Санкт-Петербургский государственный университет

Беляков Артем Алексеевич

Выпускная квалификационная работа

Новый подход к поиску доказательств через потенциалы

Уровень образования: магистратура
Направление 01.04.01 «Математика»
Основная образовательная программа ВМ.5832.2022 «Современная Математика»

Научный руководитель:
профессор, д.ф.-м.н.
Ф.В. Петров

Рецензент
профессор, д.ф.-м.н.
А.В. Гасников

Санкт-Петербург
2024

Содержание

Аннотация

В данной работе мы развиваем подход (Taylor et al. ((2018))) по поиску потенциалов. Обычно для доказательства через потенциалы требуется решить задачу полуопределенного программирования(SDP) из O(k) линейных матричных неравенств(LMIs), где k - количество итераций. В нашем же подходе мы показываем, что в определенных случаях достаточно использовать лишь константное количество неравенств. В качестве демонстрации эффективности нашего подхода мы приводим альтернативное доказательство сходимости последней итерации экстраградиентного метода для отрицательно комонотонного случая, по аналогии с (Gorbunov et al. ((2022b))).

1. Введение

Скорость сходимости различных методов первого порядка можно находить, решая так называемую Perfomance Estimation Problem(PEP), впервые сформулированную в (Drori and Teboulle ((2014))) и далее формализованную в (Taylor et al. ((2017))). Согласно (Taylor et al. ((2017))) PEP может быть переформулирована как задача полуопределенного программирования(Semidefinite programming, SDP), которая далее может быть эффективно решена на компьютере.

Скорость сходимости можно устанавливать с помощью потенциалов Ляпунова. В статьях (Taylor et al. ((2018)), Taylor and Bach ((2021))) был предложен метод поиска потенциалов Ляпунова с помощью SDP, однако он требует порядка O(k) линейных матричных неравенств(linear matrix inequalities, LMIs), где k - номер итерации. Далее мы развиваем этот подход для доказательства сублинейной скорости сходимости порядка $O(\frac{1}{k})$ и так же получаем интересные следствия для доказательств линейной скорости сходимости.

В данной работе мы главным образом фокусируемся на решении седловых задач. Они являются активным объектом исследований в силу их большого числа приложений, таких как робастная оптимизация (Ben-Tal et al. ((2009))), контроль (Hast et al. ((2013))), состязательное обучение (Goodfellow and Shlens ((2014)), Madry et al. ((2017))) и генеративно-состязательная сеть(GAN) (Goodfellow et al. ((2014))). Седловые задачи часто изучаются с помощью вариационной задачи(variational inequality problem, VIP). Здесь мы будем рассматривать неограниченный случай:

$$\text{найти } x^* \in \mathbb{R}^d \text{ такое что } F(x^*) = 0. \quad \text{(VIP)}$$

Эту задачу можно решать с помощью различных методов первого порядка таких как экстраградиентный метод (Korpelevich ((1976))) или оптимистичный градиентный метод(Popov ((1980))). Экстраградиентный метод (Extragradient, EG):

$$\forall k \geq 0$$

$$\tilde{x}^k = x^k - \gamma_1 F(x^k),$$

$$x^{k+1} = x^k - \gamma_2 F(\tilde{x}^k),$$

Оптимистичный градиентный метод (Optimistic gradient, OG) с дополнительной инициализацией $\tilde{x}^0 = x^0$:

$$\tilde{x}^k = x^k - \gamma_1 F(\tilde{x}^{k-1}), \ \forall k > 0,$$

$$x^{k+1} = x^k - \gamma_2 F(\tilde{x}^k), \ \forall k \geq 0,$$

где $\gamma_1, \gamma_2 \in \mathbb{R}$ - размеры шага.

Мы демонстрируем эффективность нашего подхода в поиске потенциалов, доказывая скорость сходимости последней итерации экстраградиентного метода в отрицательно комонотонном случае. Полученный нами результат аналогичен (Gorbunov et al. ((2022b))), но представляет интерес, так как был полностью получен на компьютере, в то время как в (Gorbunov et al. ((2022b))) он был получен на компьютере лишь частично. Далее мы будем рассматривать наш подход, решая эту задачу, но важно понимать, что наш подход имеет куда более широкое применение и не ограничен экстраградиентным методом или негативно комонотонным классом задач.

