian strategies of a player of class M is $\Psi(M) = \{r_t^s, t = 1, \dots, T, s = s_1, s_2\}$. Notice that $r_t^{s_1} : [0, 1] \times [0, 1] \rightarrow [0, 1]$ is a function of α_t and β_t and amounts to the probability of choosing action C in state s_1 in time period t . For state s_2 , $r_t^{s_2} = 0$ for any t , i.e. the strategy $r_t^{s_2}$ prescribes player of class M to choose action D in state s_2 with probability 1 in any time period. Finally, we define the set of strategies for an impatient player ν , $\ell_\nu = I$. The set of Markovian strategies of the I player in the learning phase is

$\Psi(I) = \{z_t^s, t = 1, \dots, T, s = s_1, s_2\}$, where $z_t^s : [0, 1] \times [0, 1] \rightarrow 0$ for any period t and any state s . Strategy z_t^s prescribes player of class I to choose action D in any state $s = s_1, s_2$ with probability 1 in any time period.

The player ν 's strategy in the learning phase is $\psi(\ell_\nu)$ and determines a probability of choosing action C in any time period t and any state s . It depends on player's class ℓ_ν as follows:

$$\psi(\ell_\nu) = \begin{cases} q_t^s \in [0, 1], & \text{if } \ell_\nu = P, \quad s = s_1, s_2; \\ r_t^s \in [0, 1], & \text{if } \ell_\nu = M, \quad s = s_1; \\ r_t^s = 0, & \text{if } \ell_\nu = M, \quad s = s_2; \\ z_t^s = 0, & \text{if } \ell_\nu = I, \quad s = s_1, s_2. \end{cases}$$

Based on the definition of the players' strategies in the learning phase, we may state that, if a player chooses action C in state s_2 , she has revealed her class as P , because she is the only class who may choose action C in state s_2 with positive probability. Conversely, if a player chooses action C in state s_1 , she may be identified as a P or M class because only players of these two classes may choose action C in state s_1 with positive probability.

We use Bayes rule to update beliefs α_t and β_t over time, $t = 2, \dots, T + 1$. The rule of defining beliefs for period $t + 1$ depends on the state s which appeared in period t . If state s is realized at period t , then in the next period $t + 1$, the belief that the competitor is of class P is

$$\alpha_{t+1} = \begin{cases} \alpha_{t+1}^{s_1}, & \text{if } s = s_1, \\ \alpha_{t+1}^{s_2}, & \text{if } s = s_2. \end{cases}$$

The same rule applies to β_{t+1} and γ_{t+1} .

First, consider the updating rule for state $s = s_1$. If a player chooses C in period t , she is identified as class⁷

⁷The probabilities $\alpha_{t+1}^{s_1}$ and $\beta_{t+1}^{s_1}$ are defined if $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} \neq 0$. In the case when $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} = 0$, it is impossible to observe action C in state $s = s_1$.

$$\left\{ \begin{array}{l} I \quad \text{with prob } \gamma_{t+1}^{s_1} = 0; \\ P \quad \text{with prob } \alpha_{t+1}^{s_1} = \frac{\alpha_t q_t^{s_1}}{\alpha_t q_t^{s_1} + \beta_t r_t^{s_1}}; \\ M \quad \text{with prob } \beta_{t+1}^{s_1} = 1 - \alpha_{t+1}^{s_1} = \frac{\beta_t r_t^{s_1}}{\alpha_t q_t^{s_1} + \beta_t r_t^{s_1}}. \end{array} \right. \quad (4.3.8)$$

If a player chooses D in state $s = s_1$, she is identified as class⁸

$$\left\{ \begin{array}{l} I \quad \text{with prob } \gamma_{t+1}^{s_1} = 1 - \alpha_{t+1}^{s_1} - \beta_{t+1}^{s_1} = \frac{1 - \alpha_t - \beta_t}{1 - \alpha_t q_t^{s_1} - \beta_t r_t^{s_1}}; \\ P \quad \text{with prob } \alpha_{t+1}^{s_1} = \frac{\alpha_t(1 - q_t^{s_1})}{1 - \alpha_t q_t^{s_1} - \beta_t r_t^{s_1}}; \\ M \quad \text{with prob } \beta_{t+1}^{s_1} = 1 - \alpha_{t+1}^{s_1} = \frac{\beta_t(1 - r_t^{s_1})}{1 - \alpha_t q_t^{s_1} - \beta_t r_t^{s_1}}. \end{array} \right. \quad (4.3.9)$$

Consider next the updating rule for state s_2 . As already discussed, if a player chooses C , she is identified as class P with probability 1. On the other hand, if a player chooses D , she is identified as class⁹

$$\left\{ \begin{array}{l} I \quad \text{with prob } \gamma_{t+1}^{s_2} = 1 - \alpha_{t+1}^{s_2} - \beta_{t+1}^{s_2} = \frac{1 - \alpha_t - \beta_t}{1 - \alpha_t q_t^{s_2}}; \\ P \quad \text{with prob } \alpha_{t+1}^{s_2} = \frac{\alpha_t(1 - q_t^{s_2})}{1 - \alpha_t q_t^{s_2}}; \\ M \quad \text{with prob } \beta_{t+1}^{s_2} = \frac{\beta_t}{1 - \alpha_t q_t^{s_2}}. \end{array} \right. \quad (4.3.10)$$

Cooperation phase.

We now determine how players choose strategies in the cooperation phase. The strategy of player ν in a cooperation phase is a mapping from a player's class and beliefs on the other player's class by the end of the learning phase to the set $\{\sigma_{\mathbf{c},\nu}, \sigma_{\mathbf{sc},\nu}, \sigma_{\mathbf{n},\nu}\}$. Let T be the last period of the learning phase, which we assume finite.¹⁰

At the end of the learning phase, the beliefs that the competitor is of class P or M are α_{T+1} and β_{T+1} , respectively. Thus the strategy of player ν is a function of

⁸The probabilities $\alpha_{t+1}^{s_1}$, $\beta_{t+1}^{s_1}$ and $\gamma_{t+1}^{s_1}$ are defined if $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} \neq 1$. In the case when $\alpha_t q_t^{s_1} + \beta_t r_t^{s_1} = 1$, it is impossible to observe action D in state $s = s_1$.

⁹The probabilities $\alpha_{t+1}^{s_2}$, $\beta_{t+1}^{s_2}$ and $\gamma_{t+1}^{s_2}$ are defined if $\alpha_t q_t^{s_2} \neq 1$. In the case when $\alpha_t q_t^{s_2} = 1$ ($\alpha_t = q_t^{s_2} = 1$), it is impossible to observe action D in state $s = s_2$.

¹⁰Here we do not consider the case when the learning phase last forever.

class ℓ_ν and beliefs $\alpha_{T+1}, \beta_{T+1}$ such that:

$$\sigma_\nu(\ell_\nu, \alpha_{T+1}, \beta_{T+1}) = \begin{cases} \sigma_{\mathbf{c},\nu}, & \text{if } \ell_\nu = P, \quad \alpha_{T+1} = 1, \\ \sigma_{\mathbf{sc},\nu}, & \text{if } \{\ell_\nu = P, \quad \alpha_{T+1} + \beta_{T+1} = 1, \quad \alpha_{T+1} \neq 1\} \\ & \text{or } \{\ell_\nu = M, \quad \alpha_{T+1} + \beta_{T+1} = 1\}, \\ \sigma_{\mathbf{n},\nu}, & \text{if } \{\ell_\nu \in \{P, M\}, \quad \alpha_{T+1} + \beta_{T+1} < 1\} \\ & \text{or } \{\ell_\nu = I\}. \end{cases} \quad (4.3.11)$$

From the rule (4.3.11) it follows that the impatient player always chooses strategy $\sigma_{\mathbf{n},\nu}$, a mildly patient player chooses a semi-cooperative strategy $\sigma_{\mathbf{sc},\nu}$ if he has a belief that the competitor is not an impatient player, and he chooses non-cooperative strategy $\sigma_{\mathbf{n},\nu}$ otherwise. A patient player chooses cooperative strategy $\sigma_{\mathbf{c},\nu}$ only if her belief that the competitor is patient equals to one. She chooses semi-cooperative strategy $\sigma_{\mathbf{sc},\nu}$ if believes that the competitor is not an impatient player, but her belief that the competitor of class P does not equal to one. Otherwise, a patient player chooses the non-cooperative strategy $\sigma_{\mathbf{n},\nu}$.

Payoffs and equilibrium concept.

The payoff of player $\nu = i, j$ whose class is $\ell_\nu \in \{I, M, P\}$ is the sum of her payoffs in the two phases of the game and it is a function of her class, initial beliefs and the strategies of player i and j in the learning phase:¹¹

$$\begin{aligned} \Phi_\nu(\psi_i, \psi_j, \alpha_1, \beta_1 | \ell_\nu) &= \sum_{t=1}^T \delta_\nu^{t-1} \Pi^{t-1} U_\nu(\psi_{i,t}, \psi_{j,t}) \\ &+ \delta_\nu^T \Pi^T V(\delta_\nu, (\sigma_i(\ell_i, \alpha_{T+1}, \beta_{T+1}), \sigma_j(\ell_j, \alpha_{T+1}, \beta_{T+1}))), \end{aligned} \quad (4.3.12)$$

where $\psi_\nu \in \Psi(\ell_\nu)$. The first part in the RHS of (4.3.12) is the payoff during the learning phase, while the second part is the payoff during the cooperation phase. In (4.3.12), $\sigma_i(\ell_i, \alpha_{T+1}, \beta_{T+1})$ and $\sigma_j(\ell_j, \alpha_{T+1}, \beta_{T+1})$ are defined by (4.3.11).

We are now equipped to define the equilibrium concept of the game. The equilibrium concept is perfect Bayesian equilibrium [94]. In a perfect Bayesian equilibrium

¹¹We omit players' strategies in a cooperation phase as the arguments of the function because they are uniquely defined by rule (4.3.11) given strategies ψ_i, ψ_j .

for each period $t = 1, \dots, T$ and any history of this period, the continuation strategies are a Bayesian equilibrium for the continuation game. Formally, we define

Definition 4.11. *A strategy profile $\psi^* = (\psi_i^*, \psi_j^*)$ is a perfect Bayesian equilibrium if for each player $\nu \in \{i, j\}$, any class $\ell_\nu \in \{M, P\}$, any initial beliefs $\alpha_1 \in (0, 1)$, $\beta_1 \in (0, 1)$, $\alpha_1 + \beta_1 \in (0, 1)$ and each strategy $\psi_\nu \in \Psi(\ell_\nu)$, the inequality*

$$\Phi_\nu(\psi^*, \alpha_1, \beta_1 | \ell_\nu) \geq \Phi_\nu((\psi_\nu, \psi_{-\nu}^*), \alpha_1, \beta_1 | \ell_\nu) \quad (4.3.13)$$

holds.

The strategy set of a player in the two-phase equilibrium consists of the strategy in the learning and in the cooperation phases. The solution concept is close to Markov Perfect Bayesian Equilibrium (MPBE) with the following modification. The strategy of any player is Markovian only during the learning phase when players' classes are not common knowledge, and in the cooperation phase players' strategies are behaviour from the set $\{\sigma_{\mathbf{c}, \nu}, \sigma_{\mathbf{sc}, \nu}, \sigma_{\mathbf{n}, \nu}\}$, as described in (4.3.11). To avoid confusion, we use the name of the solution concept as Partial Markov Perfect Bayesian Equilibrium¹² (PMPBE) given in [101].

§ 4.3.4 Two-phase game with cooperative and semi-cooperative outcomes

In this section we characterize equilibria according to which at the second phase of the game the cooperative or semi-cooperative strategy profiles may occur (depending on the players' classes). The rule of choosing the strategy in the cooperation phase is given by (4.3.11). In the first part of the section, we consider the case where the learning phase is limited to one period. Albeit this restriction is strong, it allows us to highlight some features of the equilibria that may be then found, in the second part, in the more general version where the length of the learning phase is endogenously determined.

One-period learning phase.

In this section we limit the length of the learning phase T to one time period. We present first this simplifying case for expositional purposes, as it helps highlighting

¹²It is called "partial" because, in the cooperation phase, the strategies are not Markovian but they are behaviour once the learning phase ends up.

the role of beliefs in this problem. Qualitatively similar results are obtained when we relax this assumption, but the analysis and the equilibrium conditions are more cumbersome. Nonetheless, the exogenous duration of the learning phase may be dictated by external conditions, e.g., the learning period may be costly and players are restricted by short duration of a learning phase.

We find the conditions when a Partially Markov Perfect Bayesian Equilibrium exists in the learning phase. We sort the equilibria by the type of equilibria adopted in the learning phase. For convenience, thresholds A_1 , A_2 , A_3 and A_4 are defined in the proof of the next theorem.

Theorem 4.8. *Consider a Prisoners' dilemma with the learning phase lasts for one time period. Let Assumptions 4.1 and 4.2 be satisfied, then there exists the following equilibria:*

1. *If the initial state is $s = s_1$:*

1.i $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ is a PMPBE for $\alpha_1 \in [A_1; A_2]$.

1.ii $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ is a PMPBE for $\alpha_1 + \beta_1 \geq A_3$.

2. *If the initial state is $s = s_2$:*

2.i $(q_1^{s_2}) = 1$ is a PMPBE for $\alpha_1 \geq A_4$.

Proof. Before proving the theorem, it is convenient to introduce the following strategy profile, as it may emerge in the case of deviation. We will next proceed with the proof.

Deviating strategy profile

Definition 4.12. A “deviating strategy profile” is denoted as $\sigma_{\mathbf{d}} = (\sigma_{\mathbf{d},i}, \sigma_{\mathbf{c},j})$, where

$\sigma_{\mathbf{d},i} = \{\sigma_{i,t}^s\}_{t=1,\dots,\infty}^{s=s_1,s_2}$ such that

$$\sigma_{i,t+1}^s(H(t)) = \begin{cases} C, & \text{if } s = s_1 \text{ and } H(t) = H_{\mathbf{c}}(t) \\ D, & \text{if } s = s_1 \text{ and } H(t) \neq H_{\mathbf{c}}(t) \cdot \\ D, & \text{if } s = s_2 \end{cases}$$

In this profile, a player j plays strategy $\sigma_{\mathbf{c},j}$ given by Definition 4.9 while player i applies strategy $\sigma_{\mathbf{d},i}$. This profile may occur when player j has a belief that the competitor i will play cooperatively while she will in fact deviate in state $s = s_2$. In turn, when player j observes a deviation from the cooperative strategy profile, she reacts with D in all stages afterwards according to strategy $\sigma_{\mathbf{c},j}$.

