

Санкт–Петербургский государственный университет

*ГУМЕНЮК Алёна Сергеевна*

Выпускная квалификационная работа

*Разработка программного комплекса для  
численной идентификации эпидемиологических  
моделей на основе данных наблюдений*

Уровень образования: бакалавриат

Направление 01.03.02 «Прикладная математика и информатика»

Основная образовательная программа СВ.5005.2016 «Прикладная  
математика, фундаментальная информатика и программирование»

Профиль «Математическое моделирование систем  
и процессов управления»

Научный руководитель:

доцент, кафедра вычислительных методов  
механики деформируемого тела,  
к.т.н. Громов Дмитрий Валерьевич

Рецензент:

доцент, кафедра информационных систем,  
к.ф.-м.н. Еремин Алексей Сергеевич

Санкт-Петербург

2020 г.

# Содержание

Введение . . . . .	3
Постановка задачи . . . . .	4
Обзор литературы . . . . .	5
<b>Глава 1. Задача динамической оптимизации . . . . .</b>	<b>7</b>
1.1. Интерполяция с помощью полиномов Лагранжа . . . . .	8
1.2. Алгебраическое представление операции дифференцирования . . . . .	10
1.3. Алгебраическое представление дифференциального уравнения . . . . .	12
1.4. Исследование сходимости метода ортогональных коллокаций для оценки производных функции . . . . .	13
1.5. Применение метода ортогональных коллокаций для решения задач динамической оптимизации . . . . .	15
<b>Глава 2. Задача идентификации . . . . .</b>	<b>18</b>
2.1. Постановка задачи . . . . .	18
2.2. Исследуемая эпидемическая модель . . . . .	19
2.3. Программная реализация . . . . .	20
2.3.1 Описание функции <code>fmincon</code> . . . . .	20
2.3.2 Входные данные для <code>fmincon</code> . . . . .	21
2.4. Результаты численных экспериментов . . . . .	23
2.4.1 Результаты численных экспериментов для системы SIR	23
2.4.2 Результаты численных экспериментов для системы SIS	28
2.5. Анализ результатов . . . . .	29
<b>Заключение . . . . .</b>	<b>31</b>
<b>Список литературы . . . . .</b>	<b>32</b>

## Введение

В современном мире развитие методов и технологий компьютерного моделирования усиливает интерес к изучению сложных систем, в частности систем, моделирующих эпидемиологические процессы.

Использование корректных моделей распространения инфекционных заболеваний позволяет прогнозировать эпидемиологическую ситуацию, анализировать масштабы распространения заболеваний, а также решать ряд задач, направленных на управление исследуемым процессом.

Модели могут быть различной сложности. В общем случае, популяция разбивается на несколько групп, различающихся по своим свойствам (восприимчивые, заразные, обладающие иммунитетом и т. д.), а затем задаются правила перехода из одной группы в другую. Динамика таких моделей обычно описывается системами дифференциальных или интегродифференциальных уравнений.

После записи системы уравнений, описывающей рассматриваемый эпидемиологический процесс, возникает вопрос определения параметров модели. Параметры являются характеристиками эпидемиологического процесса, например, вероятность заражения, скорость выздоровления и прогрессирования заболевания.

Но каким же способом искать значения параметров? Как правило, мы имеем данные о заболеваемости и смертности, связанных с исследуемым заболеванием. Эти данные позволяют получить оценки параметров, характеризующих процессы, происходящие в части популяции, которая находится под контролем медицинских служб, например, скорость выздоровления больных, получающих лечение, и смертность среди выявленных больных. Но, к сожалению, этих данных недостаточно для оценки параметров, характеризующих процесс инфицирования. Значения недостающих параметров можно выяснить из требования того, чтобы решение модели как можно точнее описывало динамику заболеваемости и смертности. Для решения этой задачи может быть использована процедура идентификации параметров, основанная на использовании методов динамической оптимизации, которая описана в данной работе.

## Постановка задачи

Целью дипломной работы является разработка программного комплекса для численной идентификации эпидемиологических моделей на основе данных наблюдений.

Для достижения поставленной цели были решены следующие задачи:

1. Сформулирована задача идентификации эпидемиологических моделей на основе данных о заболеваемости.
2. Изучена и программно реализован численный метод динамической оптимизации на основе метода ортогональных коллокаций.
3. Численно исследована задача параметрической идентификации для конкретного класса эпидемиологических моделей.
4. Проанализированы полученные результаты.

## Обзор литературы

Задача моделирования и математического исследования динамики эпидемических процессов имеет долгую историю. Среди первых статей, посвященных формальному описанию динамики эпидемических процессов можно упомянуть классическую статью [10], которая была опубликована в 1927 году. В этой статье была развита теория, которая впоследствии была названа теорией Кермака — Маккендрика, в честь авторов статьи. В основе этой теории лежит использование модели, структурирующей всю популяцию на три группы. Такая модель называется SIR. Эта аббревиатура происходит от английских слов «Susceptible—Infected—Recovered», что означает «восприимчивые — инфицированные — выздоровевшие». Одной из классических работ, посвященных современной математической эпидемиологии, является книга Андерсона и Мэя [8], в которой представляются и подробно исследуются математические модели, описывающие различные варианты взаимоотношений человека с инфекционными агентами на популяционном уровне. В ней также приведен обширный список публикаций, посвященных проблемам математической эпидемиологии.

Существует ряд статей, посвященных применению методов динамической оптимизации для решения задач математической эпидемиологии. Так, в статье [4], посвященной задаче управления распространением ВИЧ-инфекции, численные методы оптимального управления используются для определения оптимальных профилей инвестиций в различные программы профилактики и лечения. Описанный в этой статье подход опирается на теорию ортогональных коллокаций, которая основана на использовании интерполяции Лагранжа на неравномерных сетках. Некоторые математические факты, используемые в работе изложены в [7] и [9]. А именно, статья [7] описывает барицентрическую интерполяцию, которая является вариацией интерполяции полиномами Лагранжа, а книга [9] содержит описание свойств различных ортогональных многочленов, которые используются для нахождения узловых интерполяционных точек. Использование интерполяции на неравномерной сетке позволяет избежать так называемого эффекта Рунге, впервые описанного в [3]. Этот эффект представляет со-

бой нежелательную осцилляцию на границах промежутка, возникающую при интерполяции на равномерной сетке, которая значительно уменьшается при использовании специальных неравномерных сеток. Наконец, в статье [5] представлен обзор псевдоспектральных методов, основанных на коллокации в точках Лежандра—Гаусса, Лежандра—Гаусса—Радо и Лежандра—Гаусса—Лобатто, которые как раз и служат уменьшению эффекта Рунге.

# Глава 1. Задача динамической оптимизации

Перед тем как перейти к основной части работы необходимо сформулировать общую постановку задачи динамической оптимизации, которая будет использована при решении задач идентификации. Задача динамической оптимизации включает в себя формулирование целевой функции (или функционала), и определение набора линейных или нелинейных ограничений типа равенств или неравенств.

Таким образом, пусть у нас есть некоторые ограничения, образующие допустимое множество

$$X = \{x \in R^n \mid g_i(x) \geq 0, \quad q_j(x) = 0, \quad i = 1, \dots, r, \quad j = 1, \dots, l\} \subset R^n,$$

и имеется целевой функционал

$$J(x, \xi) : X \times R^m \rightarrow R, \tag{1}$$

экстремум которого необходимо найти в области допустимых значений. В этом выражении  $\xi \in R^m$  — это вектор параметров.

Решением задачи оптимизации будет называться пара

$$(x^*, \xi^*) \in X \times R^m : J(x^*, \xi^*) = \min_{\substack{x \in X \\ \xi \in R^m}} J(x).$$

В задачах динамической оптимизации состояние  $x$  подчиняется дифференциальному уравнению вида

$$\dot{x}(t) = f(t, x(t), \xi), \tag{2}$$

где  $x(t) \in X$  — вектор фазовых переменных,  $t \in [0, T]$ ,  $x(0) = x_0$ , а  $\xi$  — параметры.

Задача оптимизации состоит в поиске наилучшего значения параметров  $\xi^*$  и соответствующей им траектории  $x^*$ , которые минимизируют значение некоторой целевой функции (1) при наличии ограничений, заданных в виде дифференциальных уравнений связи (2).

В дальнейшем мы будем рассматривать целевые функционалы, записанные в интегральной форме

$$J = \int_0^T F(t, x(t), \xi) dt, \quad (3)$$

где  $F$  — заданная скалярная функция. Такая задача называется задачей Лагранжа.