1.1. Основные понятия

В контексте (VIP):

$$\text{найти } x^* \in \mathbb{R}^d \text{ такое что } F(x^*) = 0 \quad \text{(VIP)}$$

мы предполагаем, что F удовлетворяет следующим условиям:

Предположение 1. $F : \mathbb{R}^d \to \mathbb{R}^d$ является L-липшецевым оператором. То есть для любых $x, y \in \mathbb{R}^d$

$$\|F(x) - F(y)\| \leq L\|x - y\|.$$

Предположение 2. $F : \mathbb{R}^d \to \mathbb{R}^d$ является ρ-отрицательно комонотонным оператором($\rho \geq 0$). То есть для любых $x, y \in \mathbb{R}^d$

$$\langle F(x) - F(y), x - y\rangle \geq -\rho\|F(x) - F(y)\|^2.$$

1.2. Связанные работы

Сходимость последней итерации в монотонном случае. Часто (VIP) решают для оператора F липшецева и монотонного. Большенство работ при этом оценивают скорость сходимости лучшей итерации или средней (Nemirovski ((2004)), Mokhtari et al. ((2019))). Только совсем недавно (Golowich et al. ((2020a)), Golowich et al. ((2020b))) установили скорость сходимости EG и OG для последней итерации. Они доказали, что для EG и OG $\|F(x^k)\|^2 = O(\frac{1}{k})$ при дополнительном условии, что Якобиан $\nabla F(x)$ Λ-липшецев. Затем (Gorbunov et al. ((2022a))) закрыли вопрос, доказав, что в монотонном случае для EG и OG выполнено $\|F(x^k)\|^2 = O(\frac{1}{k})$, не используя никаких дополнительных предположений. Они добились этого, используя Perfomance estimation approach из (Taylor et al. ((2017))).

Сходимость последней итерации в отрицательно комонотонном случае. Так как одно из основных применений (VIP) и седловых задач относится к нейронным сетям, где рельеф далеко не является монотонным было бы хорошо ослабить это условие. Так в (Diakonikolas et al. ((2021))) впервые рассматривается отрицательно комонотонный случай, что является намного более слабым условием, чем монотонность.

В отрицательно комонотонном случае тоже было получено доказательство сходимости последней итерации для EG и OG порядка $O(\frac{1}{k})$ (Gorbunov et al. ((2022b))).

2. Поиск потенциалов

2.1. Основная теорема

В случае доказательства скорости сходимости порядка $O(\frac{1}{k})$ через потенциалы довольно естесственно ожидать потенциал следующего вида:

$$\phi_k = P_k + kQ_k$$

Действительно, это можно увидеть и в литературе. Тогда наша задача будет формулироваться так: проверить для любого $k \geq 0$, что выполнены следующие неравенства

$$P_{k+1} + (k+1)Q_{k+1} \leq P_k + kQ_k,$$

$$S_k \leq Q_k.$$

Здесь S_k - некоторая величина у которой мы хотим показать сходимость $O(\frac{1}{k})$. Для экстраградиентного метода в отрицательно комонотонном случае можно искать потенциалы в виде:

$$P_k = \begin{pmatrix} x^k - x^* \\ F(x^k) \end{pmatrix}^T \left[\begin{pmatrix} a & b \\ b & c \end{pmatrix} \otimes I_d \right] \begin{pmatrix} x^k - x^* \\ F(x^k) \end{pmatrix} \tag{1}$$

$$Q_k = \begin{pmatrix} x^k - x^* \\ F(x^k) \end{pmatrix}^T \left[\begin{pmatrix} \tilde{a} & \tilde{b} \\ \tilde{b} & \tilde{c} \end{pmatrix} \otimes I_d \right] \begin{pmatrix} x^k - x^* \\ F(x^k) \end{pmatrix} \tag{2}$$

$$S_k = \|F(x^k)\|^2$$

Сформулируем и докажем основную теорему

Теорема 2.1 (Основная теорема)**.** *Рассмотрим последовательность линейных* *потенциалов:*

$$P_{k+1} + (k+1)Q_{k+1} \leq P_k + kQ_k \tag{3}$$

для $k = 0, 1, \ldots$.

Тогда

$$Q_{k+1} \leq Q_k \tag{4}$$

$$P_{k+1} + Q_{k+1} \leq P_k \tag{5}$$

и из этих неравенст (3) *следует.*

Доказательство. Рассмотрим двойственную к (3) задачу [1]

$$M_0 + kM_1 + (k+1)M_2 + \sum \lambda_k^i A_i \succeq 0,$$

где $\lambda_k^i \geq 0$ и A_i возникают из интерполяционных условий. Если мы определим выпуклый конус $\tilde{A} \subseteq \mathbb{R}^3$, состоящий из (a_0, a_1, a_2), таких что

$$a_0M_0 + a_1M_1 + a_2M_2 + \sum \lambda^i A_i \succeq 0$$

для каких-то $\lambda^i \geq 0$, тогда этот конус в том числе будет замкнут, что следует из леммы 3.2.[2]

По условию $(1, k, k+1)$ принадлежит $\tilde{A}$, и тогда $(\frac{1}{k}, 1, 1 + \frac{1}{k}) \in \tilde{A}$. Далее по замкнутости получаем, что

$$(0, 1, 1) \in \tilde{A},$$

и это означает, что выполнено неравенство $Q_{k+1} \leq Q_k$. □

2.2. Скорость сходимости экстраградиентного метода

Рассмотрим теперь применение теоремы 2.1 к доказательству скорости сходимости экстраградиентного метода для отрицательно комонотонного случая. Если искать потенциалы вида 3 стандартным подходом из (Taylor et al. ((2018)), Taylor and Bach ((2021))) получится довольно большая задача

[1]Далее мы предполагаем, что выполнена сильная двойственность.