Denote by $V_{\mathbf{d}}(\delta_i)$ an expected payoff of deviating player i in strategy profile $\sigma_{\mathbf{d}}$. We compute the expected payoff $V_{\mathbf{d},i}(\delta_i)$ of a deviating player i which is:

$$V_{\mathbf{d}}(\delta_i) = \begin{pmatrix} V_{\mathbf{d}}^{s_1}(\delta_i) \\ V_{\mathbf{d}}^{s_2}(\delta_i) \end{pmatrix},$$

where $V_{\mathbf{d}}^s(\delta_i)$ is the payoff of player i in the subgame starting from state s . If the subgame starts from state s_1 , player i gets

$$V_{\mathbf{d}}^{s_1}(\delta_i) = a^{s_1} + \delta(\pi^{s_1}V_{\mathbf{d},i}^{s_1}(\delta_i) + (1 - \pi^{s_1})V_{\mathbf{d},i}^{s_2}(\delta_i)).$$

If the subgame starts from state s_2 , player i deviates and gets b^{s_2} . Then she will be punished by playing (D, D) in any state from the next stage until infinity. Her total payoff will be

$$V_{\mathbf{d}}^{s_2}(\delta_i) = b^{s_2} + \delta\mathbf{p}^{s_2}V_{\mathbf{n}}(\delta_i).$$

From these two equations we obtain

$$V_{\mathbf{d}}(\delta_i) = \begin{pmatrix} \frac{1}{1-\delta_i\pi^{s_1}}[a^{s_1} + \delta_i(1 - \pi^{s_1})(b^{s_2} + \delta_i\mathbf{p}^{s_2}V_{\mathbf{n}}(\delta_i))] \\ b^{s_2} + \delta\mathbf{p}^{s_2}V_{\mathbf{n}}(\delta_i) \end{pmatrix}.$$

Initial state $s = s_1$. Strategy profile $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$

Begin from a player of class P . If she does not deviate from $(1, 0)$, she gets

$$\alpha_1(a^{s_1} + \delta_P\mathbf{p}^{s_1}V_{\mathbf{c}}(\delta_P)) + (1 - \alpha_1)(c^{s_1} + \delta_P\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_P)). \quad (4.3.14)$$

If she deviates from profile $(1, 0)$ ($q_1^{s_1} = 0$), she gets:

$$\alpha_1(b^{s_1} + \delta_P\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_P)) + (1 - \alpha_1)(d^{s_1} + \delta_P\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_P)). \quad (4.3.15)$$

The deviation is not profitable if (4.3.14) is larger or equal to (4.3.15), taking into account $\delta_P \geq \delta_2^*$.

Now consider the player of class M . Her payoff in profile $(1, 0)$ is

$$\alpha_1(b^{s_1} + \delta_M\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_M)) + (1 - \alpha_1)(d^{s_1} + \delta_M\mathbf{p}^{s_1}V_{\mathbf{n}}(\delta_M)). \quad (4.3.16)$$

If she deviates from profile $(1, 0)$ (playing $r_1^{s_1} = 1$), then she gets:

$$\alpha_1(a^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{d}}(\delta_M)) + (1 - \alpha_1)(c^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)), \quad (4.3.17)$$

where $V_{\mathbf{d}}(\delta_M)$ is the payoff of class M player when she cooperates in $s = s_1$ and deviates in state $s = s_2$ (which is profitable to her according to her discount factor). The deviation is not profitable if (4.3.16) is larger or equal than (4.3.17), taking into account inequality $\delta_1^* \leq \delta_M \leq \delta_2^*$. The strategy profile $(1, 0)$ is a PMPBE when one of the following systems has a solution:

$$\begin{cases} \alpha_1 \geq \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_P \mathbf{P}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))} \\ d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \leq 0 \end{cases},$$

or

$$\begin{cases} \alpha_1 \geq \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_P \mathbf{P}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))} \\ d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) > 0 \\ \alpha_1 \leq \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M))} \end{cases}.$$

Now we need to verify the sign of expression:

$$d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)). \quad (4.3.18)$$

First, consider the difference $\delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M))$. Given Lemma 4.1, we obtain that

$$\delta_M \mathbf{P}^{s_1} V_{\mathbf{d}}(\delta_M) > \delta_M \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_M)$$

and equivalently

$$\delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) > \delta_M \mathbf{P}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \quad (4.3.19)$$

Taking into account that

$$\delta_M \mathbf{P}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \geq b^{s_1} - a^{s_1},$$

we obtain

$$d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \geq d^{s_1} - c^{s_1}.$$

Therefore, the expression (4.3.18) is positive.

Simplifying the systems and considering $\delta_P \geq \delta_2^*$, we obtain the condition:

$$\alpha_1 \in [A_1, A_2], \quad (4.3.20)$$

where

$$A_1 \equiv \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))},$$

and

$$A_2 \equiv \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}.$$

Initial state $s = s_1$. *Strategy profile* $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$

Again, we begin from a player of class P . If she does not deviate from strategy $(1, 1)$, she gets:

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)). \quad (4.3.21)$$

If she deviates from $(1, 1)$ ($q_1 = 0$), she gets:

$$(\alpha_1 + \beta_1)(b^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)) + (1 - \alpha_1 - \beta_1)(d^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)). \quad (4.3.22)$$

The deviation is not profitable if (4.3.21) is larger than or equal to (4.3.22), taking into account $\delta_P \geq \delta_2^*$.

Consider next a player of class M . Her payoff in profile $(1, 1)$ is

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)). \quad (4.3.23)$$

If she deviates from profile $(1, 1)$ ($r_1^{s_1} = 0$) she gets:

$$(\alpha_1 + \beta_1)(b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)) + (1 - \alpha_1 - \beta_1)(d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)). \quad (4.3.24)$$

The deviation is not profitable if payoff (4.3.23) is larger than or equal to (4.3.24), taking into account $\delta_1^* \leq \delta_M \leq \delta_2^*$.

Thus, the strategy profile $(1, 1)$ is a PMPBE if the following system has a solution:

$$\begin{cases} (\alpha_1 + \beta_1) [a^{s_1} - b^{s_1} - c^{s_1} + d^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))] \geq d^{s_1} - c^{s_1}, \\ (\alpha_1 + \beta_1) [a^{s_1} - b^{s_1} - c^{s_1} + d^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))] \geq d^{s_1} - c^{s_1}. \end{cases}$$

Since $\delta_M < \delta_P$, the system is equivalent to the following inequality:

$$\alpha_1 + \beta_1 \geq \frac{d^{s_1} - c^{s_1}}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}. \quad (4.3.25)$$

Initial state $s = s_2$. *Strategy profile* $(q_1^{s_2}) = (1)$

If the game starts in state s_2 , the strategy profile $(q_1^{s_2}) = (1)$ is a PMPBE if the following inequality holds:

$$\alpha_1 [a^{s_2} - b^{s_2} - c^{s_2} + d^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))] \geq d^{s_2} - c^{s_2}.$$

Since $\delta_P \geq \delta_2^*$, then $\delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P)) \geq b^{s_2} - a^{s_2}$, so that:

$$\alpha_1 \geq \frac{d^{s_2} - c^{s_2}}{a^{s_2} - b^{s_2} - c^{s_2} + d^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}. \quad (4.3.26)$$

Initial state $s = s_2$. **Profile** $(q_1^{s_2})$, $q_1^{s_2} \in (0, 1)$.

If the player of class P does not deviate from the profile $(q_1^{s_2}) = (q_1)$, his payoff will be

$$\begin{aligned} & \alpha_1 \{ (q_1)^2 (a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1 - q_1) q_1 (b^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) \\ & + (1 - q_1) q_1 (c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + (1 - q_1)^2 (d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) \} \\ & + (1 - \alpha_1) \{ q_1 (c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + (1 - q_1) (d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) \}. \end{aligned} \quad (4.3.27)$$

If he deviates from (q_1) to the pure strategy $q_1 = 0$, his payoff will be

$$\begin{aligned} & \alpha_1 \{ q_1 (b^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + (1 - q_1) (d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) \} \\ & + (1 - \alpha_1) (d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)). \end{aligned} \quad (4.3.28)$$

If he deviates from (q_1) to the pure strategy $q_1 = 1$, his payoff will be

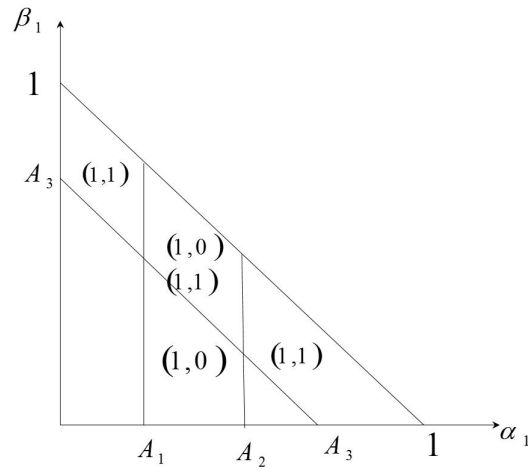
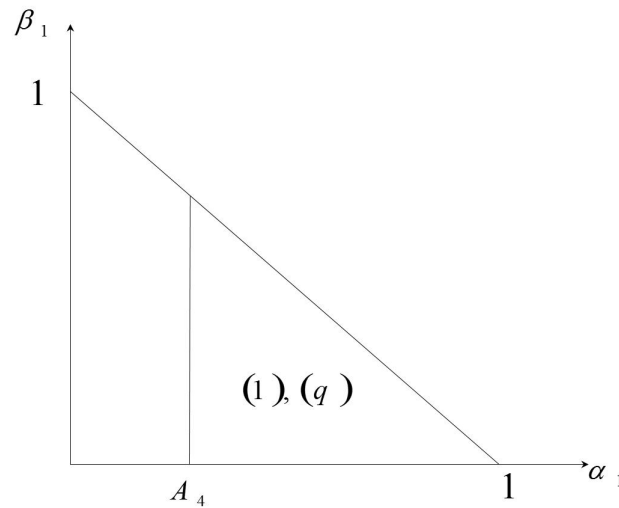
$$\begin{aligned} & \alpha_1 \{ q_1 (a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1 - q_1) (c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) \} \\ & + (1 - \alpha_1) (c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)). \end{aligned} \quad (4.3.29)$$

The deviation is not profitable if payoff (4.3.27) is larger or equal to payoffs (4.3.28) and (4.3.29). Taking into account condition $\delta_P \geq \delta_2^*$, we state that the deviation is unprofitable if

$$\alpha_1 q_1 = \frac{d^{s_2} - c^{s_2}}{d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}. \quad (4.3.30)$$

Therefore, strategy

$$q_1 = q^* \equiv \frac{d^{s_2} - c^{s_2}}{\alpha_1 (d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P)))}$$

Figure 4.8: Equilibria with initial state $s = s_1$.Figure 4.9: Equilibria with initial state $s = s_2$.

is an equilibrium if

$$\alpha_1 \geq \frac{d^{s_2} - c^{s_2}}{d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{p}^{s_2} (V_c(\delta_P) - V_n(\delta_P))}$$

or $\alpha_1 \geq A_4$. □

Figures 4.8 and 4.9 depict the regions of PMPBE for initial state s_1 and s_2 , respectively, in the space of initial beliefs (α_1, β_1) . The rule of updating beliefs in (4.3.8) and (4.3.9) helps to understand the strategy profile in the cooperation phase. Suppose, for instance, that the game starts with state s_1 and profile $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ is chosen. If action C is observed, the updated beliefs are $\alpha_2 = 1$, $\beta_2 = \gamma_2 = 0$, thus it is possible to recognize the competitor's class as P . Hence, if two players

of class P meet, the equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ leads to the cooperative strategy profile σ_c in the cooperation phase. If even one of the two players is not of class P , equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ implies that the non-cooperative strategy profile σ_n will be implemented in the cooperation phase. Indeed, since players belonging to the mildly patient and the impatient class adopt the same strategy, a patient player cannot recognize from the learning phase if the competitor is a mildly patient one, thus the semi-cooperative strategy is never used in the cooperation phase.

When the game starts at state s_1 and profile $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ is implemented, i.e., players of class M and P cooperate with probability 1, the beliefs of a competitor's class after observing C are:

$$\alpha_2 = \frac{\alpha_1}{\alpha_1 + \beta_1}, \quad \beta_2 = \frac{\beta_1}{\alpha_1 + \beta_2}, \quad \gamma_2 = 0.$$

In this case there are positive probabilities that the competitor is either P or M . Hence the strategy of the players P or M during the second phase is semi-cooperative one $\sigma_{sc,\nu}$ according to the rule (4.3.11), which allows cooperation in future states s_1 and deviation in future states s_2 . This result emerges as a player does not recognize whether the competitor is of class P or M .

When the game starts with state s_2 , a player of class M cannot be identified, since she deviates. Hence the belief β_1 does not play any role in determining the equilibrium. However, a class P competitor is identified with certainty. Hence if the players are both P , they choose cooperative strategies $\sigma_{c,\nu}$ in the cooperation phase.

The next corollary compares the equilibrium payoffs in the parameter ranges where multiple equilibria occur, as a possible refinement in the equilibrium choice.

Corollary 4.1. *Suppose the game starts from state $s = s_1$, and α_1, β_1 satisfy the conditions: $\alpha_1 \in [A_1, A_2]$ and $\alpha_1 + \beta_1 \in [A_3, 1]$. Then the payoff of a class M player in equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ is always not less than his payoff in equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$. If the game starts from state $s = s_2$ and $\alpha_1 \geq A_4$, then the payoff of class P player in equilibrium $q_1^{s_2} = 1$ is not less than the payoff in equilibrium $q_1^{s_2} = q^*$.*

Proof. The payoff of a P player in profile $(1, 0)$ is

$$\alpha_1(a^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{c}}(\delta_P)) + (1 - \alpha_1)(c^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_P)),$$

and in profile $(1, 1)$ is

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_P)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_P)).$$

The payoff of a P player in profile $(1, 1)$ is not less than his payoff in profile $(1, 0)$ if

$$\beta_1(c^{s_1} - a^{s_1} - \delta_P \mathbf{P}^{s_1} (V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))) + \alpha_1 \delta_P \mathbf{P}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P)) \leq 0,$$

or

$$\frac{\beta_1}{\alpha_1} \geq \frac{\delta_P \mathbf{P}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P))}{a^{s_1} - c^{s_1} + \delta_P \mathbf{P}^{s_1} (V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}.$$

The payoff of an M player in profile $(1, 0)$ is

$$\alpha_1(b^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)) + (1 - \alpha_1)(d^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M))$$

and in profile $(1, 1)$ is

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_M)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)).$$

The payoff of an M player in profile $(1, 1)$ is not less than his payoff in profile $(1, 0)$ if

$$\begin{aligned} \alpha_1(b^{s_1} - a^{s_1} + c^{s_1} - d^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{n}}(\delta_M) - V_{\mathbf{sc}}(\delta_M))) \\ + \beta_1(c^{s_1} - a^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{n}}(\delta_M) - V_{\mathbf{sc}}(\delta_M))) + d^{s_1} - c^{s_1} \leq 0. \end{aligned}$$

or

$$(\alpha_1 + \beta_1)(d^{s_1} - b^{s_1} + a^{s_1} - c^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))) \geq \beta_1(d^{s_1} - b^{s_1}) + (d^{s_1} - c^{s_1}). \quad (4.3.31)$$

Taking into account that $\alpha_1 + \beta_1 \geq A_3$, we may state that

$$(\alpha_1 + \beta_1)(d^{s_1} - b^{s_1} + a^{s_1} - c^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))) \geq d^{s_1} - c^{s_1}.$$

The latter inequality guarantees that (4.3.31) is satisfied because $d^{s_1} - b^{s_1} < 0$.