Существуют различные подходы к решению задач динамической оптимизации. Их можно разделить на два класса: прямые и непрямые. Непрямые методы решения задач оптимизации основаны на решении некоторой вспомогательной задачи. Классическим примером непрямого метода является принцип максимума Понтрягина [1], в котором оптимальное значение искомых параметров (управлений) определяется через решение системы  $2n$  дифференциальных уравнений специального вида. В свою очередь, прямые методы основаны непосредственно на нахождении экстремального значения функционала с использованием различных методов нелинейного программирования [2].

## 1.1 Интерполяция с помощью полиномов Лагранжа

Для решения прямой задачи динамической оптимизации необходимо найти способ эффективного численного интегрирования системы дифференциальных уравнений. В последнее время большую популярность приобрели методы, основанные на представлении дифференциального уравнения системой нелинейных алгебраических уравнений. В данной работе представлен частный случай этого класса, а именно, метод ортогональных коллокаций.

Метод ортогональных коллокаций основан на интерполяции решения дифференциального уравнения с использованием полиномов Лагранжа [4]. Пусть интересующий нас интервал  $[0, T]$  разбит на  $n_{cp} - 1$  подынтервалов  $[\tau_i, \tau_{i+1}]$ ,  $i = 1, \dots, n_{cp} - 1$ , где  $\tau_i$  - это узловые точки:  $0 = \tau_1 < \tau_2 < \dots < \tau_{n_{cp}-1} < \tau_{n_{cp}} = T$ . Обозначим значение вектора фазовых переменных  $x \in R^n$  в точке  $\tau_i$  через  $X_i = x(\tau_i)$ . Тогда интерполирующий полином

Лагранжа будет иметь вид

$$\tilde{x}(t) = \sum_{i=1}^{n_{cp}} L_i(t) X_i, \quad (4)$$

где базисные полиномы

$$L_i(t) = \prod_{\substack{j=1 \\ j \neq i}}^{n_{cp}} \frac{t - \tau_j}{\tau_i - \tau_j}.$$

Заметим, что  $L_i(t)$  являются полиномами степени  $n_{cp} - 1$ , где  $n_{cp}$  — число узлов, и удовлетворяют условию:

$$L_i(\tau_j) = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Известно, что существует единственный многочлен степени не превосходящей  $n$ , принимающий заданные значения в  $n+1$  точке. Это значит, что интерполяция полиномом Лагранжа является точной для любого полинома степени  $n \leq n_{cp}$ , заданном в  $n_{cp} + 1$  точках. Из этого следует, в частности, что  $\sum_{i=1}^{n_{cp}} L_i(t) \equiv 1$ .

Существенный недостаток интерполяции с помощью полиномов Лагранжа заключается в том, что на равномерной сетке интерполяция имеет значительную погрешность на границах промежутка. Данное явление называется феноменом Рунге [3]. Для уменьшения ошибки интерполяции принято использовать неравномерную сетку [4].

В качестве узловых точек для неравномерной сетки можно использовать коллокационные точки Лежандра — Гаусса, Лежандра — Гаусса — Лобатто или Лежандра — Гаусса — Радо, которые принадлежат открытому интервалу  $(-1, 1)$ , закрытому интервалу  $[-1, 1]$  и полу-открытому интервалу  $(-1, 1]$  (или  $[-1, 1)$ ) соответственно [5].

Пусть  $N$  - число коллокационных точек,  $P_N(x)$  — многочлен Лежанд-

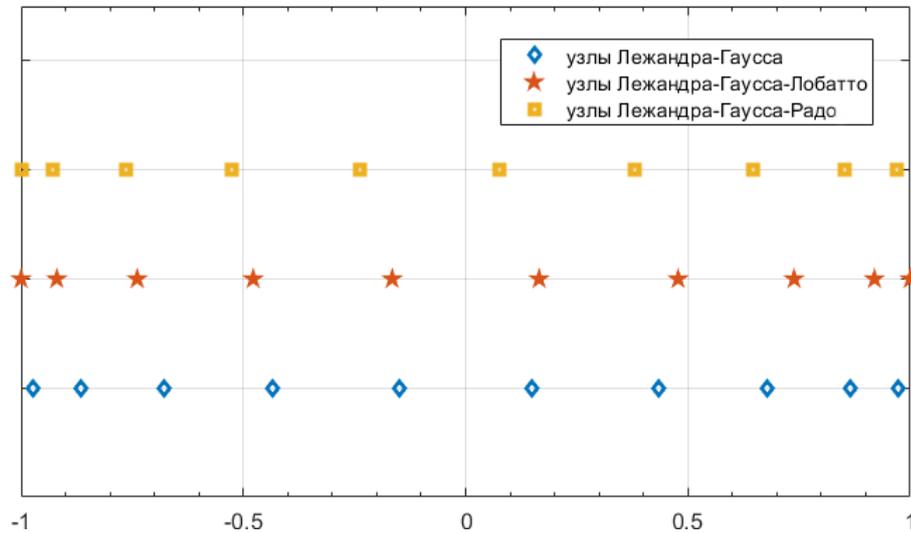
ра степени  $N$ :

$$P_N(x) = \frac{1}{2^N N!} \frac{d^N}{dx^N} (x^2 - 1)^N.$$

Тогда:

- узлы Лежандра—Гаусса — корни полинома  $P_N(x)$ ,
- узлы Лежандра—Гаусса—Лобатто — корни полинома  $\dot{P}_{N-1}(x)$  вместе с точками  $-1$  и  $1$ ,
- узлы Лежандра—Гаусса—Радо — корни полинома  $P_{N-1}(x) + P_N(x)$ .

Различия между представленными узловыми точками при  $N = 10$  представлены на рис. 1.

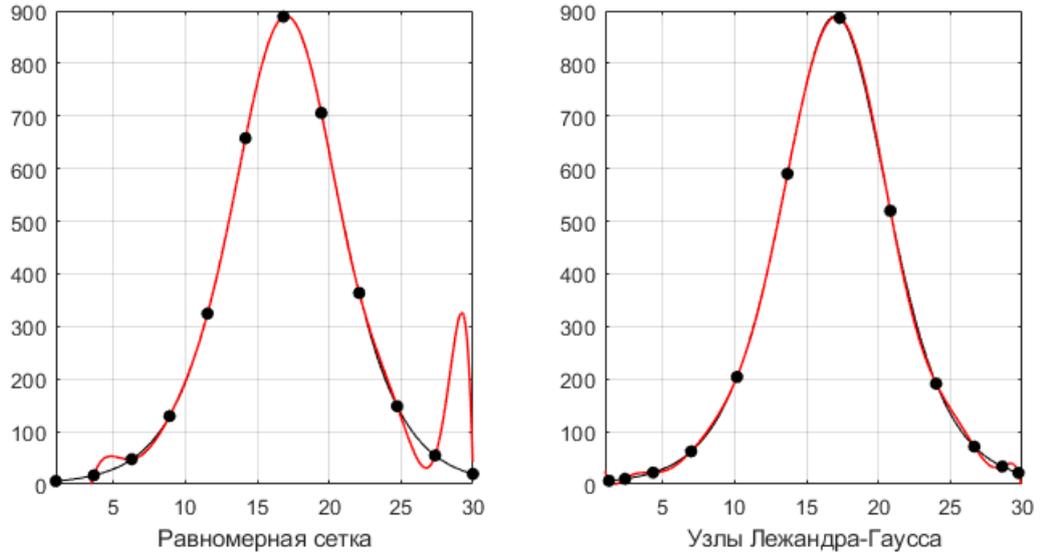


**Рис. 1:** Узловые точки Лежандра—Гаусса, Лежандра—Гаусса—Лобатто и Лежандра—Гаусса—Радо при  $N = 10$

На рис. 2 проиллюстрировано, насколько уменьшается эффект Рунге при использовании неравномерной сетки.

## 1.2 Алгебраическое представление операции дифференцирования

Представление функций с помощью интерполяционного полинома Лагранжа имеет ряд достоинств: процедуры интегрирования и дифферен-



**Рис. 2:** Интерполяция полиномами Лагранжа на равномерной и неравномерной сетках

цирования несложными преобразованиями сводятся к алгебраическим операциям. Убедимся в этом на примере дифференцирования.