[2]Это следует из того факта, что A_i возникли из интерполяционных условий. Для произвольных A_i конус может быть незамкнут.

полуопределенного программирования с $O(k)$ линейными матричными неравенствами и большим числом переменных, тоже $O(k)$. С помощью нашего подхода, с другой стороны, нужно будет проверить всего два матричных неравенства 4, 5.

В силу того, что нашу задачу и класс функций можно нормировать на произвольную константу, мы пока что можем погалать $\gamma = 1$. Воспользуемся следующей лемой из (Gorbunov et al. ((2022b))):

Лемма 2.2 (Лемма C.2 из Gorbunov et al. ((2022b)))**.** *Для EG при* $L \leq \frac{1}{2}$, $\rho \leq \frac{1}{4}$, $\gamma = 1:$

$$\|F(x^{k+1})\|^2 \leq \|F(x^k)\|^2$$

Эта лемма была доказана авторами с помощью SDP. В нашем случае возьмем Q_k пропорциональным $\|F_k(x)\|^2$, и тогда неравенство 4 можно уже считать проверенным. Теперь докажем следующую лемму:

Лемма 2.3. *Для EG при* $\gamma = 1, L \leq \frac{1}{2}, \rho \leq \frac{1}{4}$ *выполнено*

$$\|x^{k+1} - x^*\|^2 + (\frac{1}{4} - \rho)\|F(x^{k+1})\|^2 \leq \|x^k - x^*\|^2.$$

Доказательство. Рассмотрим

$$\|x^{k+1} - x^*\|^2 + (\frac{1}{4} - \rho)\|F(x^{k+1})\|^2 - \|x^k - x^*\|^2,$$

добавим интерполяционные неравенства с коэффициентами:

$$2 * (L^2\|x^k - \tilde{x}^k\|^2 - \|F(x^k) - F(\tilde{x}^k)\|^2)$$

$$2(\frac{1}{4} - \rho) * (L^2\|x^{k+1} - \tilde{x}^k\|^2 - \|F(x^{k+1}) - F(\tilde{x}^k)\|^2)$$

Здесь мы используем $\|F(x^{k+1}) - F(\tilde{x}^k)\|^2 \geq \frac{1}{2}\|F(x^{k+1})\|^2 - \|F(\tilde{x}^k)\|^2$

с коэффициентом 1

$$\langle F(x^k) - F(\tilde{x}^k), x^k - \tilde{x}^k\rangle + \rho\|F(x^k) - F(\tilde{x}^k)\|^2$$

с коэффициентом 2

$$\langle F(\tilde{x}^k), \tilde{x}^k - x^* \rangle + \rho \|F(\tilde{x}^k)\|^2$$

$\square$

Таким образом, выводим основные результаты:

Лемма 2.4. *Для EG при* $L \le \frac{1}{2}, \rho \le \frac{1}{4}, \gamma = 1$ *выполнено*

$$\|x^{k+1} - x^*\|^2 + (k+1)(\frac{1}{4} - \rho)\|F(x^{k+1})\|^2 \le \|x^k - x^*\|^2 + k(\frac{1}{4} - \rho)\|F(x^k)\|^2$$

При $\rho < \frac{1}{4}$

$$\|F(x^k)\|^2 \le \frac{\|x^0 - x^*\|^2}{k(\frac{1}{4} - \rho) + L^{-2}}$$

Лемма 2.5. *Для EG при* $4\rho \le \gamma \le \frac{1}{2L}$ *выполнено*

$$\|x^{k+1} - x^*\|^2 + (k+1)\gamma^2(\frac{1}{4} - \frac{\rho}{\gamma})\|F(x^{k+1})\|^2 \le \|x^k - x^*\|^2 + k\gamma^2(\frac{1}{4} - \frac{\rho}{\gamma})\|F(x^k)\|^2$$

При $\frac{\rho}{\gamma} < \frac{1}{4}$

$$\|F(x^k)\|^2 \le \frac{\|x^0 - x^*\|^2}{k\gamma^2(\frac{1}{4} - \frac{\rho}{\gamma}) + L^{-2}}$$

3. Постановка задачи поиска потенциалов

3.1. Постановка задачи в виде принадлежности некоторому конусу

В разделе 2 мы изложили наш подход по поиску потенциалов вида $P_k + kQ_k$ для доказательства сходимости порядка $O(\frac{1}{k})$. Далее мы проводим более глубокий анализ задачи поиска потенциалов.

Рассмотрим доказательство через потенциалы в довольно общем виде:

$$\phi_{k+1} \le \phi_k \tag{6}$$

$$c_k \psi_k \le \phi_k \tag{7}$$

Здесь ϕ_k - потенциал Ляпунова, а ψ_k - некоторая величина скорость сходимости которой мы хотим установить. Из этих неравенств мы сразу получаем, что $c_k\psi_k \leq \phi_0$, и тогда скорость сходимости $\psi_k \sim O(c_k^{-1})$. То есть $c_k \geq 0$ некоторая константа, отвечающая за скорость сходимости.