Now consider the initial state $s = s_2$. The payoff of class P player in equilibrium $(q_1^{s_2}) = 1$ is

$$\alpha_1(a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1 - \alpha_1)(c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)).$$

The payoff of class P player in equilibrium $(q_1^{s_2}) = q^*$ given by $q^* = A_4/\alpha_1$ (see Theorem 4.8) is

$$\begin{aligned} \alpha_1 q q (a^{s_2} - b^{s_2} + d^{s_2} - c^{s_2} + \delta_P \mathbf{P}^{s_2} (V_c(\delta_P) - V_n(\delta_P))) + \alpha_1 q (b^{s_2} - d^{s_2}) + q (c^{s_2} - d^{s_2}) \\ + d^{s_2} + \delta_P \mathbf{P}^{s_2} V_n(\delta_P) = d^{s_2} + \alpha_1 q (b^{s_2} - d^{s_2}) + \delta_P \mathbf{P}^{s_2} V_n(\delta_P). \end{aligned}$$

The payoff of class P player in profile $(q_1^{s_2}) = 1$ is not less than his payoff in profile $(q_1^{s_2}) = q$ if

$$\begin{aligned} \alpha_1 (a^{s_2} - c^{s_2} + \delta_P \mathbf{P}^{s_2} (V_c(\delta_P) - V_n(\delta_P))) \\ \geq (d^{s_2} - c^{s_2}) \left[1 + \frac{b^{s_2} - d^{s_2}}{d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2} (V_c(\delta_P) - V_n(\delta_P))} \right], \end{aligned}$$

which is always true for any $\alpha_1 \geq A_4$.

The payoff of player M in profile $(q_1^{s_2}) = 1$ is

$$\alpha_1 (b^{s_2} + \delta_M \mathbf{P}^{s_2} V_n(\delta_M)) + (1 - \alpha_1) (d^{s_2} + \delta_M \mathbf{P}^{s_2} V_n(\delta_M)),$$

and in profile $(q_1^{s_2}) = q$ is

$$\alpha_1 (q (b^{s_2} + \delta_M \mathbf{P}^{s_2} V_n(\delta_M)) + (1 - q) (d^{s_2} + \delta_M \mathbf{P}^{s_2} V_n(\delta_M))) + (1 - \alpha_1) (d^{s_2} + \delta_M \mathbf{P}^{s_2} V_n(\delta_M)).$$

The payoff of class M player in profile $(q_1^{s_2}) = 1$ is not less than his payoff in profile $(q_1^{s_2}) = q$ if

$$\alpha_1 (b^{s_2} - q b^{s_2} - (1 - q) d^{s_2}) \geq 0,$$

which is always true because $b^{s_2} > d^{s_2}$. □

Corollary 4.1 suggests equilibrium $(1, 1)$ as a refinement of multiple equilibria in state s_1 , and the pure strategy over the mixed one in state s_2 . In particular for state s_1 , this result intuitively suggests that, when the beliefs that the competitor is P or M are similar, it is unlikely to reach a result of full cooperation. Indeed, the outcome is a semi-cooperative strategy profile in the cooperation phase.

The next theorem summarises some comparative statics on the equilibrium payoffs with respect to beliefs.

Theorem 4.9. *The equilibrium payoffs of classes P and M players are increasing functions of α_1 . The payoffs of classes P and M players in equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ are increasing functions of β_1 .*

Proof. Consider the payoffs of the players of classes P and M as functions of parameter α_1 . By Theorem 4.8, there are three equilibria:

1. Equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$: the payoff of the player of class P is

$$\alpha_1(a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{c}}(\delta_P)) + (1 - \alpha_1)(c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)).$$

It is a linear function of α_1 with coefficient $a^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))$ which is positive because $a^{s_1} > c^{s_1}$ and $V_{\mathbf{c}}(\delta) > V_{\mathbf{n}}(\delta)$ for any $\delta \in (0, 1)$.

The payoff of the player of class M is

$$\alpha_1(b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)) + (1 - \alpha_1)(d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)).$$

It is also a linear function of α_1 with coefficient $b^{s_1} - d^{s_1}$ which is positive for any $\delta \in (0, 1)$.

2. Equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$: we begin with the player of class P . Her payoff is

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)).$$

It is a linear function of α_1 with coefficient $a^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))$ which is positive because $a^{s_1} > c^{s_1}$ and $V_{\mathbf{sc}}(\delta) > V_{\mathbf{n}}(\delta)$ for any $\delta \in (0, 1)$.

Then, the payoff of the player of class M is

$$(\alpha_1 + \beta_1)(a^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)) + (1 - \alpha_1 - \beta_1)(c^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)).$$

This is a linear function of α_1 with coefficient $a^{s_1} - c^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))$ which is positive because $a^{s_1} > c^{s_1}$ and $V_{\mathbf{sc}}(\delta) > V_{\mathbf{n}}(\delta)$ for any $\delta \in (0, 1)$.

The derivatives of the payoffs of the P and M players with respect to β_1 equal the corresponding derivatives subject to α_1 . Therefore, the payoffs are also increasing functions of β_1 .

3. Equilibrium $(q_1) = (1)$ in initial state $s = s_2$: the payoff of the player of class P is

$$\alpha_1(a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1 - \alpha_1)(c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P))$$

It is a linear function of α_1 with coefficient $a^{s_2} - c^{s_2} + \delta_P \mathbf{P}^{s_2} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))$ which is positive because $a^{s_2} > c^{s_2}$ and $V_{\mathbf{c}}(\delta) > V_{\mathbf{n}}(\delta)$ for any $\delta \in (0, 1)$.

4. Equilibrium $(q_1) = q^*$ with initial state $s = s_2$. Payoff of player of class P is

$$q^* = \frac{d^{s_2} - c^{s_2}}{\alpha_1(d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P)))}.$$

The payoff of class P is

$$\begin{aligned} & \alpha_1 \{q_1(a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{c}}(\delta_P)) + (1 - q_1)q_1(b^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) \\ & + (1 - q_1)q_1(c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + (1 - q_1)(d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P))\} \\ & + (1 - \alpha_1) \{q_1(c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + (1 - q_1)(d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P))\}. \end{aligned}$$

Substituting q^* in expression of the payoff and finding a derivative over α_1 , we obtain

$$\frac{A_4^2}{2\alpha_1^2} (b^{s_2} + c^{s_2} + s_2 \delta_P \mathbf{P}^{s_2} V_{\mathbf{n}}(\delta_P)) + \frac{A_4}{2\alpha_1^2} (d^{s_2} - c^{s_2}),$$

which is positive because $d^{s_2} > c^{s_2}$.

□

Theorem 4.9, together with Theorem 4.8, state a surprising result: a strong belief that the competitor is of class P does not lead to a cooperative strategy profile (fully collusive equilibrium) in the cooperation phase. This is immediately evident by looking at Figure 4.8. A very high α_1 gives a strong incentive to an M class player to fake patience, that is, it induces to act as a P class in order to lure the competitor to choose a cooperative strategy in the second phase. Indeed, if a cooperative strategy is played by a player of class P and state s_2 occurs at some period, then the M class player would deviate from cooperation, thus tricking her competitor. Given that players are aware of the “faking patience” effect, a semi-cooperative equilibrium occurs: cooperation in state s_1 , non-cooperation in state s_2 .

§ 4.3.5 Two-phase game with endogenous learning phase

In this section we generalise the previous results by endogenising the duration of the learning phase. Several equilibria emerge: in what follows, we aim at showing that the faking patience effect may occur for some configurations. We restrict our attention to those strategies that allow to identify the class of any player in the shortest number of periods.

Initial state $s = s_1$.

The natural structure of strategy profiles satisfying our requirement is the following. In the first time period, players of both class P and M use strategy C to be sure if there is an I player participating in the game. Then, if there are no I players, the game transits to state s_1 or s_2 in which players of class P and M use different strategies to be revealed in time period 2, i.e., their classes will be identified with probability 1. Now we find the conditions under which the described strategy profiles are PMPBE. Further we consider the strategy profile according to which in initial state s_1 players of classes P and M use different strategies in $t = 1$.

Assume that, in the first period, players of classes P and M adopt strategies $q_1^{s_1} = r_1^{s_1} = 1$. In period 2 and

- $s = s_1$, strategies are $q_2^{s_1} = 1, r_2^{s_1} = 0$;
- $s = s_2$, player P 's strategy is $q_2^{s_2} = 1$.

Using these strategies players' classes are revealed not later than in period 2. The following theorem summarises the conditions on the initial beliefs for which the described strategies form a PMPBE. To ease the exposition, coefficients A_5, A_6, A_7 and A_8 are defined in the proof of the next theorem, with $A_5, A_7 > 0$.

Theorem 4.10. *Let the initial state be $s = s_1$, and suppose that the following conditions hold:*

$$\left\{ \begin{array}{l} i. \quad A_5\alpha_1 + A_6\beta_1 \geq d^{s_1} - c^{s_1}, \\ ii. \quad A_7\alpha_1 + A_8\beta_1 \geq d^{s_1} - c^{s_1}, \\ iii. \quad \frac{\beta_1}{\alpha_1} \leq \min_{s=s_1, s_2} \left\{ \frac{\delta_P \mathbf{P}^s (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{sc}}(\delta_P)) - (b^s - a^s)}{d^s - c^s} \right\}, \\ iv. \quad \frac{\beta_1}{\alpha_1} \geq \frac{\delta_M \mathbf{P}^{s_1} (V_{\mathbf{d}}(\delta_M) - V_{\mathbf{sc}}(\delta_M)) - (b^{s_1} - a^{s_1})}{d^{s_1} - c^{s_1}}. \end{array} \right.$$

Then the following strategies are PMPBE:

$$P : (q_1^{s_1}, q_2^{s_1}, q_2^{s_2}) = (1, 1, 1), \quad M : (r_1^{s_1}, r_2^{s_1}) = (1, 0).$$

In the cooperation phase, a cooperative equilibrium occurs if two players are of class P , while a semi-cooperative equilibrium occurs if two players are of class M or one player is of class P and the other one is of class M .

Proof. Period 1, state s_1 . If player P follows the described strategy $q_1^{s_1} = 1$, his payoff will be

$$\begin{aligned} & \alpha_1 [a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{c}}(\delta_P)] \\ & + \beta_1 [a^{s_1} + \delta_P \pi^{s_1} (c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)) + \delta_P (1 - \pi^{s_1}) (c^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{sc}}(\delta_P))] \\ & + (1 - \alpha_1 - \beta_1) [c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)]. \end{aligned}$$

If he deviates to strategy $q_1^{s_1} = 0$, his class will be identified as I and his payoff will be

$$\alpha_1 [b^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)] + \beta_1 [b^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)] + (1 - \alpha_1 - \beta_1) [d^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P)].$$

Remembering that $\mathbf{p}^s = (\pi^s, 1 - \pi^s)$, the deviation of player P in period 1 is not profitable if

$$\begin{aligned} & \alpha_1 [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))] \\ & + \beta_1 [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \pi^{s_1} (c^{s_1} - a^{s_1}) \\ & + \delta_P (1 - \pi^{s_1}) (c^{s_2} - d^{s_2}) + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))] \\ & \geq d^{s_1} - c^{s_1}. \end{aligned}$$

We call

$$\begin{aligned} A_5 & \equiv [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))], \\ A_6 & \equiv [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_P \pi^{s_1} (c^{s_1} - a^{s_1}) + \delta_P (1 - \pi^{s_1}) (c^{s_2} - d^{s_2}) \\ & + \delta_P \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))]. \end{aligned}$$

Notice that $A_5 > 0$, given that $d^{s_1} - c^{s_1} > 0$, and that Theorem 1 combined with Lemma 1 implies

$$\delta_P > \frac{a^{s_1} - b^{s_1}}{\mathbf{p}^{s_1} (V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}.$$

If player M follows the described strategy $r_1^{s_1} = 1$, his payoff will be

$$\begin{aligned} & \alpha_1 [a^{s_1} + \delta_M \pi^{s_1} (b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_M)) + \delta_M (1 - \pi^{s_1}) (b^{s_2} + \delta_M \mathbf{p}^{s_2} V_{\mathbf{sc}}(\delta_M))] \\ & + \beta_1 [a^{s_1} + \delta_M \pi^{s_1} (d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)) + \delta_M (1 - \pi^{s_1}) (d^{s_2} + \delta_M \mathbf{p}^{s_2} V_{\mathbf{sc}}(\delta_M))] \\ & + (1 - \alpha_1 - \beta_1) [c^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)]. \end{aligned}$$

If he deviates to strategy $r_1^{s_1} = 0$, his class will be identified as I and his payoff will be

$$\alpha_1 [b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)] + \beta_1 [b^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)] + (1 - \alpha_1 - \beta_1) [d^{s_1} + \delta_M \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_M)].$$

The deviation of player M in period 1 is not profitable if

$$\begin{aligned} & \alpha_1 [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \\ & + \delta_M \pi^{s_1} (b^{s_1} - a^{s_1}) + \delta_M (1 - \pi^{s_1}) (b^{s_2} - d^{s_2})] \\ & + \beta_1 [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) \\ & + \delta_M \pi^{s_1} (d^{s_1} - a^{s_1})] \geq d^{s_1} - c^{s_1}. \end{aligned}$$

We call

$$\begin{aligned} A_7 \equiv & [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) + \delta_M \pi^{s_1} (b^{s_1} - a^{s_1}) \\ & + \delta_M (1 - \pi^{s_1}) (b^{s_2} - d^{s_2})], \end{aligned}$$

$$A_8 \equiv [a^{s_1} - b^{s_1} + d^{s_1} - c^{s_1} + \delta_M \mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M)) + \delta_M \pi^{s_1} (d^{s_1} - a^{s_1})].$$

Notice that $A_7 > 0$, given that $b^s > a^s > d^s > c^s$, and that 4.7 combined with Lemma 4.1 implies

$$\delta_M > \frac{a^{s_1} - b^{s_1}}{\mathbf{p}^{s_1} (V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}.$$