Рассмотрим интерполяционный полином Лагранжа (4). Дифференцируя, имеем

$$\dot{\tilde{x}}(t) |_{t=\tau_k} = \sum_{i=1}^{n_{cp}} L'_i(t) |_{t=\tau_k} X_i.$$

Данное выражение является обобщением метода конечных разностей для численного вычисления производных. Таким образом, вектор производных интерполирующей функции  $\tilde{x}(t)$  в узловых точках  $\tau_i$ ,  $i = 1, \dots, n_{cp}$ , имеет вид

$$\dot{\tilde{x}} = DX,$$

где  $D$  — матрица дифференцирования,

$$D = \begin{pmatrix} L'_1(\tau_1) & L'_2(\tau_1) & \dots & L'_{n_{cp}}(\tau_1) \\ L'_1(\tau_2) & L'_2(\tau_2) & \dots & L'_{n_{cp}}(\tau_2) \\ \dots & \dots & \dots & \dots \\ L'_1(\tau_{n_{cp}}) & L'_2(\tau_{n_{cp}}) & \dots & L'_{n_{cp}}(\tau_{n_{cp}}) \end{pmatrix}, \quad (5)$$

а  $X$  представляет собой  $(n_{cp} \times n)$ -матрицу. Производные базовых полиномов Лагранжа вычисляются по формуле [7]:

$$L'_i(\tau_j) = \begin{cases} \frac{L'(\tau_i)}{L'(\tau_j)(\tau_i - \tau_j)}, & i \neq j, \\ -\sum_{k \neq j} L'_i(\tau_k), & i = j. \end{cases}$$

Из представления базовых полиномов Лагранжа очевидно следует, что для любого фиксированного набора узловых точек матрица дифференцирования  $D$  считается однократно и может быть сохранена в памяти компьютера для дальнейшего использования.

### 1.3 Алгебраическое представление дифференциального уравнения

Рассмотрим систему дифференциальных уравнений

$$\dot{x} = F(x), \quad (6)$$

где  $F$  — достаточно гладкая векторная функция,  $F : R^n \rightarrow R^n$ . Потребуем, чтобы производные интерполяционного полинома  $\tilde{x}(t)$ , вычисленные в точках  $\tau_i$ , совпадали со значениями правых частей системы (6), вычисленных для величин  $X_i$ ,  $i = 1, \dots, n_{cp}$ :

$$\dot{\tilde{x}} = DX = \mathbf{F}(X), \quad (7)$$

где  $\mathbf{F}$  — матрица размерности  $(n_{cp} \times n)$

$$\mathbf{F} = \left( F^T(X_1), F^T(X_2), \dots, F^T(X_{n_{cp}}) \right)^T.$$

Выражение (7) представляет собой систему нелинейных алгебраических уравнений, состоящую из  $n_{cp} \times n$  уравнений и  $(n_{cp} - 1) \times n$  неизвестных, так как начальное положение известно. Для того, чтобы система не была переопределенной, удалим первые строки матриц  $D$  и  $\mathbf{F}$ , получая тем самым урезанные матрицы  $\bar{D}$  и  $\bar{\mathbf{F}}$ .

Решая систему  $\overline{D}X - \overline{F}(X) = 0$ , находим вектор  $X$ , на основе которого строится интерполяционный полином, аппроксимирующий решение дифференциального уравнения.

## 1.4 Исследование сходимости метода ортогональных коллокаций для оценки производных функции

Для того, чтобы проиллюстрировать использование метода ортогональных коллокаций мы изучим, как использование неравномерной сетки влияет на точность аппроксимации производных. В отличие от классической задачи полиномиальной интерполяции мы будем изучать, как оценка производных функции в интерполяционных точках, вычисленная с использованием матрицы дифференцирования (5), отличается от значений производных, полученных аналитически.

**Пример 1.** Рассмотрим функцию, отображающую зависимость количества смертей от времени в период эпидемии чумы в Бомбее 1905—1906 годов

$$x(t) = 890 \sec^2(0.2t - 3.4). \quad (8)$$

Производная данной функции будем иметь вид:

$$\dot{x}(t) = \frac{-356 \sinh(0.2t - 3.4)}{\cosh^3(0.2t - 3.4)}.$$

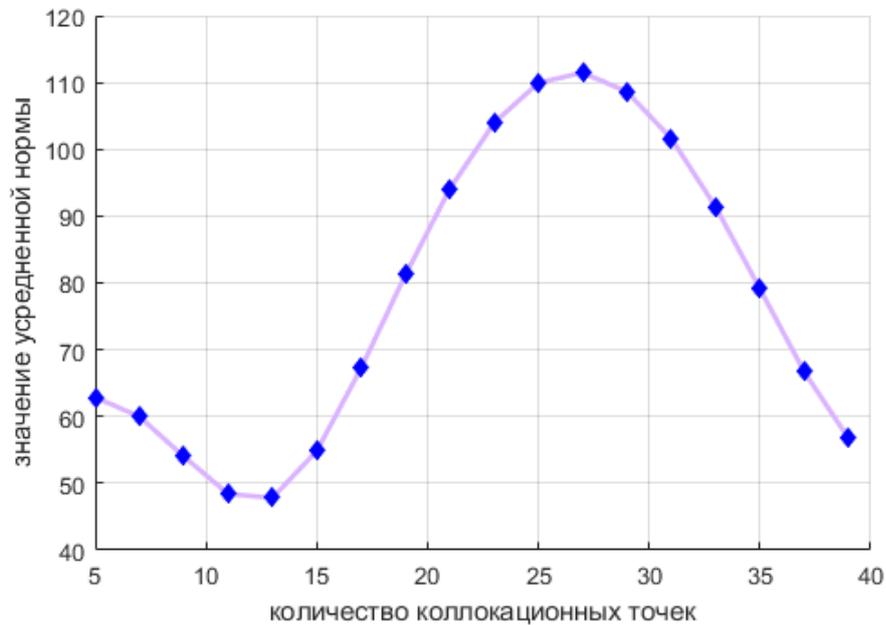
Оценим разность между приближенной, то есть полученной на основе интерполяции, и аналитической производной, вычисленной с помощью среды Matlab. Оценка проводилась с использованием следующего алгоритма:

1. Выбираем  $\tau_{cp}$  — узловые точки (равномерная сетка, узлы Лежандра — Гаусса — Радо).
2. Вычисляем на выбранной сетке матрицу производных (5) и умножаем ее на соответствующие значения функции в узлах сетки —  $D \cdot x(\tau_{cp})$ .
3. Аналитически вычисляем производную  $\dot{x}(\tau_{cp})$  на выбранной сетке.

4. Вычисляем для полученных величин значение нормы ошибки, скорректированной с учетом числа интерполяционных точек:

$$\epsilon = \frac{\|D \cdot x(\tau_{cp}) - \dot{x}(\tau_{cp})\|}{n_{cp}}.$$

В результате получаем зависимости величины невязки от числа узлов в сетке для различного выбора узловых точек, представленные на рис. 3 и рис. 4.



**Рис. 3:** Зависимость усредненной нормы  $\epsilon$  от количества узловых точек при выборе равномерной сетки для функции (8)

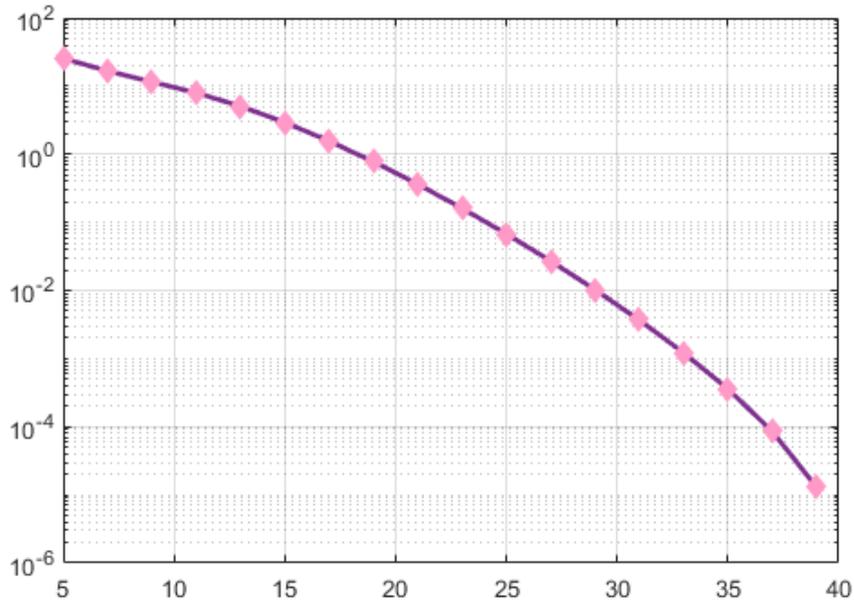
**Пример 2.** Рассмотрим функцию Рунге

$$R(x) = \frac{1}{1 + 25x^2}. \quad (9)$$

Её производная будет иметь вид:

$$\dot{R}(x) = -50 \frac{x}{(1 + 25x^2)^2}.$$

Применяя алгоритм из предыдущего примера, получаем результаты, представленные на рис. 5 и рис. 6.