В случае экстраградиентного метода в отрицательно комонотонном случае можно искать потенциалы в виде:

$$\phi_k = \begin{pmatrix} x^k - x^* \\ F(x^k) \end{pmatrix}^T \left[\begin{pmatrix} \tilde{a}_k & \tilde{b}_k \\ \tilde{b}_k & \tilde{c}_k \end{pmatrix} \otimes I_d \right] \begin{pmatrix} x^k - x^* \\ F(x^k) \end{pmatrix} \tag{8}$$

$$\psi_k = \|F(x_k)\|^2$$

Теперь рассмотрим двойственную задачу к неравенствам (6) и (7)

$$\sum_{i=1}^d a_{k+1}^i M_2^i + \sum_{i=1}^d a_k^i M_1^i + \sum \lambda_k^i A_1^i \succeq 0 \tag{9}$$

$$c_k M_0 + \sum_{i=1}^d a_k^i M_3^i + \sum s_k^i A_2^i \succeq 0 \tag{10}$$

У нас есть d последовательностей $\{a_k^i\}_{k=1}^\infty$ для $i = 1, \ldots, d$. Они соответствуют переменным в (8). В нашем случае $d = 3,\ a_k^1 = \tilde{a}_k,\ a_k^2 = \tilde{b}_k,\ a_k^3 = \tilde{c}_k$. Далее $\lambda_k^i \geq 0,\ s_k^i \geq 0, A_1^i,\ A_2^i$ - двойственные переменные и матрицы, соответствующие интерполяционным условиям. Введем следующее обозначение:

Обозначение 1. Определим $\tilde{A} \subset \mathbb{R}^{1+2d}$, как выпуклый конус, состоящий из троек вида (c, u, v), где $c \in \mathbb{R},\ u, v \in \mathbb{R}^d$ таких что

$$\sum_{i=1}^d v_i M_2^i + \sum_{i=1}^d u_i M_1^i + \sum \lambda_i A_1^i \succeq 0,$$

$$c M_0 + \sum_{i=1}^d u_i M_3^i + \sum s_i A_2^i \succeq 0$$

для некоторых $\lambda_i \geq 0,\ s_i \geq 0.$

Тогда неравенства 6 и 7 можно переформулировать в терминах принадлежности к конусу из обозначения 1:

$$(c_k, u_k, u_{k+1}) \in \tilde{A},$$

где $u_k = (a_k^1, \ldots, a_k^d)^T \in \mathbb{R}^d$. Таким образом, задача поиска потенциалов сводится к задаче поиска последовательностей $\{c_k\}_{k=1}^{\infty} \in \mathbb{R}$, $\{u_k\}_{k=1}^{\infty} \in \mathbb{R}^d$, таких что $(c_k, u_k, u_{k+1}) \in \tilde{A}$ для некоторого выпуклого конуса и нас при этом интересует асимптотика c_k.

Покажем также, что конус $\tilde{A}$ является замкнутым. Введем следующее определение:

Определение 3.1. Будем говорить, что набор матриц $\{A^i\}_{i=1}^n$ удовлетворяет условию **не положительной определенности**, если для любых $\lambda^i \geq 0$, не всех равных нулю, выполнено:

$$\sum \lambda^i A^i \not\succeq 0 \tag{11}$$

Заметим, что в нашем случае матрицы $\{A_1^i\}$, $\{A_2^i\}$ как раз удовлетворяют условию 3.1. Это следует из того, что они берутся из интерполяционных условий. Пусть $\{A^i\}$ возникли из интерполяционных условий и при этом существуют $\lambda^i \geq 0$, не все равные 0, такие что

$$\sum \lambda^i A^i \succeq 0.$$

Заметим, что A^i соответствует квадратичной форме, которая предполагается отрицательной, например, $\|F(x)-F(y)\|^2 - L^2\|x-y\|^2 \leq 0$. Тогда понятно, что $\sum \lambda^i A^i$ соответствует некоторой квадратичной форме, которая предполагается отрицательной из интерполяционных условий. Но в нашем случае она также соответствует положительно определенной матрице. Получается значения квадратичной формы всегда должны равняться 0. Мы предполагаем, что интерполяционные условия себя так вести не должны. Тогда следующая лемма устанавливает замкнутость конуса $\tilde{A}$:

Лемма 3.2. *Рассмотрим выпуклый конус $\tilde{A} \subseteq \mathbb{R}^d$, состоящий из точек $a \in \mathbb{R}^d$ таких, что:*

$$\sum_{i=1}^{d} a^i M^i + \sum \lambda^i A^i \succeq 0$$

для некотрых $\lambda^i \geq 0$. Здесь M^i - произвольные матрицы, а A^i - матрицы для которых выполнено условие не положительной определенности 3.1. Тогда этот конус замкнут.

