

В. К. Толстых

Практическая оптимизация,  
идентификация  
распределённых систем

Москва  
«НАУКА»  
2025

Толстых В. К. **Практическая оптимизация, идентификация распределённых систем.** — М.: Наука, 2025. — 440 с.

Описывается теория и примеры решения задач оптимизации, идентификации распределённых систем (систем с распределёнными параметрами) с позиции непосредственного применения к целевому функционалу численных методов поиска его экстремумов (прямой экстремальный подход) на основе градиента. Рассматривается алгоритм получения аналитического выражения бесконечномерного градиента, когда искомые параметры-управления являются функциями. Вводится новое понятие управляемости, идентифицируемости.

Предлагаются методы оптимизации в высокоразмерных и бесконечномерных пространствах. Это метод коллинеарных градиентов и методы с регулируемым направлением спуска. Последние могут обеспечивать равномерную сходимость функций управления к оптимуму. Методы основаны на оригинальной форме необходимых условий оптимальности, сформулированных не в оптимуме, а в его окрестности.

В рамках прямого экстремального подхода реализуются разнообразные виды ограничений на управление и состояние распределённой системы. Рассмотрено большое количество сложных прикладных задач оптимизации как с выпуклыми функционалами, так и с невыпуклыми, многоэкстремальными.

Для научных работников, аспирантов и магистрантов, специализирующихся в области математического моделирования и оптимизации систем, описываемых уравнениями в частных производных. Книга сопровождается доступными пояснениями и иллюстрациями многих математических понятий, поэтому её можно считать и учебным пособием.

DOI  
ISBN

# Оглавление

<b>ВВЕДЕНИЕ .....</b>	<b>6</b>
<b>1 ОСОБЕННОСТИ ОПТИМИЗАЦИИ, ИДЕНТИФИКАЦИИ РАСПРЕДЕЛЁННЫХ СИСТЕМ .....</b>	<b>11</b>
1.1 Что такое распределённая система .....	11
1.2 Бесконечномерные задачи оптимизации .....	16
1.3 Введение в прямой экстремальный подход .....	30
1.3.1. Экстремальные алгоритмы нелинейной оптимизации .....	30
1.3.2. Выбор направления минимизации, градиент .....	36
1.3.3. Длина шага минимизации .....	47
1.3.4. Помехи .....	55
1.3.5. Завершение минимизации .....	57
1.3.6. Квадратичность и выпуклость .....	60
1.3.7. Ограниченность области управления .....	68
<b>2 КОНЕЧНОМЕРНЫЕ ПАРАМЕТРЫ-УПРАВЛЕНИЯ В РАСПРЕДЕЛЁННЫХ СИСТЕМАХ .....</b>	<b>72</b>
2.1 Постановка прямой экстремальной задачи .....	72
2.2 Градиентные методы .....	73
2.3 Метод сопряжённых градиентов .....	76
2.4 Метод Ньютона и квазиньютоновские методы .....	78
2.5 Покомпонентные условия оптимальности .....	84
2.6 Метод коллинеарных градиентов .....	88
2.7 Метод с регулируемым направлением спуска .....	94
2.8 Высокоразмерные задачи. Тестирование .....	100
<b>3 БЕСКОНЕЧНОМЕРНЫЕ ПАРАМЕТРЫ-УПРАВЛЕНИЯ .....</b>	<b>106</b>
3.1 Условия оптимальности .....	106
3.2 Методы с регулируемым направлением спуска (МРНС) .....	111
3.2.1. МРНС <sub>г</sub> , регулирование относительно градиента .....	112
3.2.2. МРНС <sub>сг</sub> , регулирование относительно сопряжённых градиентов .....	117
3.3 Регуляризация .....	119
3.4 Градиент неявно заданного функционала .....	125
3.4.1. Выделение ограничений на управление .....	129
3.4.2. Линеаризация задачи .....	129
3.4.3. Отображение линеаризованных уравнений в сопряжённые пространства .....	132
3.4.4. Объединение задачи .....	143
3.4.5. Учёт краевых условий .....	144
3.4.6. Выделение градиента .....	150
3.5 Управляемость .....	153
3.5.1. Тип уравнений в частных производных .....	154
3.5.2. Тип уравнений сопряжённой задачи .....	161
3.5.3. Условия управляемости .....	164
3.6 Управление с ограничениями .....	174
3.6.1. Штрафные функции (тип 1.1) .....	177
3.6.2. Дополнительные функционалы (тип 1.2) .....	179
3.6.3. Проецирование на допустимое множество (тип 2.1) .....	180
3.6.4. Изопериметрическое условие (тип 2.2) .....	181
3.6.5. Кусочно-постоянное управление (тип 2.2) .....	184
3.6.6. Ограничение на скорость изменения управления (тип 2.2) .....	185
<b>4 ОСОБЕННОСТИ ЗАДАЧ ИДЕНТИФИКАЦИИ .....</b>	<b>188</b>
4.1 Общая постановка задачи .....	188
4.2 Критерий качества идентификации .....	193
4.3 Идентифицируемость .....	198

<b>5</b>	<b>ПРОСТЕЙШАЯ ЗАДАЧА ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ТЕПЛОВЫМИ ПРОЦЕССАМИ .....</b>	<b>202</b>
5.1	ПОСТАНОВКА ЗАДАЧИ .....	203
5.2	ГРАДИЕНТ ЦЕЛЕВОГО ФУНКЦИОНАЛА.....	207
5.2.1.	<i>Выделение ограничений на управление .....</i>	<i>207</i>
5.2.2.	<i>Линеаризация задачи .....</i>	<i>208</i>
5.2.3.	<i>Отображение линеаризованных уравнений в сопряжённые пространства .....</i>	<i>209</i>
5.2.4.	<i>Объединение задачи.....</i>	<i>209</i>
5.2.5.	<i>Учёт краевых условий.....</i>	<i>212</i>
5.2.6.	<i>Выделение градиента.....</i>	<i>213</i>
5.3	УПРАВЛЯЕМОСТЬ.....	214
5.4	ЧИСЛЕННОЕ РЕШЕНИЕ .....	219
5.4.1.	<i>Конечно-разностные аппроксимации .....</i>	<i>219</i>
5.4.2.	<i>Оценка выпуклости целевого функционала .....</i>	<i>223</i>
5.4.3.	<i>Оптимизация без ограничений.....</i>	<i>225</i>
5.4.4.	<i>Оптимизация в заданном диапазоне управлений.....</i>	<i>234</i>
5.4.5.	<i>Оптимизация с ограничением на состояние.....</i>	<i>236</i>
5.4.6.	<i>Оптимизация с изопериметрическим ограничением.....</i>	<i>236</i>
5.4.7.	<i>Оптимизация с кусочно-постоянным управлением .....</i>	<i>241</i>
5.4.8.	<i>Оптимизация с ограничением на скорость изменения управления .....</i>	<i>243</i>
5.4.9.	<i>Оптимизация с помехами .....</i>	<i>245</i>
<b>6</b>	<b>ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ ПРИ ФОРМИРОВАНИИ ОТЛИВОК.....</b>	<b>253</b>
6.1	ПОСТАНОВКА ЗАДАЧИ .....	253
6.2	ГРАДИЕНТ ЦЕЛЕВОГО ФУНКЦИОНАЛА.....	258
6.3	ИДЕНТИФИЦИРУЕМОСТЬ .....	264
6.4	ЧИСЛЕННОЕ РЕШЕНИЕ .....	271
6.4.1.	<i>Конечно-разностные аппроксимации .....</i>	<i>272</i>
6.4.2.	<i>Оценка выпуклости целевого функционала .....</i>	<i>275</i>
6.4.3.	<i>Решение тестовой задачи.....</i>	<i>277</i>
6.4.4.	<i>Решение прикладной задачи .....</i>	<i>281</i>
<b>7</b>	<b>ОПТИМАЛЬНОЕ УПРАВЛЕНИЕ ОХЛАЖДЕНИЕМ НЕПРЕРЫВНОГО СЛИТКА .....</b>	<b>284</b>
7.1	ПОСТАНОВКА ЗАДАЧИ .....	284
7.2	ГРАДИЕНТ ЦЕЛЕВОГО ФУНКЦИОНАЛА.....	293
7.3	УПРАВЛЯЕМОСТЬ.....	303
7.4	ЧИСЛЕННОЕ РЕШЕНИЕ .....	307
7.4.1.	<i>Постановка тестовой задачи синтеза термонапряжений .....</i>	<i>308</i>
7.4.2.	<i>Конечно-разностные аппроксимации .....</i>	<i>309</i>
7.4.3.	<i>Оценка выпуклости целевых функционалов.....</i>	<i>313</i>
7.4.4.	<i>Решение тестовой задачи синтеза термонапряжений .....</i>	<i>315</i>
7.4.5.	<i>Минимизация термонапряжений по функционалу J1.....</i>	<i>319</i>
7.4.6.	<i>Минимизация термонапряжений по функционалу J2.....</i>	<i>327</i>
7.4.7.	<i>Минимизация термонапряжений по функционалу J3.....</i>	<i>329</i>
<b>8</b>	<b>ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ ОТКРЫТЫХ РУСЕЛ .....</b>	<b>332</b>
8.1	ПОСТАНОВКА ЗАДАЧИ .....	332
8.1.1.	<i>Характеристическая форма исходных уравнений .....</i>	<i>337</i>
8.2	ГРАДИЕНТ ЦЕЛЕВОГО ФУНКЦИОНАЛА.....	340
8.2.1.	<i>Характеристическая форма сопряжённых уравнений.....</i>	<i>347</i>
8.3	ИДЕНТИФИЦИРУЕМОСТЬ .....	349
8.4	ЧИСЛЕННОЕ РЕШЕНИЕ .....	357
8.4.1.	<i>Конечно-разностные аппроксимации .....</i>	<i>357</i>
8.4.2.	<i>Оценка выпуклости целевого функционала тестовой задачи .....</i>	<i>362</i>
8.4.3.	<i>Решение тестовой задачи идентификации.....</i>	<i>365</i>
8.4.4.	<i>Идентификация шероховатости Каракумского канала .....</i>	<i>374</i>

<b>9</b>	<b>ОПТИМИЗАЦИЯ ФОРМЫ СОПЛА ГИДРОПУШКИ .....</b>	<b>380</b>
9.1	ПОСТАНОВКА ЗАДАЧИ .....	380
9.1.1.	<i>Характеристическая форма исходных уравнений.....</i>	<i>386</i>
9.2	ГРАДИЕНТ ЦЕЛЕВОГО ФУНКЦИОНАЛА .....	388
9.2.1.	<i>Характеристическая форма сопряжённых уравнений .....</i>	<i>406</i>
9.3	УПРАВЛЯЕМОСТЬ.....	398
9.4	АППРОКСИМАЦИЯ ЗАДАЧ ДЛЯ ЧИСЛЕННОГО РЕШЕНИЯ.....	406
9.4.1.	<i>Сетка аппроксимаций .....</i>	<i>408</i>
9.4.2.	<i>Аппроксимация исходной задачи .....</i>	<i>412</i>
9.4.3.	<i>Аппроксимация сопряжённой задачи.....</i>	<i>415</i>
9.4.4.	<i>Аппроксимация градиента .....</i>	<i>417</i>
9.5	ЧИСЛЕННОЕ РЕШЕНИЕ .....	417
9.5.1.	<i>Оценка выпуклости целевого функционала .....</i>	<i>418</i>
9.5.2.	<i>Первый локальный максимум .....</i>	<i>421</i>
9.5.3.	<i>Второй локальный максимум.....</i>	<i>424</i>
	<b>ЛИТЕРАТУРА .....</b>	<b>428</b>
	<b>ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ .....</b>	<b>432</b>
	<b>ОБОЗНАЧЕНИЯ РАЗДЕЛОВ 1–4.....</b>	<b>432</b>
	<i>Основные .....</i>	<i>437</i>
	<i>Индексы .....</i>	<i>440</i>
	<i>Пространства и множества .....</i>	<i>440</i>
	<i>Операции с множествами .....</i>	<i>442</i>
	<i>Операторы .....</i>	<i>443</i>
	<i>Греческие символы .....</i>	<i>445</i>
	<i>Специальные функции.....</i>	<i>447</i>
	<b>СОКРАЩЕНИЯ .....</b>	<b>448</b>

## Введение

В современном мире эра экстенсивного развития человечества подходит к концу. Земные, людские, финансовые и другие ресурсы не бесконечны. Наступила эра интенсивного развития, эра повсеместной экономии природных ресурсов, энергии, рабочего времени, минимизации экологического ущерба, максимизации выгоды и многого другого. В основе интенсификации развития общества лежит *оптимизация* объектов и процессов, с которыми сталкивается человечество.

Для оптимизации, оптимального управления состоянием объектов необходимо иметь их математические модели, поскольку оптимальные состояния сначала ищут на моделях, а потом переносят на реальные объекты. Для достаточно качественной оптимизации работы объектов необходимо иметь не только хорошие методы оптимизации, но и хорошие математические модели этих объектов.

Большое количество объектов и процессов описываются с желаемой точностью только моделями пространственно-распределённых систем в виде дифференциальных уравнений с частными производными. Для таких моделей особенно актуальна *идентификация*, поскольку их параметры (коэффициенты уравнений) не всегда могут измеряться непосредственно, и даже известные справочные данные не всегда могут использоваться без надлежащей адаптации. Решение задачи идентификации должно давать оптимальную математическую модель.

Оптимизация пространственно-распределённых систем — это сложная задача. Здесь искомое решение (управление), как правило, представляет собой функцию, или несколько функций, зависящих от пространственных переменных и, возможно, от времени.

Математические пространства, которым принадлежат такие управления, называются бесконечномерными.

В данной книге рассматривается и развивается прямой экстремальный подход [92] для практического численного решения оптимизационных задач с распределёнными системами. Он заключается в том, что разнообразные категории задач, которые связаны с поиском оптимальных решений (оптимальное управление, оптимальные пространственные формы, оптимальная идентификация параметров уравнений), формулируются и решаются *однообразно*, как задача на поиск экстремума заданной цели оптимизации на основе её градиента.

Заметим, что в бесконечномерных пространствах цель оптимизации (критерий качества оптимизации) — это не функция, а функционал, неявно зависящий от управления, которое содержится в дифференциальных уравнениях распределённой системы. В свою очередь градиент, как первая производная от цели оптимизации, является не вектором, а функцией. Вычисление вторых производных, которые возможно могли бы улучшить сходимость к оптимуму, в таких сложных задачах до практической реализации не доходит.

В прямом экстремальном подходе [92] для обобщения градиентных методов на бесконечномерные пространства вводится специальный параметр регулирования направления спуска (методы с регулируемым направлением спуска), которые позволяют добиваться быстрой *равномерной сходимости* к оптимуму. В приведённых в данной книге примерах решения прикладных задач демонстрируется не только работоспособность, но и высокая эффективность такого подхода.

Поскольку в некоторых случаях в распределённых системах искомые управления всё же могут быть векторами, то во втором разделе рассматривается конечномерная ситуация. Именно для таких задач оптимизации, где допускается геометрическая интерпретация,

предлагается *метод коллинеарных градиентов* и вытекающий из него *метод с регулируемым направлением спуска*. Эти методы получены на основе новой формы необходимых условий оптимальности. Условия формулируются не в точке оптимума, а в его окрестности, что имеет практическую ценность для построения итерационных приближений к оптимуму. Предлагаемые необходимые условия и сопутствующие методы обеспечивают формально градиентному алгоритму (метод первого порядка) сходимость как у метода Ньютона (метод второго порядка).

В тестовых расчётах минимизации высокоразмерных квадратичных функций предложенные методы продемонстрировали практическую нечувствительность к размерности вектора управления. Данное обстоятельство позволяет обобщить полученные алгоритмы на бесконечные задачи оптимизации, что и было сделано в третьем разделе данной книги.

При оптимизации любых систем, а особенно распределённых, очень важно знать условия *управляемости*, определяющие принципиальную возможность управления системой. Вместо традиционного понятия управляемости в виде «возможности перевода сосредоточенной системы в заданное финальное состояние» предлагается понятие «возможности управления распределённой системой на основе заданной *цели управления*».

Новое понятие управляемости описано в третьем разделе. Оно основано на теории обратных некорректных задач и связывает управляемость системы с произвольным видом целевого функционала (а не только с финальным состоянием) при весьма разнообразных пространственно-распределённых областях управления и функционала. Такое понятие управляемости, по своей сути, является условием,

требованием корректной постановки задачи оптимизации. Выполнение условий управляемости позволяет из производной целевого функционала (производной Фреше в первой вариации целевого функционала) получить именно градиент, что делает возможным применение экстремальных алгоритмов направленного поиска оптимума.

В четвёртом разделе акцентируются внимание на особенностях задач параметрической идентификации распределённых систем. Показано, что такие задачи полностью укладываются в описанную ранее схему оптимизации параметров математической модели прямым экстремальным подходом.

Начиная с пятого раздела и до конца книги обсуждаются прикладные задачи оптимизации, оптимального управления, идентификации для пространственно-распределённых систем с точки зрения разработанной теории прямого экстремального подхода. Рассмотрено больше количество актуальных, сложных прикладных задач оптимизации с разнообразными ограничениями и помехами, как с выпуклыми, так и, что особенно интересно, с невыпуклыми и даже с многоэкстремальными функционалами. Ряд этих задач ранее не имели ни теоретического, ни практического решения.

Все прикладные задачи программировались и реализовывались автором в среде Delphi 7.

Книга рассчитана на читателя, владеющего основами математического анализа и численных методов. Полезно также иметь общее представление о вариационном исчислении и функциональном анализе, а именно, что такое вариация, основные свойства линейных функциональных пространств и операторов в них. Последнее не является строго обязательным и по мере появления в излагаемом

материале подобных понятий, автор старается в доходчивой форме дать их пояснение.

Поскольку книга сопровождается, как нам кажется, простыми разъяснениями многих математических понятий и снабжена многочисленными конкретными примерами и иллюстрациями, то её можно считать и учебным пособием для студентов старших курсов и аспирантов. Для того, чтобы было удобнее работать с пособием, основные понятия излагаемого материала собраны в конце книги в обширный предметный указатель. Также в конце приведены принятые обозначения и сокращения.

# 1 Особенности оптимизации, идентификации распределённых систем

## 1.1 Что такое распределённая система

Задачи оптимизации, управления, исследования *объектов*, прогнозирование их состояния, а также многое другое требует знания достаточно точных математических *моделей* этих объектов. На моделях отыскивают оптимальные решения, которые впоследствии применяют к исследуемым, управляемым объектам. Математические модели могут представлять собой весьма разнообразные системы, что принципиально отражается на методах работы с ними, в том числе и на методах оптимизации, идентификации.

Рассмотрим объект, состоящий из множества элементов. Состояние такого объекта (скорость, температура, плотность, информация, финансы...) будет характеризоваться состоянием его элементов. Если между элементами существует понятие расстояния и состояние каждого элемента зависит от расстояния до других элементов, то такой объект или систему называют *распределённой*.

Расстояние необходимо понимать в смысле метрики<sup>1</sup> пространства, принятого в системе. Это не всегда может быть физическое пространство, расстояние в котором измеряется метрами. Например, в компьютерных сетях пространство — это сетевые узлы. Метрика здесь вводится как количество узлов, через которые проходит заданная информация.

Если состояние каждого элемента не зависит от расстояния до других элементов, то такой объект идеализируют и представляют

---

<sup>1</sup> Метрическим пространством называется непустое множество, в котором между любой парой элементов, обладающих определёнными свойствами, определено расстояние, называемое метрикой.

*сосредоточенной* системой, находящейся в некоторой точке пространства. В противном случае пространственной протяжённостью объекта пренебрегать нельзя и его рассматривают как распределённую систему.

Распределённый объект также можно считать сосредоточенным, если не ставится задача анализа его неоднородного внутреннего состояния. Например, когда в строительстве точно известно, что пространственно-распределённый объект не деформируется, то нет смысла его моделировать распределённой системой. Такой объект можно считать сосредоточенной системой и работать, например, с его центром масс.

Таким образом, распределённость или сосредоточенность математической модели объекта зависит не только от самого объекта, но и от задач, которые ставит перед собой исследователь, работая с этим объектом. Мы будем исследовать задачи, связанные с объектами, состояние которых должно описываться математическими моделями в виде распределённых систем. Более того, будем полагать, что количество элементов в таких системах можно считать бесконечно большим, а расстояние между ними — бесконечно малым. Состояние таких системы становится бесконечномерным. Поэтому такие системы также известны как *бесконечномерные системы*.

Бесконечномерные распределённые системы описываются *дифференциальными уравнениями с независимыми пространственными переменными  $x, y, z$  (или другими переменными специальной метрики)* и, возможно, временем  $t$ . Как правило, это *уравнения в частных производных*, динамические или стационарные. Коэффициенты этих уравнений, свободные члены или краевые условия могут содержать переменный параметр или вектор переменных параметров, изменяя значение которого, мы можем регулировать состояние, т.е. решение,

системы дифференциальных уравнений. Такой параметр, обычно, называют *управлением*. Он также может быть распределённым. Будем обозначать его буквой  $u$ . Состояние системы будем обозначать буквой  $v$ .

Распределённые системы в виде уравнений в частных производных называют также *системами с распределёнными параметрами*. Этот термин возник в проводной телеграфии для характеристики линий передач, длины которых сравнимы с длиной передаваемых электромагнитных волн [75]. При этом, под термином «параметр» подразумевается не искомый параметр  $u$ , а состояние системы  $v$ . Это вносит неоднозначность в терминологию задач поиска оптимальных параметров-управлений  $u$ . Поэтому, далее мы будем предпочтительно использовать термин «распределённые системы» для моделей в виде дифференциальных уравнений для поиска оптимальных параметров  $u$  в уравнениях модели.

Существующее большое разнообразие задач оптимизации с распределёнными системами можно разделить на три категории (смотрите, например, [1]):

- оптимальное управление (оптимизация во времени);
- оптимальный дизайн (оптимизация геометрической формы);
- идентификация (оптимальная оценка параметров модели).

Необходимо отметить, что идентификация — это очень важный и почти всегда обязательный этап моделирования распределённых систем, который обеспечивает задачи оптимального управления и оптимального дизайна адекватными математическими моделями.

Одной из причин возникновения разных категорий в задачах оптимизации послужило отсутствие универсального метода, подхода к решению таких задач ещё со времён их зарождения. В разных категориях

существуют и развиваются разные методы оптимизации. Если искомая переменная  $u$  зависит от времени  $t$ , то говорят об оптимальном управлении и используют необходимые условия оптимальности в виде принципа максимума Понтрягина [28, 43, 74, 77], принципа Беллмана в динамическом программировании [28, 24, 74, 42], проблемы моментов [28, 42] и другие. Если  $u$  зависит от пространственных переменных, то обычно говорят об оптимальном дизайне, например, оптимальная аэродинамическая форма крыла летательного аппарата, сопла, и используют необходимые условия оптимальности и методы, основанные на теории вариационного исчисления [56, 51] или на функциональном анализе с вариационными неравенствами [52]. Если неизвестное значение параметра  $u$  оценивается на основе минимизации невязки модельных и экспериментальных данных, то говорят об идентификации параметров модели.

В большинстве своём исторически сложившиеся подходы к решению задач оптимизации, т.е. к нахождению параметра  $u$ , являются непрямые, поскольку они сначала требуют получения необходимых условий оптимальности в той или иной форме, а только потом, из полученных условий ищут  $u$ . Как выразился академик Н. Н. Моисеев «Задача оптимального управления становится практически неразрешимой, если необходимые условия оптимальности формулируются относительно сложно» [59].

Добавим, большинство исследований в области оптимального управления распределёнными системами ограничиваются всего лишь получением необходимых условий оптимальности и не имеют практических решений. Добиться решения на основе условий оптимальности зачастую оказывается невозможным, поскольку получаемые условия не дают информации о направлении возможной коррекции приближительных значений управления до оптимального

значения. А итерационные направленные коррекции — это единственный способ достижения решений нелинейных задач, в том числе и задач оптимального управления.

Для распределённых систем параметр  $u$  может быть функцией и зависеть как от  $t$ , так и от  $x$  (а также от  $y, z$ ). Критерии оптимальности задач оптимизации — *целевые функционалы*<sup>1</sup>  $J(u)$  могут быть весьма разнообразными, включая наличие экспериментальных данных. Всё это стирает различия между категориями оптимизации. На самом деле все математические постановки таких задач в принципе одинаковы — это поиск оптимальной функции  $u_*$ , доставляющей *глобальный*<sup>2</sup> минимум (возможно максимум) нелинейному целевому функционалу  $J(u)$ . Одинаковая формулировка таких задач обуславливает и возможность применения одинаковых методов для их решения. Следуя этому утверждению, искомую неизвестную переменную  $u$  в разных задачах оптимизации мы будем называть одинаково — управление.

Представляется разумным использовать прямой экстремальный подход [92, 29, 72] для всех категорий задач оптимизации. Но на этом пути возникают свои проблемы. Пространственная распределённость систем может приводить к бесконечномерным задачам оптимизации, где классические экстремальные методы, как правило, не дают желаемых результатов. Термин *экстремальные методы* означает многообразие методов алгоритмического программирования, предназначенных для поиска *экстремального*<sup>3</sup> значения функции, функционала.

---

<sup>1</sup> Обычно, целевые функционалы представляют собой определённые интегралы с неявной зависимостью от функции  $u$ .

<sup>2</sup> т.е. наименьшее значение  $J$  для всевозможных значений  $u$ .

<sup>3</sup> Экстремум функции (функционала) — это максимальное или минимальное значение на заданном множестве своих аргументов.

## 1.2 Бесконечномерные задачи оптимизации

Если управление  $u$  распределённой системы присутствует в виде конечного набора чисел, а не функций, то мы имеем конечномерную задачу поиска оптимального управления  $u_* = \arg \min J(u)$ , где  $J(u)$  — *целевая функция*, зависящая от *вектора*  $u$ . Такие экстремальные задачи относительно хорошо исследованы. Мы же уделим основное внимание решению бесконечномерных экстремальных задач, когда управление  $u$  — это функция. При этом  $J(u)$  уже будет не функцией, а *функционалом*<sup>1</sup>.

Рассмотрим пространственно-распределённую систему в виде некоторой математической *модели*, которую символично обозначим буквой  $\mathbb{D}$ . *Состояние системы* в дифференциальных уравнениях нашей модели  $\mathbb{D}$  будем описывать функцией  $v(\tau)$ , где *пространственно-временная координата*  $\tau = \{x_1 \dots x_N, t\}$  — это множество из  $N$  пространственных координат и времени. Множество значений  $\tau$ , где заданы дифференциальные уравнения модели  $\mathbb{D}$ , — это *открытая*<sup>2</sup>, *ограниченная*<sup>3</sup> область  $\Omega$  с границей  $\partial\Omega$ .

Понятно, что состояние  $v$  определено и на *замкнутом множестве*  $\bar{\Omega} = \Omega \cup \partial\Omega$ , поскольку  $v$  — это решение дифференциальных уравнений распределённой системы, что всегда требует наличия *краевых условий*<sup>4</sup> где-либо на  $\partial\Omega$ . Таким образом, область *определения состояния*

<sup>1</sup> Уточним различие между функцией и функционалом. Если некоторому числу  $u \in U$  ставится в соответствие по определённому правилу число  $J$ , то говорят, что задана функция  $J(u)$ . Если же функции  $u(x) \in U(X)$  ставится в соответствие число  $J$ , то говорят, что задан функционал  $J(u)$ . Множество  $U$  называют *областью определения*  $J$ , а совокупность всех значений  $J$  — *областью значений*.

<sup>2</sup> Открытое множество (область) — это множество, каждая точка которого входит в него вместе с некоторой окрестностью. В нашем случае — это ограниченное множество, не включающее в себя свою границу, поскольку окрестность граничных точек не полностью входит в множество.

<sup>3</sup> Множество элементов из  $E^n$  ограниченное, если существует число  $M$ , такое, что величина (норма) всех элементов меньше  $M$ , т.е. все элементы находятся в шаре радиусом  $M$ .

<sup>4</sup> Под *краевыми условиями* мы будем подразумевать условия на границе  $\partial\Omega$ , необходимые для решения уравнений в частных производных распределённой системы, т.е. это могут быть и *граничные*, и *начальные* условия.

*системы* — это замкнутое пространственно-временное множество  $\bar{\Omega}$ , которое называют *замыканием*.

Искомые параметры (управления)  $u = u(\tau)$ ,  $\tau \in S \subset \bar{\Omega}$  могут присутствовать как в уравнениях модели  $\mathbb{D}$  распределённой системы, так и в её краевых, т.е. граничных и начальных условиях. Задача заключается в поиске функции  $u(\tau)$  такой, чтобы на некотором известном множестве  $\omega \subset \bar{\Omega}$  обеспечивалось оптимальное состояние системы в смысле заданного *критерия качества оптимизации*, записываемого в виде нелинейного *целевого функционала*, требующего *минимизации* (или *максимизации*):

$$J = \int_{\omega} I(v) d\omega \rightarrow \min,$$

где  $I(v)$  — *функция*<sup>1</sup> *цели* оптимизации, оптимального управления, идентификации. Сложная функция  $I(v(\tau))$  *определена* на множестве  $\omega$ , а её значение зависит от внутренней функции  $v$ .

Заметим, что  $J$  в такой записи всегда является функционалом состояния  $v$ . В то же время, если управление  $u$  — вектор конечной размерности, то зависимость  $J$  от управления не будет функционалом, а будет всего лишь функцией.

В задачах идентификации функция цели обычно задаётся в виде  $I(v) = (v - v_e)^2$ , где  $v_e$  — экспериментально наблюдаемое на реальном объекте состояние системы. В задачах оптимального управления функция цели может принимать самые разнообразные формы, описывающие потребление природных ресурсов, расходы финансов, качество технических изделий, уровень социальных процессов и тому подобное. Кроме того, функция цели может явно зависеть не только от

---

<sup>1</sup>  $I(v)$  — это сложно заданная от  $\tau$  функция  $I(v(\tau))$ , когда каждому значению числа  $\tau \in \omega$  через распределённую систему ставится в соответствие число  $I(v(\tau))$ .

состояния системы  $v$ , но и от управления  $u$ , т.е. иногда возможно  $I = I(v, u)$ .

Уточним задачу. Критерий качества  $J$  всегда задаётся явно от состояния управляемой системы  $v$  и *неявно* от её управления  $u$ , т.е.  $J = J(v(u))$ , даже если  $u$  каким-либо образом присутствует в функции цели, т.е. даже если  $I(v, u)$ . На самом деле целевой функционал задачи — это всегда неявный функционал  $J(u)$ , где  $u$  связано с  $v$  дифференциальными уравнениями  $\mathbb{D}$  распределённой системы. Такой характер зависимости существенно влияет на стратегию решения. Поэтому задачу *оптимизации* надо формулировать в виде:

$$J(u) = \int_{\omega} I(v, u) d\omega \rightarrow \min, \quad (1.1)$$

при условии  $\mathbb{D}(\tau, v, u)v(\tau) = 0, \quad \tau \in \bar{\Omega}$ ,

где  $\mathbb{D}$  — *дифференциальный оператор*<sup>1</sup> математической модели распределённой системы, действующий на функцию  $v(\tau)$  и зависящий от  $\tau, v, u$ . Пока полагаем, что управление системы ничем не ограничено.

Приведённый здесь оператор  $\mathbb{D}$  включает в себя не только конкретный вид *дифференциальных уравнений* распределённой системы на множестве  $\Omega$ , но и краевые условия где-либо на границе  $\partial\Omega$ . По сути, в (1.1) выражение  $\mathbb{D}v = 0$  — это *исходная задача*, решением которой (состоянием  $v$ ) мы хотим управлять посредством параметра  $u$  в операторе  $\mathbb{D}$ .

В *прямом экстремальном подходе* оптимальное значение управления  $u$  нелинейной оптимизационной задачи (1.1) ищется пошагово (итерационно) в ходе непосредственной минимизации целевого функционала  $J(u)$  экстремальными методами. Никакие

---

<sup>1</sup> Напомним понятие оператора. Если некоторой функции  $y(x)$  ставится в соответствие по некоторому правилу  $\mathbb{D}$  новая функция  $z(x)$ , то говорят, что задан оператор  $\mathbb{D}$ , значение которого  $\mathbb{D}y = z$ .

промежуточные условия, например, необходимые условия оптимальности в той или иной форме, здесь не используются.

Таким образом, задача (1.1) формулируется как *экстремальная* в виде задачи *нелинейной оптимизации* или, как ещё говорят, *нелинейного программирования*<sup>1</sup>, в бесконечномерном пространстве:

$$u_* = \arg \min J(u), \quad (1.2)$$

где  $u_*$  — это функция *оптимального управления*, или *оптимум*, или *оптималь*, или *оптимальный параметр*, доставляющий *глобальный* минимум функционалу  $J(u)$ . Все эти названия аргумента  $u_*$  в прямом экстремальном подходе являются синонимами. Сами уравнения  $\mathbb{D}v = 0$  распределённой системы в задаче (1.2) используются для определения направлений пошаговой минимизации  $J(u)$ .

В подавляющем большинстве экстремальных методов направления минимизации формируются на основе градиента  $\nabla J(u)$ . Это естественно, ведь градиент указывает направление и величину максимального роста функции, функционала. Поэтому разумно двигаться к минимуму в направлении *антиградиента*  $-\nabla J(u)$ .

Заметим, что задача *минимизации*  $\arg \min J(u)$  равносильна задаче *максимизации*  $\arg(-\max J(u))$ . Оптимальное управление в обоих случаях будет одно и то же. Поэтому далее мы будем говорить о минимизации  $J(u)$ .

В чём особенности и сложности решения бесконечномерных экстремальных задач и почему их называют бесконечномерными? Давайте начнём с конечномерных экстремальных задач и разберёмся как происходит переход от конечномерного управления к бесконечномерному.

---

<sup>1</sup> Хотелось бы сделать замечание для студентов-программистов. Программирование — это написание алгоритма вычислений для компьютера, а не перевод чьего-либо алгоритма в машинные коды.

Если ставится задача нахождения конечного множества числовых коэффициентов  $\{u_i\}_{i=1}^n$ , то их представляют в виде *вектор-столбца* с компонентами  $(u_1, \dots, u_n)$  в  $n$ -мерном *евклидовом* пространстве  $E^n$  (если  $n = 1$ , то будем писать  $E$ ), т.е. говорят об управлении  $u \in E^n$ . В этих условиях экстремальная задача (1.2) является *конечномерной задачей оптимизации* для целевой функции  $J(u)$  с областью определения  $U \subseteq E^n$ . Множество управлений  $U$  может быть *ограниченным*, т.е. может не совпадать со всем пространством  $E^n$ . Тогда говорят о *множестве допустимых управлений*  $U \subset E^n$ . Конечномерные задачи минимизации функций (выпуклых, дифференцируемых функций), обычно, не создают серьёзных проблем при их решении.

Проблемы появляются тогда, когда управление является функцией пространства и возможно времени  $u(\tau)$ ,  $\tau \in S$ , где  $S \subset \bar{\Omega}$  — это *область определения управления*. В таких условиях экстремальная задача (1.2) становится задачей *бесконечномерной оптимизации* для целевого функционала  $J(u)$  с областью определения функционала  $U(S) \subseteq L_2(S)$ , где  $L_2$  пространство функций, определённых на множестве  $S$  и интегрируемых с квадратом (свойства  $L_2$  будут уточнены немного позже). Если  $U = L_2$ , т.е. совпадает со всем пространством управлений, тогда говорят о задаче управления без ограничений, или иначе, о *задаче безусловной оптимизации*. Наличие ограничений, когда имеется *допустимое множество*  $U(S) \subset L_2(S)$ , характерно именно для задач управления, в то время как в задачах идентификации ограничения отсутствуют.

На рисунке 1.1 схематично показано как конечномерный вектор управления из пространства  $E^n$  при  $n \rightarrow \infty$  превращается в бесконечномерное управление, т.е. в функцию из  $L_2$ .

В обычной декартовой *системе координат* мы можем представить не более чем трехмерный вектор — рисунок 1.1а. Для иллюстрации значений вектора в многомерных пространствах систему координат удобно изображать в виде рисунка 1.1б. Тогда, если  $n > 3$ , то значение вектора можно представить рисунком 1.1в в виде множества из  $n$  точек на осях  $i = 1 \dots n$ . При  $n \rightarrow \infty$  все точки сливаются в одну линию, как показано на рисунке 1.1г (случай одномерной функции). При этом каждый индекс оси  $i \in \{1 \dots n\}$  превращается в точку  $\tau \in S$ , а каждая компонента вектора  $u_i$  — в локальное значение функции  $u(\tau)$  в точке  $\tau$ .

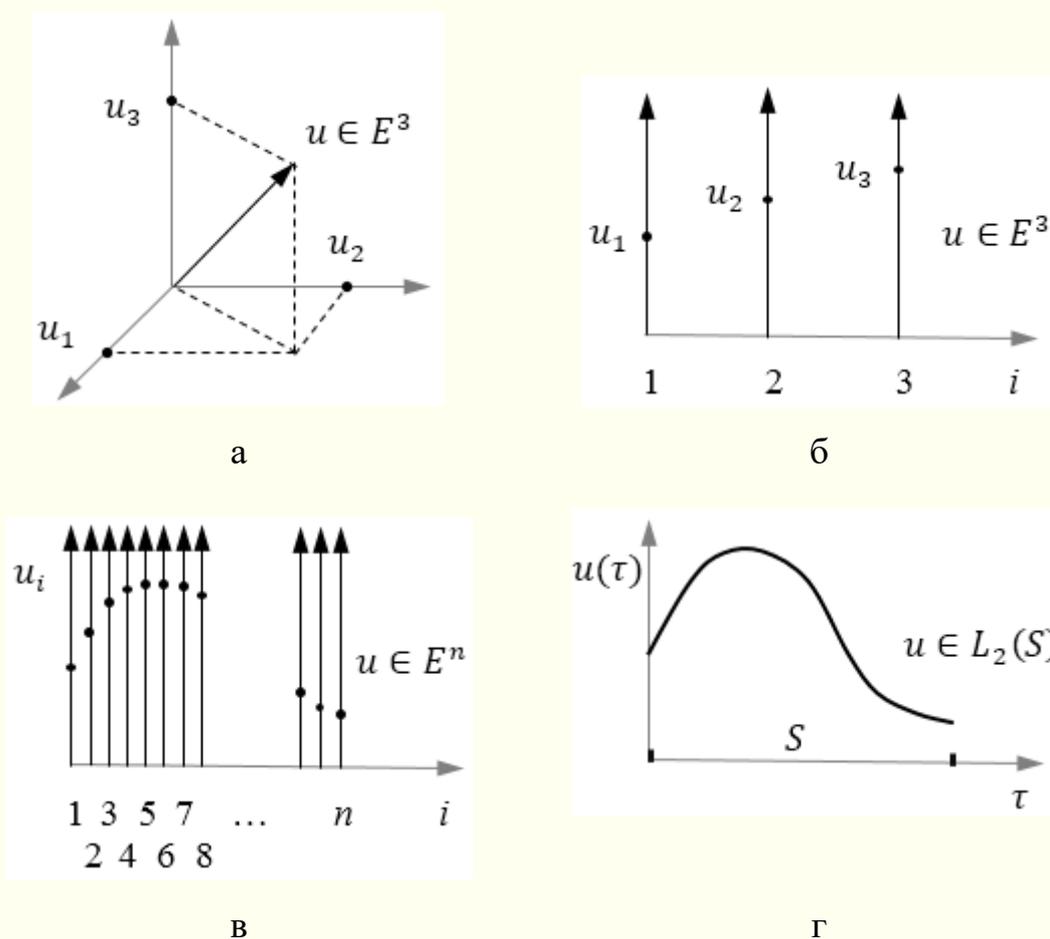


Рисунок 1.1 — Переход из конечномерного пространства в бесконечномерное

Линейные пространства<sup>1</sup>  $E^n$  и  $L_2$  являются евклидовыми<sup>2</sup>, поскольку в них определено скалярное произведение, которое обозначается угловыми скобками, возможно с индексом, уточняющим принадлежность пространству:

$$\langle a, b \rangle_{E^n} = \sum_{i=1}^n a_i b_i, \quad \langle a, b \rangle_{L_2(S)} = \int_S ab \, dS.$$

Скалярное произведение в виде характерных скобок  $\langle \dots, \dots \rangle$  подразумевает, что первый вектор — это *вектор-строка*, а второй — *вектор-столбец*. Если же перемножать вектора без указанных скобок, то необходимо явно выделять вектор-строку, что мы будем делать посредством верхнего индекса  $T$  — *транспонирование*. Например,

$$a^T b \stackrel{\text{def}}{=} \langle a, b \rangle.$$

Указанные пространства также являются банаховыми<sup>3</sup>, поскольку в них определена норма, которую записывают как

$$\|a\|_{E^n} = \sqrt{\langle a, a \rangle_{E^n}}, \quad \|a\|_{L_2(S)} = \sqrt{\langle a, a \rangle_{L_2(S)}}.$$

Далее, если будет понятно о каких из указанных пространств идёт речь, мы будем опускать поясняющие нижние индексы при скалярном произведении и норме.