Period 2. State s_1 . If in period 1 the players' classes are not revealed, i. e. only action C was observed, then the learning phase continues and the updated beliefs are

$$\alpha_2 = \frac{\alpha_1}{\alpha_1 + \beta_1}, \quad \beta_2 = \frac{\beta_1}{\alpha_1 + \beta_1}.$$

If player P uses strategy $q_2^{s_1} = 1$, his payoff will be

$$\alpha_2 [a^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{c}}(\delta_P)] + \beta_2 [c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{sc}}(\delta_P)].$$

If he deviates to strategy $q_2^{s_1} = 0$, his class will be identified as M and his payoff will be

$$\alpha_2 [b^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\text{sc}}(\delta_P)] + \beta_2 [d^{s_1} + \delta_P \mathbf{P}^{s_1} V_{\text{sc}}(\delta_P)].$$

The deviation of player P in period 2, state s_1 , is not profitable if

$$\alpha_2 [a^{s_1} - b^{s_1} + \delta_P \mathbf{P}^{s_1} (V_{\text{c}}(\delta_P) - V_{\text{sc}}(\delta_P))] + \beta_2 [c^{s_1} - d^{s_1}] \geq 0,$$

taking into account the expressions of α_2 and β_2 , we obtain condition

$$\frac{\beta_1}{\alpha_1} \leq \frac{\delta_P \mathbf{P}^{s_1} (V_{\text{c}}(\delta_P) - V_{\text{sc}}(\delta_P)) - (b^{s_1} - a^{s_1})}{d^{s_1} - c^{s_1}}.$$

If player M uses strategy $r_2^{s_1} = 0$, his payoff will be

$$\alpha_2 [b^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\text{sc}}(\delta_M)] + \beta_2 [d^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\text{sc}}(\delta_M)].$$

If he deviates to strategy $r_2^{s_1} = 1$, his class will be identified as P and his payoff will be

$$\alpha_2 [a^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\text{d}}(\delta_M)] + \beta_2 [c^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\text{sc}}(\delta_M)].$$

The deviation of player M in period 2, state s_1 , is not profitable if

$$\alpha_2 [b^{s_1} - a^{s_1} + \delta_M \mathbf{P}^{s_1} (V_{\text{sc}}(\delta_M) - V_{\text{d}}(\delta_M))] + \beta_2 [d^{s_1} - c^{s_1}] \geq 0,$$

taking into account the expressions of α_2 and β_2 , we obtain condition

$$\frac{\beta_1}{\alpha_1} \geq \frac{\delta_M \mathbf{P}^{s_1} (V_{\text{d}}(\delta_M) - V_{\text{sc}}(\delta_M)) - (b^{s_1} - a^{s_1})}{d^{s_1} - c^{s_1}}.$$

Period 2. State s_2 . If in period 2 player P uses strategy $q_2^{s_2} = 1$, his payoff will be

$$\alpha_2 [a^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\text{c}}(\delta_P)] + \beta_2 [c^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\text{sc}}(\delta_P)].$$

If he deviates to strategy $q_2^{s_2} = 0$, his class will be identified as M and his payoff will be

$$\alpha_2 [b^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\text{sc}}(\delta_P)] + \beta_2 [d^{s_2} + \delta_P \mathbf{P}^{s_2} V_{\text{sc}}(\delta_P)].$$

The deviation of player P in period 2, state s_2 , is not profitable if

$$\alpha_2 [a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2} (V_{\text{c}}(\delta_P) - V_{\text{sc}}(\delta_P))] + \beta_2 [c^{s_2} - d^{s_2}] \geq 0.$$

Taking into account the expressions of α_2 and β_2 , we obtain condition

$$\beta_1 \leq \frac{\delta_P \mathbf{P}^{s_2} (V_{\text{c}}(\delta_P) - V_{\text{sc}}(\delta_P)) - (b^{s_2} - a^{s_2})}{d^{s_2} - c^{s_2}} \alpha_1.$$

Combining all conditions in the system we prove the theorem. □

The equilibrium described in Theorem 4.10 shows the emergence of faking patience in the first period, where a player of class M cooperates and, by doing so, does not reveal herself. On the other hand, in the second period the M -class player would deviate in state s_1 , and by doing so she reveals her type and the learning phase ends afterwards, by playing the semi-cooperative equilibrium in the cooperative phase. Intuitively, the faking patience effect is also what delays entering in the cooperation phase. Suppose that a player of type M keeps playing $r_t^{s_1} = 1$ for all periods t until a change of state takes place. In this case the learning phase goes on until state s_2 occurs.

Unlike the example where the learning phase lasts one period, the conditions of Theorem 4.10 are harder to interpret. We may however take a closer look at coefficients of α_1 in conditions *i.* and *ii.*. They are unambiguously positive, which is proved above, suggesting that an increase in α_1 increases the chance that the two conditions hold. In words, the higher the belief that the competitor is patient, the higher the change of faking patience, the less likely the reaching of full cooperation.

Initial state $s = s_2$.

In state s_2 , a player of class M always chooses action D . We examine the strategy profile according to which a player of class P chooses action C , i.e., $q_1^{s_2} = 1$. Therefore, if at least one of two players chooses action C in period 1, the learning phase is over, and the cooperation phase starts from period 2. According to rule (4.3.11), if both players choose C , then in the cooperation phase players implement a cooperative strategy profile. By contrast, if one player chooses C and the other one chooses D , then players play a non-cooperative strategy profile, according to (4.3.11).

If both players choose action D , then the learning phase transmits to period 2, and players can be either of type M or I , according to the following beliefs:

$$\alpha_2 = 0, \quad \beta_2 = \frac{\beta_1}{1 - \alpha_1}, \quad \gamma_2 = \frac{1 - \alpha_1 - \beta_1}{1 - \alpha_1}.$$

Notice that these classes of players keep playing D until state s_1 is realized because of the Markovian property of strategies. Consider then the case where state s_1 occurs in the second period. A player of class I keeps playing D . On the other hand, a player of class M may choose action C (strategy $r_2^{s_1} = 1$) or action D (strategy

$r_2^{s_1} = 0$). If he uses strategy $r_2^{s_1} = 0$, the beliefs remain the same and the strategy $r_t^{s_1}$ will be equal 0 until infinity because of the Markovian property¹³. Thus we focus on the conditions for which strategy $r_2^{s_1} = 1$ is a part of PMPBE.

Theorem 4.11. *Let the game start with state $s = s_2$, and suppose*

$$\begin{cases} \beta_1 \geq \frac{(d^{s_1} - c^{s_1})(1 - \alpha_1)}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}, \\ \beta_1 \leq \frac{\alpha_1 [d^{s_2} - c^{s_2} + a^{s_2} - b^{s_2} + \delta_P \mathbf{P}^{s_2}(V_{\mathbf{c}}(\delta_P) - V_{\mathbf{n}}(\delta_P))] - (d^{s_2} - c^{s_2})}{\delta_P \mathbf{P}^{s_2}(V_{\mathbf{sc}}(\delta_P) - V_{\mathbf{n}}(\delta_P))}. \end{cases}$$

Then the following strategies are PMPBE:

- *Class P players: $q_1^{s_2} = 1$. The learning phase is over at $t = 1$ if at least one of the players is of class P.*
- *Class M players: $r_k^{s_2} = 0$ from $k = 1$ onwards until $s = s_2$; $r_t^{s_1} = 1$, once $s = s_1$ at time $t > 1$. The learning phase is over at t if at least one of the two players is of class M.*

Proof. The deviation to strategy $r_2^{s_1} = 0$ is not profitable when

$$\begin{aligned} & \beta_2(a^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{sc}}(\delta_M)) + \gamma_2(c^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)) \\ & \geq \beta_2(b^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)) + \gamma_2(d^{s_1} + \delta_M \mathbf{P}^{s_1} V_{\mathbf{n}}(\delta_M)). \end{aligned}$$

Taking into account that $\beta_2 = \frac{\beta_1}{1 - \alpha_1}$ and $\gamma_2 = \frac{1 - \alpha_1 - \beta_1}{1 - \alpha_1}$, we obtain:

$$\beta_1 \geq \frac{(d^{s_1} - c^{s_1})(1 - \alpha_1)}{d^{s_1} - c^{s_1} + a^{s_1} - b^{s_1} + \delta_M \mathbf{P}^{s_1}(V_{\mathbf{sc}}(\delta_M) - V_{\mathbf{n}}(\delta_M))}.$$

□

§ 4.3.6 Non-cooperative outcomes

We now turn to the non-cooperative results. In this case, the non-cooperative strategy profile $\sigma_{\mathbf{n}}$ is formed by rule (4.3.11) in the cooperation phase of the game. Like in the previous section, we consider first the case where the learning phase lasts one period: the results are summarised in the following theorem.

¹³Hence, this strategy of player M is a part of a PMBE with infinite learning phase.

Theorem 4.12. *Suppose the duration of learning phase is $T = 1$. Then the strategy profiles $(q_1^{s_1}, r_1^{s_1}) = (0, 0)$ and $(q_1^{s_2}) = (0)$ are PMPBE for the game with the initial state $s = s_1$ and $s = s_2$ respectively.*

Proof. Consider initial state s_1 and the strategy profile $(q_1^{s_1}; r_1^{s_1}) = (0, 0)$. A player P obtains the following payoff if she does not deviate from $(0, 0)$:

$$d^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P). \quad (4.3.32)$$

If she deviates from profile $(0, 0)$ ($q_1^{s_1} = 1$), she gets:

$$c^{s_1} + \delta_P \mathbf{p}^{s_1} V_{\mathbf{n}}(\delta_P). \quad (4.3.33)$$

Note that (4.3.32) is always greater than or equal to (4.3.33), since $d^{s_1} \geq c^{s_1}$. A deviation of a class M cannot be profitable either.

Consider initial state s_2 and the strategy profile $(q_1^{s_2}) = (0)$. A player P obtains the following payoff if she does not deviate from (0) :

$$d^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_P). \quad (4.3.34)$$

If she deviates from profile (0) ($q_1^{s_2} = 1$), she gets:

$$c^{s_2} + \delta_P \mathbf{p}^{s_2} V_{\mathbf{n}}(\delta_P). \quad (4.3.35)$$

Note that (4.3.34) is always greater than or equal to (4.3.35), since $d^{s_2} \geq c^{s_2}$. Therefore, the strategy profile $(q_1^{s_1}, r_1^{s_1}) = (0, 0)$ (the game starts from s_1) and $(q_1^{s_2}) = (0)$ (the game starts from s_2) are PMPBE. \square

We now turn to the case where the learning phase is endogenously determined. Let the initial state be s_2 . Consider the strategy $q_1^{s_2} = 0$ of player P in period 1 in state $s = s_2$. In this case, all players use action D and, after this period, the beliefs are not updated: $\alpha_2 = \alpha_1$, $\beta_2 = \beta_1$, $\gamma_2 = \gamma_1$. If in any further periods only state $s = s_2$ is realized, then the strategy of player of type P is $q_t^{s_2} = 0$ because of the Markovian property of the strategy.

The beliefs can be changed only if state $s = s_1$ is realized in the game. Let state s_1 be realized in period $t > 1$. If in this state players use strategies $q_1^{s_1} = 0$ and $r_1^{s_1} = 0$, then the beliefs do not change and again $\alpha_t = \alpha_1$, $\beta_t = \beta_1$. Therefore,

using the Markovian property, we get by induction $q_k^{s_1} = r_k^{s_1} = 0$ for any k . These strategies determine a subgame perfect equilibrium with infinite learning phase when players always adopt action D in any state. The existence of a similar PMPBE can be proved when the game starts from state s_1 and players use actions D in this state and then in the firstly appeared state s_2 they also use actions D . The ongoing discussion can be summarised as follows.

Theorem 4.13. *For any initial probabilities $\alpha_1 > 0$, $\beta_1 > 0$ such that $\alpha_1 + \beta_1 < 1$, there always exists PMPBE in which the players' strategies for both initial states s_1 and s_2 are as follows: $q_t^{s_1} = r_t^{s_1} = 0$ and $q_t^{s_2} = 0$, $t = 1, 2, \dots$ (players of all classes choose action D in any state forever). In this case the learning phase lasts forever.*

§ 4.3.7 Numerical example

Consider the game with two states represented by matrices

$$s = s_1 : \begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} (11.3, 11.3) & (3, 12) \\ (12, 3) & (5, 5) \end{pmatrix} \end{array} \quad s = s_2 : \begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} (11, 11) & (3, 16) \\ (16, 3) & (10, 10) \end{pmatrix} \end{array}$$

for which Assumption 4.1 is true. Let the probabilities to transit from state s_1 to state s_1 and state s_2 be $\pi_1 = 0.8$ and $1 - \pi_1 = 0.2$ respectively, and the probabilities to transit from state s_2 to state s_1 and state s_2 be $\pi_2 = 0.1$ and $1 - \pi_2 = 0.9$ respectively. Assumption 4.2 is also true: we obtain the discount factors $\delta_1^* \approx 0.1245$ and $\delta_2^* \approx 0.7089$. Let players M and P have the discount factors $\delta_M = 0.7$ and $\delta_P = 0.9$ respectively.

One-period learning phase.

Consider first the game with $T = 1$. Thresholds A_1 , A_2 , A_3 and A_4 :

$$A_1 = 0.0621, \quad A_2 = 0.1781, \quad A_3 = 0.1954, \quad A_4 = 0.2930.$$

By Theorem 4.8 :

- If the game starts from state $s = s_1$:

$$- (q_1^{s_1}, r_1^{s_1}) = (1, 0) \text{ is a PMPBE if } \alpha_1 \in [A_1, A_2] = [0.0621, 0.1781].$$

Initial state	$s = s_1$	$s = s_2$	
PMPBE	$(q_1^{s_1}, r_1^{s_1}) = (1, 0)$	$(q_1^{s_1}, r_1^{s_1}) = (1, 1)$	$(q_1^{s_2}) = (1)$
Player P	$72.3243 + 39.2162\alpha_1$	$72.3243 + 34.3514(\alpha_1 + \beta_1)$	$80.8378 + 29.8919\alpha_1$
Player M	$21.2418 + 7\alpha_1$	$19.2418 + 17.2353(\alpha_1 + \beta_1)$	$31.0458 + 6\alpha_1$

Table 4.17: Equilibrium payoffs for different initial states.

– $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ is a PMPBE if $\alpha_1 + \beta_1 \geq A_3 = 0.1954$.

- If the game starts from state $s = s_2$:

– $(q_1^{s_2}) = (1)$ is the PMPBE if $\alpha_1 \geq A_4 = 0.2930$.

By Theorem 4.12 we prove that the strategy profile $(q_1^{s_1}, r_1^{s_1}) = (0, 0)$ and $(q_1^{s_2}) = (0)$ for the first and the second initial states respectively are the PMPBE with infinite learning phase.