Доказательство. Пусть $a_k \in \tilde{A}$ и $a_k \to a$, покажем, что тогда a тоже лежит в $\tilde{A}$. Так как $a_k \in \tilde{A}$, то

$$\sum_{i=1}^{d} a_k^i M^i + \sum \lambda_k^i A^i \succeq 0 \tag{12}$$

для некоторых $\lambda_k^i \geq 0$. Теперь мы можем рассмотреть подпоследовательность такую, что все λ_k^i имеют конечный или бесконечный пределы, а также все $\frac{a_k^i}{\lambda_k^j}, \frac{\lambda_k^i}{\lambda_k^j}$ имеют конечные или бесконечные пределы.

Если все λ_k^i имеют конечные пределы, то мы показали, что a тоже лежит в $\tilde{A}$. Если же некоторые из них имеют бесконечные пределы, то введем следующее обозначение:

Будем говорить, что положительные последовательности $\{x_k\}$ и $\{y_k\}$ одинакового порядка, если $\lim_{k\to\infty} \frac{x_k}{y_k} = c$ для $0 < c < \infty$. Последовательность $\{x_k\}$ меньшего порядка чем $\{y_k\}$, если $\lim_{k\to\infty} \frac{x_k}{y_k} = 0$.

Тогда разделим неравенства 12 на последовательность наибольшего порядка, некоторую $\{\lambda_k^i\}$ и получим:

$$\sum c^i A^i \succeq 0,$$

где $c^i \geq 0$ и не все равны нулю. Но это протеворечит условию не положительной определенности 3.1.

□

3.2. Постановка задачи в виде динамической системы

В разделе 3.1 мы переформулировали задачу поиска потенциалов через принадлежность к некоторому замкнутому выпуклому конусу $\tilde{A} \subseteq \mathbb{R}^{1+2d}$. Теперь мы хотим записать задачу в виде некоторой динамической системы.

Рассмотрим замкнутое выпуклое множество $A = \{(u,v) \mid u,v \in \mathbb{R}^d,\ (1,u,v) \in \tilde{A}\}$. Далее мы также будем предполагать, что A компакт. [3]

[3] Это предположение действительно оправдано, так как при доказательстве скорости сходимости порядка

Предположение 3. A - компакт.

Мы хотим найти условия, которым A должно удовлетворять, чтобы $(\tilde{c}_n, u_n, u_{n+1}) \in \tilde{A}$ для $\{\tilde{c}_n\}_{n=1}^{\infty}$, $\{u_n\}_{n=1}^{\infty}$. При этом нас интересует только асимптотика $\tilde{c}_n$. Далее мы рассматриваем только последовательности с $\tilde{c}_n > 0$ для любого n. Заметим, что следующие условия

$$(\tilde{c}_k, u_k, u_{k+1}) \in \tilde{A},\ (\tilde{c}_{k+1}, u_{k+1}, u_{k+2}) \in \tilde{A}$$

равносильны этим

$$(\frac{u_k}{\tilde{c}_k}, \frac{u_{k+1}}{\tilde{c}_k}) \in A,\ (\frac{u_{k+1}}{\tilde{c}_{k+1}}, \frac{u_{k+2}}{\tilde{c}_{k+1}}) \in A.$$

При этом если мы обозначим $u = \frac{u_k}{\tilde{c}_k}$, $v = \frac{u_{k+1}}{\tilde{c}_k}$, $w = \frac{u_{k+2}}{\tilde{c}_{k+1}}$, то

$$(u, v) \in A,\ (\frac{v}{c}, w) \in A$$

где $c = \frac{\tilde{c}_{k+1}}{\tilde{c}_k}$.

Тогда рассмотрим динамическую систему на A [4] в которой можно совершать только такие переходы:

$$(u, v) \to (\frac{v}{c}, w)$$

где $u, v \in \mathbb{R}^d$, $c > 0$.

Таким образом, мы ищем последовательность $(v_0, v_1) \mapsto (\frac{v_1}{c_1}, v_2) \mapsto (\frac{v_2}{c_2}, v_3) \mapsto \cdots$ такую, что все пары лежат в A. При этом нам интересна только асимптотика $\{\prod_{i=1}^{n} c_i\}_{n=1}^{\infty}$.

$O(\frac{1}{k})$ мы ожидаем, что у потенциалов будут расти коэффициенты не быстрее чем $O(k)$.

[4] Она не является динамической системой в стандартном определении, так как из одной точки можно перейти в множество разных точек.

4. Анализ задачи по поиску потенциалов

4.1. Неподвижные точки

Точки вида $(u, cu) \in A$ мы будем называть неподвижными точками нашей динамики. [5] Нас главным образом интересует два случая: $c > 1$, $c = 1$. Неподвижные точки с $c < 1$ нам не интересны. Заметим, что если некоторая неподвижная точка (u, cu) с $c > 1$ принадлежит A, то мы немедленно получаем экспоненциальный рост $\prod_{i=1}^{n} c_i$. Для случая (u, u) дело обстоит намного сложнее.