Пространство  $L_2$  — это очень общий класс функций, не требующий не только дифференцируемости, но даже и непрерывности своих элементов-функций. Это вытекает из указанных выше свойств пространства  $L_2$ , где скалярное произведение и норма определяются только интегральными свойствами. В то же время управление  $u(\tau)$  на  $S$

<sup>1</sup> Пространство линейно, если для его элементов определены операции сложения друг с другом и умножения на число.

<sup>2</sup> Евклидово пространство — вещественное линейное пространство с введённым на нём положительно определённым скалярным произведением;

<sup>3</sup> Банахово пространство — нормированное линейное пространство, с метрикой, которая позволяет вычислять длину элементов пространства и расстояние между ними. Пространство полное по метрике.

и тем более состояние распределённой системы  $v(\tau)$  на  $\Omega$  могут иметь производные различных порядков. Как же их дифференцировать, если функции  $u$  и  $v$  являются всего лишь элементами  $L_2$ ?

Ответ на этот вопрос даёт теория обобщённых функций [79, 58, 50, 32], которая наделяет функции из  $L_2$  свойствами дифференцирования в виде *обобщённых производных*, т.е. производных, которые можно брать и от негладких, разрывных функций. Понятие обобщённых производных, в частности, в смысле Соболева [79], вводится не через классическое дифференцирование, а через интегрирование (интегральные тождества). Именно это позволяет «обходить» проблемы недифференцируемости функций. Например, *пространство Соболева*<sup>1</sup>  $W_p^r$  имеет  $r$ -раз обобщённо дифференцируемые функции, интегрируемые по модулю со степенью  $p$ . В нашем случае  $p = 2$ . Такой частный класс функций называют *пространством Гильберта*<sup>2</sup> и обозначают как  $H^r \stackrel{\text{def}}{=} W_2^r$ . Очевидно, что пространство  $H^r$  является подпространством  $L_2$  и обладает всеми его свойствами, в том числе указанной ранее нормой и скалярным произведением.

Таким образом, каковыми бы ни были функции управления и состояния распределённой системы в пространстве  $L_2$ , они могут быть дифференцируемы в обобщённом смысле необходимое количество раз и все эти производные также будут принадлежать пространству  $L_2$ . Исследованием свойств решений дифференциальных уравнений в гильбертовых пространствах, т.е. с учётом обобщённых производных, занимается математическая физика. Здесь же мы будем рассматривать практическую сторону решения задач и не будем исследовать

<sup>1</sup> Пространство Соболева — пространство, состоящее из функций, имеющих обобщённые производные. Здесь норма  $\|u\|_{W_p^r(S)} = \left( \sum_{|\alpha| \leq r} \int_S |D^\alpha u|^p dS \right)^{1/p}$ , где  $D^\alpha$  — оператор обобщённой производной по мультииндексу  $\alpha$ .

<sup>2</sup> Гильбертово пространство — обобщение евклидова пространства, допускающее бесконечную размерность.

теоретические свойства решений, такие как существование, непрерывность, дифференцируемость и т.п. Поэтому нам будет достаточно использовать пространство  $L_2$ .

При оптимизации распределённой системы, кроме возможности наличия производных от управления, могут присутствовать и другие особенности. Управление системой может осуществляться несколькими параметрами-функциями. Таких параметров-функций, на практике, бывает конечное небольшое число  $n$ . При этом они будут представлять собой некоторую *вектор-функцию*<sup>1</sup> размерности  $n$ :  $u(\tau) \in U(S) \subseteq L_2^n(S)$ . В то же время и управляемая система может состоять не из одного уравнения, а из  $m$  уравнений, т.е. состояние такой системы так же может быть *вектор-функцией*  $v(\tau) \in V(\Omega) \subseteq L_2^m(\Omega)$ . Здесь указанные пространства  $L_2^n(S)$ ,  $L_2^m(\Omega)$  — это пространства вектор-функций со скалярным произведением вида

$$\langle a, b \rangle_{L_2^n(S)} = \int_S \langle a(\tau), b(\tau) \rangle_{E^n} dS = \int_S \sum_{i=1}^n a_i(\tau) b_i(\tau) dS.$$

Мы видим, что в бесконечномерных задачах оптимизации искомое управление  $u$ , как и в конечномерных задачах, тоже может быть вектором размерности  $n$ , но компоненты у этого вектора — не числа  $u_i$ , а функции  $u_i(\tau)$ . К сожалению, практические решения задач оптимизации в векторных бесконечномерных пространствах типа  $L_2^n(S)$ , в литературе отсутствуют (во всяком случае, автору данной книги неизвестны). Проблема заключается в чрезвычайно плохой сходимости функций управления на  $S$  к оптимальному решению за большое, а точнее, очень большое и даже бесконечное число шагов (итераций).

В конечномерных пространствах для популярных алгоритмов минимизации и некоторых частных случаев целевых функций, а именно,

---

<sup>1</sup> Такая вектор-функция представляется вектор-столбцом из функций.

для выпуклых, квадратичных (данные понятия подробно разъясняются в подразделе «Квадратичность и выпуклость») имеются доказательства сходимости решений к оптимуму за конечное число итераций [30, 67, 9, 6, 7, 12, 66]. Причём в методах первого порядка, использующих первую производную целевой функции  $J'_u = \nabla J$  (*градиент*), доказывается сходимость методов в зависимости от размерности  $n$  задачи минимизации. В бесконечномерных пространствах размерность  $n = \infty$  и ожидать сходимость формальных обобщений конечномерных алгоритмов из  $E^n$  в  $L_2(S)$  за конечное число итераций, в общем случае, бессмысленно. Здесь могут быть исключения только для алгоритмов минимизации второго порядка, например, для метода Ньютона. Но применение методов второго порядка в бесконечномерных пространствах усложняет задачу, можно сказать, на порядок. Практические решения задач оптимизации, посредством вычисления вторых производных  $J''_{uu} = \nabla^2 J$  неявно заданного целевого функционала с распределённой системой, опять же, автору данной книги неизвестны.

К вопросу теоретической сходимости методов оптимизации будет уместно процитировать Б.Т. Поляка [66]: «Ещё один упрёк к теоремам о сходимости заключается в том, что они рассматривают идеализированную ситуацию, отвлекаясь от *помех*, ошибок округления, невозможности точного решения вспомогательных задач и т.п., а все эти факторы сильно влияют на поведение метода в реальных условиях». Именно поэтому в данной книге основное внимание уделяется практической эффективности методов.

Под словом *эффективность* метода оптимизации мы будем подразумевать количество вычислений целевого функционала  $J$  (вместе с уравнениями распределённой системы) и его градиента  $\nabla J$  (как правило, на основе сопряжённой распределённой системы) на всех итерациях, необходимых для приближения к решению задачи (1.2) с

заданной точностью. Именно подзадачи вычисления  $J$  и  $\nabla J$  на каждой итерации поглощают основное компьютерное время решения задачи оптимизации.

Обычно, при классической конечномерной оптимизации, вместо количества вычислений функции и градиента, говорят о количестве вычислений только градиента и используют термин количества обращений к *оракулу*<sup>1</sup> [61, 35] — подпрограмме вычисления градиента. В нашем случае, если и использовать эту терминологию, то под оракулом следует понимать подпрограмму вычисления и функционала, и градиента. Но всё же мы остановимся на термине *количество вычислений  $J$  и  $\nabla J$*  для сравнительной оценки эффективности методов оптимизации.

Иногда в бесконечномерных задачах оптимизации избавляются от функциональной зависимости  $u(\tau)$  посредством разложения искомых параметров в ряды и при этом переходят к дискретной задаче векторной оптимизации относительно коэффициентов ряда. В работе [94] на примере решения тестовой задачи оптимизации теплового потока показано, что такой переход, на самом деле, не обеспечивает решение исходной задачи.

Иногда к конечномерной оптимизации переходят посредством замены функции числовым множеством её значений в узлах пространственно-временной сетки численного решения распределённой системы:  $u_i = u(\tau_i)$ , где множество  $\{\tau_i\}_{i=1}^n$  — это выбранные на  $S$  узлы конечно-разностной сетки. Однако, если не обоснована сходимость функций  $u(\tau) \rightarrow u_*(\tau)$  в пространстве  $L_2(S)$ , то бессмысленно говорить о её сходимости в нескольких точках  $\tau_i \in S$ .

---

<sup>1</sup> Оракул проще всего представить в виде некоторого устройства, которое отвечает на последовательные вопросы численного метода. Метод оптимизации пытается решить задачу, собирая и анализируя ответы оракула [62].

Здесь можно процитировать А.И. Егорова и Л.Н. Знаменскую [42]: «Если задача оптимального управления имеет решение, то далеко не всегда «естественная» конечномерная аппроксимация задачи может служить естественным приближением исходной задачи. Иначе говоря, решения конечномерных задач могут не быть достаточно точными приближенными к решениям исходной задачи».

Когда мы работаем в бесконечномерных пространствах, то появляется ещё одна особенность, которой нет в конечномерных задачах. Это особый характер сходимости функций  $u^k(\tau) \rightarrow u_*(\tau)$ , где  $k = 0, 1, 2 \dots$  — номер итерации. Например, функции могут сходиться во всех точках  $\tau \in S$ , т.е. *поточечно*, когда в процессе итераций  $\max_{\forall \tau \in S} |u^k(\tau) - u_*(\tau)| \xrightarrow{k \rightarrow \infty} 0$ . Функции могут сходиться интегрально, в частности, с квадратом в  $L_2$ , т.е. в *среднеквадратичном*, когда  $\|u^k(\tau) - u_*(\tau)\|_{L_2(S)} \xrightarrow{k \rightarrow \infty} 0$ .

Далее мы будем говорить в большей степени о *равномерной сходимости*, которая позволяет достигать точное на  $S$  значение функции оптимального управления  $u_*(\tau)$ . Уточним это понятие.

**Определение 1.1.** Если для любого  $\varepsilon > 0$  существует такой номер  $\bar{k}$ , что для всех точек  $\tau \in S$  и всех номеров  $k > \bar{k}$  выполняется неравенство  $|u^k(\tau) - u_*(\tau)| < \varepsilon$ , то функция  $u^k(\tau)$  на  $S$  равномерно сходится к  $u_*(\tau)$ .

При равномерной сходимости имеет место и поточечная сходимость. Их различие заключается в том, что при поточечной сходимости для разных точек  $\tau \in S$  может быть свой номер  $\bar{k}$  и  $\varepsilon$ .

Чтобы лучше представить себе характер равномерной сходимости, можно обратиться к примеру рисунка 1.2, где показана сходимость

функционального ряда  $\mathcal{S}_i(x) = x^i$  на отрезке  $[0,1]$  при  $i \rightarrow \infty$ . При этом  $\lim_{i \rightarrow \infty} \|\mathcal{S}_i(x)\|_{L_2[0,1]} = 0$ , т.е. по норме в  $L_2$  ряд сходится к нулю, но  $\lim_{i \rightarrow \infty} \max_{x \in [0,1]} |\mathcal{S}_i(x)| = 1$ , т.е. равномерная сходимость на отрезке  $[0,1]$  отсутствует. Всплеск функций  $\mathcal{S}_i(x)$  вблизи правой границы истончается с ростом  $i$ , но никогда не исчезает.

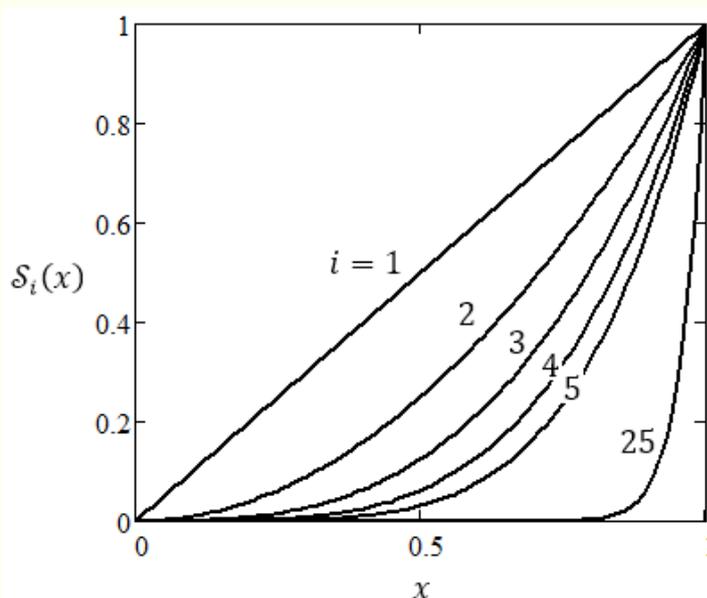


Рисунок 1.2 — Пример неравномерной сходимости

Сказанное означает, что если мы даже и будем иметь обоснование сходимости того или иного экстремального метода к решению  $u_*$  по норме в  $L_2$ , то сходимости функций  $u^k(\tau)$  к  $u_*(\tau)$  всюду на  $S$  может и не быть. Традиционные доказательства сходимости решений бесконечномерных экстремальных задач обещают нам всего лишь среднеквадратичную сходимость к решению  $u_*(\tau)$ , да и то в случае, если мы сделаем бесконечно много итераций, и без вычислительных помех.

Сделаем следующее важное замечание к вопросу о равномерной сходимости. Во-первых, с неравномерной сходимостью можно бороться

исключением из рассмотрения точек, где такая сходимость отсутствует. Ранее в примере рисунка 1.2 — это точка  $x = 1$ . Во-вторых, необходимо иметь в виду, что пространство  $L_2$  допускает существование разрывных функций. На разрывах функция может не быть определена. Поэтому далее нам и эти точки разрывов следует исключить из рассмотрения.

Естественно полагать, что множество таких «плохих» исключённых точек  $S_0 \subset S$ , где не выполняется поточечная сходимость и есть неопределённость функции, имеет меру нуль<sup>1</sup>, т.е.  $\text{mes } S_0 = \int_{S_0} dS = 0$  — мера  $S_0$  в множестве  $S$  нулевая. В таких ситуациях говорят о *сходимости почти всюду* на  $S$  [50, 48] (п.в. $S$ ). При этом, если множество  $S_0$  немного расширить до множества конечной меры  $\Delta S \supset S_0$ , где  $\Delta S \subset S$ , и далее исключить полученное  $\Delta S$  из рассмотрения, то на оставшемся множестве  $S_\Delta = S \setminus \Delta S$ , можно уже говорить о *равномерной сходимости* (теорема Егорова [50]). На этом множестве может быть достигнута равномерная сходимость функций управления к оптимали.

Возможность получения равномерной сходимости на  $S_\Delta$  зависит от размеров исключённого множества  $\Delta S$  и от особенностей алгоритмов оптимизации. Множество  $S_\Delta$  будем называть множеством *возможной равномерной сходимости*.

Казалось бы, проблема неравномерной сходимости при этом решается. Отбрасываем «плохие» множества в виде  $\Delta S$ , применяем классические экстремальные алгоритмы оптимизации и получаем прекрасную равномерную (поточечную) сходимость к оптимуму. Но это, к сожалению, не так.

---

<sup>1</sup> Например, в случае одномерной функции  $u(\tau)$ ,  $\tau \in S$  и  $S \subset E$ , множество  $S_0$  — это набор конечного количества точек. Понятно, что они имеют нулевую длину. В случае двумерной функции  $u(\tau)$ , когда  $S \subset E^2$ , множество  $S_0 \subset E$  — это набор конечного количества линий, которые имеют нулевую площадь.

Дело в том, что возможная равномерная на  $S_\Delta$  сходимость к оптимали  $u_*(\tau)$  может достигаться экстремальными алгоритмами либо при очень больших  $\Delta S$ , что делает бессмысленным саму задачу управления на оставшемся малом  $S_\Delta$ , либо — при бесконечно большом количестве итераций, что в реальной (вычислительной) жизни нереализуемо. В последнем случае при наличии вычислительных помех, сходимость к  $u_*(\tau)$  на практике завершается относительно рано, а попытка функций  $u^k(\tau)$  приблизиться к  $u_*(\tau)$ , как правило, приводит к значительным физически необоснованным искажениям управления  $u(\tau)$ .

Таким образом, очевидные преимущества прямого экстремального подхода (1.2), как мы видим, сопровождаются своими сложностями и проблемами, преодолению которых и посвящена настоящая книга.

### 1.3 Введение в прямой экстремальный подход

Обсудим здесь в общих чертах пути решения экстремальной задачи оптимизации (1.2). Пусть управление  $u(\tau) \in U = L_2(S)$ , т.е. на начальном этапе будем рассматривать задачу оптимизации без ограничений и с одной искомой функцией управления  $u(\tau)$ . Пусть состояние оптимизируемой системы также будет описываться одной неограниченной функцией  $v(\tau) \in V = L_2(\bar{\Omega})$ . Кроме этого, будем иногда обращаться и к конечномерной оптимизации, которая может иметь полезные наглядные иллюстрации для понимания общей сути экстремальных задач.

#### 1.3.1. Экстремальные алгоритмы нелинейной оптимизации

Для решения нелинейных задач, в том числе и экстремальных задач *нелинейной оптимизации*, в частности — выпуклой оптимизации (см. фундаментальную работу [62]), человечество на сегодняшний день

выработало единственный подход — это подмена нелинейной задачи набором линейных или квадратичных задач, решение которых может быть найдено аналитически. То есть, модель целевой функции или функционала  $J(u)$ , с неизвестной или сложной нелинейностью, заменяют в точках  $u^k$  моделью с известной нелинейностью (линейной или квадратичной моделью) и находят решение такой модели, которое обозначают  $u^{k+1}$ . Понятно, что в общем случае  $u^{k+1} \neq u_*$ , но предполагается, что расстояние  $\|u^{k+1} - u_*\| < \|u^k - u_*\|$ , т.е.  $u^{k+1}$  будет ближе к  $u_*$  чем предыдущее  $u^k$ . Далее, точку  $u^{k+1}$  принимают как  $u^k$ , т.е. присваивают  $k \leftarrow k + 1$  и всё повторяют до достижения желаемой близости к  $u_*$ , на сколько это будет возможным.

Оптимальное решение  $u_*(\tau)$  в прямом экстремальном подходе (1.2) ищется на множестве  $S$  (возможно на  $S_\Delta \subset S$ ) прямой, непосредственной минимизацией нелинейного целевого функционала  $J(u)$  следующим итерационным *бесконечномерным экстремальным алгоритмом*:

$$u^{k+1} = u^k + b^k d^k \quad \text{на } S, \quad k = 0, 1, \dots, \quad (1.3)$$

или в развёрнутом виде:

$$u^{k+1}(\tau) = u^k(\tau) + b^k d^k(\tau), \quad \tau \in S, \quad k = 0, 1, \dots,$$

где  $b^k$  — положительное число, называемое *шаговым множителем*, или *параметром шага*. Это число задаёт в каждой точке<sup>1</sup>  $u^k(\tau)$  глубину спуска по «поверхности» функционала вдоль *направления минимизации*  $d^k \stackrel{\text{def}}{=} d(u^k) \stackrel{\text{def}}{=} d(u^k; \tau) \in L_2(S)$ . Направление минимизации  $d^k$  — это и есть та самая «известная» модель (точнее — проекция модели на пространство управлений), для которой производится итерационная

---

<sup>1</sup> Термин «точка» в бесконечномерном пространстве означает функцию, как элемент, в данном случае, пространства  $L_2$ .

коррекция управления, подменяющая собой сложную нелинейную модель  $J(u)$  в точке  $u^k$ .

Например, линейная модель может быть построена на основе *первой производной* от  $J$ , т.е. — на основе *градиента*  $\nabla J$ , который определяет направление и скорость наибольшего изменения  $J$ . Квадратичная модель может быть построена на *второй производной*  $\nabla^2 J$ , которая определяет кривизну  $J$  (в конечномерном случае  $\nabla^2 J \equiv H$  — *матрица Гессе*).

Суть алгоритма (1.3), в одномерном случае для  $u \in E$ , можно проиллюстрировать рисунком 1.3. Здесь из точки  $u^k$  на основе линейной и квадратичной модели для функции  $J$  были сделаны шаги  $b^k d^k$  в точку  $u^{k+1}$ . Поскольку линейная модель на неограниченном множестве управлений не имеет *экстремума*, то здесь степень минимизации  $J$  из точки  $u^k$  подбирается числом  $b^k$ . В квадратичной же модели мы можем не затрачивать усилия на подбор  $b^k$ , а принять  $b^k = 1$ , именно там парабола имеет экстремум с нулевым значением производной.

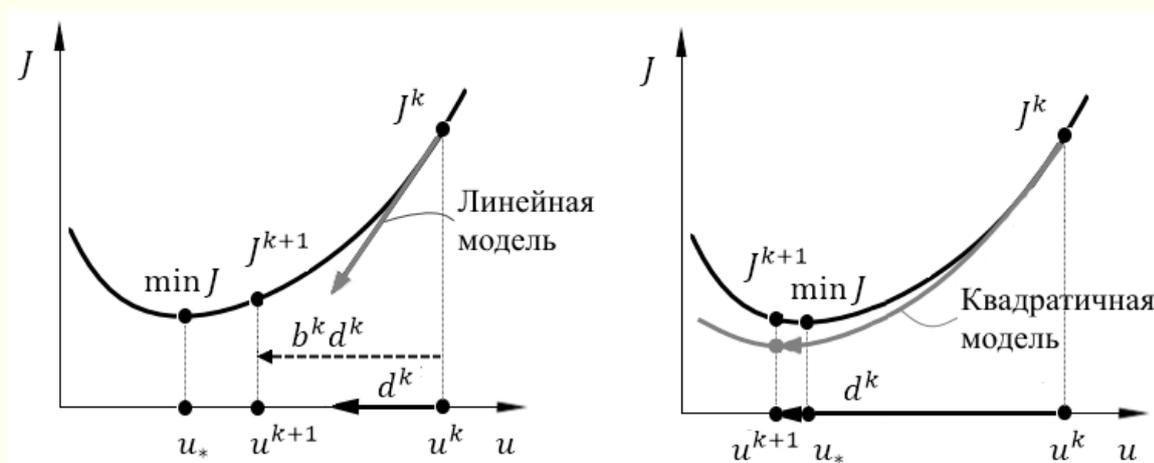


Рисунок 1.3 — Иллюстрация экстремальных алгоритмов

Видно, что квадратичная модель в виде параболы может лучше аппроксимировать нелинейную целевую функцию. Такая модель используется в алгоритмах минимизации метода Ньютона и ему подобных. Линейная же модель характерна для алгоритмов минимизации градиентными методами. Преимущества квадратичной модели очевидны, но для её построения необходимо знание второй производной  $\nabla^2 J$ , в то время как для линейной модели достаточно знания только первой производной  $\nabla J$ .

Таким образом, для прямого решения задачи оптимизации (1.1), как экстремальной (1.2), нам необходимо в каждой точке  $u^k$  знать направление шага минимизации  $d^k$  и его длину, т.е. шаговый множитель  $b^k$ .

Традиционные бесконечномерные экстремальные алгоритмы берут своё начало из конечномерных пространств, где вектор-управление  $u \in E^n$ , а алгоритм (1.3) имеет вид:

$$u_i^{k+1} = u_i^k + b^k d_i^k, \quad i = 1 \dots n, \quad k = 0, 1 \dots \quad (1.4)$$

Методы оптимизации в рамках данного алгоритма можно считать классическими. Чаще всего — это градиентные методы, где вектор  $d^k = -\nabla J^k$ , или методы сопряжённых градиентов, где вектор  $d^k = -\nabla J^k + \beta^k d^{k-1}$ ,  $\beta^k$  — соответствующие методу числа.

При обосновании сходимости конечномерных векторов  $u^k \in E^n$  к оптимуму  $u_*$  рассматривают сходимость по норме  $\|u^k - u_*\|_{E^n} \rightarrow 0$ . Эта сходимость в  $E^n$  эквивалентна<sup>1</sup> покомпонентной сходимости  $u_i^k$  к оптимуму  $u_{*i}$  для всех  $i = 1 \dots n$ . Таким образом, обосновав сходимость по норме, мы получаем сходимость  $u_i^k \rightarrow u_{*i}$  для всех  $i$ . Всё очень хорошо и однозначно.

<sup>1</sup> В  $E^n$  из сходимости векторов по норме следует их покомпонентная сходимость и наоборот:

$$\|u^k - u_*\|_{E^n} \rightarrow 0 \Leftrightarrow |u_i^k - u_{*i}|_{i=1 \dots n} \rightarrow 0.$$

В бесконечномерных пространствах дела обстоят гораздо хуже. Дело в том, что традиционные бесконечномерные методы минимизации получают формальным обобщением классических конечномерных алгоритмов на бесконечномерное пространство. При этом появляются не только новые методы, но и новые проблемы.

В бесконечномерных пространствах существует новое понятие в виде «равномерной сходимости», которое никак не коррелирует с конечномерным понятием «покомпонентной сходимости». Здесь сходимость по норме не эквивалентна поточечной (равномерной) сходимости.

Для дальнейшего описания и исследования бесконечномерных алгоритмов оптимизации нам потребуется ввести понятие *базовых направлений* минимизации  $p^k$ , которые используются традиционными методами. Это могут быть, например, упомянутые ранее направления

$$p^k = -\nabla J^k \quad \text{или} \quad p^k = -\nabla J^k + \beta^k p^{k-1}.$$

При этом, формально обобщенные бесконечномерные методы будут иметь *бесконечномерное направление минимизации*, совпадающее с базовым направлением, т.е.

$$d^k \equiv p^k \in L_2(S),$$

а *экстремальный алгоритм* (1.3), формально обобщенный на бесконечномерное пространство, будет иметь вид:

$$u^{k+1} = u^k + b^k p^k \quad \text{на } S, \quad k = 0, 1, \dots, \quad (1.5)$$

или подробнее

$$u^{k+1}(\tau) = u^k(\tau) + b^k p(u^k; \tau), \quad \tau \in S, \quad k = 0, 1, \dots$$

Например, при  $p^k = -\nabla J^k$ , мы получаем традиционный бесконечномерный градиентный метод минимизации:

$$u^{k+1} = u^k - b^k \nabla J^k \quad \text{на } S, \quad k = 0, 1, \dots$$

Формальное обобщение конечномерных алгоритмов (1.4) на бесконечномерные (1.5) меняет характер сходимости управлений, а количество итераций, необходимое для достижения оптимума даже в квадратичных задачах, может стать бесконечным.

Можно сказать, что формальное бесконечномерное обобщение (1.5) теряет чувствительность алгоритма (1.3) к поточечным значениям функции управления  $u(\tau)$ . Тонкости *равномерной* сходимости недоступны для алгоритма (1.5). Сходимость по норме  $\|u^k(\tau) - u_*(\tau)\|_{L_2(S)} \rightarrow 0$  не гарантирует равномерную сходимость на  $S$  или на  $S_\Delta$ .

В рассматриваемых в данной книге экстремальных алгоритмах оптимизации предусмотрена возможность управления сходимостью  $u^k(\tau)$  к  $u_*(\tau)$  при помощи *параметра регулирования направления спуска*  $\alpha^k(\tau)$  [92, 20, 94, 90, 86] относительно базового направления  $p^k \in L_2(S)$ . При этом *бесконечномерное направление минимизации* приобретает вид:

$$d^k = \alpha^k p^k \in L_2(S).$$

Посредством параметра  $\alpha^k$  можно контролировать в направлении  $d^k$  качество спуска и добиваться равномерной сходимости функций  $u^k(\tau) \rightarrow u_*(\tau)$  на  $S$  или  $S_\Delta$ .

Аналогичные направления могут быть построены и для случая конечномерных векторов управлений  $u \in E^n$ . При этом направление спуска будет регулироваться относительно базового вектора  $p^k \in E^n$  вектором  $\alpha^k \in E^n$ . Такой спуск может существенно ускорить сходимость к оптимуму. Здесь *конечномерное направление минимизации* будет иметь вид:

$$d^k = \overrightarrow{\alpha^k p^k} \in E^n.$$

Стрелка сверху означает *векторизацию*, где  $\overrightarrow{\alpha^k p^k} = \{\alpha_i^k p_i^k\}_{i=1}^n$ , т.е. все операции, в данном случае умножения, выполняются одновременно для всех элементов векторов.

Продолжим обсуждение бесконечномерного случая. При  $d^k(\tau) = \alpha^k(\tau)p^k(\tau)$  алгоритм (1.3) превращается в разные *методы с регулируемым направлением спуска* (МРНС) относительно разных базовых направлений  $p^k$ :

$$u^{k+1} = u^k + b^k \alpha^k p^k \quad \text{равномерно на } S_\Delta, \quad k = 0, 1, \dots, \quad (1.6)$$

*Экстремальный алгоритм МРНС* (1.6) отличается от традиционного (1.5) наличием параметра регулирования направления спуска  $\alpha$  для достижения на  $S_\Delta \subseteq S$  *равномерного* приближения функций управления к оптимуму. В предыдущем традиционном алгоритме (1.5) нет возможности управлять характером сходимости, поэтому проблемы возникающей неравномерной сходимости там даже не обсуждаются.

Таким образом, решение задачи оптимизации (1.2) экстремальным алгоритмом МРНС (1.6), в отличие от традиционного (1.5), требует нахождения подходящего параметра  $\alpha(\tau)$ , способного при интегральной (по норме) сходимости на  $S$  к оптимуму добиваться и равномерной на  $S_\Delta$  сходимости. Только в этом случае можно говорить о конечном, разумном числе итераций для удовлетворительного приближения к оптимальной функции  $u_*(\tau)$ .

### 1.3.2. Выбор направления минимизации, градиент

Рассмотренные ранее направления минимизации  $d^k$  используют *первую производную* целевого функционала (функции), т.е. градиент  $\nabla J(u)$ . Он указывает величину и направление наибольшего роста функционала (функции) в точке  $u$ . Алгоритмы оптимизации, осуществляющие движение к экстремуму функции или функционала

вдоль градиента  $\nabla J$  (антиградиента  $-\nabla J$ ) называются *градиентными методами*. Они отличаются друг от друга только стратегией выбора шагового множителя  $b^k$ .

Если мы работаем с конечномерным пространством управлений, где  $u \in E^n$ , то *градиент* определяется как *вектор-строка* из частных производных  $J'_u = \nabla J = (\nabla_1 J, \dots, \nabla_n J)^T$ , где компоненты вектора — это  $\nabla_i J(u) = \frac{\partial J(u)}{\partial u_i}$ ,  $i = 1 \dots n$ . Градиент  $\nabla J \in E^n$  в каждой точке  $u \in E^n$  можно трактовать как *вектор чувствительности* целевой функции  $J$  к множеству компонент управления  $\{u_i\}_{i=1}^n$ .

Обратимся к бесконечномерным пространствам. Здесь для каждой функции управления  $u(\tau) \in L_2(S)$  градиент  $\nabla J(u; \tau) \in L_2(S)$  представляет собой *функцию чувствительности* целевого функционала  $J$  к функции управления  $u(\tau)$  на множестве  $S$ . В функциональном анализе, для случая гладких функционалов  $J(u)$ , градиент  $\nabla J(u; \tau)$  представляет собой так называемую *производную Фреше*  $J'_u$ , или сильную производную по функции (*функциональную производную*), в данном случае, по функции  $u(\tau)$ . Далее будем рассматривать именно гладкие целевые функционалы, или иначе — функционалы, дифференцируемые по Фреше.

Алгоритм (1.5), при базовом направлении  $p^k = -\nabla J^k \in L_2(S)$ , принимает вид *бесконечномерного градиентного метода* минимизации, формально обобщённого на бесконечномерное пространство:

$$u^{k+1} = u^k - b^k \nabla J^k \quad \text{на } S, \quad k = 0, 1 \dots \quad (1.7)$$

В частности, при оптимальных шаговых множителях  $b^k$ , мы получаем *градиентный метод наискорейшего спуска* (МНС):

$$\begin{aligned} u^{k+1} &= u^k - b^k \nabla J^k \quad \text{на } S, \quad k = 0, 1 \dots, \\ b^k &= \arg \min_{b>0} J(u^k - b \nabla J^k), \end{aligned} \quad (1.8)$$

МРНС (1.6) при том же базовом направлении  $p^k = -\nabla J^k \in L_2(S)$  принимает вид:

$$u^{k+1} = u^k - b^k \alpha^k \nabla J^k \quad \text{равномерно на } S_\Delta, \quad k = 0, 1 \dots \quad (1.9)$$

Теперь обсудим проблемы, связанные с определением градиента. Напомним, целевой функционал в задачах (1.1), (1.2) зависит от искомым параметров  $u$  не явно, а через состояние системы  $v$  и это обстоятельство создаёт большие проблемы при нахождении градиента. Если бы можно было найти решение распределённой системы в виде зависимости  $v(u)$ , то это решение можно было бы подставить в целевой функционал и получить явную зависимость  $J(u)$ . В этом случае не существовало бы проблемы вычисления градиента. Но, к сожалению, для задач оптимизации с распределёнными системами сделать такое практически никогда невозможно.

Здесь задача оптимизации представляет собой множество отдельных выражений: целевой функционал; дифференциальные уравнения в частных производных; краевые условия; ограничения на управление в разнообразных формах. Проблема заключается в том, как их все объединить, записать одним выражением, например, в виде модернизированного целевого функционала и далее вычислить его градиент для решения прямой экстремальной задачи (1.2).

Эту проблему мы будем решать в два этапа: 1 — линеаризация задачи; 2 — объединение полученных линейных элементов в одинаковых пространствах и выделение градиента.

**Первый этап (линеаризация).** Разберём данный этап с простыми пояснениями и примерами. Запишем главную линейную часть приращения функционала  $J(u)$  через его градиент  $\nabla J(u)$ . При представлении этого приращения  $\Delta J$  в виде ряда Тейлора (с точностью до членов первого порядка малости) мы получим:

$$\Delta J = J(u + \delta u) - J(u) = \langle \nabla J(u), \delta u \rangle + o(\|\delta u\|),$$

где  $\delta u$  — *вариация управления* (небольшое изменение, отклонение),  $\sigma(\|\delta u\|)$  — функция  $\sigma$ -малое, обозначающая малые члены ряда выше первого порядка относительно нормы  $\|\delta u\|$ , при условии  $\|\delta u\| \rightarrow 0$ . Линейную часть приращения функционала называют *первой вариацией* или *дифференциалом Гато*<sup>1</sup>

$$\delta J = \langle \nabla J(u), \delta u \rangle. \quad (1.10)$$

Если использовать терминологию функционального анализа, то, скалярное произведение в вариации (1.10) задаёт значение линейного функционала  $\nabla J(u; \tau) \in L_2(S)$  на элементе  $\delta u(\tau) \in L_2(S)$ , и это значение равно числу  $\delta J \in E$ . Здесь линейность  $\nabla J$  подразумевается относительно вариации  $\delta u$ .

Надо различать *линейный функционал* в функциональном пространстве  $L_2$  и возможный *линейный целевой функционал* как число  $J \in E$ . В первом случае, при фиксированном значении управления  $u$ , градиент, или любая другая функция  $f(\tau) \in L_2(S)$ , называется *линейным функционалом* относительно некоторого элемента  $\delta u$ , если она *однородна*<sup>2</sup> и *аддитивна*<sup>3</sup> относительно  $\delta u$ . То есть для любых чисел  $a_1, a_2 \in E$  и любых функций  $\delta u_1, \delta u_2 \in L_2$  значение функционала  $\langle f, a_1 \delta u_1 + a_2 \delta u_2 \rangle = a_1 \langle f, \delta u_1 \rangle + a_2 \langle f, \delta u_2 \rangle$ . Во втором случае функционал  $J(u) \in E$ ,  $u \in L_2(S)$  называется *линейным*, если число  $J(a_1 u_1 + a_2 u_2) = a_1 J(u_1) + a_2 J(u_2)$ .

Линейные целевые функционалы  $J(u)$ , при оптимизации распределённых систем не используются (во всяком случае, автору данной книги они не встречались). Здесь  $J(u)$  обычно выпуклые, или

<sup>1</sup> Если функционал гладкий, то дифференциал Гато совпадает с дифференциалом Фреше, а *производная Гато* (называют слабой производной) совпадает с производной Фреше, они равны градиенту  $\nabla J$  [22, 50, 29].

<sup>2</sup> Коэффициенты внутри функции можно выносить за функцию.

<sup>3</sup> Функция суммы равна сумме функций.

квадратичные. Данные понятия будут подробно рассмотрены позже в подразделе «Квадратичность и выпуклость».

Равенство нулю первой вариации  $\delta J$  в выражении (1.10) при любых  $\delta u$  представляет собой *необходимое условие оптимальности* (НУО):

$$\langle \nabla J(u_*), \delta u \rangle = 0, \quad \forall \delta u,$$

иначе говоря, при оптимальном управлении без ограничений необходимо, чтобы

$$\|\nabla J(u)\| = 0, \quad u = u_*. \quad (1.11)$$

То есть, если некоторое управление  $u$  — оптимальное, то имеет место условие (1.11). Обратное утверждение (если выполняется (1.11), то  $u$  — оптимальное) в общем случае неверно. Оно имеет место только, если  $J(u)$  выпуклый. В этом случае необходимое условие (1.11) является и достаточным условием оптимальности.

Именно на условии  $\|\nabla J(u^k)\| \xrightarrow{k \rightarrow \infty} 0$  построены классические экстремальные алгоритмы без регулирования направления спуска.

Из выше указанного ряда Тейлора видно, что для нахождения градиента  $\nabla J$  необходимо линеаризовать, т.е. варьировать все, или почти все, выражения исходной задачи. По сути, первый этап — это классическая процедура подготовки к численному решению любой нелинейной задачи, когда нелинейную модель заменяют набором (итерациями) линейных, о чём мы говорили в предыдущем подразделе.

Если целевой функционал  $J(u)$  зависит *явно* от управления  $u$ , то для нахождения градиента  $\nabla J$  достаточно будет первого этапа. Рассмотрим такой пример.

Пусть

$$J(u) = \int_0^1 I dx, \quad I = (u(x) - 5)u(x) + 2x,$$

тогда первая вариация

$$\delta J = \langle I'_u, \delta u \rangle_{L_2(0,1)} = \langle 2u - 5, \delta u \rangle_{L_2(0,1)}.$$

Откуда, согласно (1.10), следует, что *градиент*

$$\nabla J(u; x) = 2u(x) - 5, \quad x \in (0,1).$$

Если же  $J(u)$  зависит от  $u$  неявно, то потребуется второй этап преобразований.

**Второй этап (объединение).** Полученные линейные элементы в одинаковых пространствах будем складывать (свойство аддитивности линейных пространств) — это и есть место объединения отдельных выражений задачи оптимизации. Линеаризованная и далее объединённая задача в конечном итоге принимает вид первой вариации (1.10). Откуда мы получаем градиент как некоторый линейный функционал, действующий на элемент  $\delta u$ . На этом **оба этапа** получения выражения для градиента  $\nabla J$  заканчиваются.

Давайте более предметно, более конкретно и в тоже время кратко, опишем двухэтапную процедуру определения градиента *неявно заданного* целевого функционала. Используем алгоритм определения  $\nabla J$ , описанный в прямом экстремальном подходе [20, 92]. Оба этапа укладываются в шесть пунктов. В рассматриваемом примере будем полагать, что подынтегральная целевая функция  $I(v)$  задана на  $\omega \subset \Omega$  и зависит только от *исходного состояния* системы  $v$ , управление  $u$  определено на  $S \subseteq \Omega$ . Как и ранее, размерность функций управления и состояния  $n = m = 1$ .

**1.** Если в задаче оптимизации присутствуют *ограничения* на управление, то «отбрасываем» их для учёта позже в экстремальных алгоритмах. Полагаем  $u \in U = L_2(S)$  и  $v \in V = L_2(\bar{\Omega})$ , т.е. множества управлений  $U$  и состояний  $V$  не ограничены и совпадают со своими пространствами.

2. Линеаризуем задачу (1.1) посредством варьирования искомым параметров  $\delta u$ , которые вызывают и вариацию состояния системы  $\delta v \in V$ . Получаем, во-первых, *вариацию функционала*

$$\delta J = \int_{\omega} I'_v \delta v d\omega \stackrel{\text{def}}{=} \langle I'_v, \delta v \rangle_{V^*(\omega)} \in E,$$

во-вторых, — *вариацию системы*  $\mathbb{D}v = 0$  на  $\Omega$  (без краевых условий на  $\partial\Omega$ ) в виде:

$$\mathbb{V}\delta v + \mathbb{U}\delta u = 0 \in V(\Omega),$$

где  $\mathbb{V}$ ,  $\mathbb{U}$  — некоторые *линейные операторы*<sup>1</sup>, т.е. они *аддитивные* и *однородные*. Оператор  $\mathbb{V}$  *отображает*<sup>2</sup> состояние системы из пространства  $V(\Omega)$  в него же. Оператор  $\mathbb{U}$  отображает управление из  $U(S)$  в  $V(\Omega)$ . То есть  $\mathbb{V} : V(\Omega) \rightarrow V(\Omega)$ , а  $\mathbb{U} : U(S) \rightarrow V(\Omega)$ . *Производная функции цели*  $I'_v = \partial I / \partial v$  характеризует *чувствительность* целевого функционала  $J$  к состоянию  $v$  в области  $\omega$  и представляет собой *линейный функционал* из пространства  $V^*$ , *сопряжённого*<sup>3</sup> (двойственного) к исходному пространству состояний  $V$ . Верхняя звёздочка означает сопряжённость.