**Рис. 4:** Зависимость усредненной нормы  $\epsilon$  от количества узловых точек при выборе узлов Лежандра — Гаусса — Радо для функции (8)

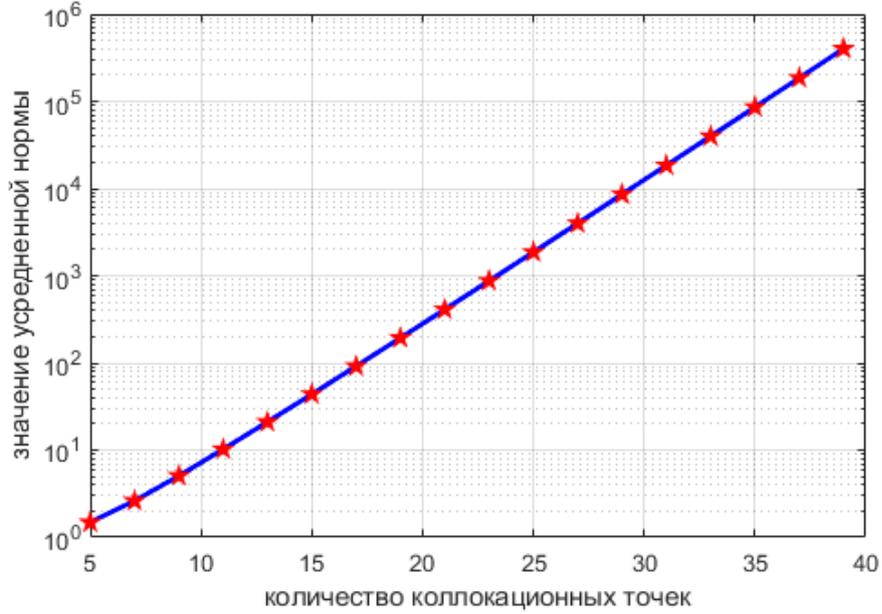
Таким образом, если интерполировать такую функцию, как (9), по  $n$  равноотстоящим узлам, то аппроксимация производных функции будет иметь ошибку, которая растет с увеличением числа интерполяционных точек. С возрастанием степени полинома, погрешность оценки, как можно заметить по рис. 5, будет стремиться к бесконечности. Однако и для более хорошей функции (8) выбор неравномерной сетки значительно улучшает результаты численной оценки значений производных функции.

## 1.5 Применение метода ортогональных коллокаций для решения задач динамической оптимизации

Используя метод ортогональных коллокаций, можно свести задачу динамической оптимизации к задаче нелинейного программирования.

**Постановка задачи.** Задачей нелинейного программирования называется задача нахождения минимума нелинейной функции многих переменных, когда на переменные накладываются ограничения типа равенств или неравенств:

$$\min \hat{f}(x_1, x_2, \dots, x_n), \quad (10)$$



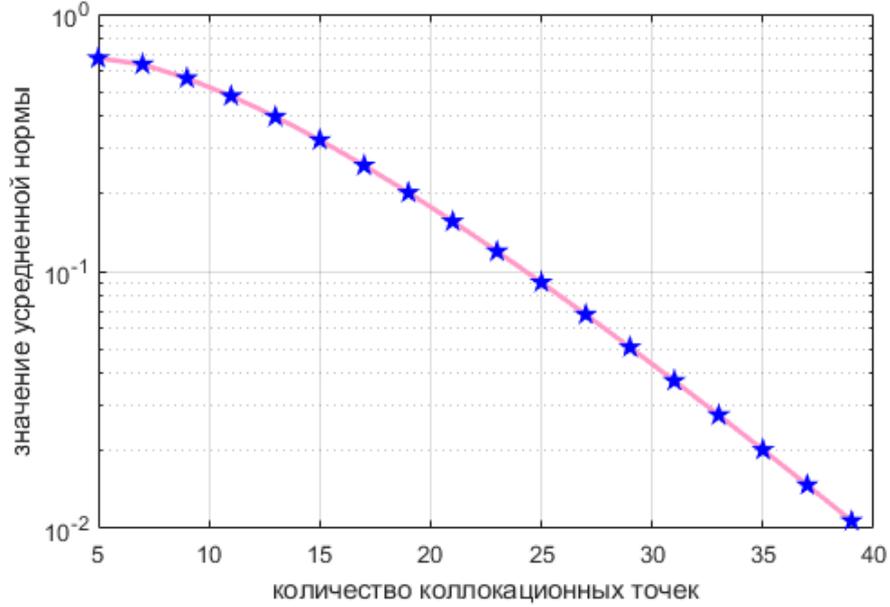
**Рис. 5:** Зависимость усредненной нормы  $\epsilon$  от количества узловых точек при выборе равномерной сетки для функции (9)

при наличии ограничений

$$\begin{aligned} g_i(x_1, x_2, \dots, x_n) &\geq 0, & i = 1, \dots, r, \\ q_j(x_1, x_2, \dots, x_n) &= 0, & j = 1, \dots, l. \end{aligned} \quad (11)$$

**I.** Рассмотрим систему дифференциальных уравнений, описывающих динамику системы (2). Пусть интересующий нас интервал  $[0, T]$  разбит на  $n_{cp} - 1$  подынтервалов узловыми точками  $\tau_i$ :  $0 = \tau_1 < \tau_2 < \dots < \tau_{n_{cp}} = T$ .  $X_i = x(\tau_i)$  — значение вектора фазовых переменных  $x \in R^n$  в точке  $\tau_i$ . Вычислим матрицу производных  $D$  (5) в узлах интерполяции. Пусть  $F(X, \xi)$  — значения правых частей системы (2). Потребуем, чтобы производные интерполяционного полинома системы совпадали с значениями правых частей системы:  $DX = F(X, \xi)$ . Избавимся от переопределенности, посредством удаления первых строк  $D$  и  $F$ . В результате система дифференциальных уравнений связи преобразуется в систему нелинейных алгебраических уравнений

$$\bar{D}X - \bar{F}(X, \xi) = 0. \quad (12)$$



**Рис. 6:** Зависимость усредненной нормы  $\epsilon$  от количества узловых точек при выборе узлов Лежандра — Гаусса — Радо для функции (9)

II. Построим квадратуру Гаусса для целевого функционала  $J$  (3):

$$J = \int_0^T f(t, x(t), \xi) dt \approx \frac{T}{2} \sum_{i=1}^n w_i f(\tau_i), \quad (13)$$

где веса  $w_i$  вычисляются по формуле [6]:

$$w_i = \frac{2}{(1 - \tau_i^2)[\dot{P}_N(\tau_i)]^2},$$

где  $P_N$  — полином Лежандра степени  $N$ .

Таким образом, нам удалось свести задачу динамической оптимизации к задаче нелинейного программирования, представив (2) в виде (13), который соответствует (10), а (3) в виде (12), который соответствует (11). Эта задача может быть решена с помощью стандартного инструмента решения нелинейных задач математического программирования, реализованного в среде Matlab — функции `fmincon`.

## Глава 2. Задача идентификации

Постановка и решение задачи параметрической идентификации моделей сложных систем существенно зависит от специфики самих моделей и особенностей представления данных измерений. Мы имеем ряд особенностей:

- фрагментарность - невозможность наблюдать все состояния системы в каждый момент времени;
- неоднородность, возникающая в силу различий в постановках экспериментов, методов и средств измерения;
- разномасштабность, состоящая в том, что данные измерений относятся к разным масштабам изменчивости.

Ниже мы покажем, как задачи параметрической идентификации может быть решена с помощью методов динамической оптимизации.

### 2.1 Постановка задачи

Рассмотрим задачу идентификации  $l$ -мерного вектора  $\xi$  параметров модели нелинейной системы дифференциальных уравнений по результатам измерений, проводимых в  $N$  точках  $\tau_i$ , заданных на некотором интервале  $[0, T]$

Пусть дифференциальное уравнение, описывающее динамику системы имеет вид

$$\dot{\tilde{x}} = \tilde{f}(\tilde{x}, \xi, t),$$

где  $\tilde{x}$  — это  $n$ -мерный вектор фазового пространства состояния системы,  $\tilde{x}(0) = \tilde{x}_0$ ,  $t \in [0, T]$ . Исходное состояние системы  $\tilde{x}_0$  полагаем заданным.