Допустим, что была найдена последовательность потенциалов $(v_0, v_1) \mapsto (\frac{v_1}{c_1}, v_2) \mapsto (\frac{v_2}{c_2}, v_3) \mapsto \cdots$. Следующая лемма показывает, при каких условиях на эту последовательность потенциалов можно гарантировать, что в A лежит некоторая неподвижная точка:

Лемма 4.1. *Определим* $c = \max\{c | \frac{\prod_{i=1}^{n} c_i}{c^n} \not\to 0\}$. *Если* $1 \leq c < \infty$, *тогда*

$$(u, cu) \in A$$

Доказательство. Рассмотрим точку $(u_n, \tilde{u}_n)$ определенную следующим образом:

$$(u_n, \tilde{u}_n) = \alpha_0(v_0, v_1) + \alpha_1(\frac{v_1}{c_1}, v_2) \cdots + \alpha_n(\frac{v_n}{c_n}, v_{n+1}) \in A$$

Здесь были выбраны следующие коэффициенты:

$$\alpha_i = \frac{\prod_{j=1}^{i} c_j}{c^i} \alpha_0,\ i = 1, \ldots, n,$$

$$\alpha_0 = (1 + \frac{c_1}{c} + \ldots + \frac{\prod_{i=1}^{n} c_i}{c^n})^{-1}$$

Тогда выполнено равенство:

$$\tilde{u}_n - cu_n = \alpha_n v_{n+1} - c\alpha_0 v_0 = \frac{\frac{\prod_{i=1}^{n} c_i}{c^n} v_{n+1} - cv_0}{1 + \frac{c_1}{c} + \cdots \frac{\prod_{i=1}^{n} c_i}{c^n}}$$

[5] Стоить однако заметить, что из точек такого вида, можно перейти не только в саму себя, но и в другие точки

Далее мы покажем существование подпоследовательности $\{n_k\}$ такой, что $\tilde{u}_{n_k} - cu_{n_k} \to 0$. По определению c : $\frac{\prod_{i=1}^n c_i}{c^n} \not\to 0$. Тогда следующий ряд рассходится $1 + \frac{c_1}{c} + \cdots + \frac{\prod_{i=1}^n c_i}{c^n} \to \infty$. Обозначим $\tilde{c}_i = \frac{c_i}{c}$. Тогда нам остается только доказать неограниченность последовательности $s_n = 1+\tilde{c}_n^{-1}+\tilde{c}_n^{-1}\tilde{c}_{n-1}^{-1}+\cdots+\tilde{c}_n^{-1}\tilde{c}_{n-1}^{-1}\cdots\tilde{c}_1^{-1}$. Допустим, что существует M такое, что $s_n \leq M$ для любого n. Легко заметить, что выполнено равенство $s_n = 1 + \frac{s_{n-1}}{\tilde{c}_n}$. Тогда

$$\tilde{c}_n = \frac{s_{n-1}}{s_n - 1}$$

$$\tilde{c}_1 * \cdots * \tilde{c}_n = \prod_{i<n} \frac{s_i}{s_i - 1}\frac{1}{s_n - 1} \geq (\frac{M}{M-1})^n \frac{1}{M-1}$$

что противоречит условию, что $\frac{\prod_{i=1}^n \tilde{c}_i}{c^n} \to 0$ для всех $c > 1$. Таким образом, мы доказали существование подпоследовательности $\{n_k\}$ такой, что $\tilde{u}_{n_k} - cu_{n_k} \to 0$. Так как A - компакт мы получаем неподвижную точку $(u, cu) \in A$, беря подпоследовательность еще раз и переходя к пределу. $\square$

Из леммы следует, что доказательство линейной сходимости методом потенциалов всегда сводится к константному количеству линейных матричных неравенств. Далее мы фокусируемся на не экспоненциальном случае. Далее мы будем предполагать, что $(u, cu) \notin A$ для любого $c > 1$. При этом неподвижная точка (u, u) лежит в A для некоторого $u \in \mathbb{R}^d$.

4.2. Динамика на прямой

Пусть $(u, u) \in A$. Далее мы исследуем динамику на прямой в окресности неподвижной точки.

Предположим, что $(u, u) + t(v, w) \in A$ для маленького t. Here we consider two cases:

- (u, u) внутри отрезка, то есть для $|t| \leq \epsilon$ $(u, u) + t(v, w) \in A$,
- (u, u) на конце отрезка, тогда t предполагается либо маленьким положительным, либо маленьким отрицательным.

Предположим, что для некоторого маленького $t > 0$ мы можем сделать шаг динамики на прямой. Это означает, что для некоторого $c \neq 0$ и некоторого s

выполнено
$$\frac{u+tw}{c} = u+sv.$$
Таким образом, мы получаем, что $tw = u(c-1)+scv$. Тогда $w = \alpha_1 u + \alpha_2 v$ для некоторых α_1, α_2.