3. Полученные линейные элементы исходной задачи принадлежат разным пространствам, а именно:  $E$  и  $V(\Omega)$ . Для их дальнейшего объединения необходимо элементы перевести (отобразить) в одинаковые пространства. Отобразим линеаризованные уравнения системы из  $V(\Omega)$  в  $E$  при помощи скалярного умножения на некоторой *линейный функционал* в виде произвольной функции  $\tilde{f}(\tau) \in V^*(\bar{\Omega})$ :

<sup>1</sup> Например, оператор  $\mathbb{U}$  будет линейным, если для любых  $a_1, a_2 \in E$  и  $\delta u_1, \delta u_2 \in L_2$  выполняется условие его однородности и аддитивности:  $\mathbb{U}(a_1 \delta u_1 + a_2 \delta u_2) = a_1 \mathbb{U} \delta u_1 + a_2 \mathbb{U} \delta u_2$ . Заметим, что для однородного оператора, нулевое значение возможно только при  $\delta u = 0$ , т.е.  $\mathbb{U}0 = 0$ .

<sup>2</sup> Отображение ставит соответствие между элементами двух множеств (или пространств) по правилу: каждому элементу первого множества соответствует один элемент второго множества.

<sup>3</sup> Поясним понятие сопряжённого пространства на примере. Если на пространстве  $V(\Omega)$  определить некоторое множество непрерывных линейных функционалов, то они образуют своё линейное пространство  $V^*(\Omega)$ , которое называется сопряжённым или двойственным к пространству  $V(\Omega)$ . Здесь  $V(\Omega)$  является пространством определения, а  $V^*(\Omega)$  — пространством существования линейных сопряжённых функционалов.

$$\langle \tilde{f}, \mathbb{V}\delta v + \mathbb{U}\delta u \rangle_{V^*(\Omega)} \in E.$$

Теперь делаем перестановки в скалярном произведении так, чтобы справа остались только вариации  $\delta v$ ,  $\delta u$ :

$$\langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\bar{\Omega})} + \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)} \in E.$$

Здесь  $\mathbb{V}^*$ ,  $\mathbb{U}^*$  — линейные сопряжённые операторы, действующие на  $\tilde{f}$ , т.е. — это отображения  $\mathbb{V}^* : V^*(\Omega) \rightarrow V^*(\bar{\Omega})$ ,  $\mathbb{U}^* : V^*(\Omega) \rightarrow U^*(S)$ . Появившееся замыкание  $\bar{\Omega}$  говорит о появлении сопряженных элементов  $\tilde{f}$  на границе  $\partial\Omega$ .

**4. Объединение задачи.** Поскольку  $\tilde{f}$ ,  $I'_v$  и операторы  $\mathbb{V}^*$ ,  $\mathbb{U}^*$  линейные и обладают свойствами аддитивности, то мы можем собрать (объединить) все части линеаризованной задачи при одинаковых вариациях. Откуда получим первую вариацию целевого функционала (пока без учёта краевых условий на  $\partial\Omega$ ):

$$\delta J = \langle \mathbb{V}^* \tilde{f} + I'_v|_{\omega}, \delta v \rangle_{V^*(\bar{\Omega})} + \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)} \in E.$$

**5. Учёт краевых условий.** Поскольку дифференциальные уравнения в  $\mathbb{D}v = 0$  обязательно сопровождаются краевыми условиями (граничными, начальными) на части  $\partial\Omega$ , то мы должны их также линеаризовать и найти из них выражения  $\delta v$ . После чего все граничные выражения  $\delta v$  следует подставить в вариацию  $\delta J$  на  $\partial\Omega$ .

**6. Выделение градиента.** Зададим значение произвольной функции  $\tilde{f} = f$  из условия

$$\mathbb{V}^* f + I'_v|_{\omega} = 0 \text{ на } \bar{\Omega}, \quad (1.12)$$

которое называется *сопряжённой задачей*, а её решение, т.е. линейный функционал  $f \in V^*(\bar{\Omega})$ , — *сопряжённым состоянием* системы. При этом *первая вариация целевого функционала* принимает вид, зависящий только от вариации управления  $\delta u$ :

$$\delta J = \langle \mathbb{U}^* f, \delta u \rangle_{U^*(S)}.$$

Откуда, согласно (1.10), следует, что *градиент* — это значение оператора  $\mathbb{U}^*$ :

$$\nabla J = \mathbb{U}^* f \in U^*(S). \quad (1.13)$$

Градиент принадлежит сопряжённому пространству управлений  $U^* = L_2^*$  и определяется через решение  $f$  сопряжённой задачи (1.12).

На этом алгоритм определения градиента заканчивается.

Для евклидова пространства  $L_2$  его сопряжённое пространство  $L_2^*$  совпадает с исходным  $L_2$  (евклидово пространство — *самосопряжённое*), поэтому далее будем писать

$$\nabla J \stackrel{\text{def}}{=} \nabla J(u; \tau) \in L_2(S).$$

Сделаем существенное замечание. Дело в том, что формально найденная первая вариация целевого функционала

$$\delta J = \langle J'_u, \delta u \rangle_{U^*(S)}$$

на самом деле содержит не градиент  $\nabla J$ , а всего лишь некоторую *производную Фреше*<sup>1</sup>  $J'_u$ . Она, как и градиент, конечно же обращается в ноль в оптимуме (при отсутствии ограничений на управление). При этом НУО (1.11) можно представить в виде:

$$J'_u(u_*; \tau) = 0, \quad \tau \in S.$$

Однако, найти корень  $u_*$  этого нелинейного бесконечномерного уравнения численными методами, применяемыми в окрестности  $u = u_* + \delta u$ , в общем случае невозможно.

Производная Фреше  $J'_u(u; \tau)$  может не быть чувствительна к управлению  $u(\tau)$  на всём  $S$  или на частях  $S$  ненулевой меры. Поэтому  $J'_u$ , в общем случае, не указывает достоверное направление коррекции управления на  $S$ . Из производной  $J'_u$  можно получить градиент  $\nabla J$  с однозначной чувствительностью к управлению  $u(\tau)$  на  $S$  только при

---

<sup>1</sup> Напомним, что градиент в бесконечномерном пространстве является производной Фреше, но обратное утверждение, о том, что всякая функциональная производная в  $\delta J$  является градиентом, неверно.

выполнении условий *управляемости*. Данные условия обеспечивают корректную постановку задачи оптимизации, в которой градиент  $\nabla J$  может использоваться для направленного поиска оптимального решения. Об управляемости мы поговорим позже.

Если целью исследования является не решение задачи оптимизации, а только получение НУО, то можно ограничиться выражением  $\|J'_u\| = 0$ . Именно этим и занимается большинство теорий оптимального управления системами с распределёнными параметрами.

Вернёмся к сопряжённой задаче (1.12). Её смысл и предназначение заключается в отображении функции  $I'_v|_\omega$  (*чувствительность*  $J$  к состоянию  $v$  на  $\omega$ ) из области  $\omega$  в область  $S$ , где определён градиент и управление. Такое отображение делается при помощи  $f \in V^*(\Omega)$  с использованием промежуточного множества  $\Omega \subset \Omega$ . Множество  $\Omega$ , согласно управляемости, задаёт корректную область определения  $V^*(\Omega)$  для оператора  $\mathbb{U}^*$  в градиенте (1.13).

То есть, сопряжённая задача, по сути, реализует отображение

$$f : V^*(\omega) \rightarrow V^*(\Omega).$$

Далее при помощи оператора

$$\mathbb{U}^* : V^*(\Omega) \rightarrow U^*(S)$$

мы из производной Фреше  $J'_u$  получаем градиент:

$$J'_u = \mathbb{U}^* f|_\Omega \rightarrow \mathbb{U}^* f|_\Omega = \nabla J \in U^*(S).$$

Область определения целевой функции  $\omega$  и управления  $S$ , в пространственно-распределённых задачах оптимизации, практически, никогда не совпадают, поэтому, такие задачи не решаются без соответствующих сопряжённых отображений из  $\omega$  в  $S$ .

Из приведённого описания процедуры нахождения аналитического выражения градиента неявно заданного целевого функционала видно, что это непростая задача. Пожалуй, именно она

требует основных усилий при решении оптимизационных задач с распределёнными параметрами. Подробнее о получении градиента мы поговорим в разделе 3.

В заключении данного подраздела приведём краткую информацию и о второй производной целевой функции, функционала.

Если приращение функционала  $\Delta J$  представить рядом Тейлора с квадратичным членом (с точностью до членов второго *порядка малости*), то получим

$$\Delta J = \langle \nabla J(u), \delta u \rangle + \frac{1}{2} \langle \nabla^2 J(u) \delta u, \delta u \rangle + o(\|\delta u\|^2),$$

где *линейный* оператор  $\nabla^2 J(u)$  — это *вторая производная* целевого функционала. Напомним, что в  $n$ -мерном пространстве оператор  $\nabla^2 J$  представляет собой  $n \times n$  *матрицу Гессе*. Если же имеет место бесконечномерное управление  $u \in L_2$ , то оператор  $\nabla^2 J$  представляет собой отображение из  $L_2$  в  $L_2$ . Линейность данного оператора означает однородность и аддитивность в заданной точке  $u$  выражения  $\nabla^2 J(u) \delta u$  относительно  $\delta u$ .

Квадратичную часть приращения называют *второй вариацией*:

$$\delta^2 J = \langle \nabla^2 J(u) \delta u, \delta u \rangle.$$

Имея вторую производную  $\nabla^2 J$  можно было бы строить экстремальные алгоритмы оптимизации на порядок лучшие чем градиентные. Можно было бы минимизировать неограниченные квадратичные целевые функции и функционалы всего за один шаг поскольку в этом случае квадратичный ряд Тейлора даёт точное разложение. Здесь остаточный член  $o(\|\delta u\|^2) = 0$ . Однако, как упоминалось ранее, практические решения подобных задач автору данной книги неизвестны.

### 1.3.3. Длина шага минимизации

Рассмотрим некоторые стратегии выбора длины шага в экстремальных алгоритмах при неограниченной минимизации функций и функционалов  $J(u)$ . Можно выделить следующие пять наиболее популярных стратегий.

**Первая стратегия** — это выбор заранее, возможно *постоянного*, шагового множителя на всех итерациях:

$$b^k = b > 0 \text{ или } b^k = \beta_1(k)b, \quad k = 0, 1, \dots,$$

где  $\beta_1(k)$  — некоторая положительная, заранее известная функция, например,  $\beta_1(k) = b/\sqrt{k+1}$  [62]. Первая стратегия нередко используется при минимизации выпуклых  $J(u)$  градиентными методами. Для таких задач оптимизации стратегия является *релаксационной*, т.е. она ослабляет, уменьшает шаги так, что  $\|b^k d^k\| \xrightarrow{k \rightarrow \infty} 0$  по мере приближения к оптимуму.

Если известна аналитическая зависимость  $J(u)$ , то можно найти оптимальное  $b$  [66]. В этом случае используются знания о выпуклости  $J(u)$ , такие как константа липшицевости градиентов, собственные числа матрицы Гессе. Понятно, что при оптимизации распределённых систем получить что-либо подобное о поведении  $J(u)$  невозможно.

К достоинствам первой стратегии следует отнести минимальность усилий в её реализации и отсутствие дополнительных вычислений  $J$  или  $\nabla J$  на каждой итерации  $k$  для расчёта шагового множителя  $b^k$ . Поэтому, данная стратегия может оказаться эффективной в некоторых случаях, при некоторых исходных данных.

**Вторая стратегия** использует *адаптивную релаксацию* шага:

$$b^k = \beta_2(k)b^{k-1}, \quad k = 1, 2, \dots,$$

где  $\beta_2(k)$  — положительная, заранее неизвестная функция, значение которой на каждой итерации определяется (адаптируется) на основании

полученной ранее информации о поведении  $J(u)$ . Например, для минимизации выпуклых  $J(u)$ , это может быть алгоритм вида [92]:

$$\left\{ \begin{array}{l} \text{Если } J^k < rJ^{k-1} \text{ тогда: } b^k = b_1 b^{k-1}, \quad u^{k+1} = u^k + b^k d^k. \\ \text{Иначе (очистить направление: } d^k = -\nabla J^k) \\ \text{повторять до } J^k < rJ^{k-1}: \\ \quad b^{k-1} \leftarrow b_2 b^{k-1}, \quad u^{k+1} = u^{k-1} - b^{k-1} \nabla J^{k-1}, \\ \quad \text{если } b^{k-1} \approx 0, \text{ тогда стоп итерации } k. \end{array} \right. \quad (1.14)$$

Здесь  $k = 1, 2, \dots$ ,  $r \geq 1$  (возможная поправка на помехи),  $b_1 \geq 1$  (усиление сходимости),  $b_2 < 1$  (ослабление сходимости),  $d$  — направление минимизации. Начальное значение параметра  $b^0$  выбирается из условия  $J^1 < J^0$ .

Алгоритм (1.14) посредством параметра  $b_1$  может усиливать сходимость, а параметром  $b_2$  — контролировать и не допускать расходимость метода из-за чрезмерно больших шагов. Обычно  $b_2 = 0.3$ . Очистка направления минимизации необходима, если  $d^{k+1}$  формируется на основе предыдущего  $d^k$ , что характерно, например, для методов сопряжённых градиентов. Однако здесь формула (1.14) далеко не всегда может оказаться эффективной, поскольку не гарантирует условия сопряжённости направлений минимизации.

Условие завершения «стоп» в цикле повторений предыдущего шага (градиентным методом) контролирует заикливание и излишнее количество чрезмерно малых шагов при уменьшении параметра  $b^{k-1}$ . Если  $b^{k-1} \approx 0$ , например, уменьшилось в тысячу раз, то следует считать, что процесс минимизации  $J(u)$  достиг своего предела и далее в сторону минимума продвинуться невозможно. Итерации оптимизации прекращаются.

Решение распределённых систем всегда связано с большими и приближёнными вычислениями. Если помехи при вычислении  $J$  на соседних итерациях  $k$  и  $k - 1$  заметно разные, то при сравнении

функционалов  $J^k < rJ^{k-1}$  необходимо брать некоторый «запас»  $r > 1$ . Обычно достаточно  $r = 1.01$ . Небольшой «запас» может предотвратить необоснованное замедление сходимости или её преждевременное завершение. С учётом этой особенности метод (1.14) можно назвать *робастным*<sup>1</sup>. Как правило, удовлетворительная сходимость достигается при  $r = 1$ , что мы и будем далее предполагать по умолчанию.

Очевидно, что алгоритм (1.14), в общем случае, для минимизации функционалов, эффективнее алгоритмов первой стратегии.

Если  $J(u)$  не выпуклый, но имеет единственный экстремум, то можно в алгоритме (1.14), при усилении сходимости, предусмотреть и возможность ослабления сходимости. Это можно сделать в рамках следующего градиентного метода ( $d^k = -\nabla J^k$ ) [85]:

$$\left\{ \begin{array}{l} \text{Если } J^k < rJ^{k-1} \text{ тогда:} \\ \quad \cdot \text{ если } \|\nabla J^k\| < \|\nabla J^{k-1}\| \text{ тогда } b^k = b_1 b^{k-1}, \\ \quad \text{иначе } b^k = \frac{b^{k-1} \|\nabla J^{k-1}\|}{b_3 \|\nabla J^k\|}; \\ \quad \cdot u^{k+1} = u^k - b^k \nabla J^k. \\ \text{Иначе повторять до } J^k < rJ^{k-1}: \\ \quad b^{k-1} \leftarrow b_2 b^{k-1}, \quad u^{k+1} = u^{k-1} - b^{k-1} \nabla J^{k-1}, \\ \quad \text{если } b^{k-1} \approx 0, \text{ тогда стоп итерации } k. \end{array} \right. \quad (1.15)$$

Здесь  $b_3 \geq 1$ . Если  $J(u)$  убывает, а норма градиента не убывает, то это означает, что текущий шаг либо пришёлся на *вогнутую* часть  $J(u)$ , либо перескочил через локальный минимум и попал в область повышенной выпуклости  $J(u)$ . В любом случае следующий шаг  $b^k \nabla J^k$  не должен быть больше предыдущего. Величина такого шага регулируется параметром  $b_3$ . Надо быть внимательным при выборе  $b_3 > 1$ , поскольку можно существенно замедлить сходимость алгоритма.

<sup>1</sup> Робастный (твёрдый, крепкий) означает устойчивость к ошибкам, способность адаптироваться к изменяющимся условиям без значительного снижения производительности.

Если  $J(u)$  не выпуклый и имеет несколько локальных экстремумов, то алгоритм (1.15) следует далее модифицировать, посредством адаптации параметра  $b_1$ . Необходимо помнить, что оптимальному управлению должен соответствовать только глобальный минимум. Для того, чтобы сходимость не завершилась в каком-либо физически неудовлетворительном локальном экстремуме (или хочется проверить существуют ли другие экстремумы), следует приближаться к нему с большим  $b_1$  — для перешагивания через экстремум. И наоборот, чтобы не перескочить через желаемый экстремум, следует приближаться к нему осторожно, с уменьшенным  $b_1$ .

Алгоритм адаптации  $b_1$  существенным образом зависит от задачи оптимизации. Данная процедура требует кропотливого участия исследователя и вряд ли может быть формализована. Тем не менее, она позволяет исследовать многоэкстремальные задачи бесконечномерной оптимизации. Обсуждаемый алгоритм выбора шагового множителя имеет вид [85]:

$$\left\{ \begin{array}{l} \text{Если } J^k < rJ^{k-1} \text{ тогда:} \\ \quad \cdot \text{ если } \|\nabla J^k\| < \|\nabla J^{k-1}\| \text{ тогда} \\ \quad \quad \{ \text{если } u^k \text{ рядом с экстремумом} - \text{изменить } b_1 \} \\ \quad \quad b^k = b_1 b^{k-1}, \\ \quad \text{иначе } b^k = \frac{b^{k-1} \|\nabla J^{k-1}\|}{b_3 \|\nabla J^k\|}; \\ \quad \cdot u^{k+1} = u^k - b^k \nabla J^k. \\ \text{Иначе повторять до } J^k < rJ^{k-1}: \\ \quad b^{k-1} \leftarrow b_2 b^{k-1}, \quad u^{k+1} = u^{k-1} - b^{k-1} \nabla J^{k-1}, \\ \quad \text{если } b^{k-1} \approx 0 \text{ тогда стоп итерации } k. \end{array} \right. \quad (1.16)$$

Условие «если  $u^k$  рядом с экстремумом» определяется исследователем из понимания сути функционирования оптимизируемого процесса. Пример использования алгоритма (1.16) приведен в разделе 9.

Для минимизации многоэкстремальных функций существуют и другие методы поиска глобально минимума, но они не применимы для минимизации функционалов в бесконечномерном пространстве.

Следующие две стратегии основаны на решении задачи линейного поиска вдоль направления минимизации  $d^k$  на каждой итерации  $k$ . Поиск осуществляется в два этапа. Сначала ищется (задаётся) отрезок, в направлении  $d^k$ , потом на этом отрезке ищется удовлетворительное  $b^k$ . Данные стратегии часто используют во многих экстремальных алгоритмах. Более того, ряд алгоритмов, например, метод сопряжённых градиентов вообще неприменим без решения задачи линейного поиска.

Заметим, что линейный поиск применим только в задачах неограниченной оптимизации. В противном случае выбор отрезка, содержащего удовлетворительное  $b^k$ , может стать невозможным. Любой «конец» отрезка, начавшегося в удовлетворительной точке  $u^k$ , может выйти за ограничения и сделать абсурдным решение исходной задачи.

**Третья стратегия** — это стратегия линейного поиска с *полной релаксацией*, когда на каждом шаге вдоль направления минимизации  $d^k$ , на отрезке заданной длины, выбирается оптимальное значение множителя

$$b^k = \arg \min_{b>0} J(u^k + bd^k), \quad (1.17)$$

то есть величина каждого шага  $b^k d^k$  определяется из решения дополнительной одномерной задачи минимизации функции  $J^k(b) = J(u^k + bd^k)$  на заданном отрезке.

Обычно минимизация делается широко известными методами *дихотомии*, — это бисекция, *золотое сечение* и т.п. Такая стратегия может давать положительный результат если изначально для  $u^k$  задаётся корректный отрезок изменения  $b$ . На этом отрезке функция  $J^k(b)$  должна быть *унимодальной*, т.е. она должна содержать единственный

экстремум, который является минимальным значением функции на рассматриваемом отрезке.

Стратегия (1.17) в градиентном методе превращает его в метод наискорейшего спуска, а для метода сопряжённых градиентов — эта стратегия обязательна. Алгоритмы третьей стратегии программируются не на много сложнее предыдущей стратегии адаптивной релаксации. Однако, чтобы найти минимум функции  $J^k(b)$  с высокой точностью, как правило, требуется слишком много вычислений  $J$ , возможно и  $\nabla J$ , на каждой итерации  $k$ .

**Четвёртая стратегия** ([97, 29, 12, 62] и многие другие) выполняет линейный поиск (1.17) не очень точно, с относительно небольшими вычислительными затратами, но при этом может обеспечить вполне удовлетворительное уменьшение  $J$ . Такие методы основываются на условиях всего лишь достаточного убывания целевой функции  $J^k(b)$  в заданном направлении  $d^k$ . Это условия Армихо, Вольфе, Гольдштейна [12, 13, 9, 25]. Они позволяют подобрать небольшой отрезок, где содержится удовлетворительное значение  $b^k$ .

Далее, на этом отрезке, можно уточнить значение  $b^k$ , например, по третьей стратегии посредством применения методов дихотомии. Можно также уточнить  $b^k$  и из условия экстремума интерполяционного многочлена для функции  $J^k(b)$ , построенного по уже имеющимся значениям функции и её производной.

Чаще всего для достаточного убывания целевой функции используется строгое условие *Вольфе*. Оно требует, чтобы шаговый множитель  $b^k$  удовлетворял двум неравенствам:

$$\begin{aligned} J(u^k + b^k d^k) &\leq J(u^k) + c_1 b^k \langle \nabla J^k, d^k \rangle, \\ |\langle \nabla J(u^k + b^k d^k), d^k \rangle| &\leq c_2 |\langle \nabla J^k, d^k \rangle|. \end{aligned}$$

Согласно первому условию, убывание  $J$  должно быть пропорционально  $b^k$  и производной по направлению  $\langle \nabla J^k, d^k \rangle$ . Согласно второму условию, шаг  $b^k d^k$  должен приближать нас к стационарной точке функции  $J^k(b)$  где производная по направлению  $\langle \nabla J^k, d^k \rangle$  близка к нулю. *Параметры в условии Вольфе*  $0 < c_1 < c_2 < 1$ , обычно задают  $c_1 = 10^{-4}$ , а при помощи  $c_2$  регулируют степень близости к минимуму функции  $J^k(b)$ .

Алгоритмы четвёртой стратегии, безусловно, эффективнее (по количеству вычислений  $J, \nabla J$ ) алгоритмов третьей стратегии. Однако, они могут быть довольно сложными, особенно в сочетании с третьей стратегией.

В рамках четвёртой стратегии мы будем использовать *метод Носедала-Райта* (Nocedal-Wright) [12]. В данном методе сначала задаётся максимально допустимый отрезок поиска шагового множителя  $b$ . Внутри отрезка делается интерполяция (мы далее выбрали квадратичную) функции  $J^k(b)$  и аналитически находится  $b^k$  из условия экстремума интерполяционной функции. Если для полученного  $b^k$  не выполняется строгое условие Вольфе, то отрезок сокращается справа или слева в зависимости от знака производной по направлению  $\langle \nabla J(u^k + b^k d^k), d^k \rangle$  и всё повторяется заново.

Метод Носедала-Райта и ему подобные содержат много изощрённостей в своих алгоритмах, поэтому для их точной реализации мы отправляем читателя к оригиналу [12].

**Пятая стратегия** — это *Trust-region* (методы доверительной области) [6, 9, 13, 12] для минимизации функций. В отличие от линейного поиска, данная стратегия определяет доверительную область (регион) радиусом  $\Delta^k$  вокруг точки  $u^k$ , где целевую функцию  $J(u^k)$  можно заменить упрощённой, обычно квадратичной *моделью*  $M^k$ , с достаточной точностью. Далее в  $\Delta^k$ -окрестности точки  $u^k$  ищется

минимум модели  $M^k$ , который считается очередным приближением  $u^{k+1}$ . Как правило, шаг Trust-region используется совместно с методом Ньютона для минимизации неквадратичных функций, когда с традиционным для метода Ньютона шаговым множителем  $b^k = 1$  нельзя обеспечить сходимость, или быструю сходимость к минимуму.

Здесь квадратичная модель  $M^k$  строится с использованием второй производной целевой функции:

$$M^k(\delta u) = J(u^k) + \langle \nabla J(u^k), \delta u \rangle + \frac{1}{2} \langle \nabla^2 J(u^k) \delta u, \delta u \rangle.$$

Минимум квадратичной модели даёт шаг

$$\delta u^k = \arg \min_{\delta u \leq \Delta^k} M^k(\delta u)$$

и соответственно новое приближение  $u^{k+1} = u^k + \delta u^k$ .

Точность модели оценивается выражением:

$$\rho^k = \frac{J(u^k) - J(u^k + \delta u^k)}{M^k(0) - M^k(\delta u^k)}.$$

Если  $\rho^k \approx 1$ , то найденный шаг  $\delta u^k$  достаточно хорошо минимизирует целевую функцию в доверительной области  $\Delta^k$  и для следующего шага радиус области  $\Delta^{k+1}$  можно увеличить. В противном случае шаг в области  $u^k$  повторяется с меньшим радиусом и возможно с другой моделью  $M^k$ .

Стратегия Trust-region хороша при минимизации неквадратичных, или даже невыпуклых функций, когда можно найти вторую производную целевой функции, или хотя бы получить достаточно точную оценку значения второй производной на каждой итерации. К сожалению, при оптимизации в бесконечномерных пространствах данную стратегию применить не представляется возможным.

### 1.3.4. Помехи

При практическом решении экстремальных задач с распределёнными системами нельзя игнорировать помехи. В таких сложных задачах всегда имеется множество мест, где помехи могут возникать и разнообразно проявлять себя вплоть до обесценивания получаемого решения.

Помехи могут возникать из-за каких-либо неточных измерений в исходных данных, неточных вычислений с вещественными числами, из-за конечно-разностных аппроксимаций интегро-дифференциальных выражений, из-за каких-либо упрощений (облегчений) уравнений и алгоритмов. В конце концов, помехи могут появляться из-за обычных ошибок в математических преобразованиях и компьютерном кодировании сложных алгоритмов. Последние ошибки являются естественным процессом компьютерного решения сложных задач. Можно упомянуть шуточный афоризм программистов: «не бывает программ без ошибок». Только одни ошибки приводят к катастрофическим решениям и обнаруживаются почти сразу, а другие не изменяют или почти не изменяют численное решение в обычных условиях и могут очень долго не обнаруживаться.

В прямом экстремальном подходе оптимальное управление ищется на основе градиента. Например, для МРНС (1.9) мы будем иметь не точное направление спуска, а

$$d^k = -\alpha(\nabla J^k + r^k) \in L_2(S),$$

где функция  $r^k(\tau)$  — это *помехи* в направлении минимизации.

Понятно, что помехи в градиенте могут приводить к искажению получаемого оптимального управления, к неустойчивости процесса оптимизации, к невозможности приближения к оптимуму с желаемой

точностью, к излишним итерациям и необоснованности завершения итерационного процесса, а также к другим негативным последствиям.

Выделяют несколько типов помех [66], связанных с причинами их появления:

1. Погрешности компьютерных вычислений и входных данных задачи оптимизации — *абсолютные детерминированные помехи*. Здесь

$$\|r^k\| \leq \varepsilon, \quad \varepsilon > 0. \quad (1.18)$$

2. Погрешности использования приближённых формул, аппроксимаций и ошибок программирования, дающих *относительные детерминированные помехи*:

$$\|r^k\| \leq \|\varepsilon \nabla J^k\|. \quad (1.19)$$

где  $\varepsilon(\tau)$  — коэффициент-функция относительных помех. Такие помехи легко преодолевает МРНС. Он устраняет их соответствующим выбором параметра  $\alpha$ , поскольку  $d^k = -\alpha(\nabla J^k + \varepsilon \nabla J^k) = -\alpha_\varepsilon \nabla J^k$ , где  $\alpha_\varepsilon = \alpha(1 + \varepsilon)$ , т.е. «старый»  $\alpha$ , при наличии рассматриваемых помех, превращается в  $\alpha_\varepsilon$  и «поглощает» в себе помехи  $\varepsilon$ .

3. Случайные погрешности характерны для измерений. Это могут быть *абсолютные случайные помехи*. Здесь математическое ожидание погрешностей  $r^k$  ограничено их дисперсией  $\sigma$ :

$$M\|r^k\|^2 \leq \sigma^2. \quad (1.20)$$

4. Случайные погрешности также могут иметь вид и *относительных случайных помех*:

$$M\|r^k\|^2 \leq \varepsilon \|\nabla J^k\|^2, \quad \varepsilon > 0.$$

Практическое решение задач оптимизации с распределёнными системами обязательно должно строиться с учётом помех. Сходимость к оптимальной функции управления  $u_*(\tau)$  в экстремальных алгоритмах должна быть устойчивой к помехам. Разумные пределы помех (помехи

физически и математически обоснованные) не должны приводить к неразумным и физически необоснованным управлениям.

Демонстрация влияния помех на результаты оптимизации подробно рассмотрены в тестовых задачах оптимального управления и идентификации в разделах 5 и 8.

### 1.3.5. Завершение минимизации

Равенство нулю градиента целевого функционала (функции), т.е. НУО (1.11)

$$\|\nabla J(u_*)\| = 0$$

является главным теоретическим *критерием завершения* итерационных процессов оптимизации.

Из-за различных *помех* в задачах оптимизации указанное НУО практически никогда не выполняется, поскольку сходимость итерационных процессов, как правило, прекращается раньше реализации данного условия. На практике НУО (1.11) заменяется некоторой *близостью нормы градиента к нулю*:

$$\|\nabla J(u^k)\| \leq \varepsilon_1, \quad (1.21)$$

где  $\varepsilon_1$  — заданное малое положительное число (здесь и далее  $\varepsilon_j$ ,  $j = 1, 2, \dots$  — множество малых положительных чисел).

Однако, очевидно, что критерий (1.21) не имеет никакого отношения к оценке *близости решения к оптимуму*  $u_*$  ни по норме (в *среднеквадратичном*) в виде

$$\|u^k - u_*\| \leq \varepsilon_2, \quad (1.22)$$

$$\|u^k - u_*\| / \|u^0 - u_*\| \leq \varepsilon_3, \quad (1.23)$$

ни тем более *поточечно* на  $S$ . Рисунок 1.4 иллюстрирует данную ситуацию на примере одномерной оптимизации.

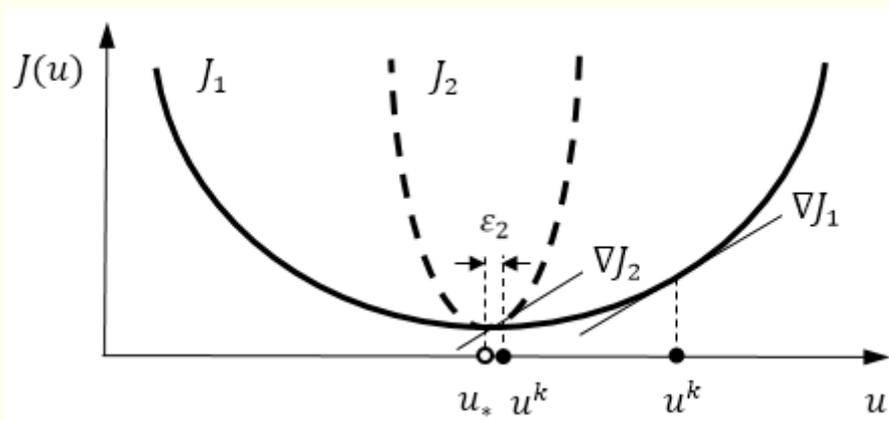


Рисунок 1.4 — Результаты минимизации функций  $J_1$  и  $J_2$  с одинаковым критерием завершения (1.21)

Касательные к функциям  $J_1$  и  $J_2$  в разных точках  $u^k$  показывают одно и то же значение величины градиента:  $\|\nabla J_1\| = \|\nabla J_2\| = \varepsilon_1$ . Из рисунка понятно, что при любой близости градиента к нулю (для любого малого  $\varepsilon_1$ ) найдётся «плохая» целевая функция (в нашем примере — это  $J_1$ ), что полученное решение  $u^k$  будет находиться далеко за пределами наперёд заданной точности оптимизации  $\varepsilon_2$ . Если к данному факту добавить ещё и наличие вычислительных погрешностей, то можно утверждать, что приближительное выражение (1.21) НУО (1.11) нельзя считать приемлемым критерием завершения процесса минимизации.

Вместо приближительной реализации НУО в форме (1.21) обычно используют следующие два критерия *практического прекращения сходимости* алгоритмов. Их формулы могут иметь вид:

$$\frac{\|u^k - u^{k-1}\|}{\|u^{k-1}\|} \leq \varepsilon_4, \quad (1.24)$$

$$\frac{|J(u^k) - J(u^{k-1})|}{J(u^{k-1})} \leq \varepsilon_5. \quad (1.25)$$

Обсудим наиболее часто используемое условие (1.24) для случая единственного оптимума. При отсутствии помех сходимость шагов

(1.24) при  $\varepsilon_4 \rightarrow 0$  влечёт за собой и сходимость к единственному оптимальному решению, т.е.  $\|u^k - u_*\| \rightarrow 0$ . Однако, данное условие ведёт себя по-разному в конечномерном и бесконечномерном случае.

Напомним, что в конечномерном пространстве  $E^n$  сходимость по норме эквивалентна покомпонентной сходимости:

$$\|u^k - u_*\|_{E^n} \rightarrow 0 \Leftrightarrow |u_i^k - u_{*i}|_{i=1\dots n} \rightarrow 0.$$

То есть в этом случае, при  $\varepsilon_4 \rightarrow 0$ , мы будем получать сходимость к оптимуму и по норме (в среднеквадратичном), и для всех номеров  $i$  компонент управления.

В бесконечномерном случае в  $L_2(S)$ , как мы знаем, из сходимости по норме не следует поточечная сходимость (равномерная сходимость) во всей области  $S$  (или даже  $S_\Delta$ ). Здесь преждевременное завершение сходимости при  $\varepsilon_4 > 0$  может приводить к недопустимо большим отклонениям функций управления  $u^k(\tau)$  от оптимального значения  $u_*(\tau)$  где-либо на  $S$ . Если добавить ещё и вычислительные помехи, то понятно, что практическое завершение сходимости (1.24) при любом малом  $\varepsilon_4 > 0$  не гарантирует достижения с желаемой точностью оптимали  $u_*(\tau)$  ни интегрально, ни тем более всюду на  $S$  (на  $S_\Delta$ ).

Когда же следует завершать итерационный процесс? К сожалению, все указанные критерии (1.21)–(1.25) являются косвенными и не отражают ту или иную близость решений  $u^k$  к оптимальному значению  $u_*$ . На практике чаще всего завершают итерации при практическом прекращении сходимости, т.е. по критерию (1.24). Этот критерий можно считать вполне приемлемыми, если мы осуществляем бесконечномерную оптимизацию с равномерной сходимостью.

### 1.3.6. Квадратичность и выпуклость

При решении экстремальных задач часто используются термины *квадратичность* и *выпуклость* целевой функции, функционала  $J(u)$ . Эти понятия тесно связаны с существованием единственного *строгого локального минимума*<sup>1</sup> для  $J(u)$ . При оптимизации на неограниченном множестве управлений, когда  $U = E^n$  или  $U = L_2$ , единственный строгий локальный минимум является и *глобальным*, т.е. здесь в локальном минимуме мы всегда получаем *оптимальное управление*  $u_*$ , и оно единственное. Будет целесообразно пояснить и проиллюстрировать на конечномерных примерах  $u \in E^n$  при  $n = 2$  суть этих терминов.

Квадратичную целевую функцию  $J(u)$  можно представить в *квадратичной форме*:

$$J(u) = \frac{1}{2} \langle Au, u \rangle, \quad (1.26)$$

где  $A$  — *симметричная матрица*  $n \times n$  действительных чисел, т.е.  $A = A^T$ . Отметим, что любую квадратичную функцию можно представить в указанной форме посредством преобразования системы координат (см., например, [97], с. 88). *Градиент квадратичной функции* определяется как  $\nabla J(u) = Au$ , а *вторая производная квадратичной функции* — это  $\nabla^2 J = A \stackrel{\text{def}}{=} H$  — *гессиан*. Мы видим, когда известна матрица  $A$  и нет сопутствующих дифференциальных уравнений, градиент определяется очень просто.

Выпуклость  $s_{conv}$  такой функции определяется знаком скалярного произведения в разных направлениях вектора  $u$ :

$$s_{conv} = \langle Au, u \rangle.$$

При этом для любого ненулевого вектора  $u$ :

---

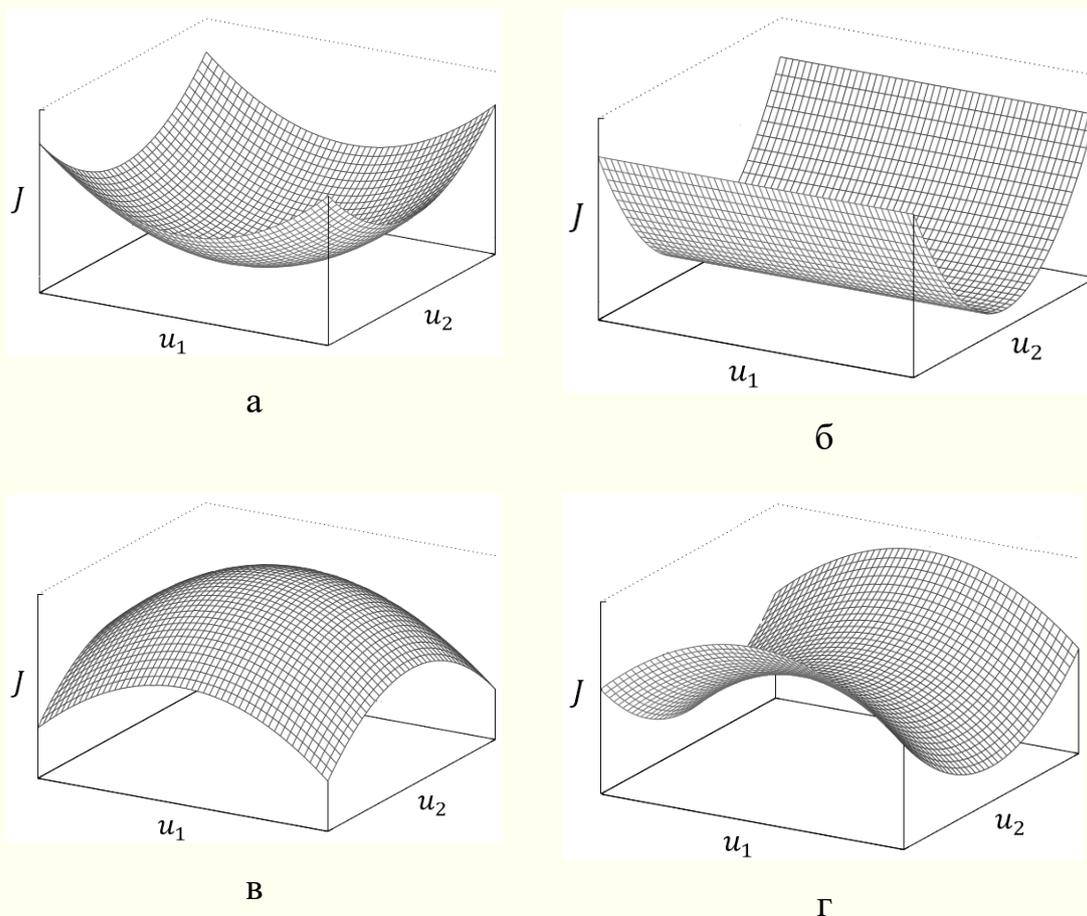
<sup>1</sup> Если  $u_*$  — единственная точка локального минимума  $J(u)$ , или говорят — точка строгого локального минимума, то для всех  $u$  в некоторой локальной окрестности  $u_*$  имеет место строгое неравенство  $J(u) > J(u_*)$  для всех  $u \neq u_*$ .