Table 4.17 summarises the expected payoffs of classes P and M players according to each equilibrium. It shows that both players' equilibrium payoffs in the game starting from state s_2 are larger than in the game starting from state s_1 with profile $(1, 0)$. On the other hand, the payoffs in equilibrium $(1, 1)$ with which players' strategies in the cooperation phase form semi-cooperative strategy profile may be larger than the corresponding payoffs in equilibrium $(1, 0)$ with which players' strategies in the cooperation phase form cooperative strategy profile. This may happen when the initial belief that the other player is of class M is sufficiently high. We demonstrate this in Fig. 4.10, case (i).

(i) The game starts with state $s = s_1$.(ii) The game starts with state $s = s_2$.

Figure 4.10: Equilibrium region.

The yellow region is the region of parameters α_1 and β_1 for which the strategy profile $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$ is PMPBE. The light blue region is the region of the

parameters for which the strategy profile $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ is an equilibrium. The region above the blue line is the region of parameters for which the expected payoff of class P player in equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ is larger than his payoff in equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$. The region above the red line is the region of the parameters for which the payoff of class M player in equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 1)$ is larger than his payoff in equilibrium $(q_1^{s_1}, r_1^{s_1}) = (1, 0)$. The range of parameters for which the strategy profile $q_1^{s_2} = 1$ given in Theorem 4.8 are PMPBE in the game starting with $s = s_2$ is depicted in Fig. 4.10, case (ii) (orange area).

Finally, consistent with Theorem 4.9, the payoffs of P or M class players in any equilibria are the linear increasing functions of probability α_1 .

Endogenous learning phase.

Consider next an example for the case where the learning phase is endogenously determined. If the game starts from state $s = s_1$, conditions of Theorem 4.10 amounts to

$$\left\{ \begin{array}{l} 32.2162\alpha_1 + 20.1154\beta_1 \geq 2, \\ 11.4673\alpha_1 + 6.70729\beta_1 \geq 2, \\ \frac{\beta_1}{\alpha_1} \leq \min \{2.08243, 0.223938\}, \\ \frac{\beta_1}{\alpha_1} \geq 0.145989, \end{array} \right.$$

then the strategy profile $q_1^{s_1} = 1$, $r_1^{s_1} = 1$ in period 1 and $q_2^{s_2} = 1$ in period 2 is PMPBE. The region of (α_1, β_1) for which the system is satisfied (yellow color) is depicted on Fig. 4.11, case (i). We may observe the “faking patience effect” which does not allow to reach a cooperation or semi-cooperation for low values of ratio β_1/α_1 . The smaller the ratio, the higher the belief that the other player is of class P and the stronger the incentive of a player of class M to fake patience.

If the game starts from state $s = s_2$, conditions of Theorem 4.11 are equivalent to

$$\left\{ \begin{array}{l} \frac{1 - \alpha_1 - \beta_1}{\beta_1} \leq 4.11765, \\ 23.8919\alpha_1 - 15.3243\beta_1 \geq 7. \end{array} \right.$$

The range of parameters for which the strategy profile $q_1^{s_2} = 1$ and $r_t^{s_1} = 1$ for any $t > 1$ in the learning phase given in Theorem 4.11 is PMPBE in the game starting



(i) The game starts with state $s = s_1$.

(ii) The game starts with state $s = s_2$.

Figure 4.11: Equilibrium region.

with $s = s_2$ is depicted in Fig. 4.11, case (ii) (orange area).

Concluding remarks.

When the learning phase is limited to one period, a player that shifts from cooperation to deviation according to the state of the world has an incentive in faking patience in the good state. Since this behaviour is expected and increases with the belief in patience, the latter loses its role in determining cooperation. In case when the length of the learning phase is endogenously determined, the faking patience effect may still emerge. Given the multiplicity of equilibria, in the latter case we limit the attention to equilibrium configuration with pure strategies and where the cooperation phase starts as soon as possible.

An interesting extension might investigate the implementation of different strategy concepts. In the present analysis, we have considered grim trigger strategies. This class of strategies is used in folk theorems to prove the existence SPNE with cooperative outcomes. We have also referred to these class of strategies as they seemed to be natural to be used in the presence of incomplete information on the other player's discount factor [130]. Future research may analyse equilibria using another trigger strategies such as tit-for-tat strategies [70], in which at every current stage the player chooses an action that the competitor played at the previous stage. In this case though, the profile of these strategies is not subgame perfect. Alternatively, the trigger strategies with limited number of punishing periods can also be used to construct the punishment of a deviating player.

Conclusions

Models of conflict-controlled systems often need accounting of random factors that affecting the optimal behavior of system participants and their payoffs. Stochastic games are a convenient tool for modeling such systems. The thesis proposes a method of constructing a cooperative version of the game when the game process is realised on a finite graph (Chapter 1), and also when the game has infinite duration (Chapter 2). Lloyd Shapley was the first to introduce initially the concept of a stochastic game with infinite duration and two participants in 1953 [210]. It was the starting point of the development of this part of game theory. The key point of my thesis is the construction of a cooperative game, when a non-cooperative stochastic game is initially defined, as well as investigating the properties of cooperative solutions to be implemented in a game by the players with cooperative behaviour. The main principles of stable cooperation for dynamic games were formulated by Leon Petrosyan in a series of papers (e. g., see [43]). Time consistency (subgame consistency), strong time consistency, strategic support, as well as a condition of irrational behavior proof (this principle was proposed by D.W.K. Yeung [223]) can be considered as basic principles. In the thesis, the principle of time consistency of cooperative solution is formulated as a principle of subgame consistency to emphasize the importance of the “position” or the node of the graph on which the game takes place, as well as the state of the node.

Chapter 3 is devoted to studying the properties of cooperative solutions in games played over event trees. In this case, the stochastic process is not affected by the players’ strategies. This class of games is widely used in practice. The event trees allow to model possible scenarios generated from the initial event. The event trees are used to define and analyze the sequences of events that are likely to occur in reality. For this class of games, we have found the conditions under which the prop-

erties of subgame consistency, strategic support and existence of an approximated subgame perfect equilibrium are satisfied.

Chapter 4 provides application models built through stochastic games, namely: three models of data transmission in wireless networks of different topologies including model with the lack of information, a model for determining d-stable coalition structure, which is the result of solving a stochastic game constructed in a special way, and finally, a two-phase game model with incomplete information about players' discount factors for which optimal behavior of players is found and it guarantees cooperation in the game.

The main results of the work are

1. A method of constructing a cooperative stochastic game in the form of a α -characteristic function based on initially given noncooperative game with finite duration is proposed. The method of cooperative solution regularization for this class of stochastic games based on the construction of a distribution procedure of an imputation in order to obtain subgame-consistent solutions is obtained. In case of nonnegative payoff functions of the players, we have proposed a method of construction of a "new" nonnegative imputation distribution procedure and the conditions of subgame consistency of a solution for this class of games. For cooperative stochastic games with finite duration, we have formulated the principle of strong subgame consistency of the core and obtained a sufficient condition of strong subgame consistency of the core.
2. A method of constructing a cooperative stochastic game based on initially given noncooperative stochastic game in case of infinite duration and a finite set of game states is proposed. The minmax approach is used to determine a characteristic function. The principles of stable cooperation for this class of stochastic games are formulated. A method of game regularization with the use of imputation distribution procedure is proposed to construct subgame-consistent cooperative solution. The conditions of the principle of strategic support of a cooperative solution in this class of stochastic games are obtained, namely, the existence of the Nash equilibrium with payoffs equal to the components of the imputation is proved. The condition of existence of subgame perfect Nash

equilibrium with the same payoffs is also obtained. The condition for existence of a strong transferable equilibrium with cooperative payoffs is got. The principle of irrational behavior-proof for stochastic games of infinite duration and a finite set of states is formulated as well as the sufficient conditions to satisfy this principle are found.

3. A method of constructing subgame-consistent core based on the imputation distribution procedure in a cooperative stochastic game played over event tree is proposed. The conditions for the existence of subgame perfect ε -equilibrium with players' payoffs equal to their cooperative payoffs in this class of stochastic games are obtained. The conditions for the existence of subgame perfect ε -equilibrium in a regularized stochastic game played over event tree are got. In this equilibrium the players' payoffs equal the corresponding components of the imputation, on the basis of which the game has been regularized. A method of constructing subgame-consistent Shapley value in a stochastic game with a random duration played over event tree is proposed.
4. Stochastic games played over binary event trees with a linear state equation and identical players is studied. In this case, the Nash equilibrium and a cooperative solution are found. By cooperative solution, we mean the strategy profile which maximizes the total expected payoff of players. The strategy profiles are obtained in an explicit form. A formula for calculating the price of anarchy for stochastic games of this type is obtained. A model of environmental pollution control is considered, for which the Nash equilibrium, the cooperative solution and the price of anarchy are found. The price of anarchy is analyzed, as well as its lower and upper bounds as functions of the model parameters. A detailed interpretation of the results is given.
5. Game theoretical models of data transmission in wireless networks of three topologies including a model with the lack of information are proposed. As a modeling tool, it is suggested that stochastic games should be used. For three models, the Nash equilibria and cooperative solutions are found, the so-called cost of cooperation rejection is calculated, and finally, the conclusions about

the necessity of coordination of the network participants' actions are made. For one data transmission model, an imputation of the total players' payoff is found when the cooperative solution is adopted. Moreover, the regularization of the imputation to satisfy the principle of subgame consistency is implemented as well as the principle of strategic support of a cooperative solution is verified.

6. A model for finding d-stable coalition structures by solving a specially defined stochastic game is provided. Two versions of the model are considered: (i) when any coalition can be formed in the game, and (ii) when there are restrictions on the set of feasible coalitions. As a collection of feasible coalitions, it is suggested that the antimatroid should be considered. The existence of d-stable coalition structure in two-person games is proved and numerical modeling of the solution of three-person game is made, all d-stable coalition structures are found.
7. A model of two-phase stochastic Bayesian two-person game "Prisoner's Dilemma" is provided. In the game, there are two states. In case of complete information about the discount factor of the players, subgame perfect equilibria are found. The equilibria are of these types: (i) players cooperate in both states, (ii) players cooperate only in one state, (iii) players never cooperate. Depending on the value of discount factors, it is proposed to divide players into three classes. In case of incomplete information about the players' discount factors, a two-phase game is constructed. In the first phase, players update the information about the class of the other player, and in the second phase, players realize one of the subgame perfect equilibria. In the thesis, Bayesian equilibria are found in case of a one-period learning phase. They lead to the full or partial cooperation of the players in the second phase of the game. Bayesian equilibria leading to non-cooperative equilibria in the second phase are also found. A comparative analysis of players' payoffs in these equilibria is performed.

Bibliography

- [1] Baranova E. M. Mnogoshagovye stohasticheskie igry so sluchajnym vremenem okonchanija [Multistage stochastic games with random duration] // Sovremennye metody teorii kraevyh zadach. Materialy Voronezhskoj vesennej matematicheskoj shkoly «Pontrjaginskie chtenija – XV». Voronezh, 2004, pp. 22–23. (in Russian)
- [2] Baranova E. M. Uslovie D.W.K. Yeung dlja stohasticheskih kooperativnyh igr [D.W.K. Yeung condition for stochastic cooperative games] // Tezisy dokladov Mezhdunarodnogo rabocheho soveshhanija «Zadachi optimal'noj ostanovki i stohasticheskogo upravlenija», Petrozavodsk, 2005, p. 8. (in Russian)
- [3] Bellman R. Dinamicheskoe programirovanie [Dynamic programming]. M.: Izd-vo inostr. lit-ry, 1960, 400 p. (in Russian)
- [4] Berzh K. Teorija grafov i ejo primenenija [Graph theory and applications]. M.: Izd-vo inostr. lit-ry, 1962, 320 p. (in Russian)
- [5] Boltjanskij V. G., Gamkrelidze R. V., Pontrjagin L. S. K teorii optimal'nyh processov [On theory of control processes] // Doklady akademii nauk, 1956, T. 110, No. 1, pp. 7–10. (in Russian)
- [6] Bure V. M., Parilina E. M. Stohasticheskie modeli peredachi dannyh v setjah s razlichnymi topologijami [Stochastic models of data transmission in networks with different configurations] // Upravlenie bol'shimi sistemami, 2017. Vol. 68. pp. 6–29. (in Russian)
- [7] Bure V. M., Parilina E. M. Teorija verojatnostej i matematicheskaja statistika [Probability theory and mathematical statistics]. SPb.: Izd-vo “Lan”, 2013. (in Russian)

- [8] Bure V. M., Parilina E. M. Igra «Mnozhestvennyj dostup» s nepolnoj informaciej [“Multiple access” game with incomplete information] // Matematicheskaja teorija igr i ee prilozhenija, 2017. Vol. 9, no. 4, pp. 3–17. (in Russian)
- [9] Vasin A. A., Morozov V. V. Teorija igr i modeli matematicheskoj ekonomiki [Game theory and models of mathematical economics]. M.: MAKS Press, 2005.
- [10] Vilkas Je. J. Optimal’nost’ v igrah i reshenijah [Optimality in games and solutions]. M.: Nauka, 1990, 256 p. (in Russian)
- [11] Vorob’ev N. N. Ustojchivye situacii v koalicionnyh igrah [Stable profiles in coalitional games] // Doklady AN SSSR, 1960, Vol. 131, pp. 493–495. (in Russian)
- [12] Vorob’ev N. N. Koalicionnye igry [Coalitional games] // Teorija verojatnosti i ejo primenenie, 1967, Vol. 12, no. 2, pp. 289–306. (in Russian)
- [13] Vorob’ev N. N. Teorija igr dlja jekonomistov-kibernetikov [Game theory for economists-cybernetist]. M.: Nauka, 1985, 272 p. (in Russian)
- [14] Gromova E. V., Petrosyan L. A. Ob odnom sposobe postroenija harakteristicheskoi funkcii v kooperativnyh differencial’nyh igrah [On a method of construction of characteristic function in cooperative differential games] // Matematicheskaja teorija igr i ejo prilozhenija, 2015, Vol. 7, no. 4, pp. 19–39. (in Russian)
- [15] Gromova E. V., Petrosyan L. A. Sil’no dinamicheski ustojchivoje kooperativnoe reshenie v odnoj differencial’noj igre upravlenija vrednymi vybrosami [Strongly time-consistent cooperative solution in a differential game of pollution control] // Upravlenie bol’shimi sistemami, 2015, Vol. 55, pp. 140–159. (in Russian)
- [16] Zakharov V. V. O reguljarizacii i dinamicheskoj ustojchivosti reshenij ierarchicheskikh differencial’nyh igr [On regularization and time consistency of solutions in hierarchical differential games] // Vestnik Leningradskogo universiteta, 1988, Ser. 1, vol. 2, no. 8, 8 p. (in Russian)
- [17] Zakharov V. V., Petrosyan L. A. Matematicheskie modeli v jekologii [Mathematical models in ecology]. — SPb.: Izd-vo SPbGU, 1997. (in Russian)