Измерениям подвергается вектор наблюдаемых величин

$$\tilde{\psi}(t) = \tilde{\psi}(\tilde{x}(t, \xi)).$$

Функции  $\tilde{f}(\tilde{x}, \xi, t)$ ,  $\tilde{\psi}(\tilde{x}(t, \xi))$  будем считать однозначными, ограниченными, непрерывными и дифференцируемыми по всем своим аргументам во всей

области их определения.

**Постановка задачи.** Таким образом, задачу идентификации можно сформулировать следующим образом:

*Минимизировать функционал*

$$I(\xi) = \sum_{i=1}^N (y_i - \tilde{\psi}[\tilde{x}(\tau_i, \xi)])^2 \quad (14)$$

*при наличии ограничений, заданных в виде дифференциальных уравнений*

$$\dot{\tilde{x}} = \tilde{f}(\tilde{x}, \xi, t).$$

*Оптимизация должна осуществляться относительно неизвестных переменных  $\xi$  и, если необходимо, относительно начальных условий.*

## 2.2 Исследуемая эпидемическая модель

Рассмотрим некоторую популяцию, состоящую из  $N$  особей, которая в каждый момент  $t$  времени подразделяется на три группы:

- $S(t)$  (susceptible) — здоровые особи, которые находятся в группе риска;
- $I(t)$  (infected) — инфицированные особи, которые являются переносчиками инфекции;
- $R(t)$  (removed) — особи, которые более не подвержены инфекции.

Таким образом, в каждый момент времени

$$S(t) + I(t) + R(t) = N.$$

Модель распространения эпидемии SIR (система Кермака — Маккенд-

рика [10]) с учетом рождаемости и смертности имеет вид

$$\begin{cases} \dot{S}' = -\frac{\beta}{N}S'I' + \mu - \mu S', \\ \dot{I}' = \frac{\beta}{N}S'I' - \gamma I' - \mu I', \\ \dot{R}' = \gamma I' - \mu R', \end{cases} \quad (15)$$

где  $\beta$  — коэффициент «заболеваемости», то есть вероятности получения болезни в случае контакта восприимчивого индивидуума с инфицированным,  $\gamma$  — коэффициент «выздоровления»,  $\mu$  — коэффициент притока и оттока популяции.

Так как  $\dot{N} = 0$ , то есть размер популяции постоянен, мы можем перейти к безразмерным величинам  $S = \frac{S'}{N}$ ,  $I = \frac{I'}{N}$ ,  $R = \frac{R'}{N}$ , которые будут обозначать доли популяции. Тогда система (15) примет вид

$$\begin{cases} \dot{S} = -\beta SI + \mu - \mu S, \\ \dot{I} = \beta SI - \gamma I - \mu I, \\ \dot{R} = \gamma I - \mu R. \end{cases} \quad (16)$$

В качестве начального состояния системы выберем  $x(0) = (0.9, 0.1, 0)^T$ . Пусть известны параметры системы  $\gamma = 3$ ,  $\mu = 1$  и также известно поведение системы в некоторые моменты времени  $tm_i$ ,  $i = 1, \dots, n_m$ . Ставим перед собой задачу определить параметр  $\beta$ .

## 2.3 Программная реализация

Для решения поставленной задачи воспользуемся методом динамической оптимизации, основанном на использовании метода ортогональных коллокаций, как описано в главе 1. Кроме этого, нам потребуется функция `fmincon` среды Matlab.

### 2.3.1 Описание функции `fmincon`

Функция `fmincon` решает задачу математического программирования и находит минимум нелинейной функции с линейными и нелинейными

ограничениями, заданными в виде равенств или неравенств.

Схематически это можно описать следующим образом:

$$\min_x f(x) \begin{cases} c(x) \leq 0, \\ c_{eq}(x) = 0, \\ A \cdot x \leq b, \\ A_{eq} \cdot x = b_{eq}, \\ lb \leq x \leq ub, \end{cases}$$

где первые две функции описывают нелинейные ограничения, заданные в виде неравенств и равенств соответственно. Следующие две строки - это линейные ограничения, заданным в виде неравенств и равенств, а последняя строка является ограничением на компоненты вектора  $x$ .

Для использования функции `fmincon` необходимо определить следующие компоненты задачи:

1. Вектор начальных значений переменных решения.
2. Целевую функцию.
3. Линейные и нелинейные ограничения.

Далее мы рассмотрим эти пункты подробнее.

### 2.3.2 Входные данные для `fmincon`

**I. Вектор переменных.** В качестве переменных выбирается вектор

$$x = \left( S_1, \dots, S_{n_{cp}}, I_1, \dots, I_{n_{cp}}, R_1, \dots, R_{n_{cp}}, \beta \right)^T,$$

где  $S_i = S(\tau_i)$ ,  $I_i = I(\tau_i)$ ,  $R_i = R(\tau_i)$  - это значения переменных  $S$ ,  $I$  и  $R$  в моменты времени  $\tau_i$ ,  $i = 1, \dots, n_{cp}$ , а  $\beta$  - неизвестный параметр системы дифференциальных уравнений (16).

**II. Целевая функция.** В нашем случае целевая функция – это функционал (14). Для заданных значений  $x$  он записывается в виде

$$I(x) = \sum_{i=1}^{n_m} (Im_i - \tilde{I}[x(\tau_i, \beta)])^2,$$

где  $Im_i$  – доля инфицированных в момент времени  $tm_i$ ,  $i = 1, \dots, n_m$ .

**III. Ограничения.** Уравнения динамики системы, а также границы значений переменных состояния формулируются в виде ограничений. В нашем случае использовались два типа ограничений: нелинейные ограничения и линейные ограничения типа равенства.

- Нелинейные ограничения.

Для задания ограничений типа  $c_{eq}(x) = 0$  воспользуемся алгоритмом, описанным в подпункте I пункта 1.5 первой главы. Нелинейные ограничения типа неравенств в данном случае не используются.

- Линейные ограничения.

Ограничения на вектор  $x$  типа  $A_{eq} \cdot x = b_{eq}$  задаются следующим образом: поскольку в каждый момент времени сумма долей популяции  $S, I$  и  $R$  должна быть равной 1, то  $b_{eq}$  будет представлять собой единичный вектор-столбец размерности  $3n_{cp} \times 1$ , а  $A_{eq}$  – матрица размерности  $3n_{cp} \times (3n_{cp} + 1)$

$$A_{eq} = \begin{pmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

Линейные ограничения типа неравенств не используются.

- Ограничения на переменные состояния.

Так как в каждый момент времени значения  $S, I$  и  $R$  не могут быть отрицательными и не могут превышать единицу, а  $\beta$  не может быть

отрицательной, то ограничения  $lb$  и  $ub$  зададим, как

$$lb = \left( 0, \dots, 0, 0 \right)^T,$$

$$ub = \left( 1, \dots, 1, \text{Inf} \right)^T.$$

## 2.4 Результаты численных экспериментов

Мы сгенерировали  $n_m = 10$  результатов измерений системы (16), соответствующих различным значениям  $\beta$ , представленные в таблицах 1-3 и выбрали  $n_{cp} = 20$  коллокационных точек типа Лежандра — Гаусса — Лобатто на промежутке  $[0, 10]$ . При осуществлении идентификации начальное значение  $\beta$  полагалось существенно отличным от модельного.

Ошибка идентификации вычислялась по формуле

$$Err = \sqrt{(I_{data} - I_{opt})^2},$$

где  $I_{data}$  — измеренная доля инфицированных, а  $I_{opt}$  — полученная в результате идентификации доля инфицированных в моменты времени  $tm_i$ ,  $i = 1, \dots, n_m$ .