- если  $s_{conv} > 0$ , то квадратичную функцию называют *строго выпуклой*, а матрицу  $A$  — *положительно определённой* (рисунок 1.5а). Здесь  $J(u)$  имеет единственную точку локального минимума (*строгий локальный минимум*), которая является оптимумом  $u_*$ ;

- если  $s_{conv} \geq 0$ , то квадратичную функцию называют *нестрого выпуклой*, или, просто, *выпуклой*, а матрицу  $A$  — *положительно полуопределённой* или *неотрицательно определённой* (например, рисунок 1.5б). Здесь  $J(u)$  имеет множество точек локального минимума, которые образуют множество оптимума  $\{u_*\}$ ;

- если  $s_{conv} < 0$ , то квадратичную функцию называют *строго вогнутой*, а матрицу  $A$  — *отрицательно определённой* (рисунок 1.5в). Здесь  $J(u)$  не имеет минимума, а имеет единственную точку максимум. При нестрогом неравенстве  $s_{conv} \leq 0$  функцию называют, просто, *вогнутой*;

- если  $s_{conv}$  не определено, т.е. для одних  $u$  оно положительно, а для других отрицательно, то квадратичная функция является *неопределённой выпуклости*, а матрица  $A$  — *неопределённой*. Например, это может быть *седловая функция*, представленная на рисунке 1.5г. Здесь  $J(u)$  не имеет ни минимума, ни максимума, хотя в центре седла выполняется НУО  $\|\nabla J(u)\| = 0$ , но такое  $u \neq u_*$ .



а — строго выпуклая функция;

б — нестрого выпуклая функция;

в — строго вогнутая функция;

г — седловая функция с неопределённой выпуклостью

Рисунок 1.5 — Примеры квадратичных функций

Выпуклость необязательно связана с квадратичностью. Формально *выпуклость* определяется следующим образом. Функция  $J(u)$  выпуклая, если для любых двух управлений  $u_a, u_b \in E^n$  и числа  $0 \leq \beta \leq 1$

$$J(\beta u_a + (1 - \beta)u_b) \leq \beta J(u_a) + (1 - \beta)J(u_b).$$

Строгая выпуклость подразумевает строгое неравенство. Для одномерных функций геометрический смысл выпуклости означает, что

график функции между любыми двумя значениями управления  $u_a, u_b \in E^2$  лежит не выше прямой, соединяющей точки значений функции  $J(u_a)$  и  $J(u_b)$ . При строгой выпуклости график функции даже не будет касаться этой прямой внутри отрезка  $[u_a, u_b]$ .

Существует также термин *сильная выпуклость*. Он означает, что график функции между любыми точками  $u_a$  и  $u_b$  лежит не выше некоторой выпуклой параболы, соединяющей  $J(u_a)$  и  $J(u_b)$ . Очевидно, что сильно выпуклая функция является и строго выпуклой. Всякая квадратичная, строго выпуклая функция является и сильно выпуклой.

Итак, строго выпуклая функция может иметь только одну точку оптимума  $u_*$ . Нестрого выпуклая функция может иметь множество оптимумов. На рисунке 1.5б такое оптимальное множество  $\{u_*\}$  представляет собой линию параллельную оси  $u_1$ .

Квадратичные функции практически всегда *овражные* (плохо обусловленные, *жесткие*<sup>1</sup>), что объясняется разной чувствительностью функции  $J(u)$  к разным компонентам управления  $u_i, i = 1 \dots n$ . На рисунке 1.6а в качестве примера изображена овражная поверхность двумерной квадратичной функции

$$J(u) = \frac{1}{2} \langle Au, u \rangle_{E^2}, \quad A = \begin{pmatrix} 5 & -20 \\ 30 & 10 \end{pmatrix}, \quad (1.27)$$

или проще,

$$J(u) = 2.5u_1^2 + 5u_1u_2 + 5u_2^2$$

с минимумом в точке  $u_* = (0,0)$ . Справа, на рисунке 1.6б, показаны линии постоянного уровня этой функции. Видно, что дно оврага прямое и проходит примерно посередине второго и четвертого квадранта. Овраг двумерной квадратичной функции всегда прямой.

<sup>1</sup> Жесткий характер поведения функции — это её резкое возрастание по одним направлениям и слабое изменение по другим направлениям. Этот термин возник в конечномерных задачах оптимизации с жесткими системами обыкновенных дифференциальных уравнений [69].

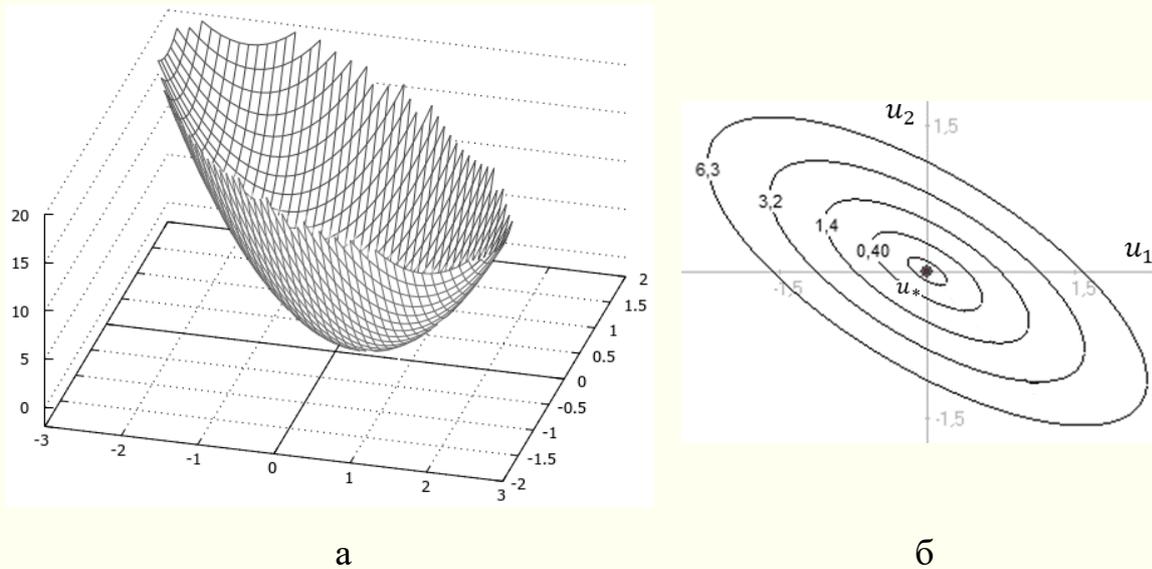


Рисунок 1.6 — Овражная квадратичная функция

Если целевая функция не квадратичная и не выпуклая, то овраг может иметь самые причудливые формы. На рисунке 1.7 изображён пример широко известной тестовой двумерной функции Розенброка (см., например, [2]):

$$J(u) = (u_1 - 1)^2 + 100(u_2 - u_1^2)^2. \quad (1.28)$$

Справа показаны линии уровня этой функции. Как мы видим, функция Розенброка имеет изогнутое дно оврага с точкой минимума в «неожиданном» месте  $u_* = (1,1)$ .

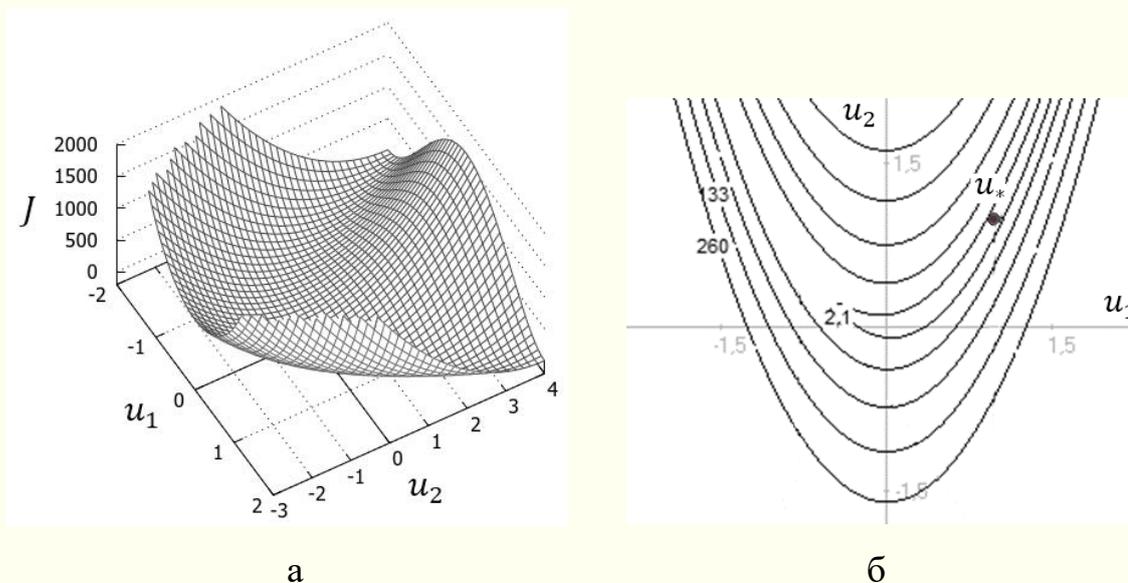


Рисунок 1.7 — Овражная неквадратичная функция Розенброка

Именно наличие оврагов и их форма может существенно осложнять поиск оптимума. Если искать минимум на овражной поверхности, двигаясь по дну оврага, то можно путешествовать очень долго, особенно в задачах высокой размерности. Здесь, приближаясь к минимуму дна в плоскости пары наиболее влиятельных компонент на значение  $J(u)$ , перед нами «раскрывается портал» в овраг относительно новой пары компонент, ставшей более влиятельной чем предыдущие. Мы попадаем в овраг в новом измерении. При этом, путешествуя в новых измерениях, мы можем незначительно перемещаться и в старом измерении, или даже когда-нибудь возвращаться в него полностью, но на новом уровне  $J$ . Это можно считать художественным сравнением с червоточинами вселенной и выходом в новые измерения. Как нам кажется, оно наглядно иллюстрирует суть многомерной овражности.

Невыпуклые функции могут иметь множество отдельных локальных минимумов. Так, например, на рисунке 1.8 показана тестовая функция Химмельблау 28 [97]:

$$J(u) = (u_1^2 + u_2 - 11)^2 + (u_1 + u_2^2 - 7)^2,$$

которая имеет четыре локальных минимума в точках, приблизительно равных  $(3.0, 2.0)$ ,  $(-2.8, 3.1)$ ,  $(-3.8, -3.3)$ ,  $(3.6, -1.8)$  и принимает в них одно и то же нулевое значение. Поэтому каждую из этих точек можно считать оптимальным управлением  $u_*$  с минимальным значением  $J(u_*) = 0$ .

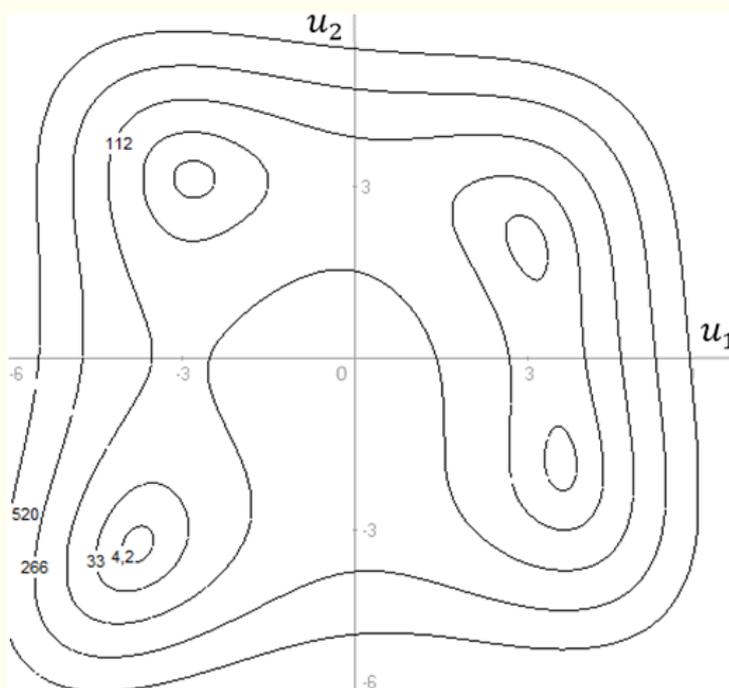


Рисунок 1.8 — Невыпуклая функция Химмельблау 28

Если говорить о функционалах, то *квадратичный функционал* определяется выражением:

$$J(u) = \Phi(u, u) + l(u) \in E,$$

где  $\Phi(u, u)$  — *билинейный функционал*;  $l(u)$  — *линейный функционал*. Бilinearность означает, что функционал  $\Phi(u, u)$  — линейный относительно первого и второго аргумента. Примером квадратичного функционала является скалярное произведение в  $J(u) = \langle u, u \rangle_{L_2}$ .

Квадратичный функционал можно представить в *квадратичной форме* аналогично (1.26), используя следующую *билинейную форму*:

$$J(u) = \frac{1}{2} \langle \mathcal{A}u, u \rangle_{L_2(S)}, \quad (1.29)$$

где  $\mathcal{A}$  — *линейный оператор*, действующий из  $L_2(S)$  в  $L_2(S)$ . Пусть далее он будет *положительным*<sup>1</sup> и *симметричным*<sup>2</sup>. Такой функционал  $J(u)$  является *строго выпуклым*, следовательно, имеет единственную точку минимума. *Градиент квадратичного функционала* (1.29) — это  $\nabla J = \mathcal{A}u$ , а *вторая производная квадратичного функционала* — это  $\nabla^2 J = \mathcal{A}$ .

В прикладных бесконечномерных задачах оптимизации распределённых систем целевой функционал  $J$  практически невозможно выразить через оператор  $\mathcal{A}$  и точно определить квадратичность или степень *выпуклости функционала*<sup>3</sup>  $J(u)$ .

О квадратичности и выпуклости функционалов обычно говорят в теоретических исследованиях для обоснования сходимости и оценки сравнительной эффективности того или иного экстремального метода.

Об овражности функционалов уже никто не говорит, поскольку здесь будет бесконечно много изогнутых в разных измерениях оврагов (измерений здесь бесконечно много) и сложность оптимизации такого функционала вырастет до бесконечности. Проще говоря, такую задачу вряд ли можно будет решить с заранее заданной точностью известными на сегодня вычислительными алгоритмами и машинами.

В бесконечномерных задачах оптимизации с распределёнными системами, как правило, предполагают из физических соображений

<sup>1</sup>  $\langle \mathcal{A}u, u \rangle_{L_2(S)} > 0, \forall u \neq 0$ .

<sup>2</sup>  $\mathcal{A}u = u\mathcal{A}$ .

<sup>3</sup> Формально выпуклость функционала определяется следующим образом. Функционал  $J(u)$  выпуклый, если для любых двух управлений  $u_a, u_b \in L_2$  и числа  $0 \leq \beta \leq 1$

$$J(\beta u_a + (1 - \beta)u_b) \leq \beta J(u_a) + (1 - \beta)J(u_b).$$

Строгая выпуклость подразумевает строгое неравенство.

существование только одного локального минимума, который, естественно, является и глобальным. Именно на его поиск направляются все усилия, и именно в этом минимуме получают оптимальное управление. Обсуждение и решение задач глобальной оптимизации распределённых систем с несколькими локальными минимумами, практически, отсутствует. Тем не менее, в данной книге имеется пример одной такой задачи [85]. Она рассмотрена и решена нами в разделе 9.

Далее в экстремальных алгоритмах оптимизации распределённых систем мы будем полагаться на строгую выпуклость и локальную квадратичность целевой функции или функционала вблизи оптимума, если не подтверждено обратное. Дадим уточняющее определение таким функционалам и функциям.

**Определение 1.2.** Будем называть функционал (функцию)  $J(u)$  в некоторой окрестности строгого локального минимума *локально квадратичным*, если  $J(u)$  в этой окрестности квадратичный, настолько, что можно применять методы оптимизации, существенно использующие квадратичность.

### 1.3.7. Ограниченность области управления

До сих пор управление  $u$ , вектор или функция, принадлежало множеству  $U$ , совпадающему со всем своим пространством, т.е.  $U = E^n$  или  $U = L_2$ . По техническим причинам не всякое оптимальное управление, полученное в неограниченных условиях, может быть реализовано на практике. Например, механизмы управления не могут слишком быстро изменять скорость управления, или не могут иметь любое положение, или они должны обеспечивать заданную суточную норму управления и т.п.

Кроме этого, управление может быть ограничено и допустимыми состояниями системы. Но этот тип неявных ограничений на управление

мы сейчас обсуждать не будем. Он подробно рассматривается в разделе 3.6 «Управление с ограничениями».

Любые *ограничения* на процесс управления, в явном или неявном виде, формируют множество *допустимых управлений*  $U$ , не совпадающее со всем пространством управлений, т.е.  $U \subset E^n$  или  $U \subset L_2$ . И если применяемые ограничения являются *активными*, то начиная с внутренней точки управления  $u^0 \in \text{int } U$ , мы придём к *краевому минимуму*  $J(u)$ , где оптимальное управление  $u_* \in \partial U$ , т.е. будет находиться на границе допустимого множества.

При оптимизации с ограничениями потенциально возникает проблема неоднозначности решения экстремальной задачи, для которой на самом деле существует единственное оптимальное решение. Эта проблема связана с возможным отсутствием *выпуклости*<sup>1</sup> множества  $U$  и произвольностью выбора в нём начального приближения  $u^0$ .

На рисунке 1.9, на примере двумерной квадратичной экстремальной задачи с целевой функцией (1.27) и *невыпуклым* допустимым множеством  $U$  (серая область), показаны две возможные траектории спуска при минимизации  $J(u)$ . Спуск, например, градиентными методами, из разных начальных точек  $u^0$  заканчивается в разных граничных точках множества  $U$  с небольшим скольжением по границе  $\partial U$ . Оба скольжения вдоль  $\partial U$  не могут привести к слиянию конечных точек, поскольку это потребует возрастания целевой функции. Получено два «оптимальных» управления. При этом левое решение минимизирует функцию до значения  $J = 0.4$ , а правое — до значения  $J = 3.2$ . Поэтому только левое решение следует считать оптимальным  $u_*$  в рассматриваемой задаче ограниченного управления.

---

<sup>1</sup> Множество  $U$  выпуклое если все точки отрезка (в соответствующем пространстве), образуемого любыми двумя точками множества  $U$ , также принадлежат множеству  $U$ . На рисунке 1.9 отрезок между двумя точками  $u^0$  не будет полностью лежать в сером множестве  $U$ , следовательно данное множество невыпуклое.

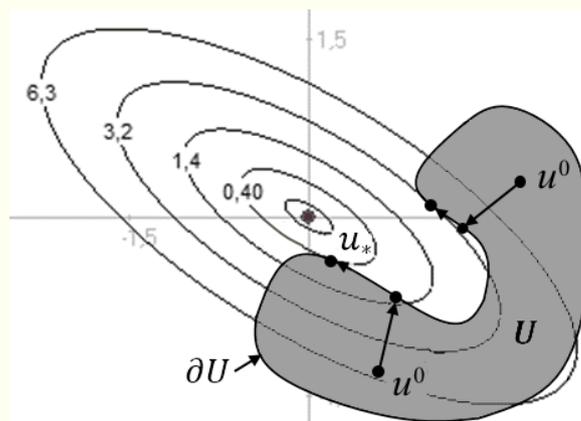


Рисунок 1.9 — Минимизация на невыпуклом множестве  $U \subset E^2$

Этот пример показывает возможную неразрешимость даже в случае выпуклой целевой функции в задаче оптимизации на невыпуклом множестве управлений. Целевую функцию (функционал)  $J(u)$  в любом случае можно минимизировать, но будет ли достигнутый краевой минимум оптимальным решением? Этот вопрос в общем случае не имеет ответа. Именно поэтому при теоретическом обсуждении вопроса достижимости оптимума в задачах с ограничениями всегда рассматривают случай *выпуклой оптимизации*, когда и функция (функционал)  $J(u)$ , и множество  $U$  — выпуклые.

В конечномерном выпуклом программировании известна следующая теорема (например, [49]). Если функция  $J(u)$  строго выпукла на выпуклом множестве  $U \subset E^n$  и точка  $u_* = \arg \min_U J(u)$ , то она единственна. То есть, для всех допустимых управлений  $u \in U$  таких, что  $u \neq u_*$  справедливо равенство  $J(u) > J(u_*)$ .

В заключении данного подраздела ещё раз напомним, что при оптимизации с ограничениями нельзя применять методы *линейного поиска* для выбора длины шага (шагового множителя  $b^k$ ) минимизации. Ограниченность управления сужает не только множество допустимых

стратегий выбора  $b^k$ , но сужает и множество допустимых методов минимизации. Например, здесь нельзя будет применять методы, использующие информацию о выпуклости  $J(u)$  с предыдущих итераций, например, метод сопряжённых градиентов, квазиньютоновские методы и т.п. Попадание на границу допустимого множества, со скольжением вдоль неё, сразу же сделает бесполезной накопленную ранее информацию о выпуклости  $J(u)$  при ранее неограниченном спуске.

## 2 Конечномерные параметры-управления в распределённых системах

### 2.1 Постановка прямой экстремальной задачи

В данном разделе будем рассматривать случаи конечномерной неограниченной оптимизации, когда управление  $u = \{u_i\}_{i=1}^n \in E^n$ . То есть, распределённое состояние системы  $v(\tau) \in L_2(\bar{\Omega})$  на пространственно-временной области  $\bar{\Omega}$  зависит, управляется *конечным* набором из  $n$  чисел, образующих ничем не неограниченный *вектор управления*  $u$ . Здесь целевой функционал  $J(u)$ , даже если и имеет вид некоторого интеграла, по своей сути является  $n$ -мерной неявно заданной *функцией* параметра  $u$ .

*Прямой экстремальный подход* (1.2) для решения задачи оптимизации (1.1) принимает конечномерной вид:

$$u_* = \arg \min J(u), \quad u \in E^n.$$

Методика решения такой задачи может оказаться относительно простой. Она полностью укладывается в рамки классических экстремальных алгоритмов.

Для реализации поставленной задачи привлечём конечномерный *экстремальный алгоритм* (1.4) в форме:

$$u_i^{k+1} = u_i^k + b^k p_i(u^k), \quad i = 1 \dots n, \quad k = 0, 1 \dots \quad (2.1)$$

Здесь *конечномерное направление минимизации*  $d^k = p^k \in E^n$ , т.е. совпадает с базовым направлением  $p^k$ , которое в той или иной форме содержит градиент  $\nabla J$ .

Для отыскания вектора *градиента*  $\nabla J = (\nabla_1 J, \dots, \nabla_n J)$ , где  $\nabla_i J(u) = \frac{\partial J(u)}{\partial u_i}$ ,  $i = 1, \dots, n$ , всегда задействуются уравнения исходной распределённой системы. В конечномерной задаче оптимизации

градиент можно находить либо аналитически, через решение сопряжённой задачи, либо посредством численного дифференцирования целевого функционала (функции).

Градиент  $\nabla J \in E^{n*}$ , т.е. он принадлежит пространству, сопряжённому с пространством управлений  $u \in E^n$ . Поскольку пространство  $E^n$  — евклидово, то сопряжённое пространство  $E^{n*}$  совпадает с исходным, оно *самосопряжённое*. Поэтому будем писать

$$\nabla J^k \stackrel{\text{def}}{=} \nabla J(u^k) \in E^n.$$

Далее, для иллюстрации сути конечномерных экстремальных алгоритмов, использующих градиент, рассмотрим кратко наиболее популярные, хорошо известные классические методы. Их понимание нам потребуется и для понимания особенностей алгоритмов бесконечномерной оптимизации.

В последующих демонстрационных тестовых расчётах будем задавать точность оптимизации критерием близости к оптимуму (1.23) с  $\varepsilon_3 = 0.01$ . Глубину шагов будем вычислялась по методу Носедала-Райта (*четвёртая стратегия*) со строгим условием Вольфе при максимальном шаге длиной  $\sqrt{n}$ . Параметры метода  $c_1 = 10^{-4}$ , а  $c_2$  в разных ситуациях будем задавать разным.

## 2.2 Градиентные методы

Зададим направление минимизации  $d^k = p^k = -\nabla J^k$ , т.е. спуск к минимуму будем осуществлять по *антиградиенту*. При этом экстремальный алгоритм (2.1) примет вид *градиентного метода*:

$$u_i^{k+1} = u_i^k - b^k \nabla_i J(u^k), \quad i = 1 \dots n, \quad k = 0, 1 \dots \quad (2.2)$$

Большое разнообразие градиентных методов различается только стратегиями выбора шагового множителя  $b^k$  [35]. Если использовать

третью или четвёртую стратегию линейного поиска  $b^k$  с условием полной релаксации (1.17)

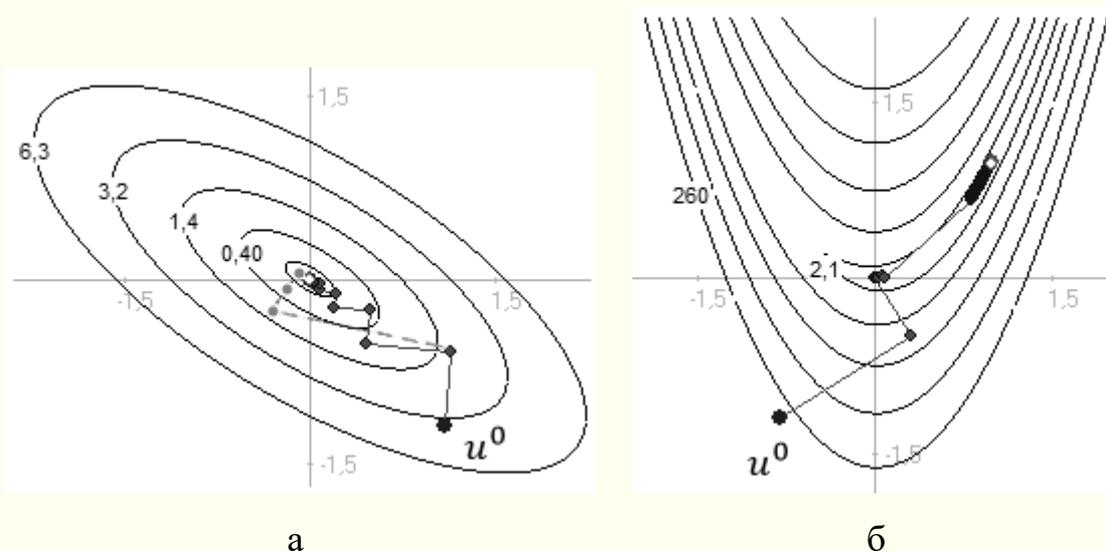
$$b^k = \arg \min_{b>0} J(u^k - b\nabla J^k),$$

то конечномерный метод (2.2) превращается в *градиентный метод наискорейшего спуска* (МНС).

Этот метод — *первого порядка*. Сильно выпуклые функции он минимизирует с *линейной скоростью сходимости*. Здесь

$$\|u^k - u_*\| \leq q \|u^{k-1} - u_*\|, \quad 0 < q < 1.$$

В качестве примера рассмотрим оптимизацию посредством МНС ранее обсуждаемых тестовых двумерных функций (1.27), (1.28). На рисунке 2.1а для первой функции (1.27) показано приближение к оптимуму  $u_* = (0,0)$  из начальной точки  $u^0 = (1.1, -1.2)$ . А для второй функции (1.28) на рисунке 2.1б показано приближение к  $u_* = (1,1)$  из  $u^0 = (-0.8, -1.2)$ . Траектории приближения изображены чёрными сплошными линиями.



а — для функции (1.27) — итераций 12, вычислений 50;  
 б — для функции (1.28) — итераций 379, вычислений 1595  
 Рисунок 2.1 — Минимизация посредством МНС

С точки зрения эффективности метода оптимизации в задачах с распределёнными параметрами, как мы уже писали, важно не столько количество итераций, сколько количество вычислений  $J, \nabla J$ . Поэтому в подписях к рисункам мы также приводим и общее количество таких вычислений.

Из рисунков видно, что МНС сначала ведёт ко дну *оврага* функции, а затем с ортогональными осцилляциями (особенно видно на квадратичной функции) ведёт вдоль дна оврага к минимуму. Эта классическая ситуация характерна для высокой точности выполнения условия (1.17) одномерной минимизации  $J^k(b)$ , что в примере было реализовано достаточно малым значением коэффициента  $c_2=0.01$  в методе Носедала-Райта.

При неточном решении задачи одномерной минимизации, картина спуска может существенно изменяться и общее количество вычислений может заметно снижаться. Так, на рисунке 2.1а серая пунктирная линия

соответствует невысокой точности одномерной минимизации с коэффициентом  $c_2=0.9$ , который и рекомендуется брать для МНС [12]. Здесь было сделано всего 5 итераций и 13 вычислений целевой функции и градиента. Мы видим, что одномерную минимизацию  $J^k(b)$  в МНС не следует выполнять слишком точно.

Градиентный метод первого порядка можно превратить в метод *нулевого порядка*, если отказаться от аналитического вычисления первых производных. Можно аппроксимировать *градиент конечными разностями* через целевую функцию  $J(u)$ , например, [6, 12]:

$$\nabla_i J \approx \frac{J(u + he_i) - J(u)}{h}, \quad i = 1 \dots n, \quad (2.3)$$

где  $e_i$  — единичные векторы базиса в пространстве  $E^n$ ,  $h$  — малое положительно число.

### 2.3 Метод сопряжённых градиентов

В *методе сопряжённых градиентов* (МСГ) направление спуска  $d^k$  задаётся из условия *сопряжённости* с предыдущим направлением  $d^{k-1}$  относительно *матрицы Гессе*  $H^k$ , т.е. из условия  $(d^k)^T H^k d^{k-1} = 0$ . Данное условие может реализовываться несколькими способами, например, это может быть направление Флетчера-Ривса<sup>1</sup>:  $d^k = p^k = -\nabla J^k + \beta^k p^{k-1}$ , где  $\beta^k = \|\nabla J^k\|^2 / \|\nabla J^{k-1}\|^2$ ,  $\beta^0 = 0$ . Итерационный процесс (2.1) приобретает вид:

$$u_i^{k+1} = u_i^k + b^k p_i(u^k), \quad i = 1 \dots n, \quad k = 0, 1 \dots \quad (2.4)$$

Глубина каждого шага определяется линейным поиском  $b^k$  из решения задачи одномерной минимизации  $J^k(b)$  с полной релаксацией (1.17). Именно оптимальность глубины шага определяет число  $\beta^k$ , обеспечивающее сопряжённость направлений  $p^k$  и  $p^{k-1}$ .

<sup>1</sup> Считается стандартным алгоритмом для квадратичных функций с положительным гессианом.

Данный метод с ростом итераций требует периодической очистки направления  $p^k$  от накапливающихся вычислительных погрешностей посредством присваивания  $\beta^k = 0$ . Такие итерации называются *моментами обновления* метода. Здесь и далее для неквадратичных функций очистка делалась после  $3n$  итераций, а для квадратичных — после заметной неортогональности градиентов [12]:

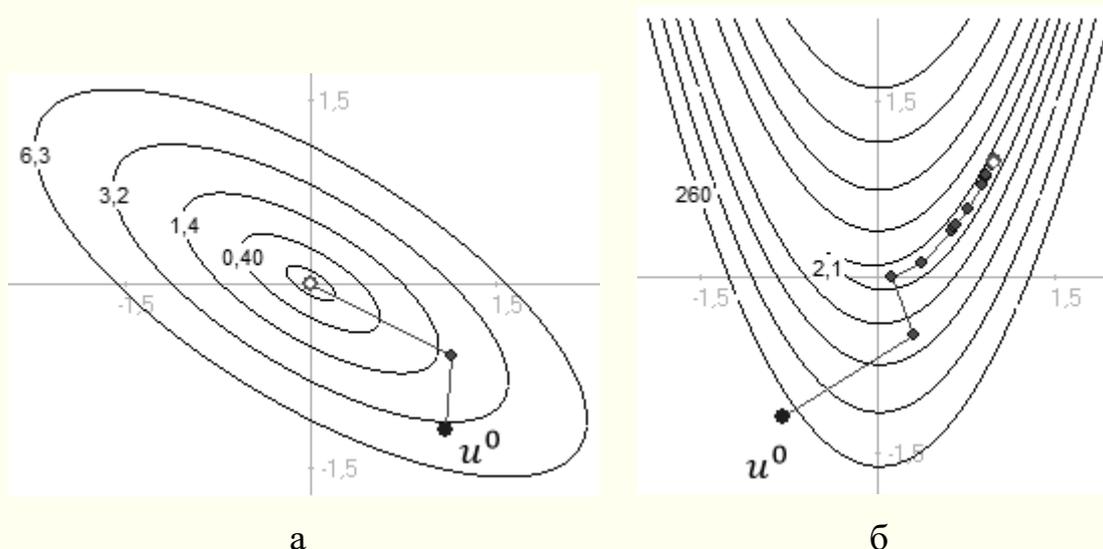
$$\frac{|\langle \nabla J^k, \nabla J^{k-1} \rangle|}{\|\nabla J^k\|^2} \geq 0.1.$$

МСГ — это метод *первого порядка*. Сильно выпуклые функции он минимизирует со *сверх линейной скоростью сходимости*:

$$\|u^k - u_*\| \leq q^k \|u^{k-1} - u_*\|, \quad \{q^k > 0\} \xrightarrow{k \rightarrow \infty} 0.$$

Для квадратичных функций имеется множество доказательств, что МСГ сходится к  $u_*$  не более чем за  $n$  шагов (если отсутствуют *помехи*).

На рисунке 2.2 для МСГ проиллюстрирован характер приближения к  $u_*$ . Глубина шагов вычислялась по методу Носедала-Райта при  $c_2 = 0.1$  [12], т.е. с достаточно высокой степенью релаксации шагов.



а — для функции (1.27) — итераций 2, вычислений 10;  
 б — для функции (1.28) — итераций 15, вычислений 82

Рисунок 2.2 — Минимизация посредством МСГ

Из рисунков видно, что МСГ сначала ведёт ко дну *оврага* функции, а затем без осцилляций, как это было в МНС, ведёт вдоль дна оврага к минимуму. Двумерную квадратичную функцию он минимизировал за два шага, что полностью согласуется с теоретическими положениями о сходимости метода. По сравнению с предыдущим МНС его достоинства очевидны.

Пожалуй, здесь так же будет уместно привести цитату [35]: «различные варианты метода сопряженных градиентов (а их насчитывается уже, как минимум, несколько десятков) работают существенно быстрее ускоренных градиентных методов с постоянными шагами и их адаптивных (универсальных) вариантов».

## 2.4 Метод Ньютона и квазиньютоновские методы

*Метод Ньютона* (МН) использует *вторые производные* целевой функции. Здесь направление спуска  $d^k = p^k = -(H^k)^{-1}\nabla J^k$ , где  $H^k \equiv \nabla^2 J^k$  — *матрица Гессе*. Итерационный процесс (2.1) принимает вид:

$$u_i^{k+1} = u_i^k + p_i(u^k), \quad i = 1 \dots n, \quad k = 0, 1 \dots \quad (2.5)$$

В данном методе шаговый множитель  $b^k = 1$ .

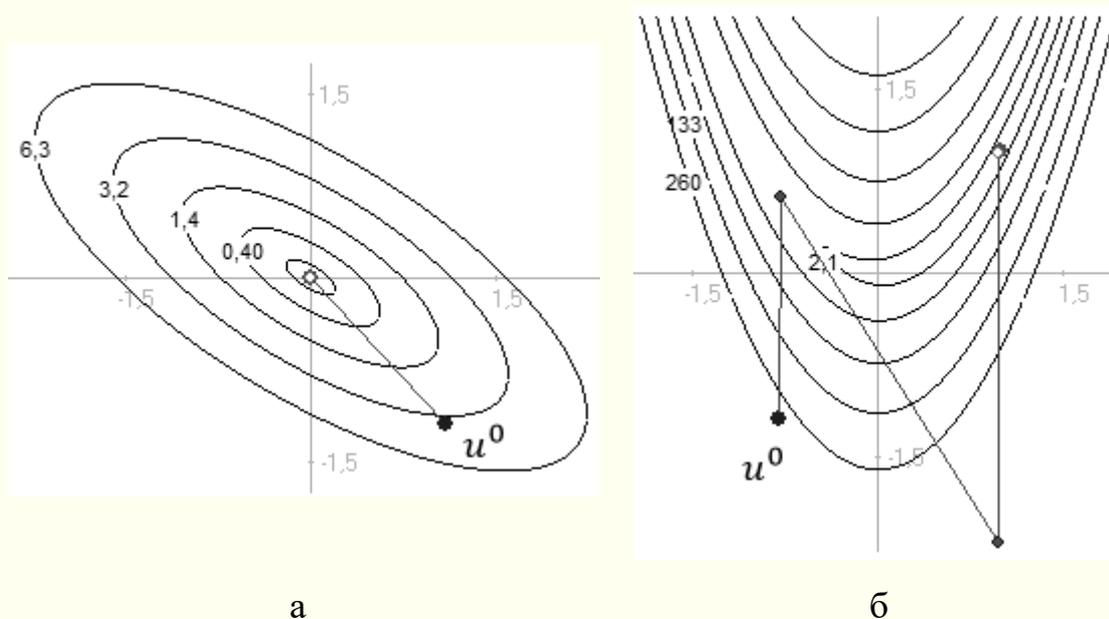
При минимизации неквадратичных функций в методе (2.5) может использоваться стратегия линейного поиска  $b^k$  или стратегия Trust-region. При приближении же к оптимуму возвращаются к  $b^k = 1$ . Такой метод называют *демпфированным* методом Ньютона или методом *Ньютона-Рафсона*.

МН — это метод *второго порядка*. Сильно выпуклые функции он минимизирует с *квадратичной скоростью сходимости*:

$$\|u^k - u_*\| \leq C \|u^{k-1} - u_*\|^2, \quad C > 0.$$

Для квадратичных функций любой размерности МН сходится к  $u_*$  за один шаг из любого начального приближения (если отсутствуют *помехи*).

На рисунке 2.3 проиллюстрирован характер приближений МН к  $u_*$ .



а — для функции (1.27) — количество итераций 1;  
 б — для функции (1.28) — количество итераций 3  
 Рисунок 2.3 — Минимизация методом Ньютона

Преимущества метода очевидны. Даже минимум сложной неквадратичной функции Розенброка (1.28) он достигает всего за три шага. Тем не менее, к недостаткам метода следует отнести то, что он требует вычисления матрицы вторых производных  $H^k$  на каждой итерации, и кроме того, при минимизации невыпуклых функций он ведёт к минимуму целевой функции только при условии  $H^k > 0$ , т.е. при *положительно определённой* матрице Гессе.

Необходимо отметить, что вместо непосредственного вычисления обратной матрицы Гессе и произведения  $(H^k)^{-1}\nabla J^k$ , для определения направления спуска  $p^k$  в МН (2.5), обычно используют *метод Гаусса* для решения системы линейных уравнений  $H^k p^k = -\nabla J^k$  относительно вектора  $p^k$ . Работа с матрицами, реализация метода Гаусса требуют относительно больших вычислительных затрат как по оперативной памяти компьютера, так и по времени вычислений.

От вычисления вторых производных можно отказаться. Например, аппроксимировать *гессиан конечными разностями* через градиент целевой функции [6, 12, 13, 66]:

$$H^k e_i \approx \frac{\nabla J(u^k + h e_i) - \nabla J(u^k)}{h},$$

где  $e_i$  — орты пространства управлений,  $h$  — шаг численного дифференцирования. В этом случае, как показали расчёты, минимизация рассматриваемых функций имеет тот же вид, что и на рисунке 2.3, а количество вычислений функций и градиентов составило для функций (1.27) и (1.28) соответственно 3 и 11. Такой *конечно-разностный метод Ньютона* (МНкр) превращается в метод *первого порядка*, практически сохраняя квадратичную скорость сходимости.

Если матрицу Гессе вычислять приблизительно через информацию о кривизне целевой функции, взятую с предыдущих итераций, то мы получим так называемые *квазиньютоновские методы* [9, 6, 12, 97] *первого порядка*. Такие методы обладают сверхлинейной скоростью сходимости немного более высокой чем у МСГ. Обратная матрица Гессе  $H^k$  здесь аппроксимируется некоторой матрицей  $C^k$ , например, в казиньютоновском методе BFGS (Бройдена-Флетчера-Гольдфарба-Шанно) для нового шага

$$C^{k+1} = (I - \rho_k s^k (y^k)^T) C^k (I - \rho_k y^k (s^k)^T) + \rho_k s^k (s^k)^T,$$

где  $I$  — единичная матрица,  $\rho_k = \frac{1}{(y^k)^T s^k}$ ,  $s^k = u^{k+1} - u^k$ ,  $y^k = \nabla J^{k+1} - \nabla J^k$ . В результате чего получаем направление минимизации  $p^k = -C^k \nabla J^k$ ,  $p^0 = -\nabla J^0$ , а вместо алгоритма (2.5) получаем (в свёрнутом виде):

$$u^{k+1} = u^k + b^k p^k, \quad k = 0, 1 \dots$$

Шаговый множитель вычисляют одномерной минимизации  $J^k(b)$  с полной релаксацией. Далее в тестах мы использовали метод Носедала-Райта при  $c_2 = 0.9$  [12].