Далее мы можем предположить, что u и v линейно независимы. Действительно, иначе в силу того, что мы можем нормировать t, мы попадаем в одну из двух ситуаций:

- $(u,u)+t(0,u) \in A$
- $(u,u)+t(u,au) \in A$

Первый случай нам не интересен, так как для любого t мы немедленно переходим в точку $t=0$. Во втором случае мы замечаем, что так как мы рассматриваем только не экспоненциальный случай, выполняются неравенства
$$\frac{1+ta}{1+t} \leq 1$$
$$a \leq 1$$
Наша динамика при этом принимает вид
$$\frac{1+at_i}{c_i} = 1+t_{i+1}$$
$$c_i = \frac{1+at_i}{1+t_{i+1}} \leq \frac{1+t_i}{1+t_{i+1}}$$
$$\prod_{i=1}^{n} c_i \leq \frac{1+t_1}{1+t_{n+1}} \leq \frac{1+\epsilon}{1-\epsilon}$$
Поэтому получаем, что последовательность $\prod c_i$ ограничена.

Далее мы предполагаем u и v линейно независимыми. При этом наша динамика принимает вид
$$\frac{u+t\alpha_1 u+t\alpha_2 v}{c} = u+sv$$

$$\begin{cases} 1 + t\alpha_1 = c \\ t\alpha_2 = sc \end{cases}$$

Далее мы можем предположить, что $\alpha_1 \neq 0$, иначе $c = 1$ и мы не можем добиться никакого роста. Мы предполагаем, что $\alpha_2 \neq 0$, так как иначе $s = 0$ и мы опять же не можем добиться никакого роста. Предположим $\alpha_1 = 1$ в силу того, что мы можем произвольно нормировать t. Тогда

$$\begin{cases} 1 + t = c \\ s = \frac{t\alpha_2}{1+t} \end{cases}$$

Здесь $t \sim 0$. Рассмотрим замену переменных $\tilde{t} = 1 + t$. Тогда $\tilde{t} \sim 1$

$$\tilde{t} = c$$

$$\tilde{s} - 1 = \frac{(\tilde{t} - 1)\alpha_2}{\tilde{t}}$$

$$\tilde{s} = (\alpha_2 + 1) - \alpha_2 \frac{1}{\tilde{t}}$$

Далее мы опустим тильду $\tilde{}$, чтобы упростить обозначения. Таким образом наша динамика выглядит так: $F(t) = (\alpha_2 + 1) - \alpha_2 \frac{1}{t}$. Нас интересует поведение этой динамики в окресности 1. Поэтому мы проверим является ли 1 аттрактором. $F'(1) = \alpha_2$. Поэтому при $|\alpha_2| < 1$ 1 притягивает динамику, если же $|\alpha_2| > 1$, то оттакливает. В случае $\alpha_2 = -1$ динамика принимает вид $F(t) = \frac{1}{t}$ и не сходится к 1.

В случае $\alpha_2 = 1$ динамика принимает вид $F(t) = 2 - \frac{1}{t}$. $F(t)$, и 1 является единственной ее неподвижной точкой, и $F(t) \leq t$. Тогда динамика сходится при $t > 1$ и рассходится при $t < 1$.

$$F(t) = (\alpha_2 + 1) - \alpha_2 \frac{1}{t}$$

Пусть $F(s_n) = s_{n+1}$. Можно заметить, что $s_n = \frac{h_{n+1}}{h_n}$, где

$$h_{n+2} = (\alpha_2 + 1)h_{n+1} - \alpha_2 h_n$$

и $s_0 = \frac{h_1}{h_0} = \frac{t}{1}$. Поэтому $h_0 = 1,\ h_1 = t$ и выполнено равенство

$$\prod_{i=0}^{n} c_i = \prod_{i=0}^{n} s_i = h_{n+1}$$

Далее мы хотим изучать асимптотику $\{h_n\}_{n=1}^{\infty}$. Рассмотрим воспроизводящую функцию для $\{h_n\}$:

$$f = \frac{1 + (t - (\alpha_2 + 1))x}{1 - (\alpha_2 + 1)x + \alpha_2 x^2} = \frac{1 + (t - (\alpha_2 + 1))x}{(1 - \alpha_2 x)(1 - x)}.$$

При $\alpha_2 \neq 1$ получаем, что $f(x) = \frac{A}{1-\alpha_2 x} + \frac{B}{1-x}$ и тогда

$$h_i = A\alpha_2^i + B.$$

Заметим, что $|\alpha_2| < 1$, тогда $\prod c_i$ ограничена.