### 2.4.1 Результаты численных экспериментов для системы SIR

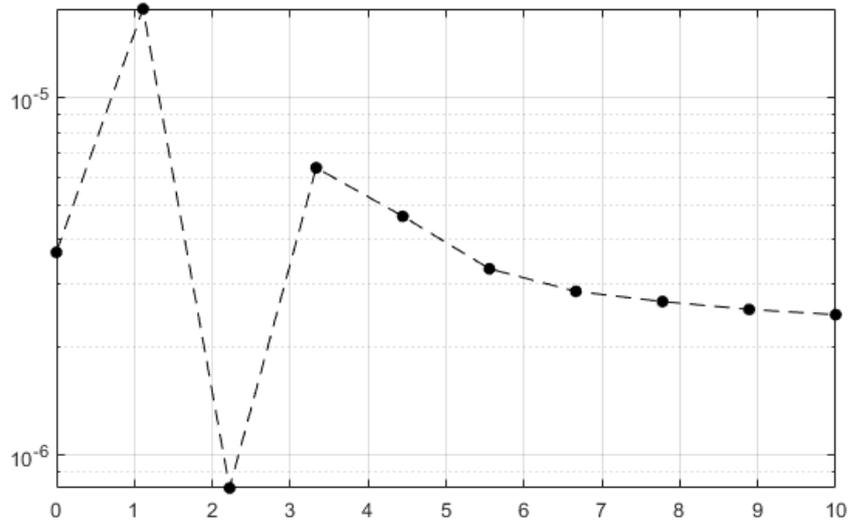
**Эксперимент 1.** В таблице 1 представлены сгенерированные значения фазовой переменной  $I$  соответствующие  $\beta = 4$  в моменты времени  $tm_i$ ,  $i = 1, \dots, n_m$ .

**Таблица 1:** Доля инфицированных в моменты времени  $tm$  в случае  $\beta = 4$

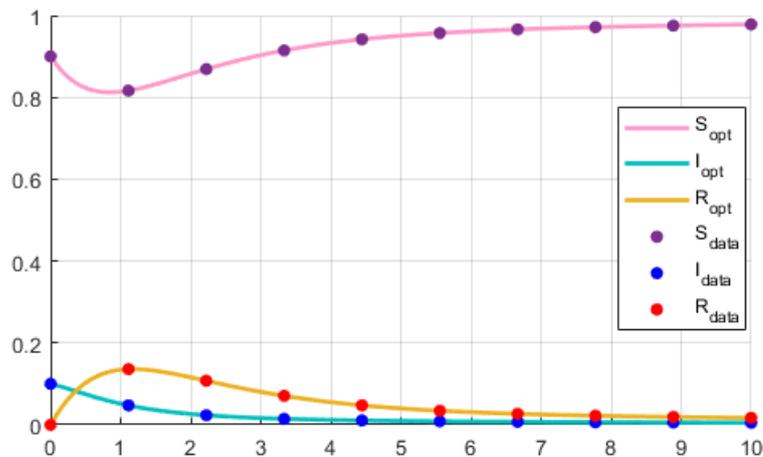
$t_m$	0	1.1111	2.2222	3.3333	4.4444
$I_{data}$	0.1000	0.0475	0.0234	0.0146	0.0107
$t_m$	5.5556	6.6667	7.7778	8.8889	10.0000
$I_{data}$	0.0085	0.0072	0.0063	0.0056	0.0050

В результате работы алгоритма за 49 итераций было получено значение  $\beta = 4.0001$ . Траектории системы, соответствующие найденному значению  $\beta$  изображены на рис. 8. Ошибка идентификации составила

$Err = 2.0701 \cdot 10^{-5}$ . На рис. 7 представлена разность между значениями  $I_{data}$  и  $I_{opt}$  в моменты времени  $tm_i, i = 1, \dots, n_m$ .



**Рис. 7:** Разность между значениями  $I_{data}$  и  $I_{opt}$  в моменты времени  $tm_i, i = 1, \dots, n_m$  для случая  $\beta = 4$



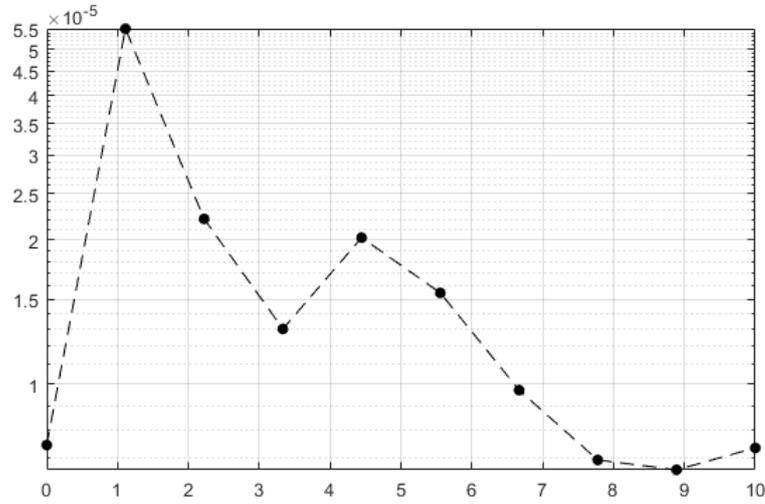
**Рис. 8:**  $S_{opt}, I_{opt}, R_{opt}$  — траектории соответствующие  $\beta = 4.0001$ , являющиеся результатом работы `fmincon`.  $S_{data}, I_{data}, R_{data}$  — измеренные данные поведения системы соответствующие  $\beta = 4$

**Эксперимент 2.** В таблице 2 представлены сгенерированные значения фазовой переменной  $I$  соответствующие  $\beta = 5$  в моменты времени  $tm_i, i = 1, \dots, n_m$ .

**Таблица 2:** Доля инфицированных в моменты времени  $tm$  в случае  $\beta = 5$

$t_m$	0	1.1111	2.2222	3.3333	4.4444
$I_{data}$	0.1000	0.0868	0.0563	0.0469	0.0465
$t_m$	5.5556	6.6667	7.7778	8.8889	10.0000
$I_{data}$	0.0484	0.0499	0.0503	0.0502	0.0501

В результате работы алгоритма за 48 итераций была получена  $\beta = 5.0002$  (рис. 10), ошибка идентификации составила  $Err = 6.8161 \cdot 10^{-5}$ . На рис. 9 представлена разность между значениями  $I_{data}$  и  $I_{opt}$  в моменты времени  $tm_i, i = 1, \dots, n_m$ .

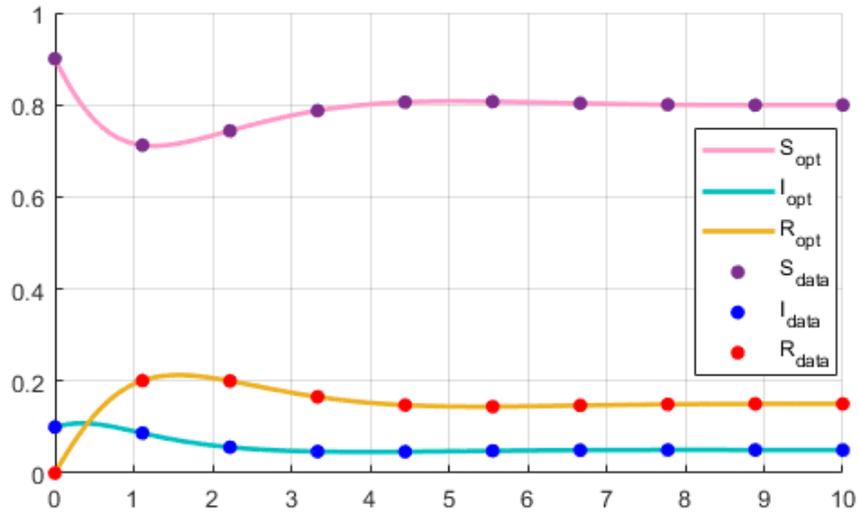


**Рис. 9:** Разность между значениями  $I_{data}$  и  $I_{opt}$  в моменты времени  $tm_i, i = 1, \dots, n_m$  для случая  $\beta = 5$

**Эксперимент 3.** В таблице 3 представлены сгенерированные значения фазовых переменных  $S, I$  и  $R$  соответствующие  $\beta = 6$  в моменты времени  $tm_i, i = 1, \dots, n_m$ .