Характер спуска квазиньютоновских методов похож на МСГ (рисунок 2.2), однако количество вычислений оказывается меньшим. Так, метод BFGS для функции Розенброка потребовал 52 вычисления функции и градиента, вместо 82 как это было в МСГ.

Метод BFGS с ростом размерности задачи из-за большого количества операций с матрицами и векторами становится весьма затратным, в первую очередь с точки зрения процессорного времени. Так, например, для квадратичной функции «повёрнутый эллипсоид»  $J(u) = \sum_{i=1}^n (\sum_{j=1}^i u_j)^2$ ,  $n = 10^3$ ,  $u^0 = -1$  он потребовал более пяти минут работы на компьютере с индексом производительности Windows равным 9.1. При этом отсутствовали решения задач с распределёнными параметрами.

Существуют более экономичные квазиньютоновские методы немного уступающие по количеству итераций методу BFGS, например, *Limited-memory* BFGS (L-BFGS) [9, 12]. Если метод BFGS хранит полную  $n \times n$  матрицу  $C^k$ , то L-BFGS хранит  $m$  пар векторов ( $m < n$ ) последних значений  $\{s^k, y^k\}$ , через которые и производятся упрощённые вычисления, обеспечивая неявную работу с некоторым приближением обратной матрицы Гессе. На первых  $k \leq m$  итерациях метод L-BFGS совпадает с BFGS.

Обычно в методе L-BFGS задают память  $m$  в диапазоне от 2 до 15. Данный метод использует компьютерные ресурсы существенно экономичнее, чем BFGS. При этом количество вычислений может отличаться от BFGS незначительно. Так, для функции Розенброка при

$m = 5$  потребовалось вычислений функции и градиента 50, вместо 52 при BFGS.

Можно также избежать явного вычисления матрицы Гессе посредством усечённых методов Ньютона (truncated Newton methods), например, *неточным методом Ньютона* (inexact Newton method) с сопряжёнными градиентами [9, 12, 29] — МНсг. Здесь решение системы линейных алгебраических уравнений  $H^k p^k = -\nabla J^k$  ищется не методом Гаусса, а посредством МСГ. Систему записывают в виде невязки

$$r(p) = H^k p + \nabla J^k \rightarrow 0 \in E^n,$$

которую можно трактовать как градиент некоторой квадратичной функции  $F(p)$ , где  $\nabla F(p) \stackrel{\text{def}}{=} r(p)$ . Экстремум функции  $F$  при НУО  $\|\nabla F\| = 0$  даёт корни системы уравнений  $r(p) = 0$ , т.е. эти задачи эквивалентны:

$$F(p) \rightarrow \text{extr} \Leftrightarrow \|r(p)\| \rightarrow 0.$$

Итерационное решение, точнее подитерации на каждой итерации  $k$ , посредством МСГ квадратичной задачи  $F(p) \rightarrow \text{extr}$  даёт направление Ньютона  $p = p^k$ .

Критерием завершения подитераций в МНсг принимают условие

$$\|r(p^k)\| \leq \eta^k \|\nabla J(u^k)\|,$$

где  $\eta^k = \min\left(c, \sqrt{\|\nabla J^k\|}\right)$  — коэффициент усиления сходимости [12, 9],  $0 < c < 1$ . Кроме этого, рекомендуется контролировать максимально допустимое количество подитераций. Именно поэтому, когда количество подитераций ограничивается (усекается) и при этом не всегда достигается точность метода Ньютона, подобные методы относят к семейству *truncated Newton method*.

Минимизация функции Розенброка (1.28) посредством МНсг имеет тот же вид, что и классический МН на рисунке 2.3, но вместо трёхкратного вычисления матрицы вторых производных, МНсг

потребовал 11 вычислений функции и градиента. Данный метод продемонстрировал хорошую эффективность, аналогичную МНкр, но при этом не использовал матрицы.

Подведём итог и дадим некоторые рекомендации. При решении задач оптимизации распределённых систем с конечномерным управлением  $u \in E^n$ , целесообразно использовать или метод сопряжённых градиентов (МСГ), легко программируемый и эффективный при небольшой размерности  $n$  (единицы, десятки), или применять неточный метод Ньютона (МНсг), если размерность  $n$  велика (сотни, тысячи).

## 2.5 Покомпонентные условия оптимальности

В классической теории экстремальных задач для реализации НУО (1.11) в пространстве  $E^n$  рассматривают такие последовательности управлений  $u^k \xrightarrow{k \rightarrow \infty} u_*$ , которые в сопряжённом пространстве обеспечивают сходимость к нулю нормы градиентов:

$$\|\nabla J(u^k)\| \xrightarrow{k \rightarrow \infty} 0. \quad (2.6)$$

Мы будем рассматривать последовательности управлений, которые обеспечивают покомпонентную сходимость градиентов, когда все компоненты  $\nabla_i J(u^k)$  изменяются одинаково *пропорционально* для всех  $i = 1 \dots n$  от итерации к итерации. Такая сходимость имеет место, когда все векторы  $\nabla J(u^k)$  при стремлении  $u^k \xrightarrow{k \rightarrow \infty} u_*$  являются *коллинеарными*<sup>1</sup>.

Впервые пропорциональная покомпонентная сходимость градиентов, для условий оптимальности в конечномерных пространствах, была представлена автором в 2000-м году на конференции [18], а в бесконечномерных пространствах — в работе [19].

---

<sup>1</sup> Два ненулевых вектора называются коллинеарными, если они лежат на параллельных прямых или на одной прямой.

Суть сходимости с коллинеарными градиентами для строго выпуклой квадратичной функции  $J(u)$ ,  $u \in E^2$  иллюстрируется на рисунке 2.4 на фоне линий уровня. Поскольку пространство, сопряжённое к евклидову пространству, совпадает с исходным, то на рисунке градиенты  $\nabla J$  нарисованы в той же плоскости, что и линии уровня. Обращаем ваше внимание, что векторы  $\nabla J$ , по своему определению, всегда расположены ортогонально линиям уровня (точнее касательным к линиям уровня) в точках своего определения.

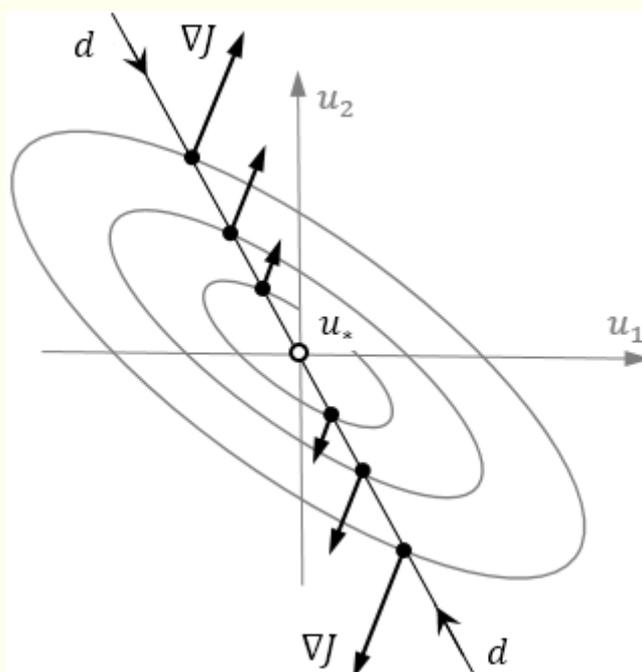


Рисунок 2.4 — Демонстрация коллинеарности градиентов

Мы видим, что векторы  $\nabla J$  строго выпуклой квадратичной функции коллинеарны друг другу во всех точках, лежащих на направлении  $d$ . Это направление — *направление на оптимум*, оно проходит через  $u_*$ .

В работе [17] условия сходимости к оптимуму с коллинеарными градиентами описаны соответствующей теоремой. Приведём её вместе с

доказательством и соответствующим определением, что будет полезным для дальнейшего понимания рассматриваемой ситуации.

**Определение 2.1.** Вектор  $\nabla J(u^k) \in E^n$  изменяется от шага  $k$  к шагу  $k + 1$  коллинеарно, если компоненты данного вектора изменяются пропорционально:

$$\frac{\nabla_i J(u^{k+1})}{\nabla_i J(u^k)} = \text{const}^k, \quad i = 1 \dots n, \quad 0 < \text{const}^k < \infty, \quad k \geq 0.$$

Если  $\nabla_i J(u^k) = 0$  для какого-либо  $i$ , то необходимо  $\nabla_i J(u^{k+1}) = 0$ . При этом отношение нулей следует считать равным  $\text{const}^k$ .

**Теорема 2.1.** Если функция  $J(u)$ ,  $u \in E^n$  квадратичная и строго выпуклая, то для сходимости к оптимуму  $u_*$  необходимо и достаточно, чтобы последовательность векторов

$$\nabla J(u^k) \xrightarrow{k \rightarrow \infty} 0 \text{ коллинеарно.} \quad (2.7)$$

**Доказательство необходимости.** Представим квадратичную целевую функцию в виде

$$J(u) = \frac{1}{2} \langle Au, u \rangle,$$

где  $A$  — положительно определённая симметричная матрица  $n \times n$  действительных чисел. Такая матрица обеспечивает строгую выпуклость целевой функции, т.е.  $J(u)$  имеет единственную точку минимума  $u_*$ . Градиент такой функции  $\nabla J(u) = Au$ , который в точке минимума обращается в ноль, т.е.  $Au_* = 0$ .

Зададим в точке  $u^k$  направление на оптимум  $u_*$  с помощью вектора  $d^k = u_* - u^k$ . Последовательность шагов  $u_* - u^k \xrightarrow{k \rightarrow \infty} 0$ , лежащих на направлении таких векторов, является множеством коллинеарных векторов. Умножив данную коллинеарную последовательность на матрицу  $-A$ , мы получим:

$$-A(u_* - u^k) = Au^k - Au_* = \nabla J(u^k) \xrightarrow{k \rightarrow \infty} 0 \text{ коллинеарно.}$$

То есть при  $u^k \xrightarrow{k \rightarrow \infty} u_*$  в направлении  $d$  необходимо выполняется условие (2.7). Понятно, что  $d$  задаёт *направление кратчайшего пути* из любой начальной точки  $u^0$  в оптимум  $u_*$ . Длина этого пути равна расстоянию  $\|u_* - u^0\|$ .

**Доказательство достаточности.** Направление  $d$  является единственным местом определения коллинеарных градиентов. Действительно, если последовательность  $\nabla J(u^k) \xrightarrow{k \rightarrow \infty} 0$  коллинеарно, то умножив её на  $-A^{-1}$  (обратная матрица существует, т.к.  $A > 0$ ), получаем:

$$-A^{-1}\nabla J(u^k) = -A^{-1}(Au^k - Au_*) = u_* - u^k \xrightarrow{k \rightarrow \infty} 0 \text{ коллинеарно.}$$

Следовательно, последовательность коллинеарных векторов  $d^k = u_* - u^k$ , входящих в одну точку  $u_*$ , будет лежать на одной прямой, являющейся направлением  $d$ . Это означает, что если  $J(u)$  строго выпуклая, квадратичная и выполняется (2.7), то этого достаточно, чтобы для произвольной точки  $u^k$  существовало единственное направление  $d$  на оптимум  $u_*$ .

**Конец доказательства.** ■

Если учесть, что  $A$  — это гессиан  $H$ , то указанную чуть выше операцию слева можно представить как  $-H^{-1}\nabla J = p$ , где  $p$  — направление минимизации в МН, которое, как известно, указывает на  $u_*$  строго выпуклой квадратичной функции. С другой стороны (справа), мы имеем последовательность  $u_* - u^k$ , лежащую в направлении  $d$  на точку  $u_*$ . Поэтому из теоремы 2.1 вытекает следующее утверждение.

**Следствие 2.1.** Для строго выпуклой квадратичной функции  $J(u)$  условие оптимальности (2.7) в каждой точке  $u^k$  задаёт направление  $d^k$  на минимум, используя коллинеарные градиенты, также как МН задаёт направление  $p^k$ , используя вторые производные:

$$p^k \stackrel{\text{def}}{=} -(H^k)^{-1} \nabla J^k = b^k d^k,$$

где число  $b^k = \|(H^k)^{-1} \nabla J^k\| / \|d^k\|$  — шаговый множитель, выравнивающий длину векторов  $p^k$  и  $d^k$ .

Реализация классического НУО (2.6) может обеспечить сходимость к  $u_*$  всего лишь с уменьшением длины вектора  $\nabla J$ , игнорируя при этом характер изменения компонент вектора  $\nabla J$ , т.е., игнорируя направление приближения к  $u_*$ . В то же время, реализация НУО в форме (2.7) требует приближения к  $u_*$  с уменьшением длины вектора  $\nabla J$  по жёстко заданному направлению, а именно, по направлению коллинеарности градиентов, т.е. с пропорциональным уменьшением всех компонент вектора  $\nabla J$ . В частности, для квадратичных функций — это движение по прямой (кратчайшему пути) к оптимуму. В этом и есть принципиальное отличие обеих условий оптимальности (2.6) и (2.7).

## 2.6 Метод коллинеарных градиентов

Новые условия оптимальности в форме (2.7) теоремы 2.1 дают дополнительные возможности для конструирования эффективных экстремальных алгоритмов. Так, в работе [17] предложен и подробно описан *метод коллинеарных градиентов* (МКГ). Здесь мы вкратце остановимся на его сути. Данный метод не только показывает очень хорошие результаты конечномерной оптимизации, но и позволяет сделать обобщения на бесконечномерные задачи.

Начнём со строго выпуклой квадратичной целевой функции  $J(u)$ . Согласно теореме 2.1 имеет место утверждение: для двух любых коллинеарных градиентов строго выпуклой квадратичной функции может быть построено единственным образом направление  $d$ , проходящее через оптимум  $u_*$ .

Это означает, что в окрестности любого  $u^k$  на основе коллинеарных градиентов мы можем найти  $u^{k*}$  и построить направление  $d^k = (u^{k*} - u^k)$ , проходящее через  $u_*$  как показано на рисунке 2.5. Обратите внимание, что точек  $u^{k*}$  может быть две: передняя (относительно направления на минимум, когда  $d^k$  — направление минимизации, шаговой множитель должен быть  $b^k > 0$ ) и задняя (здесь получаемое  $d^k$  — направление максимизации, соответственно необходимо  $b^k < 0$ ).

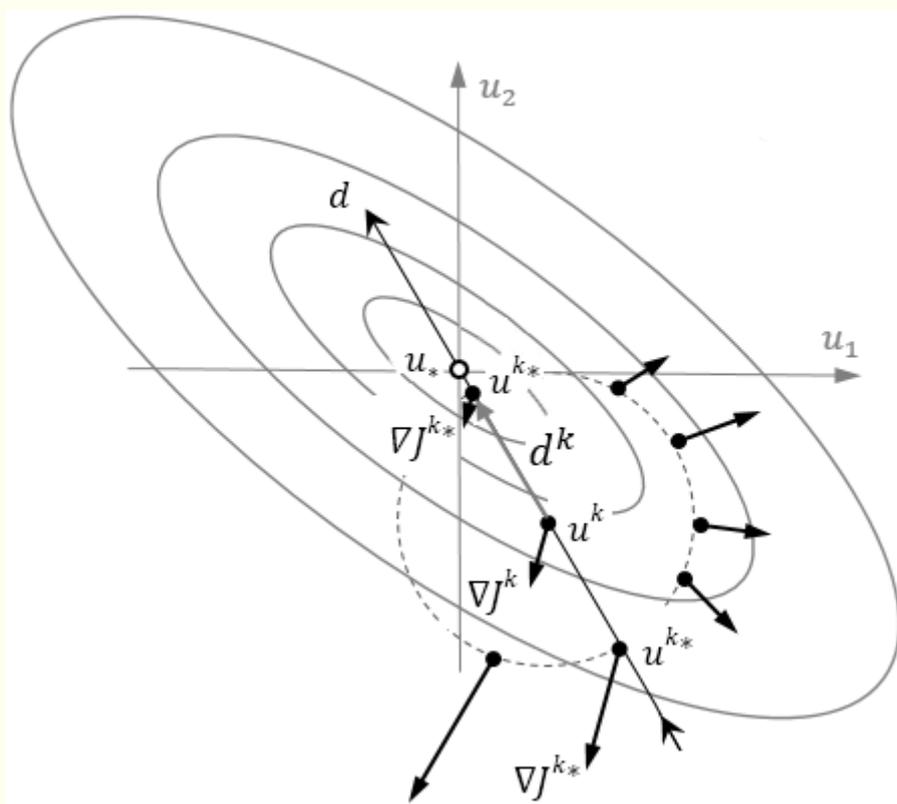


Рисунок 2.5 — Построение направления минимизации в окрестности (пунктирная линия) точки  $u^k$

Имея направление  $d^k$  можно за один шаг (*одношаговый МКГ*) минимизировать строго выпуклую квадратичную функцию из любого начального приближения  $u^0$ :

$$u_* = u^0 + b^0 d^0. \quad (2.8)$$

Если целевая функция не квадратичная, то МКГ можно использовать для определения  $d^k$  на каждой очередной итерации в предположении, что функция  $J(u)$  в каждой точке  $u^k$  *локально квадратичная*. Тогда вместо одношагового метода получим *многошаговый МКГ*:

$$u^{k+1} = u^k + b^k d^k, \quad k = 0, 1 \dots \quad (2.9)$$

Параметр шага  $b^k$  (включая  $k = 0$ ) легко находится из условия экстремума параболы  $J(b) = J(u^k + b d^k)$ . Для известных градиентов в точках  $u^{k*}$  и  $u^k$  мы получаем, *оптимальный шаговый множитель*:

$$b^k = \left( 1 - \frac{\langle \nabla J(u^{k*}), d^k \rangle}{\langle \nabla J(u^k), d^k \rangle} \right)^{-1}. \quad (2.10)$$

Знак и величина  $b^k$  полностью задаются *формулой шага* (2.10). Однако, при невыпуклой минимизации формула (2.10) может привести к максимуму  $J(b)$ , поскольку формула получена всего лишь из необходимого условия экстремума параболы. Чтобы такого не происходило мы должны уточнять направление шага  $b^k d^k$ . Возможно, полученное направление — это направление роста функции, вогнутой в окрестности  $u^k$ . То есть, следует контролировать направление минимизации по углу  $\gamma$  между векторами  $\nabla J^k$  и  $d^k$ . Если  $\cos \gamma = \frac{\langle \nabla J^k, d^k \rangle}{\|\nabla J(u^k)\| \|d^k\|} \geq 0$ , то  $d^k$  — это направление максимизации и в (2.9) следует брать  $b^k$  с обратным знаком.

Для шагового множителя (2.10), полученного из условия квадратичности функции  $J(u^k)$  в окрестности радиусом  $\|d^k\|$ , мы

получаем шаг МКГ, совпадающий с шагом метода Ньютона. При этом Следствие 2.1 принимает следующий вид.

**Следствие 2.2.** Для строго выпуклой квадратичной функции  $J(u)$  условие оптимальности (2.7) в каждой точке  $u^k$  задаёт направление шага  $b^k d^k$  на минимум, используя *коллинеарные градиенты*, также как МН задаёт направление  $p^k$ , используя *вторые производные*:

$$p^k \stackrel{\text{def}}{=} -(H^k)^{-1} \nabla J^k = b^k d^k,$$

где число  $b^k$  — шаговый множитель (2.10).

Остаётся открытым вопрос выбора точки  $u^{k*}$  для построения направления  $d^k = (u^{k*} - u^k)$ ,  $k \geq 0$ .

В соответствии с условием оптимальности (2.7) теоремы 2.1, *необходимо* на каждом шаге  $k$  находить такую точку  $u$  в окрестности  $u^k$ , где градиент  $\nabla J(u)$  будет коллинеарным к  $\nabla J(u^k)$ . Такое значение  $u \stackrel{\text{def}}{=} u^{k*}$  и на его основе может быть сделан очередной шаг  $k + 1$ . Если при этом функция  $J(u)$  квадратичная и строго выпуклая, то, согласно следствию 2.2, этого будет *достаточно* для достижения оптимума  $u_*$  за один шаг из  $u^k$  методом (2.8) с оптимальным шаговым множителем (2.10).

Степень коллинеарности можно оценивать по расхождению (*невязке*) ортов градиентов [18] в некоторой окрестности точки  $u^k$ :

$$r^k(u) \stackrel{\text{def}}{=} \frac{\nabla J(u)}{\|\nabla J(u)\|} s - \frac{\nabla J(u^k)}{\|\nabla J(u^k)\|}. \quad (2.11)$$

где  $s = 1$  если  $\text{sgn}\langle \nabla J(u), \nabla J(u^k) \rangle \geq 0$ , иначе  $s = -1$ . Знак  $s$  позволяет одинаково оценивать коллинеарность как сонаправленных, так и противоположно направленных градиентов.

Приведённая невязка (2.11) — это система  $n$  уравнений, корень которых  $u = u^{k*}$  мы должны найти:

$$r^k(u) = 0 \in E^n.$$

Подобные системы, обычно, решаются посредством МСГ с подитерациями  $u^{kl}$ ,  $l = 1, 2 \dots$ , как это было показано ранее при рассмотрении *неточного метода Ньютона* — МНсг. Подробное описание алгоритма подитераций МКГ можно найти в [17, 88].

При этом, МКГ имеет существенные отличия от МНсг:

- МНсг ищет направление  $p^k$ , когда  $p^k H^k = -\nabla J^k$ , а МКГ ищет направление  $d^k$ , где градиенты коллинеарные. Эти методы обнуляют разные невязки посредством подитераций МСГ;
- МНсг применяется в точке  $u^k$ , а МКГ — в относительно большой окрестности точки  $u^k$ ;
- МНсг при неположительной выпуклости  $J(u)$  вынужден делать антиградиентные шаги, в то время как МКГ всегда делает шаги по направлению с коллинеарными градиентами;
- МКГ не прибегает к выбору  $b^k$ , требующего затратных вычислений кроме простой формулы (2.10).

В качестве визуальной иллюстрации работы МКГ рассмотрим оптимизацию строго выпуклой квадратичной функции Schwefel 1.2 [15] «повёрнутый эллипсоид»:

$$J(u) = \sum_{i=1}^n \left( \sum_{j=1}^i u_j \right)^2, \quad n = 2, \quad u_* = (0, 0).$$

На рисунке 2.6 для случая трёх стартовых точек  $u^0$  (чёрные точки) показаны результаты одношаговой оптимизации алгоритмом (2.8), (2.10) с подитерациями  $u^{0l}$  (серые точки). Первый шаг подитераций  $u^{01} - u^0$  осуществляется в сторону роста функции *под углом*  $45^\circ$  ко всем осям координат из точки  $u^0$  в окрестности радиусом  $\delta = 0.5$  (пунктирные окружности). Такой шаг задавался по формуле

$$u_i^{01} = u_i^0 + \frac{\delta}{\sqrt{n}} \operatorname{sgn} \nabla_i J^0, \quad i = 1 \dots n.$$

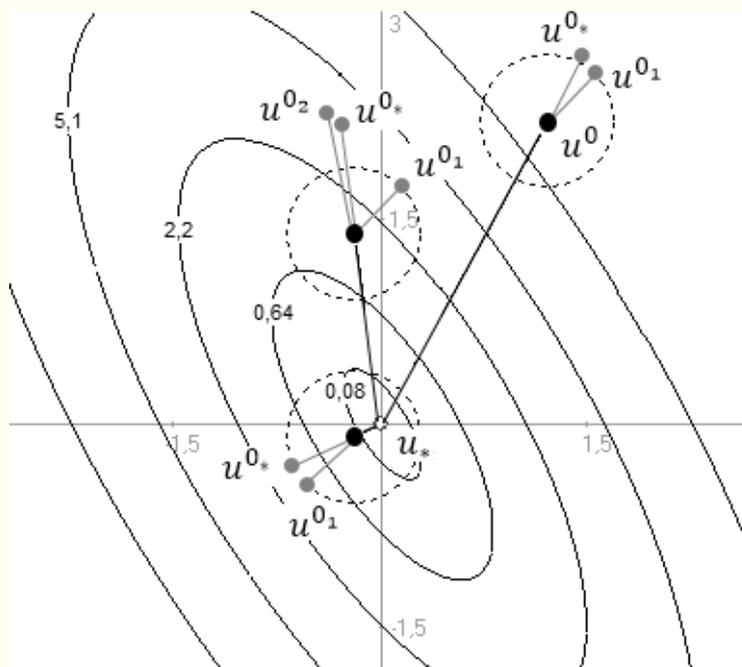


Рисунок 2.6 — Подитерации одношагового МКГ

Для всех начальных точек  $u^0$  было сделано не более трёх подитераций  $u^{0l}$  ( $1 \leq l \leq 3$ ), которые дали нам точки  $u^{0*}$  с направлениями максимизации  $d^0 = u^{0*} - u^0$ . Далее во всех случаях МКГ (2.8) сделал по одному шагу  $b^0 d^0$  в точку минимума  $u_*$  с параметрами  $b^0 < 0$ , полученным по формуле (2.10). Сделанные шаги соответствуют МН, что и следовало ожидать из утверждений следствия 2.2.

Таким образом, выполнение НУО теоремы 2.1 (градиенты в точках  $\nabla J(u^0)$  и  $\nabla J(u^{0*})$  коллинеарные) реализовало и достаточное условие оптимальности (направление шага  $b^0 d^0$  проходит через оптимум  $u_*$ ), что позволило нам решить поставленную задачу за одну итерацию, подобно МН.

В работах [17, 88] приведено обширное тестирование МКГ на невыпуклых целевых функциях разной размерности. Метод показал эффективность не хуже, а в некоторых случаях и лучше, чем МН, МНкр,

МНсг, L-BFGS, МСГ. Напомним, МКГ не использует вторые производные целевой функции, матрицу Гессе и ей подобные.

К вопросу о скорости сходимости МКГ необходимо отметить следующее. Во-первых, МКГ (2.8) имеет квадратичную скоростью сходимости, как и МН, что утверждается следствием 2.2 соответствующей теоремы и, что продемонстрировали тестовые расчёты. Во-вторых, для неквадратичных  $J(u)$  или при больших  $n$ , алгоритм приближительного вычисления коллинеарных градиентов из системы (2.11), превращает МКГ в приближительный метод Ньютона, относящийся к семейству *truncated Newton method*.

МКГ является очень экономичным по количеству вычислений  $J$  и  $\nabla J$ . Благодаря формуле (2.10), он не требует затратных вычислений шагового множителя  $b^k$  посредством линейного поиска (например, методом золотого сечения и т.п.).

## 2.7 Метод с регулируемым направлением спуска

Теперь рассмотрим некоторый «упрощённый» вариант МКГ. Поиск направления минимизации из условия обнуления невязки (2.11) на каждой итерации при большой размерности задачи может оказаться весьма затратной процедурой, а в бесконечномерном случае, на который мы хотим сделать обобщение, — вообще нереализуемой. Поэтому далее построим экстремальные алгоритмы на эвристических способах приближительной реализации НУО (2.7) теоремы 2.1.

Запишем *метод с регулируемым направлением спуска* — МРНС (1.9), где базовое направление  $p^k = -\nabla J^k$ , в конечномерной форме. Мы получим конечномерный метод с регулируемым направлением спуска *относительно градиента* — МРНСг:

$$u^{k+1} = u^k - b^k \overline{\alpha^k \nabla J^k}, \quad k = 0, 1, \dots, \quad (2.12)$$

или подробнее:

$$u_i^{k+1} = u_i^k - b^k \alpha_i^k \nabla_i J^k, \quad i = 1 \dots n, \quad k = 0, 1 \dots$$

Здесь  $b^k > 0$ .

Если в качестве базового направления  $p^k$  взять вектор сопряжённых градиентов, то мы получим метод с регулируемым направлением спуска относительно сопряжённых градиентов — МРНСг. Но сейчас мы ограничимся рассмотрением только МРНСг.

Давайте сравним приведённый МРНСг (2.12) с многошаговым МКГ (2.9) и выясним, что в них общего и в чём разница. Можно было бы сказать, что МКГ с алгоритмом  $u^{k+1} = u^k + b^k d^k$  (здесь направление минимизации  $d^k = (u^{k*} - u^k)$ ) совпадает с методом (2.12) (здесь направление минимизации  $d^k = -\overrightarrow{\alpha^k \nabla J^k}$ ), если параметр  $\alpha^k$  определить как:

$$\alpha_i^k = \frac{(u_i^{k*} - u_i^k)}{-\nabla_i J^k}, \quad i = 1 \dots n. \quad (2.13)$$

Легко заметить, что при этом, о совпадении алгоритмов можно говорить только в случае, если для всех  $i, k$  значение  $\nabla_i J^k$  не обращается в ноль.

Это означает, что МРНСг заведомо является ограниченным методом по сравнению с МКГ. Тем не менее, для конкретного круга задач (квадратичные высокой размерности) он оказывается очень эффективным и легко обобщаемым на бесконечномерные ситуации. Обсудим выбор параметра регулирования направления спуска  $\alpha^k$  в МРНСг (2.12).

Если допустить, что в какой-либо конкретной задаче  $\nabla_i J^k \neq 0 \forall i, k$  и параметр  $\alpha^k$  можно было бы вычислять по формуле (2.13), то всё равно остаётся открытым вопрос выбора *передней* (поскольку  $b^k > 0$ ) точки  $u^{k*}$ , которая формирует для МРНСг направление минимизации  $d^k = (u^{k*} - u^k)$  с коллинеарными градиентами. Нам придётся задавать

$u^{k*}$  каким-то эвристическим, приближительным способом, а значит и сам параметр  $\alpha^k$  не сможет гарантировать высокую точность коллинеарности градиентов  $\nabla J(u^{k*})$  и  $\nabla J(u^k)$ .

Слишком грубое приближение  $\alpha^k$  к значению (2.13), когда угол между антиградиентом  $-\nabla J^k$  и направлением спуска  $d^k = -\overline{\alpha^k \nabla J^k}$  будет более  $90^\circ$ , приведёт к потере направления минимизации. Для гарантии получения направления минимизации, при приближительной реализации выражения (2.13), наложим следующее ограничение на  $\alpha^k$ :

$$0 < \alpha_i^k < \infty, \quad \forall i, k. \quad (2.14)$$

То есть *параметр регулирования направления спуска* в МРНСг — это вектор из ограниченного положительного *полупространства*:

$$\alpha^k \in E_+^n,$$

где  $E_+^n = \{\alpha \in E^n \mid 0 < \alpha_i < \infty \quad \forall i\}$ . При этом скалярное произведение  $\langle -\nabla J^k, d^k \rangle > 0$ , т.е. угол между этими векторами всегда будет менее  $90^\circ$ .

Итак, вместо точного направления  $d^k = (u^{k*} - u^k)$ , мы будем использовать некоторое приближительное направление  $\tilde{d}^k = (\tilde{u}^k - u^k)$ , где  $\tilde{u}^k \approx u^{k*}$ . В итоге из выражения (2.13) при условии (2.14) получим следующую *формулу для параметра регулирования* направления спуска:

$$\alpha_i^k = \left| \frac{\tilde{d}_i^k}{\nabla_i J^k} \right|, \quad \nabla_i J^k \neq 0, \quad \forall i, \quad k = 0, 1 \dots \quad (2.15)$$

Теперь рассмотрим возможные способы практической реализации этой формулы для минимизации *локально квадратичных* функций  $J(u)$ . Будем использовать эвристическую идею пробных *шаблонных приближений*  $\tilde{u}^0$  на первой итерации, что ранее применялось для решения бесконечномерных задачах [92, 94, 19, 90].

Эта идея базируется на приближительном выполнении условий оптимальности теоремы 2.1, которое мы будем реализовывать следующим образом. *Необходимо*, чтобы шаблонный шаг  $\tilde{d}_i^0 = \tilde{u}_i^0 - u_i^0$

был заметным для всех  $i$  и приводил к заметным изменениям всех компонент градиента  $\nabla_i J(\tilde{u}^0)$  относительно  $\nabla_i J(u^0)$ . При этом, если повезёт, то изменения компонент градиента могут оказаться даже пропорциональными. В этом случае шаблон  $\tilde{u}^0 = u^{0*}$  будет формировать направление  $d^0 = (u^{0*} - u^0)$  именно с коллинеарными градиентами  $J(u^{0*})$  и  $\nabla J(u^0)$ , что достаточно для одношаговой минимизации строго выпуклой квадратичной функции. Далее мы не будем полагаться на такую удачу.

Если шаблонный шаг  $\tilde{u}^0$  уже задан, то из выражения (2.15) легко находится параметр регулирования направления спуска  $\alpha^0$ :

$$\alpha_i^0 = \left| \frac{\tilde{u}_i^0 - u_i^0}{\nabla_i J^0} \right|, \quad \nabla_i J^0 \neq 0, \quad i = 1 \dots n. \quad (2.16)$$

На последующих итерациях алгоритма МРНСГ (2.12) будем принимать

$$\alpha^k = \alpha^0 \stackrel{\text{def}}{=} \alpha, \quad k = 1, 2 \dots$$

То есть, для минимизации локально квадратичных  $J(u)$  будем формировать параметр регулирования направления спуска на первой итерации в относительно небольшой окрестности строгого локального минимума, где  $J(u)$  можно считать квадратичной. Заметим, что шаблонный шаг можно делать как в сторону роста, так и в сторону убывания функции  $J(u)$ . В любом случае, согласно (2.16), направление  $\overrightarrow{\alpha \nabla J^k}$  алгоритма МРНСГ (2.12), где  $b^k > 0$ , будет минимизирующим.

В итоге МРНСГ принимает вид:

$$u^{k+1} = u^k - b^k \overrightarrow{\alpha \nabla J^k}, \quad k = 0, 1 \dots \quad (2.17)$$

Или в развёрнутом виде:

$$u_i^{k+1} = u_i^k - b^k \alpha_i \nabla_i J^k, \quad i = 1 \dots n, \quad k = 0, 1 \dots$$

Глубину шагов, т.е. шаговый множитель  $b^k$ , можно вычислять по любой стратегии, начиная с первого шага, когда  $k = 0$ .

Практическое использование МРНСг следует начинать с параметра регулирования спуска  $\alpha = \{1\}_{i=1}^n$ . То есть надо начинать решать задачу оптимизации с тестирования ситуации посредством градиентных методов, например, МНС. Возможно, вам сразу повезёт (удачное начальное приближение  $u^0$ ) — достаточно хорошо выполнятся НУО (2.7) теоремы 2.1, и не понадобится больше прилагать никаких усилий. В этом случае шаблонное приближение  $\tilde{u}^0$  будет иметь вид:

$$\tilde{u}_i^0 = u_i^0 - \delta \frac{\nabla_i J^0}{\|\nabla J^0\|}, \quad i = 1 \dots n,$$

что соответствует шагу глубиной  $\delta$  вдоль антиградиента.

Теперь рассмотрим несколько *эвристических шаблонов*  $\tilde{u}^0$ , которые могут быть вполне достаточными для решения широкого круга задач квадратичной (локально квадратичной) оптимизации. Приведём эти шаблоны с соответствующими *формулами параметра регулирования* направления спуска

❖ Зададим первый шаг глубиной  $\delta$  «под 45°» ко всем осям координат пространства управлений, например, в сторону убывания целевой функции:

$$\tilde{u}_i^0 = u_i^0 - \frac{\delta}{\sqrt{n}} \operatorname{sgn} \nabla_i J^0, \quad i = 1 \dots n. \quad (2.18)$$

Этот шаг, согласно (2.16) и с учётом  $\operatorname{sgn} \nabla_i J^0 = \nabla_i J^0 / |\nabla_i J^0|$ , задаёт

$$\alpha_i = \frac{\delta / \sqrt{n}}{|\nabla_i J^0|}, \quad \nabla_i J^0 \neq 0, \quad i = 1 \dots n.$$

❖ Следующий шаблон зададим как изменение начальных компонент вектора управления пропорционально своему же начальному значению с некоторым коэффициентом пропорциональности  $0 < \lambda < 1$ :

$$\tilde{u}_i^0 = \lambda u_i^0, \quad u_i^0 \neq 0, \quad i = 1 \dots n. \quad (2.19)$$

Это соответствует

$$\alpha_i = (1 - \lambda) \left| \frac{u_i^0}{\nabla_i J^0} \right|, \quad \nabla_i J^0 \neq 0, \quad u_i^0 \neq 0, \quad i = 1 \dots n.$$

❖ В особых случаях, кроме указанных шаблонов, исследователь может задать какой-либо другой эвристический вектор  $\varphi \in E^n$ :

$$\tilde{u}_i^0 = \varphi_i, \quad i = 1 \dots n \quad (2.20)$$

и соответственно получить

$$\alpha_i = \left| \frac{\varphi_i - u_i^0}{\nabla_i J^0} \right|, \quad \nabla_i J^0 \neq 0, \quad i = 1 \dots n.$$

Обычно одного первого шаблона (2.18) оказывается вполне достаточно для успешной реализации МРНСг.

Какой из перечисленных шаблонов лучше? Ответ на этот вопрос, в общем случае, можно получить только при сравнительной минимизации целевой функции, проделав не менее пяти-десяти итераций с разными шаблонами. Заметим, что выбор шаблона из условия наилучшей коллинеарности градиентов (наиболее точная реализация определения 2.1) на начальном шаге — это некорректная процедура. Дело в том, что для достижения коллинеарности  $\nabla J(\tilde{u}^0)$  и  $\nabla J(u^0)$  может потребоваться  $\alpha \in E^n$ . Но, в силу ограничения (2.14), в МРНСг вектор  $\alpha \in E_+^n$ , и он не может принадлежать всему пространству  $E^n$ . Такое ограничение сужает возможность выбора коллинеарных градиентов. Это, как мы ранее указывали, делает алгоритм МРНСг ограниченным по сравнению с МКГ.

Итак, подведём некоторые итоги полученных результатов и предварительные рекомендации для использования конечномерного МРНСг для минимизации *локально квадратичных* функций:

- направление первого шага минимизации всегда известно, оно определяется шаблоном  $\tilde{u}^0$ ;
- параметр регулирования направления спуска  $\alpha$  определяется на первом шаге и не меняется на последующих итерациях;

- для определения  $\alpha$  начальное приближение управления  $u^0$  должно удовлетворять условию  $\nabla_i J^0 \neq 0 \forall i$  и, возможно даже,  $u_i^0 \neq 0$  (в случае шаблона (2.19);
- лучшее значение  $\alpha$  (если в этом есть необходимость) определяется численным экспериментом с разными шаблонами по результатам попытки минимизации целевой функции на нескольких первых итерациях;

Теперь перейдём к тестированию полученных алгоритмов минимизации на функциях большой размерности. Оценим эффективность описанного МРНСГ с предлагаемыми шаблонами и целесообразность обобщения метода на бесконечномерные задачи.

## 2.8 Высокоразмерные задачи. Тестирование

Исследуем поведение МРНСГ на тестовых задачах оптимизации со строго выпуклыми квадратичными целевыми функциями большой размерности. Оценим его эффективность в зависимости от размерности  $n$  задачи, сравним с аналогичными показателями других популярных методов.

Зададим первую функцию QF1 [2] с её точкой минимума и рекомендуемым начальным приближением:

$$J(u) = \frac{1}{2} \sum_{i=1}^n i u_i^2 - u_n, \quad (2.21)$$

$$u_* = \left(0, 0, \dots, 0, \frac{1}{n}\right), \quad u^0 = (1, \dots, 1).$$

Вторая функция — это HilbertA [5]:

$$J(u) = \sum_{i=1}^n u_i \sum_{j=1}^n \frac{u_j}{i+j-1}, \quad (2.22)$$

$$u_* = (0, \dots, 0), \quad u^0 = (1, \dots, 1).$$

Построим следующую тестовую функцию на базе функции Staircase2 [2], но экстремальный вектор  $u_*$  зададим не единичным, а в виде непостоянных значений компонент. Назовём рассматриваемую функцию Staircase2sin:

$$J(u) = \sum_{i=1}^n \left( \sum_{j=1}^i (u_j - a_j) - i \right)^2, \quad a_j = \frac{1}{2} \sin \left( 2\pi \frac{j-1}{n-1} \right) \quad (2.23)$$

$$u_* = \left\{ 1 + \frac{1}{2} \sin \left( 2\pi \frac{j-1}{n-1} \right) \right\}_{j=1}^n, \quad u^0 = (0, \dots, 0).$$

Следующая функция — Tridia [2]:

$$J(u) = (u_1 - 1)^2 + \sum_{i=2}^n i(2u_i - u_{i-1})^2, \quad (2.24)$$

$$u_* = \left( 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16} \dots \right), \quad u^0 = (1, \dots, 1).$$

Функция Tolstykh:

$$J(u) = (u_1 + u_n)^2 + \sum_{i=2}^n (u_{i-1} - iu_i)^2, \quad (2.25)$$

$$u_* = (0, \dots, 0), \quad u^0 = (1, \dots, 1).$$

Особенность последней функции (2.25) в том, что при указанном  $u^0$  начальный градиент имеет нулевую вторую компоненту:  $\nabla_2 J^0 = 0$ .