- [18] Zakharov V. V., Shhegrjaev A. N. Ustojchivaja kooperacija v dinamičeskikh zadachah marshrutizacii transporta [Stable cooperation in dynamic transport routing] // Matematičeskaja teorija igr i ejo prilozhenija, 2012, 4 (2), pp. 39–56. (in Russian)
- [19] Kolokoltsov V. N., Malafeev O. A. Matematičeskoe modelirovanie mnogoagentnyh sistem konkurencii i kooperacii. Teorija igr dlja vseh [Mathematical modeling of multiagent systems of competition and cooperation. Game theory for everyone]. — SPb.: Izd-vo “Lan”, sec. ed., 2012, 624 p. (in Russian)
- [20] Kornienko E. A. Reshenija kooperativnyh dinamičeskikh igr [Solutions of cooperative dynamic games]. PhD thesis na soisk. uch. st. kand. fiz.-mat. n.: 01.01.09 // SPb.: SPbSU, 2003. (in Russian)
- [21] Kuzhutin D. V. Dinamičeskaja ustojchivost’ reshenij dlja nekotoryh klassov igr s nepolnoj informaciej [Time consistency of solutions for some classes of games with incomplete information]. PhD thesis na soisk. uch. st. kand. fiz.-mat. n.: 01.01.09 // SPb.: SPbSU, 1993. (in Russian)
- [22] Mazalov V. V. Matematičeskaja teorija igr i prilozhenija [Mathematical game theory and applications]. — SPb.: Sankt-Peterburg-Moskva-Krasnodar, Izd-vo “Lan”, 2nd. ed., 2016, 446 p. (in Russian)
- [23] Malafeev O. A. O sushhestvovanii situacii ε -ravnovesija v dinamičeskikh igrakh s zavisimymi dvizhenijami [On existence of ε -equilibrium in dynamic games with dependent motions]. // Zh. vychisl. matem. i matem. fiz., 1974, 14:1, pp. 88–98. (in Russian)
- [24] Malafeev O. A. O sushhestvovanii situacij ε -ravnovesija v smeshannyh strategijah v beskoalicionnyh differencial’nyh igrakh [On existence of ε -equilibrium in mixed strategies in non-coalitional differential games] // Differenc. uravnenija, 1979, Vol. 15, no. 4, pp. 609–613. (in Russian)
- [25] Markovkin M. V. Linejno-kvadratičnye kooperativnye differencial’nye igry [Linear-quadratic cooperative differential games]. Dissertacija na soisk. uch. st. kand. fiz.-mat. n.: 01.01.09 // SPb.: SPbSU, 2006. (in Russian)

- [26] Naumova N. I. Vektor Shepli i ego obobshhenija. Uchebnoe posobie [The Shapley value and its extensions]. — SPb.: Izd-vo VVM, 2017. (in Russian)
- [27] Parilina E. M. Kooperativnaya stohasticheskaya igra s konechnym chislom igrovyh ehlementov [Cooperative stochastic game with finite number of game elements] // Tezisy dokladov mezhdunarodnogo kongressa «Nelinejnyj dinamicheskij analiz-2007» — SPb.: Izd-vo SPbGU, 2007, p. 337. (in Russian)
- [28] Parilina E. M. Kooperativnaja igra peredachi dannyh v besprovodnoj seti [Cooperative game of data transmission in a wireless network] // Matematicheskaja teorija igr i ee prilozhenija, 2009, vol. 1, no. 4, pp. 93–110. (in Russian)
- [29] Parilina E. M. Ustojchivaja kooperacija v stohasticheskikh igrah [Stable cooperation in stochastic games] // Matematicheskaja teorija igr i ee prilozhenija, 2010, vol. 2, no. 3, pp. 21–40. (in Russian)
- [30] Parilina E. M. Stohasticheskie igry v setevykh zadachah [Stochastic games in network problems] // Sbornik tezisov konf. «Ustojchivost' i processy upravlenija» SPb, 2010, pp. 170–171. (in Russian)
- [31] Parilina E. M. Kooperativnaja igra peredachi dannyh v besprovodnoj seti [Cooperative game of data transmission in a wireless network] // Upravlenie bol'shimi sistemami, 2010, vol. 31-1, pp. 191–209. (in Russian)
- [32] Parilina E. M. Kooperativnye stohasticheskie igry, uchityvajushhie odnoshenie igrokov k risku [Risk-sensitive cooperative stochastic games] // Materialy mezhdunarodnoj konf. «Matematika, jekonomika, menedzhment: 100 let so dnja rozhdenija L.V. Kantorovicha», Sankt-Peterburg, 2012, p. 114. (in Russian)
- [33] Parilina E. M. Strategicheskaja ustojchivost' odnotochechnyh principov optimal'nosti v kooperativnykh stohasticheskikh igrah [Strategic consistency of single-point optimality principles in cooperative games] // Matematicheskaja teorija igr i ee prilozhenija, 2014, vol. 6, no. 1, pp. 56–72. (in Russian)
- [34] Parilina E. M., Petrosjan L. A. Sil'no pozicionno sostojatel'noe c-jadro v stohasticheskikh igrah [Strongly subgame-consistent core in stochastic games] //

Matematicheskaja teorija igr i ee prilozhenija, 2017, vol. 9, no. 2, pp. 39–61.
(in Russian)

- [35] Petrosjan L. A. Ustojchivost' reshenij v differencial'nyh igrakh so mnogimi uchastnikami [Time consistency of solutions in differential games with many participants] // Vestnik Leningradskogo universiteta. Serija 1: matematika, mehanika, astronomija. 1977. Vol. 19. pp. 46–52. (in Russian)
- [36] Petrosjan L. A. Postroenie sil'no-dinamicheski ustojchivyh reshenij v kooperativnyh differencial'nyh igrakh [Construction of strongly time-consistent solutions in cooperative differential games] // Vestnik Leningradskogo universiteta. Serija 1: matematika, mehanika, astronomija. 1992. Vol. 2, pp. 33–38. (in Russian)
- [37] Petrosjan L. A. Sil'no dinamicheski ustojchivye principy optimal'nosti v mnogokriterial'nyh zadachah optimal'nogo upravlenija [Strongly time-consistent optimality principles in multicriteria problems of optimal control] // Tehnicheskaja kibernetika. 1993. No. 1, pp. 169–174. (in Russian)
- [38] Petrosjan L. A. Polukooperativnye igry [Semi-cooperative games] // Vestnik Sankt-Peterburgskogo Universiteta. Serija 1: Matematika, mehanika, astronomija, 1998. No. 2, pp. 57–63. (in Russian)
- [39] Petrosjan L. A., Baranova E. M. Stohasticheskie igry so sluchajnoj prodolzhitel'nost'ju [Stochastic games with random duration]// Trudy XXXIV nauchnoj konferencii aspirantov i studentov "Processy upravlenija i ustojchivost'", SPb, 2003, pp. 456–462. (in Russian)
- [40] Petrosjan L. A., Baranova E. M. Kooperativnye stohasticheskie igry [Cooperative stochastic games] // Tezisy dokladov Mezhdunarodnogo seminaru "Teorija upravlenija i teorija obobshhennyh reshenij uravnenij Gamil'tona-Jakobi", Ekaterinburg, 2005, pp. 33–35. (in Russian)
- [41] Petrosjan L. A., Baranova E. M. Kooperativnye stohasticheskie igry v stacionarnyh strategijah [Cooperative stochastic games in stationary strategies]//

Sbornik trudov mezhdunarodnoj konferencii “Ustojchivost’ i processy upravlenija”, SPb, 2005. Vol. 1, pp. 495–503. (in Russian)

- [42] Petrosjan L. A., Baranova E. M., Shevkopljas E. V. Mnogoshagovye kooperativnye igry so sluchajnoj prodolzhitel’nost’ju [Multistage cooperative games with random duration] // Trudy instituta matematiki i mehaniki. Optimal’noe upravlenie i differencial’nye igry. Sbornik statej. 2004. Vol. 10, no. 2, pp. 116–130. (in Russian)
- [43] Petrosjan L. A., Danilov N. A. Ustojchivye reshenija neantagonisticheskikh differencial’nyh igr s tranzitivnymi vyigryshami [Time-consistent solutions of non-antagonistic differential games with transferable payoffs] // Vestnik LGU. 1979. No. 1, pp. 46–54. (in Russian)
- [44] Petrosjan L. A., Zenkevich N. A. Principy ustojchivoj kooperacii [Principles of stable cooperation] // Matematicheskaja teorija igr i ee prilozhenija, 2009. Vol. 1, no. 1, pp. 106–123. (in Russian)
- [45] Petrosjan L. A., Zenkevich N. A., Shevkopljas E. V. Teorija igr [Game theory]. Sankt-Peterburg: BHV – Peterburg, 2012. 480 p. (in Russian)
- [46] Petrosjan L. A., Kuzjutin D. V. Igry v razvernutoj forme: optimal’nost’ i ustojchivost’ [Games in extensive form: optimality and consistency]. — SPb: Izd-vo S.-Peterburgskogo universiteta. 2000. (in Russian)
- [47] Petrosjan L. A., Murzov N. V. Teoretiko-igrovye zadachi mehaniki [Game-theoretical problems of mechanics] // Litovskij matematicheskij sbornik, 1966, Vol. 6, pp. 423–432. (in Russian)
- [48] Petrosjan L. A., Sedakov A. A., Sjurin A. N. Mnogoshagovye igry s koalicionnoj strukturoj [Multistage games with coalition structure] // Vestnik Sankt-Peterburgskogo universiteta. Serija 10: prikladnaja matematika, informatika, processy upravlenija, 2006, no. 4, pp. 97–110. (in Russian)
- [49] Petrosjan L. A., Shevkopljas E. V. Kooperativnye differencial’nye igry so sluchajnoj prodolzhitel’nost’ju [Cooperative differential games with random

- duration] // Vestnik Sankt-Peterburgskogo Universiteta. Serija 1: Matematika, mehanika, astronomija, 2000, no. 4, pp. 14–18. (in Russian)
- [50] Petrosjan O. L., Gromova E. V., Pogoshev S. V. O sil'no dinamičeski ustojčivom podmnozhestve C -jadra v kooperativnyh differencial'nyh igrah s predpisannoju prodolzhitel'nost'ju [On strong time-consistent subset of the core in cooperative differential games with prescribed duration] // Matematičeskaja teorija igr i ee prilozhenija. 2016. Vol. 8, no. 4, pp. 79–106. (in Russian)
- [51] Pečerskij S. L., Janovskaja E. B. Kooperativnye igry: reshenija i aksiomy [Cooperative games: solutions and axioms]. SPb.: Izd-vo Evrop. univ-ta v S.-Peterburge, 2004, 459 p. (in Russian)
- [52] Sedakov A. A. O sil'noj dinamičeskoj ustojčivosti c -jadra [On strongly time consistency of the core] // Matematičeskaja teorija igr i ee prilozhenija. 2015. Vol. 7, no. 2, pp. 69–84. (in Russian)
- [53] Tur A. V. Linejno-kvadratičnye stohastičeskie diskretnye igry so sluchajnoj prodolzhitel'nost'ju [Linear-quadratic stochastic discrete time games with random duration] // Matematičeskaja teorija igr i ee prilozhenija. 2014. Vol. 6, no. 3, pp. 76–92. (in Russian)
- [54] Tur A. V. Kooperacija v diskretnyh linejno-kvadratičnyh igrah. [Cooperation in discrete-time linear-quadratic games] Dissertacija na soisk. uch. st. kand. fiz.-mat. n.: 01.01.09 // SPb.: SPbSU, 2015. (in Russian)
- [55] Shevkopljas E. V. Ustojčivaja kooperacija v differencial'nyh igrah so sluchajnoj prodolzhitel'nost'ju [Stable cooperation in differential games with random duration] // Matematičeskaja teorija igr i ee prilozhenija, 2010, vol. 2, no. 3, pp. 162–190. (in Russian)
- [56] Chistjakov S. V. Programmnye iteracii i universal'nye ε -optimal'nye strategii v pozicionnoj differencial'noj igre [Program iterations and universal ε -optimal strategies in positional differential game], Dokl. AN SSSR, 1991, 319:6, pp. 1333–1335. (in Russian)

- [57] Chirkova Ju. V. Cena anarhii v igre balansa zagruzki sistemy obsluzhivaniya s tremja mashinami [Price of anarchy for machine load balancing game with 3 machines] // *Matematicheskaja Teorija Igr i ee Prilozhenija*, 2014, vol. 6, no. 4, pp. 85–96. (in Russian)
- [58] Algaba, E., Bilbao, J.M., van den Brink, R., Jiménez-Losada, A. Axiomatizations of the Shapley value for cooperative games on antimatroids // *Mathematical Methods of Operations Research*, 2003, 57, 49–65.
- [59] Altman E., Avrachenkov K., Garnaev A. A jamming game in wireless networks with transmission cost // *Lecture Notes in Computer Science*, 2007, 4465 LNCS, pp. 1-12.
- [60] Altman E., Avrachenkov K., Garnaev A. Jamming in wireless networks under uncertainty // *Mobile Networks and Applications*, 2011, 16 (2), pp. 246–254.
- [61] Altman E., Avrachenkov K., Miller G., Prabhu B. Discrete power control: Cooperative and non-cooperative optimization, *Proceedings - IEEE INFOCOM*, 2007, art. no. 4215595, pp. 37–45.
- [62] E. Altman, D. Barman, R. El Azouzi and T. Jimenez, “A game theoretic approach for delay minimization in slotted ALOHA,” 2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577), Vol.7, pp. 3999–4003, 2004.
- [63] Amir R. Stochastic games in economics: The lattice-theoretic approach. In A. Neyman and S. Sorin (eds.) // *Stochastic Games and Applications*, NATO Science Series C, Mathematical and Physical Sciences, 2003, vol. 570, pp. 443–453.
- [64] Anshelevich E., Dasgupta A., Kleinberg J., Tardos, E., Wexler, T. and Roughgarden, T. The price of stability for network design with fair cost allocation // In 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2004, pp. 59–73.