В случае $\alpha_2 = 1,\ t > 1$

$$f = \frac{1 + (t-2)x}{(1-x)^2} = \frac{2-t}{1-x} + \frac{t-1}{(1-x)^2}$$

Заметим, что $\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + \cdots$ и мы получаем линейный рост. Тогда

$$h_n = (n+1)(t-1) + (2-t)$$

$$\prod_{i=0}^{n} c_i = h_{n+1} = n(t-1) + t$$

Лемма 4.2. *Рассматривая динамику на прямой $(u,u) + t(v,w)$ в окресности неподвижной точки (u,u) мы можем получить только либо линейный рост, либо отсутствие роста. Линейный рост может быть получен только с помощью такой динамики*

$$(u,u) + t(v, u+v),$$

где $t > 0$.

$$\prod_{i=0}^{n} c_i = nt + (1+t)$$

4.3. Прямой недостаточно

Пусть

$$A = \overline{conv}\{(u,u) + t(v, u+v) + f(t)(h,h) + \phi(t)(0,h)\}$$

для маленького $t \geq 0$. Здесь мы берем $f(t) = -t^2$, $\phi(t) = \frac{t^3}{1+t}$, а u, v, h предполагаются линейно независимыми. Тогда динамика принимает вид

$$\frac{u + t(u+v) + (f(t) + \phi(t))h}{c} = u + sv + f(s)h$$

$$\begin{cases} c = 1 + t \\ s = \frac{t}{1+t} \\ \frac{f(t)+\phi(t)}{1+t} = f(\frac{t}{1+t}) \end{cases}$$

Функции $f(t)$ и $\phi(t)$ были выбраны так, чтобы третье равенство было выполнено и $\phi(t) > 0$, $t > 0$. Рассмотрим множество $Diff = \{\tilde{w} - w \mid (w, \tilde{w}) \in A\} = \overline{conv}\{tu + \phi(t)h\}$. Таким образом, для любого $z \in Diff \quad z = au + bh$, где $a, b > 0$, кроме $z = 0$. Заметьте, что $\phi(t)$ выпукла и возрастающая для маленьких t.

Таким образом в A не может быть прямой вида 4.2.

5. Заключение

В работе мы предложили альтернативный метод поиска потенциалов, который в отличии от стандартного подхода (Taylor et al. ((2018)), Taylor and Bach ((2021))) требует проверки всего константного количества линейных матричных неравенств. С помощью данного подхода мы установили альтернативное доказательство скорости сходимости последней итерации экстраградиентного метода в комонотонном случае. Также мы провели анализ задачи поиска потенциалов и связали задачу поиска потенциалов вида $P_k + kQ_k$ с динамикой на прямой в некотором компактном выпуклом множестве.

Тем не менее до сих пор остается множество открытых вопросов. Так с помощью нашего подхода непонятно как работать со скоростями сходимости по типу $O(k^2)$. Также в качестве будущих направлений для исследований мы отмечаем дальнейший анализ динамики сформулированной для задачи поиска потенциалов. Тут однако может пригодиться серьезное вовлечение в алгебраическую геометрию и компьютерную алгебру.

Список литературы

A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton University Press, 2009.

J. Diakonikolas, C. Daskalakis, and M. Jordan. Efficient methods for structured nonconvex-nonconcave min-max optimization. 2021.

Y. Drori and M. Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. 2014.

N. Golowich, S. Pattathil, and C. Daskalakis. Tight last-iterate convergence rates for no- regret learning in multi-player games. 2020a.

N. Golowich, S. Pattathil, C. Daskalakis, and A. Ozdaglar. Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems. 2020b.

I. Goodfellow and J. Shlens. Explaining and harnessing adversarial examples. 2014.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. 2014.

E. Gorbunov, N. Loizou, and G. Gidel. Extragradient method: O(1/k) last-iterate convergence for monotone variational inequalities and connections with cocoercivity. 2022a.

E. Gorbunov, A. Taylor, S. Horváth, and G. Gidel. Convergence of proximal point and extragradient-based methods beyond monotonicity: the case of negative comonotonicity. 2022b.

M. Hast, K. J. Astrom, B. Bernhardsson, and S. Boyd. Pid design by convex-concave optimization. 2013.

G. M. Korpelevich. The extragradient method for finding saddle points and other problems. 1976.

A. Madry, A. Makelov, L. Shmidt, and D. Tsipras. Towards deep learning models resistant to adversarial attacks. 2017.

A. Mokhtari, A. Ozdaglar, and S. Pattathil. Proximal point approximations achieving a conver- gence rate of o(1/k) for smooth convex-concave sad- dle point problems: Optimistic gradient and extra- gradient methods. 2019.

A. Nemirovski. Prox-method with rate of con- vergence o(1/t) for variational inequalities with lip- schitz continuous monotone operators and smooth convex-concave saddle point problems. 2004.

L. Popov. A modification of the arrow-hurwicz method for search of saddle points. 1980.

A. Taylor and F. Bach. Stochastic first-order methods: non-asymptotic and computer-aided analyses via potential functions. 2021.

A. Taylor, J. M. Hendrickx, and F. Glineur. Smooth strongly convex interpolation and exact worst-case performance of first-order methods. 2017.

A. Taylor, B. V. Scoy, and L. Lessard. Lyapunov functions for first-order methods: Tight automated convergence guarantees. 2018.