Теперь давайте протестируем МРНСг (2.17) с предложенными эвристическими способами выбора параметра  $\alpha$ . Глубину шагов во всех тестах будем рассчитывать методом Носедала-Райта с максимальной глубиной шага  $2\sqrt{n}$  и параметром Вольфе  $c_2 = 0.1$ . Критерием завершения итераций для всех функций примем условие относительной близости к оптимуму (1.23) с  $\varepsilon_3 = 0.01$ .

Рассматривались следующие три варианта шаблонов для выбора параметра  $\alpha$  регулировки направления спуска:

Вариант 0. Без шаблона, т.е. по МНС;

Вариант 1. Шаблон «под 45°» (2.18) с  $\delta = 0.5\|u^0 - u_*\|$ ;

Вариант 2. Пропорциональный шаблон (2.19) с  $\lambda = 0.3$ ;

В таблице 2.1 приведены результаты минимизации для тестовых функций с указанием наилучшего варианта регулировки направления спуска. Значение эффективности минимизации (количество вычислений функции  $J$  и её градиента  $\nabla J$ ) для разных функций и их начальных приближений может зависеть от значения параметра Вольфе  $c_2$ . Поэтому в скобках дополнительно указано количество вычислений с наилучшим  $c_2$ , отличным от принятого 0.1. Так, для функции Staircase2sin были выполнены расчёты при  $c_2 = 0.1$  и в скобках — при  $c_2 = 0.9$ , а для функции Tridia выполнены при  $c_2 = 0.1$  и в скобках — при  $c_2 = 0.5$ .

Таблица 2.1 — Количество вычислений МРНСГ с наилучшими вариантами регулирования спуска при начальном приближении из (2.21)–(2.25)

$n$	Функция, вариант регулирования для $\alpha$				
	QF1, варианты 1 и 2	HilbertA, вариант 1	Staircase2sin, вариант 1	Tridia, варианты 1 и 2	Tolstykh вариант 0
10	13	5	716 (245)	258 (68)	639
$10^2$	13	5	698 (444)	156 (68)	14789
$10^3$	13	5	421 (414)	56 (39)	885 414
$10^4$	13	5	744 (432)	23 (36)	—

При минимизации функции Tolstykh с начальным приближением  $u^0 = (1, \dots, 1)$  нельзя применить варианты регулировок 1 или 2, поскольку градиент  $\nabla_2 J^0 = 0$ . Здесь применялся нулевой вариант, т.е.

МНС, а не МРНСГ, и соответственно были получены неудовлетворительные результаты в последней колонке таблицы 2.1.

Поскольку начальное приближение  $u^0$  для каждой функции может принципиально влиять на выбор шаблона  $\tilde{u}^0$ , то, кроме рекомендованных в задачах (2.21)–(2.25) начальных приближений, были проведены дополнительные тестовые расчёты с иным приближением, общим для всех функций:

$$u^0 = (-1, \dots, 1) = \left\{ -1 + \frac{2(i-1)}{n-1} \right\}_{i=1}^n. \quad (2.26)$$

Результаты таких расчётов размещены в таблице 2.2.

Таблица 2.2 — Количество вычислений МРНСГ с наилучшими вариантами регулирования спуска при начальном приближении (2.26)

$n$	Функция, вариант регулирования для $\alpha$				
	QF1, вариант 2	HilbertA, вариант 1	Staircase2sin, вариант 0	Tridia, вариант 2	Tolstykh, вариант 2
10	14	375 (83)	71	68	102
$10^2$	6	265 (55)	71	68	118
$10^3$	6	257 (57)	71	39	86
$10^4$	6	257 (57)	71	36	46

Для функции HilbertA в скобках указано количество вычислений при  $c_2 = 0.9$ . Для функции Staircase2sin начальное приближение (2.26) оказалось очень удачным для МНС (вариант 0).

Сравним результаты минимизации тестовых функций со следующими широко известными, традиционными методами первого порядка:

1. МСГ — метод сопряжённых градиентов Флетчера-Ривса, с обновлением направления спуска при существенном

нарушении ортогональности градиентов согласно условию [12]

$$|\langle \nabla J^k, \nabla J^{k-1} \rangle| / \|\nabla J^k\|^2 \geq 0.1;$$

2. МНКр — конечно-разностный метод Ньютона, с шагом дифференцирования  $h = 10^{-8}$ ;
3. МНсг — неточный метод Ньютона с сопряжёнными градиентами, где коэффициент  $c$  в  $\eta^k$  подбирался наилучшим в диапазоне  $10^{-1}$ – $10^{-15}$ . Количество подитераций МСГ не ограничивалось.
4. L-BFGS — квазиньютоновский метод с памятью в 5 векторов;

Шаговый множитель  $b^k$  определялся с параметром Вольфе  $c_2 = 0.1$  для МСГ и  $c_2 = 0.9$  для L-BFGS. Для МНКр, МНсг шаговой множитель отсутствует, здесь  $b^k = 1$ .

В таблице 2.3 приведены результаты расчётов для двух значений размерности  $n$ , показывающих картину зависимости количества вычислений указанными выше методами от размерности задач (2.21)–(2.25). Данные для МНКр при функции HilbertA отсутствуют, поскольку здесь получаемая матрица Гессе (матрица Гильберта) была отрицательно определённой. Вычислительные помехи оказались чрезвычайно высокими, что естественно для плохо обусловленной матрицы Гильберта. Для МНсг коэффициент  $c = 10^{-3}$  был наилучшим почти для всех функций кроме HilbertA, здесь было принято  $c = 10^{-14}$ , и для функции Staircase2sin  $c = 10^{-5}$ .

Таблица 2.3 — Количество вычислений традиционными методами

	QF1	HilbertA	Staircase2sin	Tridia	Tolstykh
$n$	<b>МСГ</b>				
10	34	244	23	57	44
$10^3$	218	405	1236	940	23730
	<b>МНкр</b>				
10	11	—	11	11	11
$10^3$	1001	—	1001	1001	1001
	<b>МНсг</b>				
10	14	9	9	11	11
$10^3$	99	27	265	194	1513
	<b>L-BFGS</b>				
10	17	50	21	33	37
$10^3$	127	169	1495	359	3581

Мы видим падение эффективности (рост количества вычислений) у всех методов с ростом размерности задачи оптимизации.

Из сравнения таблиц 2.1, 2.2 с таблицей 2.3 можно сделать вывод, что минимизация квадратичных функций посредством МРНСг, в отличие от традиционных методов (МСГ, МНкр, МНсг, L-BFGS), не ведёт к росту количества вычислений функции и её градиента с ростом размерности  $n$  задачи оптимизации. То есть *эффективность* МРНСг не зависит от размерности квадратичной задачи оптимизации. Получен принципиально новый результат по сравнению с традиционными экстремальными методами первого порядка.

Приведенные тесты являются подтверждением возможности и целесообразности обобщения МРНСг на задачи бесконечной размерности.

### 3 Бесконечномерные параметры-управления

В данном разделе мы обобщим полученные ранее конечномерные результаты на бесконечномерную ситуацию, когда управление  $u(\tau) \in L_2(S)$ . Рассмотрим более подробно и саму задачу, и пути её решения.

Итак, *прямой экстремальный подход* для распределённых систем может формулировать задачу оптимизации (1.1) как бесконечномерную экстремальную (1.2):

$$u_* = \arg \min J(u), \quad u \in L_2(S).$$

В общем случае возможна и более сложная формулировка, когда управление будет вектор-функцией  $u \in L_2^n(S)$ ,  $n > 1$ . Возможны и ограничения на управление, но мы их рассмотрим позже.

Напомним, что оптимальное решение полученной экстремальной задачи ищется либо традиционными алгоритмами (1.5), формально обобщённым на бесконечномерное пространство, —

$$u^{k+1} = u^k + b^k p^k \quad \text{на } S, \quad k = 0, 1, \dots,$$

либо алгоритмами с регулируемым направлением спуска (1.6) —

$$u^{k+1} = u^k + b^k \alpha^k p^k \quad \text{равномерно на } S_\Delta, \quad k = 0, 1, \dots,$$

где  $p^k$  — *базовое направление* минимизации. Первые алгоритмы на множестве  $S$  могут обеспечить только интегральную *среднеквадратичную сходимость* к оптимуму. Вторые же могут обеспечивать *равномерную* на  $S_\Delta \subseteq S$  *сходимость*.

#### 3.1 Условия оптимальности

Сделаем необходимые обобщения конечномерных покомпонентных условий оптимальности теоремы 2.1 на бесконечномерные задачи оптимизации.

По-прежнему начнём исследование задачи с квадратичной формы функционала. Здесь эта форма будет иметь вид *квадратичного строго выпуклого* целевого функционала

$$J(u) = \frac{1}{2} \langle \mathcal{A}u, u \rangle_{L_2(S)},$$

где  $\mathcal{A}$  — линейный *положительный* и *симметричный* оператор, действующий из  $L_2(S)$  в  $L_2(S)$ . При этом  $J(u)$  имеет единственную точку<sup>1</sup> минимума  $u_*$ , а градиент  $\nabla J = \mathcal{A}u$ , который в точке минимума обращается в ноль, т.е.  $\mathcal{A}u_* = 0$ .

Далее нам придётся работать со значениями функций в каждой пространственно-временной точке  $\tau \in S$ . При этом возможна ситуация, когда равномерная сходимость управлений  $u^k(\tau)$  к оптимуму  $u_*(\tau)$  на  $S$  отсутствует, или когда управление разрывное, неопределённое на некоторых подмножествах нулевой меры в  $S$ . В этих случаях будем сужать  $S$  до множества *возможной равномерной сходимости*  $S_\Delta \subset S$ , где указанные проблемы отсутствуют. Подобные вопросы мы обсуждали в начале книги (раздел 1.2).

Выделение множества  $S_\Delta$  не означает, что любая последовательность, обладающая интегральной сходимостью  $\|u^k(\tau) - u_*(\tau)\|_{L_2(S_\Delta)} \xrightarrow{k \rightarrow \infty} 0$ , будет сходиться к нулю ещё и равномерно на  $S_\Delta$ . Множество  $S_\Delta$  просто не содержит областей, где в принципе, т.е. теоретически, нельзя добиться поточечной сходимости.

Теперь уточним понятие коллинеарности бесконечномерных градиентов посредством определения пропорциональности изменения их значений. Также уточним и понятие направления на оптимум  $u_*$ .

---

<sup>1</sup> Напоминаем, термин «точка» в бесконечномерном пространстве означает функцию, как элемент, в данном случае, пространства  $L_2$ .

**Определение 3.1.** Функция  $\nabla J(u^k; \tau) \in L_2(S_\Delta)$  изменяется от шага  $k$  к шагу  $k + 1$  коллинеарно если поточечные значения данной функции изменяются пропорционально на  $S_\Delta$ :

$$\frac{\nabla J(u^{k+1}; \tau)}{\nabla J(u^k; \tau)} = \text{const}^k, \quad \tau \in S_\Delta, \quad 0 < \text{const}^k < \infty, \quad k \geq 0.$$

Если  $\nabla J(u^k; \tau) = 0$  для какого-либо  $\tau$ , то необходимо  $\nabla J(u^{k+1}; \tau) = 0$ . При этом отношение нулей следует считать равным  $\text{const}^k$ .

**Определение 3.2.** Будем называть направлением на оптимум  $u_*$  из точки  $u^0$  множество функций

$$d^0(\tau) \in \{b^k \cdot (u_*(\tau) - u^0(\tau)) \mid b^k > 0 \quad \forall k\}, \quad \tau \in S_\Delta,$$

а шаги управлений  $u^k - u^0 = b^k(u_* - u^0)$  — лежащими вдоль направления  $d^0$ .

Увеличивая  $b^k$  от приблизительно нулевого значения до единицы, мы получим вдоль  $d^0$  коллинеарные (пропорциональные на  $S_\Delta$ ) шаги  $u^k - u^0$  вплоть до оптимального управления  $u^k = u_*$ . Таким образом, во-первых, направление  $d^0$  — это *направление кратчайшего пути* из  $u^0$  в  $u_*$ . Во-вторых, при стремлении  $u^k \rightarrow u_*$  вдоль направления  $d^0$  будет иметь место сходимость  $u^k \xrightarrow{k \rightarrow \infty} u_*$  пропорционально на  $S_\Delta$ .

Теперь сформулируем бесконечномерный аналог теоремы 2.1. Алгоритм доказательства этой теоремы такой же, как и у конечномерной, поэтому здесь мы приведём теорему без доказательств.

**Теорема 3.1 (Очень сильное условие оптимальности).** Если функционал  $J(u)$ ,  $u \in L_2(S)$  квадратичный и строго выпуклый, то для сходимости к оптимуму  $u_*$  необходимо и достаточно, чтобы функциональная последовательность

$$\nabla J(u^k) \xrightarrow{k \rightarrow \infty} 0 \quad \text{пропорционально на } S_\Delta. \quad (3.1)$$

*Замечание 1.* Пропорциональная сходимость градиентов (3.1) сопровождается пропорциональной сходимостью управлений по кратчайшему пути вдоль  $d^0$ :

$$u^k - u_* \xrightarrow{k \rightarrow \infty} 0 \text{ пропорционально на } S_\Delta.$$

*Замечание 2.* Очевидно, что имеет место утверждение аналогичное следствию 2.2. Приведём такое утверждение.

**Следствие 3.1.** Для строго выпуклого квадратичного функционала  $J(u)$  условие оптимальности (3.1) в каждой точке  $u^k$  формирует шаг  $b^k d^k$ , совпадающий с шагом  $p^k$  бесконечномерного метода Ньютона:

$$p^k \stackrel{\text{def}}{=} -(\mathcal{A}^k)^{-1} \nabla J^k = b^k d^k \in L_2(S_\Delta).$$

где шаговый множитель  $b^k = \arg \min_{b>0} J(u^k + b d^k)$ .

Сформулированное условие оптимальности — это очень сильное требование к бесконечно большому количеству поточечных значений функции  $\nabla J(u; \tau)$ . Очевидно, что выполнить его или даже проконтролировать на  $S_\Delta$  невозможно. Поэтому, нам придётся ослабить условия теоремы 3.1 от *пропорциональной до равномерной сходимости* на  $S_\Delta$ . Отказываясь от реализации теоремы 3.1, мы отказываемся и от реализации возможного бесконечномерного аналога метода коллинеарных градиентов (МКГ), и от утверждений следствия 3.1.

Это означает, что мы не будем иметь точное направление  $d^0$  на оптимум и не получим одношагового алгоритма наподобие МКГ (2.8). Мы будем говорить не о сходимости вдоль  $d^0$ , а по некоторой ломаной траектории  $\sum_{k=0}^{\infty} \tilde{d}^k \approx d^0$ , где  $\tilde{d}^k = u^{k+1} - u^k$ . Вид такой траектории определяется множеством шагов  $\tilde{d}^k$  для которых имеет место равномерная на  $S_\Delta$  сходимость функций управления  $u^k(\tau)$  к оптимуму  $u_*(\tau)$ . Близость траекторий  $\sum_{k=0}^{\infty} \tilde{d}^k$  и  $d^0$  определяется близостью равномерной сходимости к пропорциональной сходимости.

Понятно, если равномерная сходимость к  $u_*$  оказывается и пропорциональной, то получим  $\sum_{k=0}^{\infty} \tilde{d}^k = d^0$ .

Далее будем обсуждать только необходимые условия оптимальности.

**Теорема 3.2 (Сильное НУО).** Если строго выпуклый квадратичный функционал  $J(u)$ ,  $u \in L_2(S)$  имеет минимум в точке  $u_*$ , а последовательность управлений  $u^k$  такова, что  $|u^k - u_*| \xrightarrow{k \rightarrow \infty} 0$  равномерно на  $S_{\Delta}$ , то необходимо:

$$|\nabla J(u^k)| \xrightarrow{k \rightarrow \infty} 0 \text{ равномерно на } S_{\Delta}. \quad (3.2)$$

**Доказательство.** Применив линейный положительный оператор  $\mathcal{A}$  к равномерно сходящейся на  $S_{\Delta}$  последовательности  $|u^k - u_*| \xrightarrow{k \rightarrow \infty} 0$ , получим

$$\mathcal{A}|u^k - u_*| = |\mathcal{A}u^k - \mathcal{A}u_*| = |\nabla J(u^k)| \xrightarrow{k \rightarrow \infty} 0 \text{ равномерно на } S_{\Delta}.$$

Что и следовало доказать. ■

Приведём также традиционное классическое НУО.

**Теорема 3.3 (Слабое НУО).** Если строго выпуклый квадратичный функционал  $J(u)$ ,  $u \in L_2(S)$  имеет минимум в точке  $u_*$ , а последовательность управлений  $u^k$  такова, что  $\|u^k - u_*\|_{L_2(S)} \xrightarrow{k \rightarrow \infty} 0$ , то необходимо:

$$\|\nabla J(u^k)\|_{L_2(S)} \xrightarrow{k \rightarrow \infty} 0. \quad (3.3)$$

Заметим, что в классическом НУО для бесконечномерных задач всегда говорят о сходимости на всём множестве  $S$ , а не на  $S_{\Delta}$ , как это было в предыдущих теоремах 3.1 и 3.2. Это объясняется тем, что множество возможных «плохих» точек  $S_0 \subset S$ , где нет поточечной сходимости, имеет меру нуль в  $L_2(S)$ . Это мы обсуждали в конце раздела 1.2.

Очевидно, что наиболее сильной (*очень сильной сходимостью*) из указанных трёх сходимостей является пропорциональная сходимость (3.1), а наиболее *слабой сходимостью* — среднеквадратичная сходимость (3.3). Между ними находится поточенная *сильная сходимость* (3.2).

Если проводить ассоциацию с бесконечномерными векторами, то условие слабой сходимости (3.3) может обеспечить сходимость к  $u_*$  всего лишь с уменьшением нормы бесконечномерного вектора  $\nabla J$ . При этом игнорируется направление (разнообразие функций  $\nabla J(\tau)$  с одной и той же нормой) приближения к  $u_*$ . Очень сильное условие сходимости (3.1) требует приближения к  $u_*$  в наилучшем направлении  $d^0$ , которое для квадратичных функционалов является направлением кратчайшего пути с коллинеарными градиентами. А сильное условие (3.2) из всех возможных направлений (3.3) отбирает только те направления, которые относительно недалеко отклоняются от наилучшего направления  $d^0$ .

### 3.2 Методы с регулируемым направлением спуска (МРНС)

В бесконечномерном пространстве параметров-управлений, когда  $u \in L_2(S)$ , метод с регулируемым направлением спуска — МРНС относительно базовых направлений минимизации  $p^k \in L_2(S)$  имеет вид (1.6):

$$u^{k+1} = u^k - b^k \alpha^k p^k \quad \text{равномерно на } S_\Delta, \quad k = 0, 1, \dots$$

Будем далее полагать, что начальные приближения управления  $u^0$ , для применения данного метода, находятся в области *локальной квадратичности* целевого функционала, или целевой функционал квадратичный. Это означает, что параметр регулирования направления спуска  $\alpha^k$  в данном случае, по аналогии с конечномерной ситуацией, можно задавать только на начальной итерации, т.е.

$$\alpha^k = \alpha^0 \stackrel{\text{def}}{=} \alpha \quad \text{на } S_\Delta, \quad k = 1, 2 \dots$$

Рассмотрим далее два варианта базовых направлений. Это направление антиградиента и направление сопряжённых градиентов. Поскольку общие идеи их применения и выбора параметра  $\alpha^k$  будут одинаковы, то мы ограничимся подробным рассмотрением одного первого варианта. Для второго приведём итоговые результаты.

### 3.2.1. МРНСг, регулирование относительно градиента

В качестве базового направления зададим направление антиградиента, т.е.  $p^k(\tau) = -\nabla J(u^k; \tau)$ . При этом МРНС (1.9), для минимизации локально квадратичных функционалов, примет вид следующего метода с регулируемым направлением спуска *относительно градиента* МРНСг:

$$u^{k+1} = u^k - b^k \alpha \nabla J^k \quad \text{равномерно на } S_\Delta, \quad k = 0, 1 \dots \quad (3.4)$$

или в развёрнутом виде:

$$u^{k+1}(\tau) = u^k(\tau) - b^k \alpha(\tau) \nabla J(u^k; \tau) \quad \text{равномерно на } S_\Delta, \quad k = 0, 1 \dots$$

Здесь *параметр регулирования направления спуска*  $\alpha(\tau) \in C_+^1(S_\Delta) = \{\alpha \in C^1(S_\Delta) \mid 0 < \alpha(\tau) < \infty \quad \forall \tau \in S_\Delta\}$  задаётся из условия равномерной сходимости к оптимуму и принадлежит ограниченному положительному полупространству гладких функций. Именно он определяет, как будет реализовываться сильное НУО (3.2). Для выбора шагового множителя  $b^k$  можно использовать любую стратегию, описанную в разделе 1.3.3.

Надо не забывать и про более сложные задачи, когда искомое управление может быть *вектор-функцией*  $u \in L_2^n(S)$ , а градиент — соответственно вектор-функцией  $\nabla J \in L_2^n(S_\Delta)$ . В этом случае мы будем иметь параметр  $\alpha$  в виде вектор-функции  $\alpha(\tau) \in C_+^{1,n}(S_\Delta)$ , а сам МРНСг (3.4) примет вид:

$$u^{k+1} = u^k - b^k \overrightarrow{\alpha \nabla J^k} \quad \text{равномерно на } S_\Delta, \quad k = 0, 1 \dots \quad (3.5)$$

Здесь *векторизация* (стрелка сверху) применяется ко всем  $i$ -тым компонентам  $\alpha_i(\tau)$  и  $\nabla_i J(u^k; \tau)$  вектор-функций  $\alpha$  и  $\nabla J^k$  размерности  $n$ .

При этом возникают дополнительные проблемы учёта степени влияния каждой компоненты  $u_i(\tau)$  на целевой функционал  $J(u) = J(u_1(\tau), \dots, u_n(\tau))$ . Кроме этого, необходимо иметь в виду, что областью определения каждого управления может быть своё множество  $S_{\Delta i} \in \{S_{\Delta i}\}_{i=1}^n \stackrel{\text{def}}{=} S_\Delta$ . В такой ситуации для разных  $i$ -тых управлений, при общем шаговом множителе  $b^k$ , может потребоваться выравнивание глубины  $i$ -тых шагов с учётом разной чувствительности  $J$  к разным  $u_i$ .

Такое выравнивание, например, можно делать посредством добавление весовых коэффициентов  $\beta_i$ . При этом множество функций  $\alpha_i(\tau)$  должно обеспечивать равномерность движения к оптимуму на  $S_{\Delta i}$ , а множество весов  $\beta_i$  — выравнивать скорость сходимости к оптимуму для разных компонент  $u_i$ . При этом, МРНСГ (3.4) примет вид:

$$u^{k+1} = u^k - b^k \overrightarrow{\beta^k \alpha \nabla J^k} \quad \text{равномерно на } S_\Delta, \quad k = 0, 1 \dots$$

Поскольку целевой функционал считается локально квадратичным, то  $\beta_i$  можно выбирать аналогично конечномерному МРНСГ (2.17), например, способом (2.18) с шагом под «45°». В соответствующей формуле после (2.18) заменяем обозначения  $\alpha_i$  на  $\beta_i$  и числа  $\nabla_i J^0$  на числа  $\|\alpha_i \nabla_i J^0\|$ . Получаем для шага глубиной, например, 0.1 по каждой компоненте  $u_i^0$  следующие весовые коэффициенты:

$$\beta_i = \frac{0.1 \|\alpha_i^0\|_{L_2(S_\Delta)}}{\|\alpha_i \nabla_i J^0\|_{L_2(S_\Delta)}}, \quad \|\nabla_i J^0\|_{L_2(S_\Delta)} \neq 0, \quad i = 1 \dots n.$$

Указанный способ выбора  $\beta_i$  можно считать приемлемым, если  $J$  отличается чувствительностью ко всем  $n$  компонентам управления хотя бы не более одного порядка.

Если коэффициенты  $\beta_i$  внести в функции  $\alpha_i(\tau)$ , то МРНСГ сохранит вид (3.5). Выбор  $\alpha$  как  $n$ -мерного вектора — это отдельная задача и мы её здесь рассматривать не будем. Далее, чтобы не усложнять рассмотрение бесконечномерного алгоритма МРНСГ будем полагать  $n = 1$ .

Для выбора параметра-функции  $\alpha(\tau) \in C_+^1(S_\Delta)$  возьмём за основу результаты опять-таки конечномерного МРНСГ (2.17). При этом, шаблонный шаг из начального приближения  $u^0(\tau)$  в  $\tilde{u}^0(\tau)$ , по аналогии с (2.16), будет задавать параметр  $\alpha(\tau) = \left| \frac{\tilde{u}^0(\tau) - u^0(\tau)}{\nabla J(u^0; \tau)} \right|$ ,  $\nabla J(u^0; \tau) \neq 0$ ,  $\tau \in S_\Delta$ . Здесь следует обсудить условие  $\nabla J(u^0; \tau) \neq 0$ . Его нельзя контролировать всего лишь в расчётных узлах  $\{\tau_i\}$  сетки аппроксимации градиента  $\nabla J(u^0; \tau_i)$ . Совпадение нулевого значения градиента с каким-либо узлом  $\tau_i$  сетки скорее всего никогда не произойдёт. Поэтому на практике надо контролировать постоянство знака дискретно заданной расчётной функции  $\nabla J(u^0; \tau_i)$  на множестве  $S_\Delta$ . Действительно, ведь смена знака непрерывной функции может происходить только в окрестности её нулевого значения, которое вряд ли попадёт в какой-нибудь расчётный узел  $\tau_i$ . Поэтому, далее вместо  $\nabla J(u^0; \tau) \neq 0$  будем требовать *постоянство знака градиента*:  $\text{sgn } \nabla J(u^0; \tau) = \text{const}$ .

Получаем в бесконечномерном случае следующий *параметр регулирования направления спуска* [92, 94, 19, 90]:

$$\alpha(\tau) = \left| \frac{\tilde{u}^0(\tau) - u^0(\tau)}{\nabla J(u^0; \tau)} \right|, \quad \text{sgn } \nabla J(u^0; \tau) = \text{const}, \quad \tau \in S_\Delta. \quad (3.6)$$

Постоянство знака начального градиента, естественно, накладывает некоторые ограничения на выбор приемлемого  $u^0(\tau)$ .

Будем задавать в формуле (3.6) *шаблонные приближения*  $\tilde{u}^0(\tau)$ , удовлетворяющие на шаге  $k = 0$  сильному НУО теоремы 3.2. Для этого

на пробном шаге от  $u^0$  к  $\tilde{u}^0$  необходимо добиться равномерных на  $S_\Delta$  изменений управления и градиента:

$$\begin{aligned} u^0(\tau) &\xrightarrow{\text{равномерный шаг}} \tilde{u}^0(\tau), \quad \tau \in S_\Delta; \\ \nabla J(u^0; \tau) &\xrightarrow{\text{равномерное изменение}} \nabla J(\tilde{u}^0; \tau), \quad \tau \in S_\Delta; \end{aligned}$$

Проще говоря, шаблонный шаг  $|\tilde{u}^0 - u^0|$  должен быть заметным для всех  $\tau \in S_\Delta$  и он должен приводить к заметным на  $S_\Delta$  изменениям градиента  $|\nabla J(\tilde{u}^0) - \nabla J(u^0)|$ .

При этом, если повезёт, то изменения градиента могут оказаться даже пропорциональными как в (3.1), т.е. удовлетворяющими *очень сильному* НУО теоремы 3.1. В этом случае окажется, что шаблон  $\tilde{u}^0 = u^{0*}$  формирует наилучшее направление спуска  $d^0 = \alpha \nabla J^0 = u^{0*} - u^0$ , где градиенты  $\nabla J(u^0)$  и  $\nabla J(u^{0*})$  коллинеарные, что *достаточно* для одношаговой минимизации строго выпуклого квадратичного функционала посредством МРНСг (3.4) с параметром  $\alpha$  (3.6). Далее мы не будем полагаться на такую удачу и рассмотрим некоторый формализм, позволяющий задать удовлетворительные шаблонные приближения.

По аналогии с конечномерной ситуацией решение задачи оптимизации посредством МРНСг (3.4) следует начинать с параметром  $\alpha = 1$ , т.е. с тестирования ситуации посредством градиентного метода. Градиентный метод, формально обобщенный на бесконечномерное пространство, — это алгоритмом (1.5) с направлением  $p^k = -\nabla J \in L_2(S_\Delta)$ . При этом шаблонное приближение  $\tilde{u}^0$  можно представить как:

$$\tilde{u}^0 = u^0 - \delta \frac{\nabla J(u^0)}{\|\nabla J(u^0)\|} \text{ на } S_\Delta.$$

Далее будем исходить из худшего и более вероятного случая, когда на практике градиентный метод, например, МНС не обеспечивает сильной сходимости (3.2) теоремы 3.2. Поэтому рассмотрим другие

возможные варианты шаблонов с соответствующими *параметрами регулирования* сходимости.

❖ Аналогично конечномерному выражению (2.18) зададим шаблон в виде шага «под 45°»:

$$\tilde{u}^0(\tau) = u^0(\tau) \pm \delta, \quad \tau \in S_\Delta, \quad (3.7)$$

где положительное число  $\delta$  определяет величину смещение начальной функции  $u^0(\tau)$ . Знак перед  $\delta$ , т.е. смещение относительно  $u^0$ , не играет никакой роли для дальнейшего определения  $\alpha$  из формулы (3.6):

$$\alpha(\tau) = \delta / |\nabla J(u^0; \tau)|, \quad \text{sgn } \nabla J(u^0; \tau) = \text{const}, \quad \tau \in S_\Delta.$$

❖ Приведём шаблон *пропорционального* изменения начального приближения в соответствии с конечномерным аналогом (2.19):

$$\tilde{u}^0(\tau) = \lambda u^0(\tau), \quad u^0(\tau) \neq 0, \quad \tau \in S_\Delta, \quad (3.8)$$

где коэффициент пропорциональности  $0 < \lambda < 1$ . Такому шаблону соответствует

$$\alpha(\tau) = (1 - \lambda) \left| \frac{u^0(\tau)}{\nabla J(u^0; \tau)} \right|, \quad u^0(\tau) \neq 0, \\ \text{sgn } J(u^0; \tau) = \text{const}, \quad \tau \in S_\Delta.$$

Здесь появляется новое, дополнительное ограничение на начальное управление  $u^0(\tau)$ . Оно не должно быть нулевым ни в одной точке  $S_\Delta$ . Практическая реализация данного ограничения, также, как и в случае градиента, должна контролироваться постоянством знака начальной ненулевой функции  $u^0(\tau)$  на  $S_\Delta$ .

❖ Кроме указанных шаблонов исследователь может *эвристически* задать какую-либо другую функцию  $\varphi(\tau) \in L_2(S_\Delta)$ :

$$\tilde{u}^0(\tau) = \varphi(\tau), \quad \tau \in S_\Delta. \quad (3.9)$$

Соответственно

$$\alpha(\tau) = \left| \frac{\varphi(\tau) - u^0(\tau)}{\nabla J(u^0; \tau)} \right|, \quad \text{sgn } J(u^0; \tau) = \text{const}, \quad \tau \in S_\Delta.$$

Обычно первого шаблона (3.7) оказывается вполне достаточно для решения широкого круга задач *локально квадратичной* оптимизации посредством МРНСг. При наличии работающего бесконечномерного градиентного метода, оптимизация МРНСГ с любым шаблоном (3.7)–(3.9) реализуется довольно просто. Если все результаты вычислений с шаблонами (3.7), (3.8) оказываются недостаточно эффективными по количеству вычислений, то следует подумать о собственной эвристической функции  $\varphi$  в (3.9), либо сменить начальное приближение.

В отличие от конечномерных задач, мы столкнулись с появлением бесконечномерных особенностей. Во-первых, — это равномерная на  $S_\Delta$ , т.е. *сильная сходимость* к оптимуму (данное понятие сходимости отсутствует в конечномерных задачах), во-вторых, — требование *знакопостоянства начального градиента*  $\nabla J(u^0; \tau)$  на  $S_\Delta$ .

Если шаблонные приближения и соответствующий им параметр  $\alpha$  будут выбраны удачно, вплоть до выполнения очень сильного условия оптимальности теоремы 3.1, то, согласно следствию 3.1, мы будем иметь МРНСГ первого порядка с *квадратичной скоростью сходимости* как МН второго порядка. Если же мы выберем шаблонные приближения не столь удачно, скорость сходимости МРНСГ немного снизится, но всё равно будет выше, чем у обычного градиентного метода.

### 3.2.2. МРНСг, регулирование относительно сопряжённых градиентов

В качестве базового направления зададим, например, наиболее часто используемое направление сопряжённых градиентов Флетчера-Ривса:

$$p^k = -\nabla J^k + \beta^k p^{k-1}, \quad \beta^k = \|\nabla J^k\|^2 / \|\nabla J^{k-1}\|^2, \quad \beta^0 = 0.$$

При этом, формальное обобщение *метода сопряжённых градиентов* — МСГ (2.4) на бесконечномерное пространство будет иметь вид:

$$u^{k+1} = u^k + b^k p^k \quad \text{на } S, \quad k = 0, 1 \dots \quad (3.10)$$

Напомним, что в конечномерном случае, когда  $u \in E^n$ , МСГ может минимизировать квадратичные функции не более чем за  $n$  шагов. Это на много лучше, чем МНС. В бесконечномерном случае, при  $n = \infty$ , МСГ может терять свою эффективность и даже работоспособность.

По аналогии с (3.4), для минимизации локально квадратичных функционалов, рассмотрим в качестве МРНС метод с регулируемым направлением спуска *относительно сопряжённых градиентов* МРНСсг:

$$u^{k+1} = u^k + b^k \alpha p^k \quad \text{равномерно на } S_\Delta, \quad k = 0, 1 \dots, \quad (3.11)$$

Для выбора шагового множителя  $b^k$  следует использовать *третью* или *четвёртую стратегию* линейного поиска. Именно в этом случае направления градиентов могут быть сопряжёнными.

Поскольку с ростом итераций сопряжённое направление  $p^k$  требует очистки от накапливающихся вычислительных погрешностей, то для метода (3.11) необходимо делать *моменты обновления*, т.е. периодически присваивать  $\beta^k = 0$ . Например, при квадратичной оптимизации, обновления можно делать после заметной неортогональности градиентов [12]:

$$\frac{|\langle \nabla J^k, \nabla J^{k-1} \rangle|}{\|\nabla J^k\|^2} \geq 0.1.$$

Так же, как и в предыдущем подразделе, для выбора параметра регулирования направления спуска  $\alpha(\tau) \in C_+^1(S_\Delta)$ , будем задавать шаблонные приближения  $\tilde{u}^0$ , удовлетворяющие НУО теоремы 3.2 о равномерной сходимости на  $S_\Delta$ . Шаблонные приближения для (3.11) подбираются на первом шаге  $k = 0$ , где направление минимизации  $p^0 =$

$-\nabla J^0$ , как и в МРНСг. Поэтому здесь могут использоваться описанные ранее варианты шаблонов (3.7)–(3.9).

Практические результаты оптимизации посредством МРНСг чаще всего показывают существенно ускоренную, по сравнению с МРНСг, равномерную на  $S_\Delta$  сходимость к оптимуму  $u_*$ .

### 3.3 Регуляризация

У задач оптимизации есть одна особенность, все они являются обратными. Наиболее явно это проявляется при идентификации, когда по наблюдаемому состоянию системы  $v_e$  находят оптимальные значения параметров  $u$  в уравнениях модели, а не наоборот, как это происходит в прямых задачах. Обратные задачи, как правило, некорректны в классическом смысле и их решение требует регуляризации. Разберёмся подробнее в этом вопросе и выясним, как может обеспечиваться регуляризация решения в экстремальных алгоритмах оптимизации.

Запишем задачу оптимизации (1.1) отдельно в виде символического операторного уравнения исходной прямой задачи

$$\mathbb{D}(\tau, v, u)v(\tau) = 0, \quad \tau \in \bar{\Omega} \quad (3.12)$$

и целевого функционала

$$J(u) = \int_{\omega} I(v, u)d\omega \rightarrow \min. \quad (3.13)$$

Пусть функционал  $J(u)$  будет строго выпуклый, что обеспечит существование единственной точки оптимума  $u_*$ . Это требование несколько не ограничивает дальнейшие рассуждения и выводы, связанные с регуляризацией, а всего лишь упрощает их демонстрацию и понимание.

Предполагается, что прямая задача (3.12) обладает *классически корректным* (корректным по Адамару<sup>1</sup>) решением, которое характеризуется *прямым отображением* пространства управлений в пространство состояний системы:

$$U \rightarrow V. \quad (3.14)$$

То есть, задавая любое управление в пространстве  $U$ , мы однозначно и устойчиво попадаем в единственный элемент пространства состояний  $V$ .

Задача оптимизации (3.13) с уравнениями (3.12) характеризуется *обратным отображением*, когда требует из пространства состояний  $V$  вернуться в точку  $u_*$  пространства управлений:

$$V \xrightarrow{\min J} u_* \in U. \quad (3.15)$$

Приведённые схемы иллюстрируют особенности задач оптимизации. Кроме обратного характера отображений, мы видим, что отображения являются ещё и условными, — при условии  $\min J$ . В отличие от прямого отображения (3.14), где пространство управлений переводится в пространство состояний, в обратном условном отображении (3.15) пространство состояний переводится в одну точку  $u_*$  пространства управлений.

*Обратные задачи*, включая (3.15), как мы уже говорили, в большинстве своём являются некорректными. В нашем случае, например, на какой-нибудь итерации  $k$  экстремального метода, мы из состояния системы  $v^k$  хотим при условии  $J \rightarrow \min$  перейти к очередному приближению  $u^{k+1}$  так, чтобы  $\|u^{k+1} - u_*\| < \|u^k - u_*\|$ . Но из-за слишком большого «неосторожного» шага, мы можем не только нарушить данное неравенство, но получить и физически нереальное состояние  $v^{k+1}$  исходной задачи (3.12). Дальнейшие итерации на основе

---

<sup>1</sup> Согласно Ж. Адамару корректные математические задачи с уравнениями в частных производных должны иметь следующие свойства: решение существует; решение единственно; решение устойчиво (непрерывно зависит от данных задачи).

подобных состояний, скорее всего, будут приводить к катастрофической неустойчивости и расходимости решения обратной задачи (3.15).

При исследовании некорректных задач А. Н. Тихонов ввёл понятие *условной корректности* [84], заключающееся в требовании ограничения пространства решений (в данном случае пространства управлений  $U = L_2$ ) до *компактного множества*<sup>1</sup>  $\mathcal{U} \subset U$  существования, единственности и устойчивости.

Это значит, что, во-первых, исследователю необходимо ограничить пространство решений до замкнутого множества  $U \subset L_2$ . Во-вторых, необходимо чтобы такое *допустимое множество* управлений  $U = \mathcal{U}$ . При этом, функции  $u(\tau) \in U$  должны быть «разумными», т.е. равномерно непрерывными (не очень сильно колебаться). Требование разумности, которое далее будем считать правилом управления, делает возможным сравнение множеств  $\mathcal{U}$  и  $U$ .

Для сужения множества классически некорректных решений до *компакта корректности*<sup>2</sup> используют регуляризирующие алгоритмы, например, [30, 53, 81, 82, 29]. Описанные ранее МРНС и другие градиентные методы обладают регуляризирующими свойствами (смотрите, например, [90, 36]). Покажем, как такие экстремальные алгоритмы могут обеспечить сходимость к решению в компакте корректности  $\mathcal{U}$ .

Пусть мы имеем приближение управления  $u^k \in \mathcal{U}$ . Задача заключается в том, чтобы очередное приближение управления не вышло из компакта  $\mathcal{U}$ , т.е. необходимо обеспечить  $u^{k+1} \in \mathcal{U}$ .

Рассматриваемый в данной книге прямой экстремальный подход состоит из двух, практически самостоятельных, подзадач. Первая — это

<sup>1</sup> Простыми словами компакт  $\mathcal{U}$  — это такое ограниченное, замкнутое множество функций, которые не слишком сильно колеблются (равномерно непрерывны, не сильно изгибаются).

<sup>2</sup> Здесь мы используем этот термин как синоним компактного множества.

определение градиента целевого функционала  $\nabla J$ , и вторая — прямая минимизация целевого функционала  $J$  на основе полученного градиента.

Первая из этих подзадач будет подробно рассмотрена в разделах 3.4 «Градиент неявно заданного функционала» и 3.5.3 «Условия управляемости». Здесь мы лишь напомним тот факт, что  $\nabla J$  принадлежит сопряжённому пространству управлений, т.е. пространству  $U^*$  и находится через решение  $f$  сопряжённой задачи типа (1.12), которая представляет собой прямую задачу в сопряжённом пространстве  $V^*$ :

$$V^* \rightarrow U^* \ni \nabla J. \quad (3.16)$$

Это означает, что процедура поиска  $\nabla J = U^* f$  в выражении (1.13) относится к проблемам классически корректных задач. Предположим, что данная подзадача решена.