- [65] Avrachenkov K., Cottatellucci L., Maggi L. Cooperative Markov decision processes: Time consistency, greedy players satisfaction, and cooperation maintenance // *International Journal of Game Theory*, 2013, 42 (1), pp. 239–262.
- [66] Aumann R. J. Acceptable points in general cooperative n-person games. // In: Tucker, A. W., Luce, R. D. (eds.) *Contributions to the Theory of Games IV*. 1959. Princeton: Princeton University Press.
- [67] Aumann R.J. Subjectivity and correlation in randomized strategies // *Journal of Mathematical Economics*, 1974, Vol. 1, pp. 67–96.
- [68] Aumann, R.J., Dreze, J.H., Cooperative games with coalition structure // *International Journal of Game Theory*, 1974, Vol. 3, pp. 217–237.
- [69] Aumann R.J., Peleg, B., 1960. Von Neumann-Morgenstern Solutions to Cooperative Games without Side Payments // *Bulletin of the American Mathematical Society*, 1960, vol. 66, pp. 173–179.
- [70] Axelrod, R, and Hamilton, W. D. The evolution of cooperation // *Science*, 1981, vol. 211, pp. 1390–1396.
- [71] Baranova E. M. The Condition for Keeping Cooperation in Stochastic Cooperative Games // *Proceedings of the Russian-Finnish Graduate School Seminar “Dynamic Games and Multicriteria Optimization”*, edited by V.V.Mazalov, Petrozavodsk, 2006, p. 54–58.
- [72] Baranova, E. M., Petrosjan, L. A. Cooperative Stochastic Games in Stationary Strategies // *Game Theory and Applications*, 2006, Vol. 11, pp. 1–7.
- [73] Basar, T. and Zhu, Q. Price of Anarchy, Information, and Cooperation in Differential Games // *Dynamic Games and Applications*, 2011, vol. 1(1), pp. 50–73.
- [74] Bazenkov N. I. Double best response dynamics in topology formation game for ad hoc networks // *Automation and Remote Control*. 2015. 76 (2). Pp. 323–335.

- [75] Bellman R. Dynamic programming. Princeton: Princeton University Press, New Jersey, 1957.
- [76] Bogomolnaia, A., Jackson, M.O. The stability of hedonic coalition structures // *Games and Economic Behavior*, 2002, 38 (2), pp. 201–230.
- [77] Breton M., Haurie A., Filar J. A. On the computation of equilibria in discounted stochastic dynamic games // *Journal of Economic Dynamics and Control*, 1986, Vol. 10, Is. 1–2, pp. 33–36.
- [78] Breton M., L'Écuyer P. Approximate solutions to continuous stochastic games. In: Hämmäläinen R.P., Ehtamo H.K. (eds) *Differential Games — Developments in Modelling and Computation*. Lecture Notes in Control and Information Sciences, 1991, vol. 156, pp. 257-264. Springer, Berlin, Heidelberg.
- [79] Bure V., Parilina E. Stochastic models of Data Transmission with Specific Network Topologies // *Collected abstracts of papers presented on the 11th International Conference Game Theory and Management* Ed. L.A. Petrosyan and N.A. Zenkevich.- SPb.: Saint Petersburg State University, 2017, p. 125.
- [80] Buttyan L., Hubaux J.-P. *Security and Cooperation in Wireless Networks: Thwarting Malicious and Selfish Behavior in the Age of Ubiquitous Computing* // Cambridge University Press New York. - 2007. - NY, USA.
- [81] Chander, P. and Tulkens, H. The Core of an Economy with Multilateral Environmental Externalities // *International Journal of Game Theory*, 1997, 23, 379–401.
- [82] Chirkova, Yu. V. Price of anarchy in machine load balancing game // *Automation and Remote Control*, 2015, Vol. 76, Is. 10. Pp. 1849–1864.
- [83] Chistyakov, S., Petrosyan, L. Strong Strategic Support of Cooperative Solutions in Differential Games // *Contributions to game theory and management*, 2011, Vol. IV, p. 105–111.
- [84] Cole, R., Dodis, Y. and Roughgarden, T. How much can taxes help selfish routing? // *Journal of Computer and System Sciences*, 2006, 72, pp. 444–467.

- [85] Correa, J.R., Schulz, A.S. and Stier Moses, N.E. Selfish Routing in Capacitated Networks // *Mathematics of Operations Research*, 2004, 29(4), pp. 961–976.
- [86] Correa, J.R., Schulz, A.S. and Stier Moses, N.E. Fast, Fair and Efficient Flows in Networks // *Operations Research*, 2007, 55(2), pp. 215–225.
- [87] Driessen T. , Muto S. ,Nakayama M. A cooperative Game of Information Trading: the Core, the Nucleolus and the Kernel // *ZOR Methods and Models of Operations Research*, 1992, vol. 36, no 1, pp 55–72.
- [88] Dutta, P., 1995. A folk theorem for stochastic games. // *Journal of Economic Theory*, 1995, Vol. 66, pp. 1–32.
- [89] Filar, J., Petrosjan, L. Dynamic Cooperative Games // *International Game Theory Review*, 2000, 2:1, 47–65.
- [90] Filar, J., Vrieze, K. *Competitive Markov Decision Processes*. N.Y.: Springer-Verlag New York, 1997.
- [91] Fink, A. M., Equilibrium in a stochastic n -person game // *Journal of Science of the Hiroshima University*, 1964, A-I 28, pp. 89–93.
- [92] Flesh, J., Kuipers, J., Mashiah-Yaakovi, A., Shmaya, E., Shoenmakers, G., Solan, E., Vrieze, K. Nonexistence of subgame-perfect-equilibrium in perfect information games with infinite horizon // *International Journal of Game Theory*, 2014, vol. 43, pp. 945–951.
- [93] Flesh, J., Predtethinski, A. On refinements of subgame perfect ϵ -equilibrium // *International Journal of Game Theory*, 2015, DOI 10.1007/s00182-015-0468-8.
- [94] Fudenberg, D. and Tirole, J. 1991. Perfect Bayesian equilibrium and sequential equilibrium // *Journal of Economic Theory*, vol. 53, no. 2, pp. 236–260.
- [95] Fudenberg, D., Yamamoto, Y. Repeated games where the payoffs and monitoring structure are unknown // *Econometrica*, 2010, vol. 78, pp. 1673–1710.

- [96] Gene T., Reynolds S. S., Sen S. Dynamic Oligopolistic Games Under Uncertainty: A Stochastic Programming Approach. // *Journal of Economic Dynamics & Control*, 2007, 31, 55–80.
- [97] Gene T., Sen S. An Analysis of Capacity and Price Trajectories for the Ontario Electricity Market Using Dynamic Nash Equilibrium Under Uncertainty. // *Energy Economics*, 2008, vol. 30, p. 173–191.
- [98] Germain M. Toint P., Tulkens H. and de Zeeuw A. Transfers to Sustain Dynamic Core-Theoretic Cooperation in International Stock Pollutant Control // *Journal of Economic Dynamics & Control*, 2003, Vol. 28 (1), pp. 79–99.
- [99] Gillies D. B. Solutions to general non-zero-sum games. In Tucker, A. W.; Luce, R. D. *Contributions to the Theory of Games IV*. (Annals of Mathematics Studies 40). Princeton: Princeton University Press. 1959, pp. 47–85.
- [100] González-Díaz, J., Sánchez-Rodríguez, E. Towards an axiomatization of the core-center // *European Journal of Operational Research*, Vol. 195, Is. 2, 2009, Pp. 449–459.
- [101] Harrington Jr. J. E., Zhao W. Signaling and tacit collusion in an infinitely repeated Prisoners' Dilemma // *Mathematical Social Sciences*, 2012, 64, p. 277–289.
- [102] Hart S., Kurz M. Endogenous formation of coalitions // *Econometrica*, 1983, 51 (4), pp. 1047–1064.
- [103] Haurie A., Krawczyk J. B., Zaccour G. *Games and Dynamic Games*, Singapore: Scientific World, 2012.
- [104] Haurie A., Zaccour G., Smeers Y. (1990). Stochastic Equilibrium Programming for Dynamic Oligopolistic Markets // *Journal of Optimization Theory and Applications* 66(2), 243–253.
- [105] P. Jean-Jacques Herings and R. Peeters, “Homotopy methods to compute equilibria in game theory,” *Econ Theory*, vol. 42, pp. 119–156, 2010.

- [106] Hörner J., Rosenberg D., Solan E., Vieille N. On a Markov game with one-sided information // *Operations Research*, 2010, vol. 58 (4 PART 2), pp. 1107–1115.
- [107] Hörner J., Sugaya T., Takahashi S., Vieille N. Recursive Methods in Discounted Stochastic Games: An Algorithm for $\delta \rightarrow 1$ and a Folk Theorem // *Econometrica*, 2011, vol. 79, pp. 1277–1318.
- [108] Hörner J., Takahashi S., Vieille N. Truthful Equilibria in Dynamic Bayesian Games // *Econometrica*, 2015, vol. 79, pp. 1277–1318.
- [109] Jørgensen S., Martín-Herrán G., Zaccour G. Dynamic Games in the Economics and Management of Pollution // *Environmental Modeling and Assessment*, 2010, Vol. 15, pp. 433–467.
- [110] Jáskiewicz A., Nowak A. S. Stationary almost markov perfect equilibria in discounted stochastic games // *Mathematics of Operations Research*, 2016, vol. 41 (2), pp. 430–441.
- [111] Kalai E. Bounded rationality and strategic complexity in repeated games // *Game theory and applications*, Pro. Int. Conf., Columbus/OH (USA), p. 131–157.
- [112] Kamhoua C., Pissinou N. Mitigating selfish misbehavior in multi-hop networks using stochastic game theory // In *Proceedings of the 2010 IEEE 35th Conference on Local Computer Networks (LCN '10)*. IEEE Computer Society, Washington, DC, USA. - 2010. - Pp. 232-235.
- [113] Kohlberg E. On the Nucleolus of a Characteristic Function Game // *SIAM Journal of Applied Mathematics*, 1971, vol. 20 , pp. 62–66.
- [114] Kohlberg E. The Nucleolus as a Solution to a Minimization Problem // *SIAM Journal of Applied Mathematics*, 1972, vol. 23, no. 1, pp. 34–39.
- [115] Kohlberg E., Neyman A., 2015. The cooperative solution of stochastic games. Harvard Business School. Working Paper, No. 15-071, March 2015.
- [116] Konishi H., Ray D. Coalition formation as a dynamic process // *Journal of Economic Theory*, 2003, 110, p. 1–41.

- [117] Koutsoupias E., Papadimitriou C. H. Worst-case equilibria // In Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS), 1999, pp. 404–413.
- [118] Kozlovskaya N., Zenkevich N. Stable cooperation under environmental constraints, *International Game Theory Review*, 2010, 12 (4), pp. 453–470.
- [119] Kreps V., Domansky V. Repeated games with asymmetric information modeling financial markets with two risky assets // *RAIRO Recherche Operationnelle*, 2013, vol. 47 (3), pp. 251–272.
- [120] Kuhn H. W. Extensive Games // *Proceedings of National Academy of Sciences of the USA*, 1950, vol. 36, pp. 570–576.
- [121] Kuhn H. W. Extensive Games and the Problem of Information // *Annals of Mathematics Studies*, 1953, no. 28, pp. 193–216.
- [122] Lehrer E., Scarsini M. On the Core of Dynamic Cooperative Games // *Dynamic games and applications*, 2013, vol. 3, p. 359–373.
- [123] C. E. Lemke and J. T. Howson, “Equilibrium Points of Bimatrix Games,” *J Soc Indust Appl Math*, vol. 12, pp. 413–423, 1964.
- [124] Liu Y., Garnaev A., Trappe W. Maintaining throughput network connectivity in ad hoc networks // *Proceedings of ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*. - 2016. Art. no. 7472905. Pp. 6380-6384.
- [125] Long N.V. Applications of dynamic games to global and transboundary environmental issues: a review of the literature. // *Strategic Behavior and the Environment*, 2012, 2(1), pp. 1-59.
- [126] WOLFRAM MATHEMATICA // <https://www.wolfram.com/mathematica/>
- [127] Maggi L., Avrachenkov K., Cottatellucci L. Stochastic games for cooperative network routing and epidemic spread, 2011, *IEEE International Conference on Communications*, art. no. 5963532.

- [128] Mailath G. J., Postlewaite A., Samuelson L. Contemporaneous perfect epsilon-equilibria // *Games and Economic Behavior*, 2005, Vol. 53, pp. 126–140.
- [129] Mailath G. J., Samuelson L. *Repeated Games and Reputations: Long-Run Relationships*. Oxford: Oxford University Press, 2006.
- [130] Maor, C. and Solan, E. Cooperation under incomplete information on the discount factors // *International Journal of Game Theory*, 2015, Vol. 44, pp. 321–346.
- [131] S. Marbán, P. van de Ven, P. Borm, and H. Hamers “ALOHA networks: a game-theoretic approach,” *Math Meth Oper Res*, vol. 78, is. 2, pp. 221–242, 2013.
- [132] Marín-Solano J., Shevkoplyas E. V. Non-constant discounting and differential games with random time horizon, *Automatica*, 2011, Vol. 47, Is. 12, p. 2626–2638.
- [133] Markovkin M. V., D. W. K. Yeung’s Condition for Linear Quadratic Differential Games, In: *Dynamic Games and Their Applications*, eds. L. A. Petrosyan, A. Y. Garnaev, St. Petersburg State University, St. Petersburg, 2006, 207–216.
- [134] Maschler M., Solan E., Zamir S. *Game Theory*. Cambridge University Press, 2013.
- [135] Maplesoft. <http://www.maplesoft.com/>
- [136] Mathworks, 2017, <https://se.mathworks.com>
- [137] Mazalov V. V., Rettieva A. N. Fish wars and cooperation maintenance // *Ecological Modelling*, 2010, vol. 221 (12), pp. 1545–1553.
- [138] Mazalov V. V., Rettieva A. N. Fish wars with many players // *International Game Theory Review*, 2010, 12 (4), pp. 385–405.
- [139] Meinhardt H.I. Graphical Extensions of the Mathematica Package TU Games // <http://library.wolfram.com/infocenter/MathSource/5709/TuGamesView3D.pdf>.

- [140] Mertens J-F, Neyman A. Minimax Theorems for Undiscounted Stochastic Games // *Game Theory and Mathematical Economics*, 1981, p. 83–87.
- [141] Mertens J. F., Neyman A., *Stochastic Games*. // *International Journal of Game Theory*, 1981, Vol. 10, pp. 53–66.
- [142] Mertens J-F., Parthasarathy T. P. Equilibria for discounted stochastic games // In: Neyman A, Sorin S, eds., *Stochastic Games and Applications*, Kluwer Academic Publishers, 2003, pp. 131–172.
- [143] Mertens J. F., Zamir S. Formulation of Bayesian analysis for games with incomplete information, *International Journal of Game Theory*, 1985, Vol. 14, No. 1, pp. 1–29.
- [144] Montero M. On the nucleolus as a Power Index // *Homo Oeconomicus*, 2005, vol. 22(4), pp. 551–567.
- [145] Naumova N. Solidary Solutions to Games with Restricted Cooperation // *Contributions to Game Theory and Management*, vol. 6, 2013, pp. 316–337.
- [146] von Neumann J., Morgenstern O. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.
- [147] Neyman A. Existence of Optimal Strategies in Markov Games with Incomplete Information // *International Journal of Game Theory*, 2008, 37(4), p. 581–596.
- [148] Neyman A. Stochastic Games with Short-Stage Duration // *Dynamic Games and Applications*, 2013, Vol.3, p. 236–278.
- [149] Neyman A., Sorin, S. *Stochastic Games and Applications*. Dordrecht: Kluwer Academic Press, 2003.
- [150] Njilla L.Y., Pissinou N. Dynamics of data delivery in mobile ad-hoc networks: A bargaining game approach // *Proceedings of 2015 IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2015* - art. no. 7208634. - 2015. - Pp. 98-103.