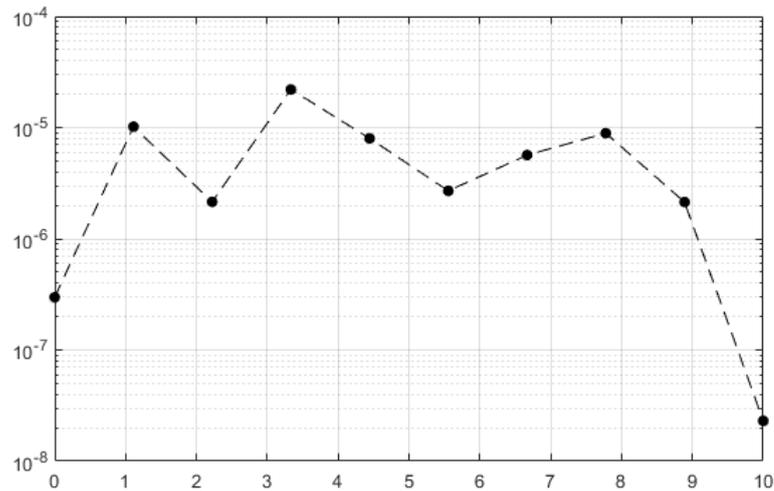
**Таблица 3:** Доля инфицированных в моменты времени  $tm$  в случае  $\beta = 6$

$t_m$	0	1.1111	2.2222	3.3333	4.4444
$I_{data}$	0.1000	0.1295	0.0834	0.0765	0.0817
$t_m$	5.5556	6.6667	7.7778	8.8889	10.0000
$I_{data}$	0.0843	0.0838	0.0833	0.0832	0.0833

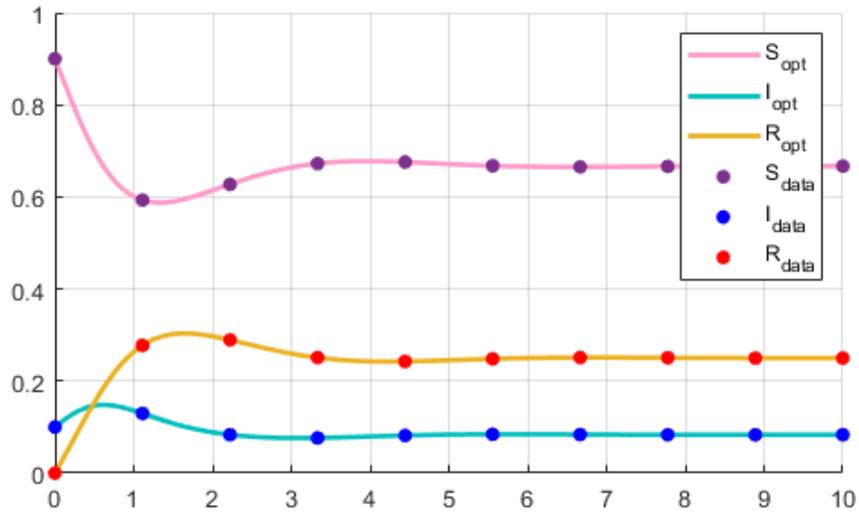


**Рис. 10:**  $S_{opt}$ ,  $I_{opt}$ ,  $R_{opt}$  — траектории соответствующие  $\beta = 5.0001$ , являющиеся результатом работы `fmincon`.  $S_{data}$ ,  $I_{data}$ ,  $R_{data}$  — измеренные данные поведения системы соответствующие  $\beta = 5$

В результате работы алгоритма за 40 итераций была получена  $\beta = 5.9999$  (рис. 12), ошибка идентификации составила  $Err = 2.7918 \cdot 10^{-5}$ . На рис. 11 представлена разность между значениями  $I_{data}$  и  $I_{opt}$  в моменты времени  $tm_i$ ,  $i = 1, \dots, n_m$ .

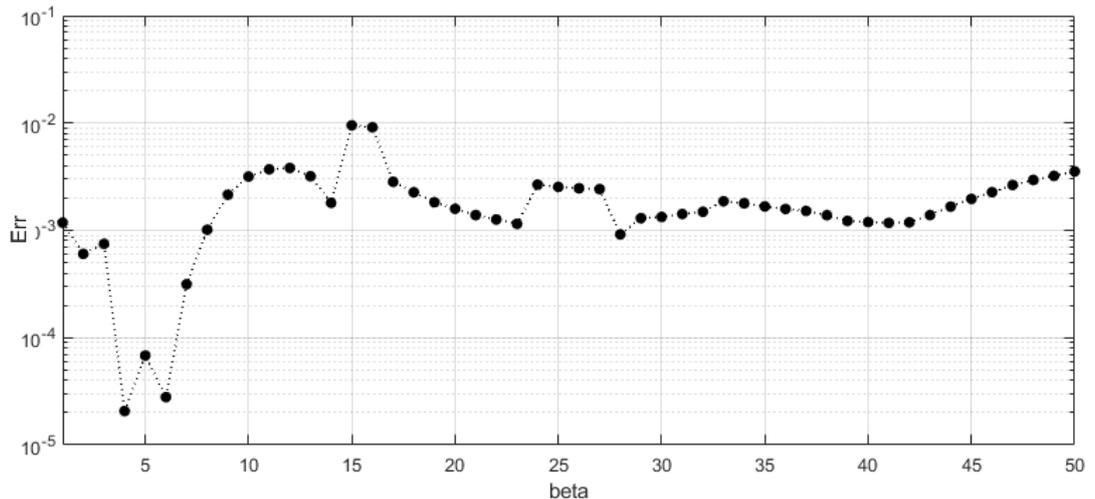


**Рис. 11:** Разность между значениями  $I_{data}$  и  $I_{opt}$  в моменты времени  $tm_i$ ,  $i = 1, \dots, n_m$  для случая  $\beta = 6$ .



**Рис. 12:**  $S_{opt}$ ,  $I_{opt}$ ,  $R_{opt}$  — траектории соответствующие  $\beta = 6.0001$ , являющиеся результатом работы `fmincon`.  $S_{data}$ ,  $I_{data}$ ,  $R_{data}$  — измеренные данные поведения системы при  $\beta = 6$  из таблицы 3.

**Случай других значений  $\beta$  для системы (16).** При рассмотрении других значений  $B_1 = \{\beta \in \mathbb{Z} \mid \beta \in [1, 50]\}$ , было выявлено, что ошибка идентификации при использовании представленного алгоритма не превышает  $10^{-2}$  (рис. 13).



**Рис. 13:** График ошибки идентификации для данных, соответствующих  $B_1$ .

## 2.4.2 Результаты численных экспериментов для систем SIS

Некоторые инфекции, например, грипп, не дают какого-либо длительного иммунитета и каждой особь популяции осуществляется переход из разряда восприимчивых в разряд инфицированных, а затем вновь в разряд восприимчивых. Для моделирования таких процессов существует SIS модель: параметр  $\beta$  отвечает за скорость перехода из разряда восприимчивых в разряд инфицированных, а параметр  $\gamma$  обратно:

$$\begin{cases} \dot{S} = -\beta SI + \gamma I, \\ \dot{I} = \beta SI - \gamma I. \end{cases} \quad (17)$$

Протестируем работу алгоритма на представленной системе. Пусть нам известно начальное состояние системы  $x(0) = (0.99 \ 0.01)^T$  и известно состояние системы в некоторые моменты времени  $tm_i$ ,  $i = 1, \dots, n_m$ ,  $n_m = 10$ . Рассмотрим промежуток  $[0, 10]$ , число коллокационных точек  $n_{cp} = 20$ . Проведем идентификацию обоих параметров  $\beta$  и  $\gamma$ .

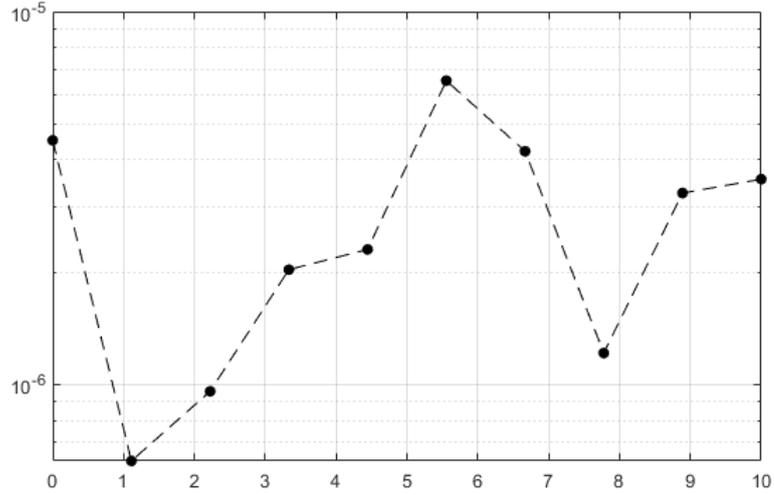
**Эксперимент 4.** В таблице 4 представлены сгенерированные значения фазовых переменных  $S, I$  и  $R$  соответствующие  $\beta = 5$  и  $\gamma = 4$  в моменты времени  $tm_i$ ,  $i = 1, \dots, n_m$ .