Перейдём ко второй подзадаче, т.е. к минимизации функционала  $J(u)$  на основе градиента  $\nabla J$ , например, посредством МРНСг. Рассмотрим случай вектор-функции управления  $u \in \mathcal{U} \subset L_2^n(S)$ .

Для регуляризации решения на каждой итерации, согласно теории некорректных задач [81], привлечём квадратичную форму стабилизирующего функционала:

$$\Phi = \frac{1}{2} \int_S \delta u^T \delta u dS, \quad (3.17)$$

где  $\delta u \in L_2^n(S)$  — вариация управления.

Функционал  $\Phi$  является *штрафным функционалом*, который необходимо добавлять к минимизируемому функционалу  $J(u)$  в каждой точке  $u^k \in \mathcal{U}$  с некоторым положительным коэффициентом, называемым *параметром регуляризации*  $W^k$ . В общем случае, при  $n > 1$ , на каждую компоненту  $\delta u_i$ ,  $i = 1 \dots n$  должен действовать свой параметр регуляризации. Поэтому,  $W^k$  — это некоторая диагональная матрица размерности  $n \times n$  коэффициентов регуляризации  $w_i > 0$ ,  $i =$

1 ... n. В общем случае  $w = w(\tau) \in C_+^{1,n}(S)$ . Смысл стабилизирующего функционала  $\Phi$  заключается в штрафовании решения, т.е. нового приближения  $u^{k+1}$ , при значительном отклонении  $\delta u$  от  $u^k$ , когда решение может выходить из компакта корректности  $\mathcal{U}$ . Здесь отклонение

$$\delta u = u^{k+1} - u^k.$$

Итак, на каждой итерации к целевому функционалу (3.13) будем добавлять стабилизирующий функционал (3.17) при помощи параметра регуляризации  $W^k$ . При этом, с учётом (3.17), мы получаем *сглаживающий функционал*

$$T = J + \frac{1}{2} \int_S \delta u^T W \delta u dS \rightarrow \min. \quad (3.18)$$

Согласно теории некорректных задач, минимизация сглаживающего функционала (3.18) при удовлетворительном параметре регуляризации  $W^k$  гарантирует для  $u^k \in \mathcal{U}$  корректное (регуляризованное) приближение  $u^{k+1}$ . То есть получаем  $u^{k+1} \in \mathcal{U}$ .

Необходимым условием минимума функционала (3.18) в точке  $u^k$  является равенство нулю его первой вариации относительно  $\delta u$ :

$$\delta T^k = \delta J^k + \int_S (W^k \delta u)^T \delta u dS = \langle \nabla J^k + W^k \delta u, \delta u \rangle_{L_2^n(S)} = 0,$$

В силу произвольности вариации  $\delta u$  необходимо:

$$\nabla J^k + W^k \delta u = 0 \text{ на } S.$$

С учётом того, что  $\delta u = u^{k+1} - u^k$ , и с учётом пожелания равномерной сходимости на  $S_\Delta$ , получаем следующий градиентный алгоритм с регуляризованной коррекцией спуска:

$$u^{k+1} = u^k - (W^k)^{-1} \nabla J^k \text{ равномерно на } S_\Delta, \quad k = 0, 1 \dots$$

Если сделать следующие преобразования:

$$(W)^{-1} \stackrel{\text{def}}{=} \text{diag}\{w_1^{-1}, w_2^{-1}, \dots, w_n^{-1}\} = b \cdot \text{diag}\{\alpha_1, \alpha_2, \dots, \alpha_n\},$$

где  $w_i^{-1}(\tau) = b\alpha_i(\tau)$ ,  $b > 0$ , то полученный алгоритм принимает вид МНРСГ типа (3.5) (когда  $\alpha$  не постоянный на итерациях):

$$u^{k+1} = u^k - b^k \overline{\alpha^k \nabla J^k} \quad \text{равномерно на } S_\Delta, \quad k = 0, 1, \dots,$$

где *векторизация* заменяет собой диагональную матрицу, а *параметры регуляризации*  $W^k(\tau)$  оказываются (с точностью до числа  $b^k$ ) *параметрами регулирования направления спуска*  $\alpha^k(\tau) \in C_+^{1,n}(S_\Delta)$ . При  $n = 1$  и локальной квадратичности  $J(u)$  метод приобретает одномерный вид МНРСГ (3.4):

$$u^{k+1} = u^k - b^k \alpha \nabla J^k \quad \text{равномерно на } S_\Delta, \quad k = 0, 1, \dots$$

Таким образом, градиентные методы, включая методы с регулируемым направлением спуска, являются регуляризирующими. Параметр регулирования направления спуска  $\alpha^k$  совместно с шаговым множителем  $b^k$  являются параметрами регуляризации решения. Степень регуляризации на каждой итерации определяется степенью регулировки направления шага и его глубины.

Если допустимое начальное приближение управления является вполне разумным и обладает свойством  $u^0 \in \mathcal{U}$ , то все последующие приближения  $u^{k+1}$ , например, в МНРСГ (3.4), при правильном выборе параметров  $b^k \alpha(\tau)$ , не выйдут из компакта корректности и в конечном итоге мы получим корректное оптимальное решение  $u_* \in \mathcal{U}$ . Данное утверждение будет верным, если будет обоснована и корректность определения градиента  $\nabla J^k$ . Эту задачу мы обсудим немного позже, она связана с проблемой управляемости.

Напомним, что допустимое множество  $U = \mathcal{U}$  предполагает не только его замкнутость, но и разумность управлений в нём, когда функции  $u^k(\tau)$  внутри  $U$  не слишком сильно колеблются. Именно такие управления мы и будем подразумевать далее по умолчанию.

### 3.4 Градиент неявно заданного функционала

Для реализации прямого экстремального подхода (1.2) посредством МРНСГ (3.4) или МРНСсг (3.11) необходимо уметь находить градиент  $\nabla J(u; \tau)$  целевого функционала  $J(u)$ , который зависит от управления  $u(\tau)$  не явно, а через пространственно-распределённую систему. Рассмотрим алгоритм [20] определения градиента<sup>1</sup> для *неявно* заданного целевого функционала. Для начала подробно сформулируем задачу оптимизации.

Во-первых, постановку, математическую формулировку *задачи оптимизации* надо начинать с цели оптимизации. Зададим цель оптимизации функционалом (*целевой функционал*):

$$J(u) = \int_{\omega} I(v, u) d\omega \rightarrow \min.$$

Здесь  $I(v, u)$  — *функция цели* оптимизации, явно зависящая от внутренних функций в виде состояния системы  $v(\tau)$  и, возможно, от управления  $u(\tau)$ . *Областью определения функции цели*  $I$  является пространственно-временное множество  $\omega \subset \bar{\Omega}$ , которое, как правило, не совпадает со всей областью  $\bar{\Omega}$  функционирования распределённой системы, а является только её частью. В то же время *областью определения функционала* является множество (пространство) управлений  $U$ .

Если какая-либо функция цели задаётся только на внутренней части  $\Omega$  или на части границы  $\partial\Omega$ , то её обычно определяют на всей области  $\Omega$  или  $\partial\Omega$ . При этом используют *оператор сужения*  $\cdot|_{\omega}$  в виде

<sup>1</sup> точнее — производной Фреше  $J'_u$ , которая при выполнении условий управляемости становится градиентом  $\nabla J$ . Считаем здесь, что условия управляемости выполняются.

вертикальной черты с множеством, на которое надо сузить действие функции. Например,  $I(v)$ , заданную на  $\omega$ , можно определить на всём множестве  $\bar{\Omega}$  как

$$I(v, u)|_{\omega} = \begin{cases} I(v, u) & \text{если } \tau \in \omega, \\ 0 & \text{если } \tau \in \bar{\Omega} \setminus \omega. \end{cases}$$

При этом целевой функционал принимает вид:

$$J(u) = \int_{\Omega} I(v, u)|_{\omega} d\Omega + \int_{\partial\Omega} I(v, u)|_{\omega} d\partial\Omega. \quad (3.19)$$

Обычно в задачах оптимизации присутствуют частные случаи (3.19), когда имеется только первый или только второй интеграл.

Во-вторых, необходимо определиться с математической моделью оптимизируемого объекта, чётко разделить множество условий, уравнений задачи и явно выделить:

- непосредственно дифференциальные уравнения модели распределённой системы;
- уравнения краевых условий для дифференциальных уравнений модели;
- возможные ограничения на функционирование системы в виде дополнительных условий на состояние и управление (подробнее смотрите раздел 3.6).

Запишем *исходную прямую задачу* для модели распределённой системы в следующем операторном виде:

$$\begin{aligned} \mathbb{D}(\tau, v, u)v + F(\tau, v, u) &= 0 \quad \text{на } \Omega, \\ G(\tau, v, u) &= 0 \quad \text{на } \Gamma \subset \partial\Omega. \end{aligned} \quad (3.20)$$

Здесь  $\mathbb{D}$  — это дифференциальный оператор, формирующий *дифференциальные уравнения*, как правило, в частных производных. Он задаёт производные, действующие на вектор-функцию  $v \in L_2^m(\bar{\Omega})$  *исходного состояния системы*. Оператор  $\mathbb{D}$  в общем случае зависит от  $\tau, v, u$ . Свободный член  $F \in L_2^m(\Omega)$  также может зависеть от  $\tau, v, u$ .

Начальные и граничные (*краевые условия*) исходной системы (3.20) определяются функцией  $G$  на множестве  $\Gamma$ , которое состоит из области граничных условий  $\Gamma_{\partial} = \partial X \times (t_0, t_1)$  и области начальных условий  $\Gamma_0 = X \times t_0$ , т.е.  $\Gamma = \Gamma_{\partial} \cup \Gamma_0$ .

Область  $\bar{\Omega}$  будем рассматривать как *прямое произведение* замкнутой *пространственной* области  $X \subset E^N$ ,  $1 \leq N \leq 3$  и времени  $t \in [t_0, t_1]$ . Таким образом

$$\bar{\Omega} = X \times [t_0, t_1] \subset E^{N+1}.$$

Например, одномерный *дифференциальный оператор*  $\mathbb{D}$  порядка  $r$  может иметь вид:  $\mathbb{D} \cdot = A(x, v, u) \frac{\partial^r}{\partial x^r}$ , где  $A$  — матрица  $m \times m$ ,  $x$  — пространственная координата. Точка в производной означает место элемента на который действует оператор  $\mathbb{D}$ . Далее в примерах ограничимся случаем, когда целые числа  $r \in \{1, 2\}$ , т.е. ограничимся первыми, максимум вторыми производными.

Из (3.20) видно, что распределённая система, её дифференциальные уравнения, функционируют на ограниченной, но открытой пространственно-временной области  $\Omega$ , а *область определения состояния системы*  $v$ , всё-таки, немного шире — это *замкнутая* область  $\bar{\Omega} = \Omega \cup \partial\Omega$ . Поэтому в дифференциальных уравнениях следует писать  $v \in V(\Omega)$ , а для всей задачи, включая граничные условия, имеем  $v \in V(\bar{\Omega})$ .

Количество уравнений системы (3.20), определяется количеством компонент вектор-функции  $v$ , т.е. равно  $m$ . Количество граничных условий в  $G$  определяется порядком  $r$  дифференциального оператора  $\mathbb{D}$  и размерностью системы  $m$ . С управляющими параметрами немного сложнее.

Управления могут быть функциями времени  $u(t)$ , пространства, например, одномерного —  $u(x)$ , пространства-времени  $u(x, t)$  и обычными векторами  $u$ . При этом все они могут быть определены внутри пространственно-временной области  $\Omega$ , на пространственной границе  $\Gamma_\partial$ , в начальный момент<sup>1</sup> на  $\Gamma_0$ . В общем случае мы можем иметь управление в виде *вектор-функции*  $u(x, t) \in U(S) \subseteq L_2^n(S)$ , где компоненты  $n$ -мерного вектора  $u$  будут функциями, определёнными на различных частях  $\bar{\Omega}$ . Кроме того, возможен многомерный случай  $x \in \{x_1, \dots, x_N\}$ .

Обычно область определения управления  $S = \Omega$ , или  $S = \Gamma_\partial$ , или в виде частей указанных множеств, или проекций  $P_{\tau_i}(\Omega)$ ,  $i \in \{1, \dots, N + 1\}$  этих множеств на пространственно-временные оси координат  $\tau_i \in \{x_1, \dots, x_N, t\}$ . Здесь  $P_{\tau_i}(\cdot)$  — *оператор проецирования* на ось  $\tau_i$ .

Перейдём к подробному описанию алгоритма получения градиента  $\nabla J$  целевого функционала (3.19), который неявно, через распределённую систему (3.20), зависит от управления  $u$ . Для этого нам потребуется найти выражение первой вариации (1.10) в виде:

$$\delta J = \langle \nabla J, \delta u \rangle_{U^*(S)}. \quad (3.21)$$

Рассматриваемый далее алгоритм представляет собой модификацию классического метода множителей Лагранжа в вариационном исчислении. Множители Лагранжа также называют и сопряжёнными переменными, а подобные подходы — «adjoint approach» (например, [11]). Методы классического вариационного исчисления для распределённых систем хорошо описаны, например, в работе [56]. Предлагаемая здесь модификация метода множителей Лагранжа [20, 92]

---

<sup>1</sup> Характерно для *ретроспективных* задач восстановления начальных условий. Мы их рассматривать не будем.

использует элементы функционального анализа, и была уже нами рассмотрена вкратце в разделе 1.3.2.

### 3.4.1. Выделение ограничений на управление

В разделе 3.6 «Управление с ограничениями» мы подробно обсудим возможные характерные ограничения на управление в распределённых системах. Они могут учитываться отдельно в экстремальных алгоритмах. Поэтому здесь считаем, что  $V = L_2^m$  и  $U = L_2^n$  и далее, при определении градиента  $\nabla J$ , рассматривать ограничения не будем.

### 3.4.2. Линеаризация задачи

Это первый этап получения градиента  $\nabla J$ . Он является подготовкой ко второму этапу, где мы будем объединять линейные элементы задачи.

Согласно (3.21), для определения градиента в какой-либо точке  $u \in U$ , необходимо выписать главную линейную часть приращения целевого функционала  $J(u)$ . Для этого в точке  $u$  зададим вариацию управления  $\delta u$ . Вариации управления  $\delta u \in U(S)$  приводят к вариациям состояния системы  $\delta v \in V(\bar{\Omega})$ .

При этом *вариация непосредственно только целевого функционала* (3.19) приводит к выражению вида

$$\delta J(u) = \langle I'_v|_{\omega}, \delta v \rangle_{V^*(\Omega)} + \langle I'_v|_{\omega}, \delta v \rangle_{V^*(\partial\Omega)} + \langle I'_u|_{\omega}, \delta u \rangle_{U^*(S)} \in E, \quad (3.22)$$

что можно трактовать как значение *линейных функционалов*  $I'_v(\tau) \in V^*(\bar{\Omega})$  на элементе  $\delta v(\tau) \in V(\bar{\Omega})$  и  $I'_u(\tau) \in U^*(S)$  на элементе  $\delta u(\tau) \in U(S)$ .

Линейный функционал  $I'_v = \left( \frac{\partial I}{\partial v_1}, \dots, \frac{\partial I}{\partial v_m} \right)^T$  — это вектор-строка с  $m$  компонентами *частных производных* целевой функции  $I$  по вектор-

состоянию  $v$ . Линейный функционал  $I'_v$  принадлежит пространству  $V^*$ , сопряжённому к пространству исходных состояний  $V$ . Линейный функционал  $I'_u = \left( \frac{\partial I}{\partial u_1}, \dots, \frac{\partial I}{\partial u_n} \right)^T$  принадлежит пространству  $U^*$ , сопряжённому к пространству управлений  $U$ .

*Вариация управляемой системы (3.20) приводит к выражению*

$$\delta(\mathbb{D}v + F) = \mathbb{V}\delta v + \mathbb{U}\delta u = 0 \in V(\Omega), \quad (3.23)$$

где  $\mathbb{V}, \mathbb{U}$  — *линейные операторы*, действующие из  $V(\Omega)$  в  $V(\Omega)$  и из  $U(S)$  в  $V(\Omega)$ , соответственно. Линейные операторы  $\mathbb{V}, \mathbb{U}$  — это частные производные от уравнений системы (3.20) по состоянию  $v$  и управлению  $u$ .

Рассмотрим пример указанных операторов. Этот пример будет с некоторой «завышенной» сложностью, но мы его считаем показательным.

Пусть мы имеем упоминаемый ранее одномерный матричный дифференциальный оператор  $\mathbb{D} \cdot = A(x, v, u) \frac{\partial^r \cdot}{\partial x^r}$  и свободный член  $F(v, u)$ . Область определения управления  $S = \Omega$ . После варьирования получим линейный (относительно  $\delta v$ ) матричный дифференциальный оператор, действующий на  $\delta v$ , с соответствующим *отображением*:

$$\mathbb{V} \cdot = A \frac{\partial^r \cdot}{\partial x^r} + \mathbb{C}_r \cdot + F'_v \cdot : V(\Omega) \rightarrow V(\Omega).$$

Первое слагаемое — это результат варьирования по  $v$  производной  $\frac{\partial^r v}{\partial x^r}$  при матрице  $A$ , второе слагаемое  $\mathbb{C}_r$  — это результат варьирования матрицы  $A$  при производной  $\frac{\partial^r v}{\partial x^r}$  (в соответствии с правилами вычисления производной от произведения), третье слагаемое  $F'_v$  — это результат варьирования по  $\delta v$  свободного члена  $F$ :

$$F'_v = \begin{pmatrix} \frac{\partial F_1}{\partial v_1} & \dots & \frac{\partial F_1}{\partial v_m} \\ \dots & \dots & \dots \\ \frac{\partial F_m}{\partial v_1} & \dots & \frac{\partial F_m}{\partial v_m} \end{pmatrix}.$$

Операторы  $\mathbb{C}_r$  и  $F'_v$  представляют собой матрицы размерности  $m \times m$ , они не являются дифференциальными операторами относительно  $\delta v$ . Нижний индекс  $r$  при  $\mathbb{C}_r$  означает всего лишь то, что оператор  $\mathbb{C}_r$  получен из слагаемого, содержащего производные порядка  $r$ , а именно,  $\frac{\partial^r v}{\partial x^r}$ .

Давайте уточним вид оператора  $\mathbb{C}_r$ . Вариация  $\mathbb{C}_r \delta v = \left( \frac{\partial A}{\partial v} \delta v \right)^T \frac{\partial^r v}{\partial x^r}$ , что означает умножение вектор-строки  $\left( \frac{\partial A}{\partial v} \delta v \right)^T$  на вектор-столбец  $\frac{\partial^r v}{\partial x^r}$ . Производная с вариацией  $\frac{\partial A}{\partial v} \delta v = \sum_{i=1}^m \frac{\partial A}{\partial v_i} \delta v_i$ . То есть, в  $\mathbb{C}_r$  находятся  $m$  матриц, полученных дифференцированием матрицы  $A$  (элементов  $a_{ij}$  матрицы) по всем компонентам состояния при соответствующих вариациях  $\delta v_i$ . При этом оператор  $\mathbb{C}_r$  можно представить как

$$\mathbb{C}_r \cdot = \left( \frac{\partial A}{\partial v} \cdot \right)^T \frac{\partial^r v}{\partial x^r} \equiv \sum_{i=1}^m \frac{\partial A}{\partial v_i} \cdot |_{i} \frac{\partial^r v}{\partial x^r}.$$

При вариации  $\delta u$ , в рассматриваемом примере, получаем линейный матричный оператор с соответствующим *отображением*:

$$\mathbb{U} \cdot = \left( \frac{\partial A}{\partial u} \cdot \right)^T \frac{\partial^r v}{\partial x^r} + F'_u \cdot \equiv \sum_{i=1}^n \frac{\partial A}{\partial u_i} \cdot |_{i} \frac{\partial^r v}{\partial x^r} + F'_u \cdot : U(S = \Omega) \rightarrow V(\Omega).$$

Это результат варьирования матрицы  $A$  и свободного члена  $F$  по управлению  $u$ . Оператор  $\mathbb{U}$  не являются дифференциальным относительно  $\delta u$ . Последняя матрица может быть неквадратной:

$$F'_u = \begin{pmatrix} \frac{\partial F_1}{\partial u_1} & \cdots & \frac{\partial F_1}{\partial u_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial u_1} & \cdots & \frac{\partial F_m}{\partial u_n} \end{pmatrix}.$$

### 3.4.3. Отображение линеаризованных уравнений в сопряжённые пространства

Давайте запишем линеаризованные уравнения (3.23) в тех же пространствах, что и линеаризованный целевой функционал (3.22). Для этого необходимо взять произвольный *линейный функционал*  $\tilde{f}(\tau)$  из сопряжённого пространства  $V^*(\bar{\Omega}) = L_2^m(\bar{\Omega})$ . Напомним, что любую функцию можно называть линейным функционалом, если она однородна и аддитивна, а сопряжённое евклидово пространство  $L_2^{m*} = L_2^m$  — *самосопряжённое*.

Запишем значение  $\tilde{f}$  на вариации системы (3.23):

$$\langle \tilde{f}, \delta(\mathbb{D}v + F) \rangle_{V^*(\Omega)} = \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} + \langle \tilde{f}, \mathbb{U}\delta u \rangle_{V^*(\Omega)} = 0 \in E. \quad (3.24)$$

Заметим, что полученные таким образом уравнения системы, из-за произвольности  $\tilde{f}$ , не содержат связи между векторами  $\delta v$ ,  $\delta u$  и их компонентами. Если исходная система (3.20) имела  $n$  независимых параметров  $\{u_j\}_{j=1..n}$ , то система в вариациях (3.24) имеет  $m + n$  *независимых вариаций*  $\{\delta v_i, \delta u_j\}_{i=1..m, j=1..n}$ .

Поскольку вариации  $\delta v$ ,  $\delta u$  теперь независимы, то уравнения (3.24) можно и следует преобразовать к значениям линейных функционалов непосредственно на элементах (компонентах элементов)  $\delta v$ ,  $\delta u$ . Это преобразование иногда называют тождеством Лагранжа. Оно использует теорему Грина, которая преобразует, например, поверхностный интеграл к контурному. Итак, мы получаем из (3.24) следующее тождество:

$$\begin{aligned}
\langle \tilde{f}, \mathbb{V} \delta v \rangle_{V^*(\Omega)} + \langle \tilde{f}, \mathbb{U} \delta u \rangle_{V^*(\Omega)} & \quad (3.25) \\
& \equiv \langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\Omega)} + \langle \bar{\mathbb{V}}^* \tilde{f}, \delta v \rangle_{V^*(\partial\Omega)} \\
& + \langle \bar{\bar{\mathbb{V}}}^* \tilde{f}, \frac{\partial \delta v}{\partial \tau} \rangle_{V^*(\partial\Omega)} + \dots + \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)} = 0 \in E,
\end{aligned}$$

где  $\mathbb{V}^*$ ,  $\mathbb{U}^*$  — линейные сопряжённые операторы, причём  $\mathbb{V}^*$  действует из  $V^*(\Omega)$  в  $V^*(\Omega)$ , а  $\mathbb{U}^*$  действует из  $V^*(\Omega)$  в сопряжённое пространство управлений  $U^*(S)$ . Операторы  $\bar{\mathbb{V}}^*$ ,  $\bar{\bar{\mathbb{V}}}^*$  — граничные сопряжённые операторы, действующие на границе (на контуре)  $\partial\Omega$ , где значения линейных функционалов  $\bar{\mathbb{V}}^* \tilde{f}$  и  $\bar{\bar{\mathbb{V}}}^* \tilde{f}$  определяются на вариациях  $\delta v$  (одна черта над  $\bar{\mathbb{V}}^*$ ) и её производных  $\frac{\partial \delta v}{\partial \tau}$  (две черты  $\bar{\bar{\mathbb{V}}}^*$ ) соответственно.

Многоточие в (3.25) имеет место, если порядок  $r$  дифференцирования состояния  $v$  по  $\tau$  распределённой системы более чем второй, при этом на границе, согласно теореме Грина, появятся дополнительные линейные функционалы, действующие на производные вариаций до порядка  $\frac{\partial^{r-1} \delta v}{\partial \tau^{r-1}}$ . Символическую запись *производной*  $\frac{\partial \cdot}{\partial \tau}$  по вектору пространственно-временных переменных  $\tau$  размерности  $N + 1$  (пространство плюс время) следует понимать как сумму производных по всем компонентам  $\tau$ , т.е.  $\frac{\partial \cdot}{\partial \tau} = \sum_{j=1}^{N+1} \frac{\partial \cdot}{\partial \tau_j}$ . Далее, для облегчения понимания сути преобразований, мы не будем в общих выражениях подробно расписывать эту производную.

Реализация тождества Лагранжа (3.25) — это, пожалуй, самая трудоёмкая часть работы по получению аналитического выражения градиента целевого функционала, да и всей задачи оптимизации.

Давайте подробнее остановимся на сути преобразований (3.25). Разберём их на примере предыдущего оператора  $\mathbb{D} \cdot = A(x, v, u) \frac{\partial^r \cdot}{\partial x^r}$ . Пусть  $v \in L_2^m$ , где  $m > 1$ ,  $u \in L_2(S)$   $r \in \{1, 2\}$ , что может соответствовать одномерной стационарной системе  $m$  уравнений с первыми или вторыми

производными от состояния  $v$ . Область  $\Omega=(x_a, x_b) = S$ , граница  $\partial\Omega = x_a \cup x_b$ .

При этом первое слагаемое (относительно  $\delta v$ ) в левой части тождества Лагранжа (3.25) принимает вид:

$$\langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} = \int_{x_a}^{x_b} \tilde{f}^T \left[ A \frac{\partial^r \delta v}{\partial x^r} + \mathbb{C}_r \delta v + F'_v \delta v \right] dx.$$

Необходимо на  $\Omega$  избавиться от производной вариации  $\frac{\partial^r \delta v}{\partial x^r}$ . Для этого можно использовать *формулу Грина* в подходящем виде, а если говорить проще, то достаточно использовать правило *дифференцирования произведений*. Возьмём первое слагаемое, занесём его под производную и вычтем лишние члены:

$$\begin{aligned} \int_{x_a}^{x_b} \tilde{f}^T A \frac{\partial^r \delta v}{\partial x^r} dx & \quad (3.26) \\ &= \int_{x_a}^{x_b} \left[ \frac{\partial}{\partial x} \left( \tilde{f}^T A \frac{\partial^{r-1} \delta v}{\partial x^{r-1}} \right) - \frac{\partial \tilde{f}^T}{\partial x} A \frac{\partial^{r-1} \delta v}{\partial x^{r-1}} \right. \\ &\quad \left. - \tilde{f}^T \frac{\partial A}{\partial x} \frac{\partial^{r-1} \delta v}{\partial x^{r-1}} \right] dx \\ &= \tilde{f}^T A \frac{\partial^{r-1} \delta v}{\partial x^{r-1}} \Big|_{x_a}^{x_b} \\ &\quad + \int_{x_a}^{x_b} \left[ - \frac{\partial \tilde{f}^T}{\partial x} A - \tilde{f}^T \frac{\partial A}{\partial x} \right] \frac{\partial^{r-1} \delta v}{\partial x^{r-1}} dx. \end{aligned}$$

Если  $r = 1$ , то на этом мы и остановимся. Результат очевидного интегрирования производной дал нам граничные члены на  $x_a$  и  $x_b$ . Именно здесь и появляются в тождестве Лагранжа (3.25) слагаемые на  $\partial\Omega$ .

Итак, для  $r = 1$ , оценим полученные результаты. С учётом необходимости *транспонирования* векторов и матриц при их перестановке, например, для векторов  $a$ ,  $b$  и матрицы  $Z$ , когда

$$a^T Z = (Z^T a)^T \text{ или } \langle a, Zb \rangle = \langle Z^T a, b \rangle,$$

и с учётом того, что «нулевая» производная  $\frac{\partial^0 \delta v}{\partial x^0} = \delta v$ , мы получаем первое слагаемое в тождестве Лагранжа (3.25) относительно  $\delta v$  в виде:

$$\begin{aligned} & \langle \tilde{f}, \mathbb{V} \delta v \rangle_{V^*(\Omega)} \\ &= \langle -A^T \frac{\partial \tilde{f}}{\partial x} + \left( -\frac{\partial A^T}{\partial x} + \mathbb{C}_1^* + F_v'^T \right) \tilde{f}, \delta v \rangle_{V^*(\Omega)} \\ &+ \langle A^T \tilde{f}, \delta v \rangle_{V^*(x_b)} - \langle A^T \tilde{f}, \delta v \rangle_{V^*(x_a)}, \end{aligned} \quad (3.27)$$

где  $\mathbb{C}_1^*$  — недифференциальный матричный оператор, сопряжённый с  $\mathbb{C}_1$ , действующий на  $\tilde{f}$ . Напомним, что исходный оператор  $\mathbb{C}_1$  действует на  $\delta v$  так, что  $\mathbb{C}_1 \delta v = \sum_{i=1}^m \frac{\partial A}{\partial v_i} \delta v_i \frac{\partial v}{\partial x}$ . Операция  $\mathbb{C}_1^* \tilde{f}$  не приводит к появлению производных от сопряжённой переменной  $\tilde{f}$ .

Вид сопряжённого оператора  $\mathbb{C}_1^*$  определяется тождеством

$$\langle \tilde{f}, \mathbb{C}_1 \delta v \rangle_{V^*(\Omega)} \equiv \langle \mathbb{C}_1^* \tilde{f}, \delta v \rangle_{V^*(\Omega)}.$$

Вычисление  $\mathbb{C}_1^*$  целесообразно выполнять непосредственно с данными конкретной задачи. Общий вид этого оператора не несёт в себе какой-либо теоретической, смысловой нагрузки. Однако ради любопытства и дальнейшего его использования, найдём вид  $\mathbb{C}_1^*$ .

В нашем случае, при  $m$ -мерной размерности вектор-функций  $\tilde{f}$ ,  $v$ , мы получаем:

$$\begin{aligned}
\langle \tilde{f}, \mathbb{C}_1 \delta v \rangle &= \int_{\Omega} \left[ \tilde{f}^T \left( \frac{\partial A}{\partial v_1} \frac{\partial v}{\partial x} \delta v_1 + \dots + \frac{\partial A}{\partial v_m} \frac{\partial v}{\partial x} \delta v_m \right) \right] d\Omega \\
&= \int_{\Omega} \left[ \left( \frac{\partial A}{\partial v_1} \frac{\partial v}{\partial x} \right)^T \tilde{f} \delta v_1 + \dots + \left( \frac{\partial A}{\partial v_m} \frac{\partial v}{\partial x} \right)^T \tilde{f} \delta v_m \right] d\Omega \\
&= \left\langle \left( \left( \frac{\partial A}{\partial v_1} \frac{\partial v}{\partial x} \right)^T \tilde{f}, \dots, \left( \frac{\partial A}{\partial v_m} \frac{\partial v}{\partial x} \right)^T \tilde{f} \right), \delta v \right\rangle \\
&= \left\langle \begin{pmatrix} \left( \frac{\partial A}{\partial v_1} \frac{\partial v}{\partial x} \right)^T \\ \dots \\ \left( \frac{\partial A}{\partial v_m} \frac{\partial v}{\partial x} \right)^T \end{pmatrix} \tilde{f}, \delta v \right\rangle \equiv \langle \mathbb{C}_1^* \tilde{f}, \delta v \rangle.
\end{aligned}$$

Следовательно, сопряжённый оператор  $\mathbb{C}_1^*$  — это матрица

$$\mathbb{C}_1^* = \begin{pmatrix} \left( \frac{\partial A}{\partial v_1} \frac{\partial v}{\partial x} \right)^T \\ \dots \\ \left( \frac{\partial A}{\partial v_m} \frac{\partial v}{\partial x} \right)^T \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^m \frac{\partial a_{1j}}{\partial v_1} \frac{\partial v_j}{\partial x} & \dots & \sum_{j=1}^m \frac{\partial a_{mj}}{\partial v_1} \frac{\partial v_j}{\partial x} \\ \dots & \dots & \dots \\ \sum_{j=1}^m \frac{\partial a_{1j}}{\partial v_m} \frac{\partial v_j}{\partial x} & \dots & \sum_{j=1}^m \frac{\partial a_{mj}}{\partial v_m} \frac{\partial v_j}{\partial x} \end{pmatrix}, \quad (3.28)$$

где каждая  $i$ -тая строка матрицы — это вектор-строка  $\left( \frac{\partial A}{\partial v_i} \frac{\partial v}{\partial x} \right)^T$ .

Обращаем внимание, что при вычислении *производной*  $\frac{\partial A^T}{\partial x}$  в (3.27) необходимо учитывать зависимость  $A = A(x, v(x), u(x))$ . При этом производная

$$\begin{aligned}
\frac{\partial A^T(x, v, u)}{\partial x} &= \sum_{i=1}^m \frac{\partial A^T}{\partial v_i} \frac{\partial v_i}{\partial x} + \sum_{i=1}^m \frac{\partial A^T}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial A^T}{\partial x} \Big|_{v, u = \text{const}} \\
&\equiv \left( \frac{\partial A^T}{\partial v} \right)^T \frac{\partial v}{\partial x} + \left( \frac{\partial A^T}{\partial u} \right)^T \frac{\partial u}{\partial x} + \frac{\partial A^T}{\partial x} \Big|_{v, u}.
\end{aligned}$$

Если сравнить приведенное выражение с матрицей оператора  $\mathbb{C}_1^*$ , то можно понять, что удобнее находить не отдельно оператор  $\mathbb{C}_1^*$ , а

совместно с производной  $\frac{\partial A^T}{\partial x}$ . При этом многие слагаемые в  $\mathbb{C}_1^*$  и  $\frac{\partial A^T}{\partial x}$  сократятся.

Теперь давайте сравним выражение (3.27) с (3.25) и увидим, что в рассматриваемом примере *сопряжённый оператор*

$$\mathbb{V}^* \cdot = -A^T \frac{\partial \cdot}{\partial x} - \frac{\partial A^T}{\partial x} \cdot + \mathbb{C}_1^* \cdot + F_v'^T \cdot, \quad \text{для } r = 1,$$

а граничные операторы

$$\bar{\mathbb{V}}^* \cdot = A^T \cdot |_{x_b} - A^T \cdot |_{x_a}, \quad \bar{\bar{\mathbb{V}}}^* = 0.$$

Теперь возьмём  $r = 2$ , когда управляемая система содержит только вторые производные от состояния  $v$ . Чтобы избавиться от вторых производных вариаций  $\frac{\partial^2 \delta v}{\partial x^2}$  нам придётся либо дважды воспользоваться правилом дифференцирования произведений, либо изначально избавиться от вторых производных, введя дополнительные переменные (по сути, уравнения) в виде первых производных от состояний  $v$  [33, 64]. Выберем далее первое.

Ранее, применив один раз правило *дифференцирования произведений* к слагаемому с производной  $\frac{\partial^r \delta v}{\partial x^r}$ , мы получили выражение (3.26) с  $\frac{\partial^{r-1} \delta v}{\partial x^{r-1}}$ . При  $r = 2$  оно принимает вид:

$$\int_{x_a}^{x_b} \tilde{f}^T A \frac{\partial^2 \delta v}{\partial x^2} dx = \tilde{f}^T A \frac{\partial \delta v}{\partial x} \Big|_{x_a}^{x_b} + \int_{x_a}^{x_b} \left[ -\frac{\partial \tilde{f}^T}{\partial x} A - \tilde{f}^T \frac{\partial A}{\partial x} \right] \frac{\partial \delta v}{\partial x} dx.$$

Теперь избавимся от первой производной вариации  $\frac{\partial \delta v}{\partial x}$  под интегралом справа, применив очередной раз правило дифференцирования произведений. В итоге получим:

$$\begin{aligned}
\int_{x_a}^{x_b} \tilde{f}^T A \frac{\partial^2 \delta v}{\partial x^2} dx &= \tilde{f}^T A \frac{\partial \delta v}{\partial x} \Big|_{x_a}^{x_b} \\
&+ \int_{x_a}^{x_b} \left[ \frac{\partial}{\partial x} \left( \left( -\frac{\partial \tilde{f}^T}{\partial x} A - \tilde{f}^T \frac{\partial A}{\partial x} \right) \delta v \right) \right. \\
&\left. + \left( \frac{\partial^2 \tilde{f}^T}{\partial x^2} A + \frac{\partial \tilde{f}^T}{\partial x} \frac{\partial A}{\partial x} + \frac{\partial \tilde{f}^T}{\partial x} \frac{\partial A}{\partial x} + \tilde{f}^T \frac{\partial^2 A}{\partial x^2} \right) \delta v \right] dx.
\end{aligned}$$

При этом, часть тождества Лагранжа (3.25) при  $\delta v$  принимает вид:

$$\langle \tilde{f}, \mathbb{V} \delta v \rangle_{V^*(\Omega)} \quad (3.29)$$

$$\begin{aligned}
&= \langle A^T \frac{\partial^2 \tilde{f}}{\partial x^2} + 2 \frac{\partial A^T}{\partial x} \frac{\partial \tilde{f}}{\partial x} \\
&+ \left( \frac{\partial^2 A^T}{\partial x^2} + \mathbb{C}_2^* + F_v'^T \right) \tilde{f}, \delta v \rangle_{V^*(\Omega)} \\
&+ \langle A^T \tilde{f}, \frac{\partial \delta v}{\partial x} \rangle_{V^*(x_b)} - \langle A^T \tilde{f}, \frac{\partial \delta v}{\partial x} \rangle_{V^*(x_a)} \\
&+ \langle -A^T \frac{\partial \tilde{f}}{\partial x} - \frac{\partial A^T}{\partial x} \tilde{f}, \delta v \rangle_{V^*(x_b)} \\
&- \langle -A^T \frac{\partial \tilde{f}}{\partial x} - \frac{\partial A^T}{\partial x} \tilde{f}, \delta v \rangle_{V^*(x_a)}.
\end{aligned}$$

где  $\mathbb{C}_2^*$  — это матричный оператор, сопряжённый с  $\mathbb{C}_2 := \sum_{i=1}^m \frac{\partial A}{\partial v_i} \cdot |i \frac{\partial^2 v}{\partial x^2}$ .

По аналогии с  $\mathbb{C}_1^*$  он имеет вид:

$$\mathbb{C}_2^* := \begin{pmatrix} \left( \left( \frac{\partial A}{\partial v_1} \frac{\partial^2 v}{\partial x^2} \right)^T \right) \\ \dots \\ \left( \left( \frac{\partial A}{\partial v_m} \frac{\partial^2 v}{\partial x^2} \right)^T \right) \end{pmatrix}.$$

При вычислении второй производной  $\frac{\partial^2 A^T}{\partial x^2}$  в (3.29) опять же необходимо учитывать зависимость  $A = A(x, v(x), u(x))$ .

Сравнивая выражение (3.29) с (3.25), мы видим, что в рассматриваемом примере сопряжённый оператор

$$\mathbb{V}^* \cdot = A^T \frac{\partial^2 \cdot}{\partial x^2} + 2 \frac{\partial A^T}{\partial x} \frac{\partial \cdot}{\partial x} + \frac{\partial^2 A^T}{\partial x^2} \cdot + \mathbb{C}_2^* \cdot + F_v'^T \cdot, \quad \text{для } r = 2,$$

а граничные операторы

$$\bar{\mathbb{V}}^* \cdot = \left( -A^T \frac{\partial \cdot}{\partial x} - \frac{\partial A^T}{\partial x} \cdot \right) \Big|_{x_b} + \left( \frac{\partial \cdot}{\partial x} + \frac{\partial A^T}{\partial x} \cdot \right) \Big|_{x_a},$$

$$\bar{\bar{\mathbb{V}}}^* \cdot = A^T \cdot |_{x_b} - A^T \cdot |_{x_a}.$$

При практической численной реализации оператора  $\mathbb{V}^*$  в области  $\Omega$ , возможно, будет удобнее его реализовывать совместно с линейным функционалом  $\tilde{f}$ , тогда получим более компактное выражение для оператора  $\mathbb{V}^*$ :

$$\mathbb{V}^* \tilde{f} = \frac{\partial^2 A^T \tilde{f}}{\partial x^2} + \mathbb{C}_2^* \tilde{f} + F_v'^T \tilde{f}.$$

Немного позже нам пригодится вид сопряжённого оператора  $\mathbb{V}^*$  от смешанной второй производной для двумерного случая, когда  $\tau = (x_1, x_2)$ , т.е. исходный оператор  $\mathbb{D} \cdot = A(\tau, v, u) \frac{\partial^2 \cdot}{\partial x_1 \partial x_2}$ . Давайте в этом примере найдём его в прямоугольной области  $\Omega = (x_{1a}, x_{1b}) \times (x_{2a}, x_{2b})$ .