- [151] Novikov D.A. Games and networks // Automation and Remote Control. - 2014. - 75 (6). - Pp. 1145-1154.
- [152] Nowak A. S. Existence of equilibrium stationary strategies in discounted non-cooperative stochastic games with uncountable state space // Journal of Optimization Theory and Applications, 1985, 45 (4), pp. 591–602.
- [153] Nowak A. S. Sensitive equilibria for ergodic stochastic games with countable state spaces // Mathematical Methods of Operations Research, 1999, vol. 50 (1), pp. 65–76.
- [154] Nowak A. S., Radzik T. A solidarity value for n-person transferable utility games // International Journal of Game Theory, 1994, 23 (1), pp. 43–48.
- [155] Owen G. Values of games with a priori unions // In: Henn, R., Moeschlin, O. (eds.) Essays in mathematical economics and game theory, Springer-Verlag, Berlin, 1977, p. 76–88.
- [156] Parilina E. M. Subgame Consistency of Optimality Principles in Cooperative Stochastic Games // Collected abstracts of papers presented on International Conference Game Theory and Management, SPb, 2007, p. 104.
- [157] Parilina E. M. Game-theoretic Approach to Resource Sharing Management // Proceedings of The Second International Conference on Game Theory and Applications — World Academic Union (World Academic Press), 2007, pp. 49-52.
- [158] Parilina E. M. Subgame Consistency of Shapley Value in Cooperative Data Transmission Game in Wireless Network // Contributions to Game Theory and Management, SPb, 2008, Vol. 1, p. 381–994.
- [159] Parilina E. Principles of Stable Cooperation in Stochastic Games // Collected abstracts of papers presented on GTM 2011, SPb, 2011., стр. 176–177.
- [160] Parilina E. M. Principles of Stable Cooperation in Stochastic Games // Contributions to game theory and management, Vol. 5, 2012. P. 243–256.

- [161] Parilina E. M. Cooperation in Dynamic Games with Stochastic Payoffs // Constructive non-smooth analysis and related topics. Abstracts of papers of the international conference, Saint Petersburg, 2012, p. 131–132.
- [162] Parilina E. M. Dynamic Scenarios of Data Transmission in Wireless Networks // Extended abstracts of the papers presented at the international meeting “Network games and management”, Petrozavodsk, 2012, pp. 51–53.
- [163] Parilina E. M. Stable cooperation in stochastic games // Automation and Remote Control. 2015. Vol. 76, pp. 1111–1122.
- [164] Parilina E. Strategic Support of the Shapley Value in Stochastic Games // Contributions to Game Theory and Management, 2016. Vol. IX, p. 246–265.
- [165] Parilina E. A Survey on Cooperative Stochastic Games with Finite and Infinite Duration // Contributions to Game Theory and Management, Vol. 11, 2018, pp. 129–196.
- [166] Parilina E. M., Petrosyan L. A. Strongly Subgame-Consistent Core in Stochastic Games // Automation and Remote Control, Vol. 79, Is. 8, pp. 1515–1527.
- [167] Parilina E., Sedakov A. Stable coalition structure as solution of stochastic game // Collected abstracts of papers presented on 8th International Conference GTM — 2014. SPb: Graduate School of Management SpbU, 2014, p.140–141.
- [168] Parilina E., Sedakov A. Stochastic Approach for Determining Stable Coalition Structure // International Game Theory Review, 2015. Vol. 17, № 4, p. 155009-1–155009-22.
- [169] Parilina E., Sedakov A., Zaccour G. Price of Anarchy in a Linear-State Stochastic Dynamic Game // European Journal of Operational Research, 2017. Vol. 258, no. 2, p. 790–800.
- [170] Parilina E., Tampieri A. Stability and Cooperative Solution in Stochastic Games // Theory and Decision, <https://doi.org/10.1007/s11238-017-9619-7>.
- [171] Parilina E., Zaccour G. Node-Consistent Core for Games Played over Event Trees // Automatica. 2015. Vol. 53. P. 304–311.

- [172] Parilina E., Zaccour G. Approximated cooperative equilibria for games played over event trees // *Operations Research Letters*, 2015, Vol. 43, p. 507–513.
- [173] Parilina E., Zaccour G. Strategic Support of Node-Consistent Cooperative Outcomes in Dynamic Games Played over Event Trees // *International Game Theory Review*, 2016. Vol. 18, № 2, art.no. 1640002, 16 pages.
- [174] Parilina E., Zaccour G. Node-Consistent Shapley Value for Games Played over Event Trees with Random Terminal Time // *Journal of Optimization Theory and Applications*, 2017. Volume 175, Issue 1, pp 236–254.
- [175] Patrone F., Dinar A., Moretti S. Application of Stochastic Cooperative Games in Water Resources // "Frontiers in Water Resource Economics", Edited by R. Goetz and D. Berga (Kluwer Academic Publishers), Series: Natural Resource Management and Policy, Vol. 29:XXI + 275, 2006.
- [176] Peleg B., Sudhölter P. Introduction to the theory of cooperative games. Berlin, Heidelberg, New York: Springer Science + Business Media B. V. 328 p. (Theory and Decision Library Series C, Vol. 34).
- [177] Peski M. Repeated games with incomplete information and discounting // *Theoretical Economics*, 2014, Vol. 9, p. 651–694.
- [178] Petrosjan L. A. Cooperative Stochastic Games // *Advances in Dynamic Games. Annals of the ISDG. Application to Economics, Engineering and Environmental Management*, ed. by A. Haurie, S. Muto, L. A. Petrosjan, T.E.S. Raghavan. 2006. P. 139-146.
- [179] Petrosjan L. A., Baranova E. M. Cooperative Stochastic Games in Stationary Strategies // *Proceedings of the Fifth International ISDG Workshop International Society of Dynamic Games, Segovia (Spain)*, 2005, p. 225–234.
- [180] Petrosjan L. A., Grauer L. V. Strong Nash equilibrium in multistage games // *International Game Theory Review*, 2002, Vol. 4, № 2, p. 255–264.
- [181] Petrosjan L., Mamkina S. Dynamic games with coalitional structures // *International Game Theory Review*, 2006, vol. 8 (2), pp. 295–307.

- [182] Petrosjan L., Zaccour G. Time-consistent Shapley value allocation of pollution cost reduction // *Journal of Economic Dynamics & Control*, 2003, 27, pp. 381–398.
- [183] Petrosjan L. A., Zenkevich N. A., Conditions for Sustainable Cooperation // *Automation and Remote Control*, 2015, Vol. 76(10), pp. 1894–1904.
- [184] Petrosjan L. A., Zenkevich N. A. *Game Theory*, Singapore-London-New York: World Scientific Publishing Co, 2016, 552 p.
- [185] Petrosyan L., Chistyakov S. Strategic support of Cooperative Solutions in 2-Person Differential Games with Dependent Motions // *Contributions to Game Theory and Management*, 2013. Vol. 6, p. 388–394.
- [186] Petrosyan L., Sedakov A. One-way flow two-stage network games // *Вестник Санкт-Петербургского университета. Серия 10: прикладная математика, информатика, процессы управления*, 2014, № 4, с. 72–81.
- [187] Petrosyan L., Sedakov A. Strategic support of cooperation in dynamic games on networks // "Stability and Control Processes" in Memory of V.I. Zubov (SCP), 2015 International Conference, 2015, pp. 256–260.
- [188] Petrosyan L., Sedakov A. The Subgame-Consistent Shapley Value for Dynamic Network Games with Shock // *Dynamic Games and Applications*, 2016, Vol. 6, No 4, pp. 520–537.
- [189] Pineau P.-O., Murto P. An Oligopolistic Investment Model of the Finnish Electricity Market // *Annals of Operations Research*, 2003, vol. 121, p. 123–148.
- [190] Pineau P.-O., Rasata H., Zaccour G. Impact of some Parameters on Investments in Oligopolistic Electricity Markets // *European Journal of Operational Research*, 2011, vol. 213 (1), p. 180–195.
- [191] Predtetchinski A. The strong sequential core for stationary cooperative games // *Games and Economic Behavior*, 2007, vol. 61, p. 50–66.

- [192] Radner R. Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives // *Journal of Economic Theory*, 1980, vol. 22, p. 136–154.
- [193] Reactor Safety Study. An Assessment of Accident Risks in U. S. Commercial Nuclear Power Plants. Executive Summary: Main Report. [PWR and BWR]. United States: 1975. Web. doi:10.2172/7134131.
- [194] Reddy P. V., Shevkoplyas E., Zaccour G. Time-Consistent Shapley Value for Games Played over Event Trees // *Automatica*, 2013, 49 (6), p. 1521–1527.
- [195] Reddy P. V., Zaccour G. A friendly computable characteristic function // *Mathematical Social Sciences*, 2016, Vol. 82, pp. 18–25.
- [196] Perakis G., Roels G. The price of anarchy in supply chains: Quantifying the efficiency of price-only contracts // *Management Science*, 2007, 53 (8), pp. 1249–1268.
- [197] Raghavan T. E. S. A stochastic game model of tax evasion. In: *Advances in dynamic games*, Birkhauser Boston, Boston, MA, 2006, pp. 397–420.
- [198] T. E. S. Raghavan and J. A. Filar, “Algorithms for Stochastic Games — A Survey,” *ZOR — Methods and Models of Operations Research*, vol. 35, pp. 437–472, 1991.
- [199] Rettieva A. N. Stable coalition structure in bioresource management problem // *Ecological Modelling*, 2012, 235-236, pp. 102–118.
- [200] Rosen J. B. Existence and Uniqueness of Equilibrium Points for Concave n-Person Games // *Econometrica*, 1965, Vol. 33 (3). pp. 520–534.
- [201] Roughgarden T. *Selfish Routing and the Price of Anarchy*, MIT Press, 2005.
- [202] Roughgarden T. The price of anarchy in games of incomplete information // *ACM Transactions on Economics and Computation* 3 (1), Article 6 (March 2015).

- [203] Roughgarden T., Schoppmann, F. Local smoothness and the price of anarchy in splittable congestion games // *Journal of Economic Theory*, 2015, 156, pp. 317–342.
- [204] Roughgarden T., Tardos, E. How Bad is Selfish Routing? // *Journal of the ACM*, 2002, 49(2), pp. 236–259.
- [205] Sagduyu Y. E., Ephremides A. A game-theoretic look at simple relay channel // *Wireless Networks*. - 2006. - Vol. 12, No. 5. - Pp. 545–560.
- [206] Schmeidler D. The Nucleolus of a Characteristic Function game // *SIAM Journal of Applied Mathematics*, 1969, vol. 17, pp. 1163–1170.
- [207] Schneider R. *Convex bodies: the Brunn–Minkowski theory*, Cambridge Univ. Press, 1993.
- [208] Selten R. Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrage-
träägheit // *Zeitschrift für die gesamte Staatswissenschaft*, 1965, Vol. 121, pp. 301–24 and 667–89.
- [209] Selten R. Reexamination of the perfectness concept for equilibrium points in extensive games // *International Journal of Game Theory*, 1975, 4, pp. 25–55.
- [210] Shapley L. S. Stochastic Games // *Proceedings of National Academy of Sciences of the USA*, 1953, vol. 39, pp. 1095–1100.
- [211] Shapley L. S. A Value for n-person Games. In H.W. Kuhn, A.W. Tucker, eds. *Contributions to the theory of games II*, Princeton: Princeton Univ. Press, 1953, pp. 307–317.
- [212] Solan E. Discounted Stochastic Games // *Mathematics of Operations Research*, 1998, vol. 23, pp. 1010–1021.
- [213] Solan E. Stochastic Games, 2009, in *Encyclopedia of Database Systems*, Springer.
- [214] Solan E., Vieille N. Correlated Equilibrium in Stochastic Games // *Games and Economic Behavior*, 2002, vol. 38, pp. 362–399.

- [215] Solan E., Vieille N. Deterministic multi-player Dynkin games // *Journal of Mathematical Economics*, 2003, vol. 39, p. 911–929.
- [216] Solan E., Vieille N. Stochastic games // *Proceedings of the National Academy of Sciences of the United States of America*, 2015, Vol. 112 (45), pp. 13743–13746.
- [217] Takahashi M., Stochastic games with infinitely many strategies // *Journal of Science of the Hiroshima University*, 1964, A-I 28, pp. 95–99.
- [218] <http://mmiras.webs.uvigo.es/TUGlab/>
- [219] Tur A. V. The irrational behavior proof condition for linear-quadratic discrete-time dynamic games with nontransferable payoffs // *Contributions to Game Theory and Management*, 7 (2014), 384–392.
- [220] Vieille N. Solvable states in n-player stochastic games // *SIAM Journal on Control and Optimization*, 2000, vol. 38 (6), pp. 1794–1804.
- [221] Vieille N. Chapter 48 Stochastic games: Recent results // *Handbook of Game Theory with Economic Applications*, 2002, 3, pp. 1833–1850.
- [222] Xu N., Veinott Jr. A., Sequential stochastic core of a cooperative stochastic programming game. *Operations Research Letters*, 2013, vol. 41, p. 430–435.
- [223] Yeung D.W.K., 2006. An irrational-behavior-proof condition in cooperative differential games // *International Game Theory Review*, 2006, Vol. 8, pp. 739–744.
- [224] Yeung D.W.K., Petrosyan L. *Cooperative Stochastic Differential Games*. Springer, New York, NY, 2005
- [225] Yeung D.W.K., Petrosyan L.A. Subgame Consistent Cooperative Solution of Dynamic Games with Random Horizon // *Journal of Optimization Theory and Applications*, 2011, 150 (1), pp. 78–97.
- [226] Zaccour G. *Théorie des jeux et marchés énergétiques: marché européen de gaz naturel et échanges d'électricité*, Ph.D. Thesis, HEC Montréal, 1987.

- [227] Zaccour G. Time Consistency in Cooperative Differential Games: A Tutorial // *INFOR*, 2008, 46 (1), p. 81–92.
- [228] Zakharov V., Dementieva M. Multistage cooperative games and problem of time consistency // *Int. Game Theory Rev.* 2004, vol. 06, is. 01, pp. 1–14.
- [229] Zyskowski M., Zhu Q. Price and Variance of Anarchy in Mean-Variance Cost Density-Shaping Stochastic Differential Games, 52nd IEEE Conference on Decision and Control, December 10-13, 2013, Florence, Italy.