**Таблица 4:** Доля инфицированных в моменты времени  $tm$  в случае  $\beta = 5$ ,  $\gamma = 4$

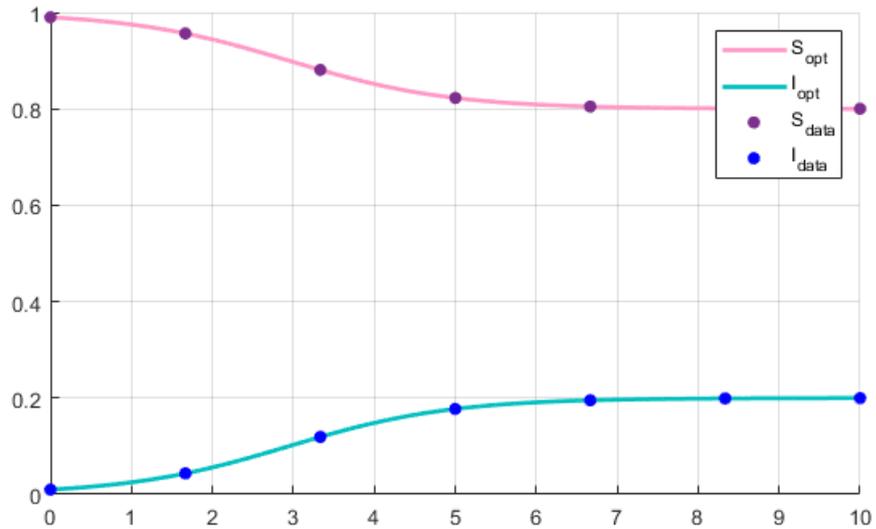
$t_m$	0	1.6667	3.3333	5.0000	6.6667	8.3333	10.0000
$I_{data}$	0.0100	0.0436	0.1192	0.1773	0.1953	0.1991	0.1998

В результате работы алгоритма за итераций была получена  $\beta = 5.0001$  и  $\gamma = 4.0001$  (рис. 14), ошибка идентификации составила  $Err = 1.0780 \cdot 10^{-5}$ . На рисунке 15 представлена разность между значениями  $I_{data}$  и  $I_{opt}$  в моменты времени  $tm_i$ ,  $i = 1, \dots, n_m$ .

**Случай других значений  $\beta$  и  $\gamma$  для системы (17).** При рассмотрении других значений  $B_2 = \{\beta \in Z \mid \beta \in [1, 10]\}$  и  $\Gamma = \{\gamma \in Z \mid \gamma \in [1, 10]\}$ , было выявлено, что ошибка идентификации при использовании представленного алгоритма не превышает  $10^{-1}$  (Рис. 16).



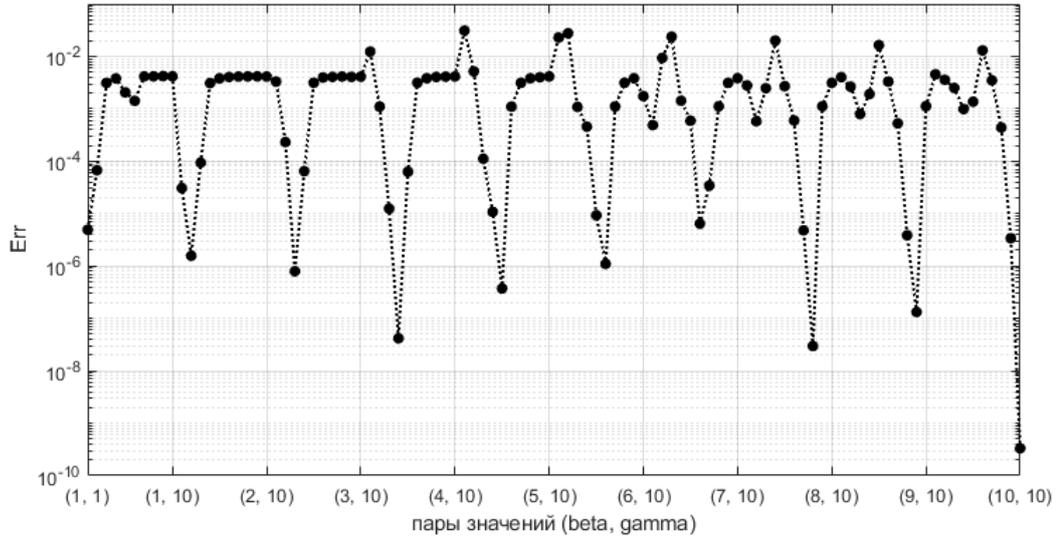
**Рис. 14:** Разность между значениями  $I_{data}$  и  $I_{opt}$  в моменты времени  $tm_i$ ,  $i = 1, \dots, n_m$  для случая  $\beta = 5$ ,  $\gamma = 4$



**Рис. 15:**  $S_{opt}$ ,  $I_{opt}$  — траектории, соответствующие  $\beta = 5.0001$  и  $\gamma = 4.0001$ , являющиеся результатом работы `fmincon`.  $S_{data}$ ,  $I_{data}$  — измеренные данные поведения системы при  $\beta = 5$  и  $\gamma = 4$  из таблицы 4.

## 2.5 Анализ результатов

В результате проделанной работы удалось решить задачу параметрической идентификации для нескольких типов систем дифференциальных уравнений, моделирующих эпидемиологический процесс. При этом максимальная ошибка идентификации составила  $Err \leq 10^{-2}$  в случае систе-



**Рис. 16:** График ошибки идентификации для данных, соответствующих  $B_2 \times \Gamma$

мы (16) и  $Err \leq 10^{-1}$  в случае системы (17).

Также мы выяснили, что алгоритм подходит для систем дифференциальных уравнений разных размерностей и может идентифицировать не только один из параметров системы, но и совокупности параметров, как представлено в эксперименте 4 и обобщении для системы (17).

## Заключение

Основной задачей представленной выпускной квалификационной работы было разработать программный комплекс, направленный на численную идентификацию параметров эпидемиологических моделей на основе данных наблюдений. В наше время вопрос идентификации систем дифференциальных уравнений, моделирующих распространение инфекционных заболеваний, является крайне актуальным. Решение этой задачи, представленное в данной ВКР, состояло из нескольких частей.

Во-первых, была сформулирована задача динамической оптимизации, а затем с помощью псевдоспектрального метода ортогональных коллокаций, в рамках которого численная аппроксимация решения дифференциально уравнения была получена как решение системы алгебраических уравнений, сведена к задаче нелинейного программирования.

Во-вторых, на основе первого пункта, была проведена постановка задачи параметрической идентификации эпидемиологических моделей на основе данных о заболеваемости.

Поставленная задача была переформулирована таким образом, чтобы ее решение можно было получить, используя функцию `fmincon` среды Matlab.

Были получены и проанализированы результаты для разных эпидемиологических моделей на основе тестовых данных.

## Список литературы

- [1] Л.С. Понтрягин, В.Г. Болтянский, Р.В. Гамкрелидзе, Е.Ф. Мищенко. Математическая теория оптимальных процессов. — Москва: Наука, 1969. 393 с.
- [2] Rao, A.V. A survey of numerical methods for optimal control. // Advances in the Astronautical Sciences. 2009. 135(1), P. 497–528.
- [3] Runge K. Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. // Zeitschrift für Mathematik und Physik. 1901. Bd. 46., P. 224–243.
- [4] Gromov D.V., Bulla I., Serea O.S., Romero-Severson E.O. Numerical optimal control for HIV prevention with dynamic budget allocation. // Mathematical Medicine and Biology. 2018. Vol. 35, No 4, P. 469–491.
- [5] Garg D., Patterson M., Hager W., Rao A., Benson D., Huntington G. An overview of three pseudospectral methods for the numerical solution of optimal control problems. // 2017, P. 2–5.
- [6] Иванова Т.В. Численные методы в оптике. Учебное пособие. — СПб: Университет ИТМО, 2017. 84 с.
- [7] Berrut J-P., Trefethen L.N. Barycentric Lagrange interpolation. // SIAM Review. 2004. Vol.46, No 3, P. 501–517.
- [8] Андерсон Р., Мэй Р. Инфекционные болезни человека. Динамика и контроль. — Москва: Научный мир, 2004. 784с.
- [9] Суетин П.К. Классические ортогональные многочлены. — Москва: ФИЗМАТЛИТ, 2005. 480с.
- [10] Kermack, W. O. and McKendrick, A. G. A Contribution to the Mathematical Theory of Epidemics. // Proc. Roy. Soc. Lond. 1927, P. 700–721.