Вариация указанного оператора  $\mathbb{D}$  относительно  $\delta v$  в выражении (3.23) приводит к слагаемому  $\mathbb{V} \delta v$ , где

$$\mathbb{V} \cdot = A \frac{\partial^2 \cdot}{\partial x_1 \partial x_2} + \mathbb{C} \cdot + F_v', \quad \mathbb{C} = \sum_{i=1}^m \frac{\partial A}{\partial v_i} \cdot |_{i} \frac{\partial^2 v}{\partial x_1 \partial x_2}.$$

Значение линейного функционала  $\tilde{f}$  на  $\mathbb{V} \delta v$  принимает вид:

$$\langle \tilde{f}, \mathbb{V} \delta v \rangle_{V^*(\Omega)}$$

$$= \iint_{x_{1a}, x_{2a}}^{x_{1b}, x_{2b}} \tilde{f}^T \left[ A \frac{\partial^2 \delta v}{\partial x_1 \partial x_2} + \sum_{i=1}^m \frac{\partial A}{\partial v_i} \delta v_i \frac{\partial^2 v}{\partial x_1 \partial x_2} + F_v' \delta v \right] dx_1 dx_2.$$

Преобразуем первое слагаемое со смешанной производной вариации  $\delta v$  и избавимся от производной  $\frac{\partial \delta v}{\partial x_1}$ :

$$\begin{aligned}
& \iint_{x_{1a}, x_{2a}}^{x_{1b}, x_{2b}} \tilde{f}^T A \frac{\partial^2 \delta v}{\partial x_1 \partial x_2} dx_1 dx_2 \\
&= \iint_{x_{1a}, x_{2a}}^{x_{1b}, x_{2b}} \left[ \frac{\partial}{\partial x_1} \left( \tilde{f}^T A \frac{\partial \delta v}{\partial x_2} \right) - \frac{\partial \tilde{f}^T}{\partial x_1} A \frac{\partial \delta v}{\partial x_2} - \tilde{f}^T \frac{\partial A}{\partial x_1} \frac{\partial \delta v}{\partial x_2} \right] dx_1 dx_2 \\
&= \int_{x_{2a}}^{x_{2b}} \tilde{f}^T A \frac{\partial \delta v}{\partial x_2} \Big|_{x_{1a}}^{x_{1b}} dx_2 \\
&+ \iint_{x_{1a}, x_{2a}}^{x_{1b}, x_{2b}} \left[ -\frac{\partial \tilde{f}^T}{\partial x_1} A - \tilde{f}^T \frac{\partial A}{\partial x_1} \right] \frac{\partial \delta v}{\partial x_2} dx_1 dx_2.
\end{aligned}$$

Теперь избавимся от производной  $\frac{\partial \delta v}{\partial x_2}$  в последнем интеграле и получим

$$\begin{aligned}
& \iint_{x_{1a}, x_{2a}}^{x_{1b}, x_{2b}} \tilde{f}^T A \frac{\partial^2 \delta v}{\partial x_1 \partial x_2} dx_1 dx_2 \\
&= \int_{x_{2a}}^{x_{2b}} \tilde{f}^T A \frac{\partial \delta v}{\partial x_2} \Big|_{x_{1a}}^{x_{1b}} dx_2 \\
&+ \iint_{x_{1a}, x_{2a}}^{x_{1b}, x_{2b}} \left[ \frac{\partial}{\partial x_2} \left( \left( -\frac{\partial \tilde{f}^T}{\partial x_1} A - \tilde{f}^T \frac{\partial A}{\partial x_1} \right) \delta v \right) \right. \\
&+ \left( \frac{\partial^2 \tilde{f}^T}{\partial x_1 \partial x_2} A + \frac{\partial \tilde{f}^T}{\partial x_1} \frac{\partial A}{\partial x_2} + \frac{\partial \tilde{f}^T}{\partial x_2} \frac{\partial A}{\partial x_1} \right. \\
&\left. \left. + \tilde{f}^T \frac{\partial^2 A}{\partial x_1 \partial x_2} \right) \delta v \right] dx_1 dx_2.
\end{aligned}$$

В итоге соответствующая часть тождества Лагранжа (3.25) принимает вид:

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{V} \delta v \rangle_{V^*(\Omega)} \\
&= \langle A^T \frac{\partial^2 \tilde{f}}{\partial x_1 \partial x_2} + \frac{\partial A^T}{\partial x_2} \frac{\partial \tilde{f}}{\partial x_1} + \frac{\partial A^T}{\partial x_1} \frac{\partial \tilde{f}}{\partial x_2} \\
&+ \left( \frac{\partial^2 A^T}{\partial x_1 \partial x_2} + \mathbb{C}^* + F_v'^T \right) \tilde{f}, \delta v \rangle_{V^*(\Omega)} \\
&+ \langle A^T \tilde{f}, \frac{\partial \delta v}{\partial x_2} \rangle_{V^*(x_{1b})} - \langle A^T \tilde{f}, \frac{\partial \delta v}{\partial x_2} \rangle_{V^*(x_{1a})} \\
&+ \langle -A^T \frac{\partial \tilde{f}}{\partial x_1} - \frac{\partial A^T}{\partial x_1} \tilde{f}, \delta v \rangle_{V^*(x_{2b})} \\
&- \langle -A^T \frac{\partial \tilde{f}}{\partial x_1} - \frac{\partial A^T}{\partial x_1} \tilde{f}, \delta v \rangle_{V^*(x_{2a})}.
\end{aligned}$$

Итак, в последнем примере со смешанной производной мы получили в  $\Omega$  сопряжённый оператор

$$\mathbb{V}^* \cdot = A^T \frac{\partial^2 \cdot}{\partial x_1 \partial x_2} + \frac{\partial A^T}{\partial x_1} \frac{\partial \cdot}{\partial x_2} + \frac{\partial A^T}{\partial x_2} \frac{\partial \cdot}{\partial x_1} + \left( \frac{\partial^2 A^T}{\partial x_1 \partial x_2} + \mathbb{C}^* + F_v'^T \right) \cdot.$$

Если его вычисление реализовывать совместно с линейным функционалом  $\tilde{f}$ , то получим более компактное выражение:

$$\mathbb{V}^* \tilde{f} = \frac{\partial^2 A^T \tilde{f}}{\partial x_1 \partial x_2} + \mathbb{C}^* \tilde{f} + F_v'^T \tilde{f}.$$

Все полученные сопряжённые дифференциальные операторы  $\mathbb{V}^*$  задают отображение:

$$\mathbb{V}^* : V^*(\Omega) \rightarrow V^*(\Omega).$$

Мы рассмотрели не совсем простые примеры операторов  $\mathbb{D}$ , когда в матрице  $A$  имеет место зависимость  $A(x, v, u)$ . Выкладки стали бы на много проще и короче, если бы мы использовали наиболее частые в практике зависимости типа  $A(x)$  или даже  $A = \text{const}$ . Тем не менее, рассмотренные примеры следует считать полезными и показательными. Они нам ещё пригодятся.

Вид сопряжённого оператора  $\mathbb{V}^*$  определяет тип сопряжённых дифференциальных уравнений, что очень важно для их анализа (в том числе анализа управляемости) и построения алгоритмов численного решения. Форма, в которой мы получили разные виды оператора  $\mathbb{V}^*$ , позволяет относительно легко провести необходимый анализ.

Теперь обратимся к тождеству Лагранжа (3.25) с вариацией  $\delta u$ , которое справа содержит линейный сопряжённый оператор  $\mathbb{U}^*$ . Реальный вид  $\mathbb{U}^*$  следует вычислять аналогично  $\mathbb{C}_r^*$  непосредственно из тождества

$$\langle \tilde{f}, \mathbb{U}\delta u \rangle_{V^*(\Omega)} = \langle \mathbb{U}^* \tilde{f}, \delta v \rangle_{V^*(S)}.$$

Здесь множество  $S$  может совпадать со всей областью  $\Omega$ , или быть её частью, её проекцией на какие-либо оси координат. Например, в одномерной нестационарной задаче возможно управление  $u(x) \in L_2(S)$ , которое должно быть определено на всей пространственно-временной области  $\Omega$ . Если  $\Omega = (x_a, x_b) \times (t_0, t_1)$ , то  $S = P_x(\Omega) = (x_a, x_b)$ , т.е.  $S$  — это проекция прямоугольника  $\Omega$  на ось  $x$ .

Наиболее часто встречаются следующие частные случаи сопряжённого оператора  $\mathbb{U}^*$ .

Первый случай. Если рассматриваемая ранее матрица  $A$  зависит от управления  $u(\tau) \in L_2^n(S = \Omega)$ , а размерность  $n = 1$ , то оператор  $\mathbb{U}^*$  (для первых двух примеров оператора  $\mathbb{D}$ , без смешанных производных) имеет вид:

$$\mathbb{U}^* = \left( \frac{\partial A}{\partial u} \frac{\partial^r v}{\partial x^r} \right)^T + F'_u{}^T.$$

Второй случай, когда матрица  $A$  не зависит от управления, а указанное управление находится только в свободном члене  $F$ . Имеем

$$\mathbb{U}^* = F'_u{}^T.$$

В любом случае сопряжённый оператор  $\mathbb{U}^*$  реализует *отображение* из пространства сопряжённых состояний в сопряжённое пространство управлений, где может существовать градиент  $\nabla J$ :

$$\mathbb{U}^* : V^*(\Omega) \rightarrow U^*(S).$$

Третий случай, — это когда управление входит в краевые условия, но об этом мы поговорим чуть позже, когда будем рассматривать именно краевые условия.

#### 3.4.4. Объединение задачи

Переходим ко второму этапу получения градиента  $\nabla J$ , к объединению полученных линейных элементов задачи.

Линеаризованные выражения (3.22) и (3.25) задачи оптимизации записаны в одном и том же пространстве  $E$  и представляют собой значения линейных функционалов из сопряжённых пространств  $V^*$  и  $U^*$  на элементах  $\delta v$ ,  $\frac{\partial^r \delta v}{\partial \tau^r}$ ,  $\delta u$ . Поскольку линейные функционалы обладают свойством аддитивности, то мы можем их объединить в соответствующих пространствах на элементах, к которым они относятся.

Таким образом, мы получаем выражение вариации целевого функционала (без учёта краевых условий):

$$\begin{aligned} \delta J(u) = & \langle \mathbb{V}^* \tilde{f} + I'_v|_{\omega}, \delta v \rangle_{V^*(\Omega)} + \langle \bar{\mathbb{V}}^* \tilde{f} + I'_v|_{\omega}, \delta v \rangle_{V^*(\partial\Omega)} \\ & + \langle \bar{\mathbb{V}}^* \tilde{f}, \frac{\partial \delta v}{\partial \tau} \rangle_{V^*(\partial\Omega)} + \dots + \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)}, \end{aligned} \quad (3.30)$$

где

$$\mathbb{U}^* \cdot = \mathbb{U}^* \cdot + I'_u|_{\omega}.$$

линейный, но *неоднородный* оператор, который отличается от «традиционно» линейного  $\mathbb{U}^*$  добавлением к нему слагаемого  $I'_u|_{\omega}$ , независящего от  $\tilde{f}$ . Неоднородный оператор  $\mathbb{U}^*$ , вместо однородного  $\mathbb{U}^*$ ,

появляется в тех задачах, где функция цели  $I$  зависит явно от параметра  $u$ .

Добавим ещё. Особенностью линейных неоднородных операторов, в частности  $\mathbb{U}_\emptyset^*$ , является то, что они принимают нулевые значения на *ненулевых* элементах своей области определения, т.е.  $\mathbb{U}_\emptyset^* 0 \neq 0$ . Нижний индекс  $\emptyset$  как раз означает отсутствие нулевого ядра у линейного оператора. В то же время, при оптимальном управлении найдётся  $\tilde{f} \neq 0$  при котором значение  $\mathbb{U}_\emptyset^* \tilde{f} = 0$ . Подробнее об этом и зачем нам это, будет рассмотрено позже при обсуждении проблем управляемости и идентифицируемости.

### 3.4.5. Учёт краевых условий

Учесть *краевые*, т.е. граничные и начальные условия, можно следующими двумя способами.

Первый способ. Мы должны сделать все шаги, описанные выше: линеаризовать краевые условия  $G$  в (3.20) посредством варьирования, ввести дополнительные линейные функционалы (множители Лагранжа) на  $\Gamma$  и получить сопряженные операторы для независимых краевых вариаций  $\delta v$  и  $\delta u$ . После этого мы можем объединить полученные граничные и начальные элементы с уже имеющимися в (3.30) элементами на границе  $\partial\Omega$  и найти значения дополнительных множителей Лагранжа, которые будут учитывать начальные и граничные условия исходной задачи.

Второй способ [20] проще и естественнее. Рекомендуем использовать именно его. С одной стороны, выражение (3.30) для первой вариации  $\delta J$  содержит *независимые вариации*  $\{\delta v_i, \delta u_j\}_{i=1..m, j=1..n}$  на  $\Omega$  и, что важно сейчас — на границе  $\partial\Omega$ . С другой стороны, после *варьирования краевых условий*  $G$  исходной задачи (3.20), мы по-

прежнему будем иметь в них *зависимые вариации* в виде нескольких уравнений, которые запишем в следующей символической форме:

$$\delta G = G'_v \delta v + G'_u \delta u + G'_{v'} \frac{\partial \delta v}{\partial x} = 0 \in V(\Gamma). \quad (3.31)$$

В этом примере и далее мы ограничимся рассмотрением одномерной нестационарной распределённой системы с одним уравнением ( $m = 1$ ) в области  $\Omega = (x_a, x_b) \times (t_0, t_1)$  и производными порядка  $r \leq 2$ . При этом область краевых условий системы  $\Gamma = \Gamma_0 \cup \Gamma_\partial$ , где линия начальных условий  $\Gamma_0 = [x_a, x_b] \times t_0$ , область граничных условий  $\Gamma_\partial = \Gamma_a \cup \Gamma_b$ , при этом  $\Gamma_a = x_a \times (t_0, t_1)$ ,  $\Gamma_b = x_b \times (t_0, t_1)$ . Вся граница области  $\Omega$  — это прямоугольник  $\partial\Omega = \Gamma \cup \Gamma_1$ , где  $\Gamma_1 = [x_a, x_b] \times t_1$  — линия терминальных условий.

Необходимо из множества уравнений (3.31) на разных частях  $\Gamma$  выразить значения вариаций  $\delta v$ ,  $\delta u$ ,  $\frac{\partial \delta v}{\partial x}$  и подставить их в ранее полученную первую вариацию  $\delta J$  целевого функционала (3.30) в соответствующие части на  $\partial\Omega$ . Таким образом мы учтём начальные и граничные условия исходной задачи в первой вариации  $\delta J$ . В общем виде это сделать будет громоздко и неочевидно, кроме того, какого-либо теоретического смысла такие обобщения не имеют. Поэтому мы рассмотрим вкратце несколько типичных примеров, иллюстрирующих суть данного способа.

**Пример 1.** Управление в граничных условиях на  $\Gamma_\partial$  отсутствует, т.е.  $G'_u \delta u = 0$ . Это означает, что управление  $u$  может быть определено на  $S = \Omega$ , как мы это и предполагали ранее в вариации  $\delta J$  (3.30).

Если нестационарная распределённая система имеет первую производную по времени и вторую по пространству, то, например, при граничном условии *первого рода*<sup>1</sup> на левой границе  $\Gamma_a$  и граничном

<sup>1</sup> Граничное условие *первого рода* (условие Дирихле) — это условие существования известного решения на границе, например,  $v(x_a, t) = g_a(x_a, t)$ .

условии *второго рода*<sup>1</sup> на правой границе  $\Gamma_b$ , мы будем иметь следующую функцию  $G$ :

$$G = \begin{cases} G|_{\Gamma_a} = v - g_a = 0, \\ G|_{\Gamma_b} = \frac{\partial v}{\partial x} - g_b = 0, \\ G|_{\Gamma_0} = v - g_0 = 0, \end{cases}$$

где  $g_a(x_a, t)$ ,  $g_b(x_b, t)$  — заданные функции граничных условий на  $\Gamma_a$ ,  $\Gamma_b$ , и  $g_0(x, t_0)$  — функция известных начальных условий.

При этом вариация краевых условий  $\delta G$  примет вид:

$$\begin{aligned} \delta G|_{\Gamma_a} = \delta v = 0, \quad G'_v = 1, G'_u = 0, G'_{v'} = 0, \\ \delta G|_{\Gamma_b} = \frac{\partial \delta v}{\partial x} = 0, \quad G'_v = 0, G'_u = 0, G'_{v'} = 1, \\ \delta G|_{\Gamma_0} = \delta v = 0, \quad G'_v = 1, G'_u = 0, G'_{v'} = 0. \end{aligned}$$

Теперь, полученные значения вариаций  $\delta v$ ,  $\frac{\partial \delta v}{\partial x}$  надо подставить в граничные элементы на соответствующих частях  $\partial\Omega$  полученной ранее в (3.30) вариации  $\delta J$ .

В выражении (3.30) произойдут следующие изменения. Граничные слагаемые из члена  $\langle \bar{V}^* \tilde{f} + I'_v|_{\omega}, \delta v \rangle_{V^*(\partial\Omega)}$ , при вариации краевых условий  $\delta v = 0$  на  $\Gamma_0$  и  $\Gamma_a$ , исчезнут. Аналогично, из слагаемого  $\langle \bar{V}^* \tilde{f}, \frac{\partial \delta v}{\partial x} \rangle_{V^*(\partial\Omega = \Gamma_a \cup \Gamma_b)}$  исчезнет вариация  $\frac{\partial \delta v}{\partial x} = 0$  на  $\Gamma_b$ . В итоге, краевые вариации состояния на  $\partial\Omega$  в выражении (3.30) примут вид:

$$\langle \bar{V}^* \tilde{f} + I'_v|_{\omega}, \delta v \rangle_{V^*(\Gamma_1 \cup \Gamma_b)} + \langle \bar{V}^* \tilde{f}, \frac{\partial \delta v}{\partial x} \rangle_{V^*(\Gamma_a)}.$$

То есть, вместо вариации  $\delta v$  на  $\Gamma_0$  и  $\Gamma_a$ , а также вариации  $\frac{\partial \delta v}{\partial x}$  на  $\Gamma_b$  мы получили выражения при вариации  $\delta v$  на  $\Gamma_1$ ,  $\Gamma_b$  и при вариации  $\frac{\partial \delta v}{\partial x}$  на  $\Gamma_a$ .

<sup>1</sup> Граничное условие *второго рода* (условие Неймана) — это условие существования известного значения первой производной решения на границе, например,  $\frac{\partial v}{\partial x}(x_b, t) = g_b(x_b, t)$ .

Внешне это выглядит так, как будто области определения краевых вариаций исходной задачи в сопряжённых пространствах *переворачиваются*. Далее будем говорить о сопряжённых областях  $\Gamma_a^* \equiv \Gamma_b$ ,  $\Gamma_b^* \equiv \Gamma_a$ ,  $\Gamma_0^* \equiv \Gamma_1$ , на которых после «переворота» остаются в  $\delta J$  краевые вариации.

**Пример 2.** Рассмотрим левое граничное условие *первого рода с управлением* на  $\Gamma_a$ , когда функция  $G$  имеет вид:

$$G|_{\Gamma_a} = \lambda v - u = 0 \in V(\Gamma_a),$$

где  $\lambda$  — некоторое вещественное число,  $\Gamma_a = S \equiv \Gamma_{\partial 1}$  — область определения управления в граничном условии первого рода (дополнительный нижний индекс 1). Предполагаем, что других управлений с иной областью определения в задаче нет.

Поскольку для реальных задач оптимизации распределённых систем не характерно присутствие целевого функционала на той же границе, что и управление, то в нашем примере давайте в вариации (3.30) уберём слагаемые  $I'_v|_{\omega}$ ,  $I'_u|_{\omega}$ , формально попадающие на границу  $\Gamma_a$ .

Подставим вариацию граничного состояния  $\delta v = \lambda^{-1} \delta u$  в вариацию (3.30) и получим на границах:

$$\langle \bar{V}^* \tilde{f} + I'_v|_{\omega}, \delta v \rangle_{V^*(\partial\Omega=\Gamma_a \cup \Gamma_b)} = \langle \bar{U}^* \tilde{f}, \delta u \rangle_{V^*(S)} + \langle \bar{V}^* \tilde{f} + I'_v|_{\omega}, \delta v \rangle_{V^*(\Gamma_b)},$$

где оператор  $\bar{U}^* = \lambda^{-1} \bar{V}^*$ .

Мы видим, что после учёта вариации  $\delta v$  на левой границе остались члены при  $\delta v$  только справа. То есть, область определения граничной вариации  $\delta v$  перевернулась с  $\Gamma_a$  на  $\Gamma_a^* \equiv \Gamma_b$ . В то же время на левой границе  $\Gamma_a \equiv \Gamma_{\partial 1}$  в вариации  $\delta J$  появился новый член в виде линейного функционала  $\bar{U}^* \tilde{f}$  при вариации управления  $\delta u$ . Мы получили на левой границе оператор с *отображением* в сопряжённую область управлений:

$$\bar{U}^* : V^*(\Gamma_{\partial 1}) \rightarrow U^*(S).$$

**Пример 3.** Обратимся к граничному условию *второго рода с управлением* на правой границе  $\Gamma_b$ , когда функция  $G$  имеет вид:

$$G|_{\Gamma_b} = \lambda \frac{\partial v}{\partial x} - u = 0 \in V(\Gamma_b),$$

где  $\Gamma_b = S \equiv \Gamma_{\partial 2}$  — область определения управления в граничном условии второго рода (дополнительный нижний индекс 2). По-прежнему полагаем, что других управлений в задаче нет и  $I'_v|_{\omega}, I'_u|_{\omega}$  отсутствуют.

Если подставить производную вариации  $\frac{\partial \delta v}{\partial x} = \lambda^{-1} \delta u$  в соответствующие члены на  $\Gamma_b$  в вариации (3.30), то мы получим:

$$\langle \bar{\bar{V}}^* \tilde{f}, \frac{\partial \delta v}{\partial x} \rangle_{V^*(\partial \Omega = \Gamma_a \cup \Gamma_b)} = \langle \bar{\bar{V}}^* \tilde{f}, \frac{\partial \delta v}{\partial x} \rangle_{V^*(\Gamma_a)} + \langle \bar{\bar{U}}^* \tilde{f}, \delta u \rangle_{V^*(S)},$$

где оператор  $\bar{\bar{U}}^* = \lambda^{-1} \bar{\bar{V}}^*$ .

Видно, что опять-таки область определения граничной вариации, точнее производной от вариации состояния  $\frac{\partial \delta v}{\partial x}$  на границе, переворачивается с  $\Gamma_b$  на  $\Gamma_b^* \equiv \Gamma_a$ , а на старом месте  $\Gamma_b \equiv \Gamma_{\partial 2}$  появится линейный функционал  $\bar{\bar{U}}^* \tilde{f}$  при  $\delta u$ . Мы получили *отображение*

$$\bar{\bar{U}}^* : V^*(\Gamma_{\partial 2}) \rightarrow U^*(S).$$

Таким образом, в вариации  $\delta J$  могут появляться вариации управления  $\delta u$  с оператором  $\bar{\bar{U}}^*$  или  $\bar{\bar{V}}^*$  на соответствующей части границы  $\partial \Omega$ , где определено управление  $u$ . В наших примерах, — это множества  $\Gamma_{\partial 1}$  и  $\Gamma_{\partial 2}$  граничных условий первого и второго рода.

**Пример 4.** Рассмотрим ещё один пример, когда состояние распределённой системы описывается несколькими уравнениями, что характерно для *гиперболических систем*. Пусть  $m = 2$  и имеется левое граничное условие на  $\Gamma_a$  в виде *связи компонент состояния* системы:

$$av_1 - bv_2 = 0.$$

Здесь  $a, b$  некоторые вещественные константы или известные функции, не зависящие от  $v, u$ . При этом в граничных условиях вариаций (3.31) на

$\Gamma_a$  мы имеем  $G'_u = 0$ ,  $G'_{v'} = 0$ , а  $m \times m$  матрица  $G'_v$  превращается в вектор-строку  $G'_v = (a, -b)^T$ .

Вариация граничных условий примет вид:

$$\delta G|_{\Gamma_a} = G'_v \delta v = a\delta v_1 - b\delta v_2 = 0.$$

Эти граничные условия для  $\delta v$  необходимо учесть в вариации  $\delta J$  (3.30). Очевидно, что найти все компоненты вектора  $\delta v = \{\delta v_1, \delta v_2\}$  из одного уравнения невозможно. Поэтому далее будем находить не целый вектор  $\delta v$ , а его некоторые компоненты. Например, мы можем записать  $\delta v_1 = \frac{b}{a}\delta v_2$ , т.е. вектор-столбец  $\delta v = \left(\frac{b}{a}\delta v_2, \delta v_2\right)$ , или наоборот, выразить  $\delta v$  через  $\delta v_1$ .

Теперь в вариации  $\delta J$  на границе  $\Gamma_a$  независимые вариации некоторых компонент  $\{\delta v_i\}_{i=1..m}$  станут зависеть друг от друга. На  $\Gamma_a$  в (3.30) мы потеряем одну компоненту  $\delta v_1$  (или наоборот  $\delta v_2$ ). В итоге

$$\langle \bar{V}^* \tilde{f} + I'_v|_{\omega}, \delta v \rangle_{V^*(\Gamma_a)} = \langle \bar{V}^* \tilde{f} + I'_v|_{\omega}, \left(\frac{b}{a}\delta v_2, \delta v_2\right) \rangle_{V^*(\Gamma_a)},$$

т.е. на границе  $\Gamma_a$  у нас исчезает линейный функционал при  $\delta v_1$ , а при  $\delta v_2$  останется (или наоборот). В то же время на противоположной границе  $\Gamma_a^* \equiv \Gamma_b$  все компоненты вариаций  $\delta v_1, \delta v_2$  останутся.

Если же на границе  $\Gamma_a$  задать не одно уравнение, а  $m$  уравнений, то, очевидно, что в вариации  $\delta J$  на  $\Gamma_a$  совсем ничего не останется при  $\delta v$ , а на противоположной границе  $\Gamma_a^* \equiv \Gamma_b$  всё сохранится. То есть и здесь вариации граничных условий «переворачиваются».

В конце концов, после учёта в  $\delta J$  всех краевых зависимостей вариаций  $\delta v, \delta u$ , мы получаем вместо (3.30) следующее принципиальное выражение первой вариации целевого функционала в области  $\Omega$  и на её границе  $\partial\Omega$ :

$$\begin{aligned}
& \delta J(u) \tag{3.32} \\
& = \langle \mathbb{V}^* \tilde{f} + I'_v|_\omega, \delta v \rangle_{V^*(\Omega)} + \langle \bar{\mathbb{V}}^* \tilde{f} + I'_v|_\omega, \delta v \rangle_{V^*(\Gamma_{\dots})} \\
& + \langle \bar{\bar{\mathbb{V}}}^* \tilde{f}, \frac{\partial \delta v}{\partial \tau} \rangle_{V^*(\Gamma_{\dots})} + \dots \\
& + \langle \mathbb{U}_\emptyset^* \tilde{f}, \delta u \rangle_{U^*(S=\Omega)} + \langle \bar{\mathbb{U}}^* \tilde{f}, \delta u \rangle_{U^*(\Gamma_{\partial 1})} + \langle \bar{\bar{\mathbb{U}}}^* \tilde{f}, \delta u \rangle_{U^*(\Gamma_{\partial 2})},
\end{aligned}$$

где  $\Gamma_{\dots}$  — это множества определения «перевёрнутых» краевых вариаций состояния системы,  $\Gamma_{\partial 1}$ ,  $\Gamma_{\partial 2}$  — множества, соответствующие вариациям управления в граничных условиях первого рода (связь  $v$  и  $u$  на  $\Gamma_{\partial 1}$ ) и второго рода (связь  $\frac{\partial v}{\partial \tau}$  и  $u$  на  $\Gamma_{\partial 2}$ ), если таковые имеются.

Понятно, что вид всех слагаемых, содержащих граничные элементы, может отличаться от приведённого примера (3.32) в зависимости от типа граничных условий и области определения управления. Однако, особенности различных случаев краевых условий не влияют на рассмотренный способ их учёта. Поэтому далее будем обращаться к выражению первой вариации целевого функционала в принципиальном виде (3.32).

### 3.4.6. Выделение градиента

Необходимым условием оптимальности для выпуклого целевого функционала является обращение в ноль его первой вариации:

$$\delta J = 0.$$

Первая вариация  $\delta J$  (3.32) имеет  $m + n$  независимых вариаций компонент  $\delta v_{i=1\dots m}$ ,  $\delta u_{i=1\dots n}$  от вектор-функций  $v \in L_2^m$ ,  $u \in L_2^n$ . На данном этапе преобразований необходимо восстановить связь, т.е. взаимную зависимость вариаций всех компонент  $\delta v$ ,  $\delta u$ . Это можно легко сделать, если задать конкретное значение  $\tilde{f} = f$ , такое, чтобы все линейные функционалы при  $\delta v$  и  $\frac{\partial \delta v}{\partial \tau}$  обратились в ноль:

$$\begin{aligned} \mathbb{V}^* f + I'_v|_\omega &= 0 \in V^*(\Omega), \\ \bar{\mathbb{V}}^* f + I'_v|_\omega &= 0 \in V^*(\Gamma_{\dots}^*), \quad \bar{\bar{\mathbb{V}}}^* f = 0 \in V^*(\Gamma_{\dots}^*), \end{aligned} \quad (3.33)$$

Полученные выражения называют *сопряжённой задачей*, а её решение  $f$  — *сопряжённым состоянием*. Сопряжённая задача — это система из  $m$  линейных дифференциальных уравнений, характеризуемая линейным оператором  $\mathbb{V}^*$ , с набором краевых условий, характеризуемых линейными операторами  $\bar{\mathbb{V}}^*$  и  $\bar{\bar{\mathbb{V}}}^*$ . Линейность определяется относительно  $f$ .

Обратим внимание, что в нестационарной сопряжённой задаче вместо начальных условий на  $\Gamma_0$  будут присутствовать *терминальные условия* на  $\Gamma_0^* \equiv \Gamma_1$ . То есть такая задача всегда решается в обратном по времени направлении (начальные условия переворачиваются и превращаются в терминальные). Вид и род сопряжённых граничных условий где-либо на  $\Gamma_{\dots}^*$  (в предыдущих примерах — на  $\Gamma_a^* \equiv \Gamma_b$ ,  $\Gamma_b^* \equiv \Gamma_a$ , где имел место *переворот* граничных вариаций состояния) для каждой задачи оптимизации целесообразно выявлять в процессе конкретных преобразований.

При значении сопряжённого состояния, удовлетворяющего условиям (3.33), первая вариация (3.32) приобретает вид:

$$\delta J(u) = \langle \mathbb{U}_\emptyset^* f, \delta u \rangle_{U^*(\Omega)} + \langle \bar{\mathbb{U}}^*, \delta u \rangle_{U^*(\Gamma_{\partial 1})} + \langle \bar{\bar{\mathbb{U}}}^* f, \delta u \rangle_{U^*(\Gamma_{\partial 2})}. \quad (3.34)$$

Мы видим, что вариация целевого функционала, с точностью до граничных членов, принимает искомую нами форму (3.21), где  $\delta J = \langle \nabla J, \delta u \rangle_{U^*(S)}$ . Откуда следует, что *градиент целевого функционала* — это

$$\nabla J(u) = \left\{ \begin{array}{l} \mathbb{U}_\emptyset^* f \in U^*(\Omega) \\ \bar{\mathbb{U}}^* f \in U^*(\Gamma_{\partial 1}) \\ \bar{\bar{\mathbb{U}}}^* f \in U^*(\Gamma_{\partial 2}) \end{array} \right\} \in U^*(S). \quad (3.35)$$

Здесь верхняя строка соответствует задаче оптимизации с управлением, определённым внутри области  $\Omega$ . Вторая строка соответствует

граничному управлению в условии первого рода на границе  $\Gamma_{\partial 1}$ . Последняя нижняя строка соответствует граничному управлению в условии второго рода на  $\Gamma_{\partial 2}$ .

Если  $I = I(v)$ , т.е. целевая функция явно не зависит от управления  $u$ , определённого на  $\Omega$ , то градиент (3.35) будет иметь вид:

$$\nabla J(u) = \left\{ \begin{array}{l} \mathbb{U}^* f \in U^*(\Omega) \\ \bar{\mathbb{U}}^* f \in U^*(\Gamma_{\partial 1}) \\ \bar{\bar{\mathbb{U}}}^* f \in U^*(\Gamma_{\partial 2}) \end{array} \right\} \in U^*(S). \quad (3.36)$$

Выражение (3.36), в отличие от (3.35), содержит однородный линейный оператор  $\mathbb{U}^*$  вместо  $\mathbb{U}_\emptyset^*$ . Именно такая ситуация встречается в большинстве случаев задач оптимизации.

Таким образом, функция градиента  $\nabla J(u; \tau)$  определяется в пространстве управлений на множестве  $S$  через сопряжённое состояние (линейный функционал)  $f$ , значение которого зависит от производной  $I'_v|_\omega$ , характеризующей чувствительность целевого функционала  $J$  к исходному состоянию  $v$  на множестве  $\omega$ . Для того, чтобы найти значение градиента на множестве  $S$  необходимо решить две прямые задачи: исходную (3.20) и сопряжённую (3.33).

Ещё раз отметим, что при использовании других краевых условий и области определения управления, полученные выражения для градиента немного изменятся, но своей принципиальной сути не потеряют.

Теперь самое время вспомнить о чём мы писали в разделе 1.3.2 по поводу градиента. Полученные производные Фреше из первой вариации  $\delta J = \langle J'_u, \delta u \rangle_{U^*(S)}$  будут градиентами в выражениях (3.35), (3.36) только при выполнении условий управляемости. Давайте визуальное акцентируем на этом внимание. Например, для управления, определённого внутри области  $\Omega$ , будем иметь:

$$J'_u = \mathbb{U}_\emptyset^* f \xrightarrow{\text{условия управляемости}} \nabla J.$$

Поэтому далее следует обсудить проблему управляемости.

### 3.5 Управляемость

Знание и понимание управляемости необходимо для корректной постановки и решения любой задачи оптимального управления. Требование корректности только исходной прямой задачи здесь недостаточно, поскольку мы имеем дело с обратной, классически некорректной задачей управления.

В самом общем понимании *управляемость* — это совокупность условий, определяющих принципиальную<sup>1</sup> возможность управления состоянием системы. Если система не управляема в принципе (задача управления поставлена некорректно), то нет смысла искать оптимальное управление, оптимизировать систему.

Анализ управляемости должен начинаться с определения типа дифференциальных уравнений исходной и полученной сопряжённой задач. Это необходимо не только для корректного решения обеих задач, но и для выявления условий принципиальной возможности решения всей задачи оптимизации. Именно эти условия приводят к понятию управляемости, которое мы подробно обсудим ниже в разделе 3.5.3 «Условия управляемости».

Для уравнений разных типов совершенно по-разному ставятся краевые задачи, привлекаются разные методы решения и, естественно, по-разному формулируются условия управляемости. Подробно с

---

<sup>1</sup> Принципиальная возможность управления ещё не гарантирует возможность практического решения задачи оптимального управления. Можно иметь положительные принципиальные условия, но применять неподходящие экстремальные алгоритмы. Например, при бесконечномерной оптимизации не использовать алгоритмы равномерного спуска к оптимуму. В итоге может оказаться невозможным достижение оптимального решения. Другой пример, когда задача оптимизации решается непрямыми подходами, могут быть получены необходимые условия оптимальности в такой форме, из которой будет невозможно выразить оптимальное управление.

разными типами уравнений можно ознакомиться в обширной литературе по уравнениям в частных производных, например, [65, 83, 76, 44, 47]. Здесь мы рассмотрим несколько наиболее распространённых типичных примеров.

Начнём с исходной задачи, тип которой определяется оператором  $\mathbb{D}$ . Далее рассмотрим сопряжённую задачу, тип которой определяется оператором  $\mathbb{V}^*$ . Выясним взаимосвязь типов этих задач [20].

### 3.5.1. Тип уравнений в частных производных

Ограничимся рассмотрением основных типов, а именно, *гиперболических*, *эллиптических* и *параболических* уравнений в частных производных для системы  $\mathbb{D}v = 0$  в задаче (3.20) на примере двух независимых переменных. При этом возможно несколько характерных случаев. Разберём их по порядку.

Начнём со *стационарной* задачи в виде одного уравнения с производными второго порядка. Здесь состояние  $v(\tau) \in L_2(\bar{\Omega})$ , а вектор  $\tau = (x_1, x_2)$  определяется двумя пространственными координатами. *Дифференциальный оператор*  $\mathbb{D}$  имеет вид:

$$\begin{aligned} \mathbb{D} \cdot &= a_{11} \frac{\partial^2 \cdot}{\partial x_1^2} + 2a_{12} \frac{\partial^2 \cdot}{\partial x_1 \partial x_2} + a_{22} \frac{\partial^2 \cdot}{\partial x_2^2} + \dots \\ &= \sum_{i,j=1}^2 a_{ij} \frac{\partial^2 \cdot}{\partial x_i \partial x_j} + \dots, \end{aligned} \tag{3.37}$$

где  $a_{ij} = a_{ji}$ . Многоточие означает члены оператора  $\mathbb{D}$ , содержащие производные менее второго порядка, включая члены без производных. Они не влияют на тип уравнения.

Дифференциальное уравнение  $\mathbb{D}v = 0$  является *квазилинейным*, поскольку все высшие производные входят в оператор  $\mathbb{D}$  линейно. Если

коэффициенты  $a_{ij}$  при высших производных не являются функциями  $v$ , то дифференциальное уравнение называется *полулинейным*, а если и остаточные члены (в многоточии) зависят линейно от  $v$ , то мы получаем *линейное* дифференциальное уравнение.

Оператор  $\mathbb{D}$ , в произвольной фиксированной точке  $\tau \in \Omega$ , представляет собой *квадратичную форму* с операторами набла:

$$\mathbb{D} \cdot = (\nabla^T A \nabla) \cdot + \dots,$$

где  $A$  — симметричная матрица с элементами  $a_{ij}$ . Такой квадратичной форме  $(\nabla^T A \nabla)v$  можно формально поставить в соответствие «обычную» квадратичную форму  $\tau^T A \tau = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2$ , которая в зависимости от значений коэффициентов принимает вид эллипса, параболы или гиперболы, что и объясняет используемую терминологию типов уравнений, операторов.

Тот или иной тип оператора  $\mathbb{D}$  может определять различные семейства (множества) линий  $\xi(x_1, x_2) \in \bar{\Omega}$  (в трёхмерном пространстве — это поверхности), которые называются *характеристиками*. Особенность этих линий заключается в следующем. Если они есть, то решение  $v$  может быть найдено только вдоль этих линий, поэтому, для отыскания решения на любой линии  $\xi$  она должна выходить из области известных решений, например, из соответствующей части краевых условий. Можно сказать иначе: если какая-либо характеристика  $\xi$  не выходит из области известных решений, то решение во всех точках  $\xi$  найти будет невозможно. Характеристики формируют область влияния краевых условий на решение, поэтому их понимание очень важно. Разные типы операторов  $\mathbb{D}$  (гиперболический, эллиптический, параболический) отличаются разными семействами характеристик и описывают существенно разные физические процессы.

Чтобы найти характеристики уравнения  $\mathbb{D}v = 0$  необходимо обратиться к соответствующему *характеристическому уравнению*  $a_{11}(dx_1)^2 - 2a_{12}dx_1dx_2 + a_{22}(dx_2)^2 = 0$ . Это обычное квадратное уравнение относительно производной  $dx_2/dx_1$ . Его корни представляют собой два *характеристических направления* (касательные к характеристикам):

$$\frac{dx_2}{dx_1} = \frac{a_{12} + \sqrt{d}}{a_{11}}, \quad \frac{dx_2}{dx_1} = \frac{a_{12} - \sqrt{d}}{a_{11}},$$

где дискриминант  $d = a_{12}^2 - a_{11}a_{22}$ . Интегрируя характеристические направления с точностью до констант  $C_1, C_2$ , мы получаем два типа линий в виде уравнений, связывающих  $x_2$  и  $x_1$ . Эти линии представляют собой два семейства характеристик:

$$\xi_1(x_1, x_2) = C_1 \text{ для } +\sqrt{d}, \quad \xi_2(x_1, x_2) = C_2 \text{ для } -\sqrt{d}.$$

Можно заметить, что семейства характеристик, а значит и тип уравнения, определяются дискриминантом  $d$ .

Если  $d > 0$ , то уравнение  $\mathbb{D}v = 0$ , будет *гиперболического* типа с различными характеристиками  $\xi_1, \xi_2$ , которые отражают направление и скорость распространения возмущений из областей влияния до заданной точки  $\tau \in \bar{\Omega}$ , где и находят решение  $v(\tau)$ . Такие уравнения описывают распространение колебаний (волн, возмущений) в сплошных средах. Они могут быть как стационарные (установившиеся), так и *нестационарные*. В последнем случае независимая переменная  $x_1$  трактуется как время  $t$ , а  $x_2 \stackrel{\text{def}}{=} x$ . Например, распространение плоской волны вдоль оси  $x$  с конечной скоростью  $a$  можно описать оператором

$$\mathbb{D} \cdot = \frac{\partial^2 \cdot}{\partial t^2} - a^2 \frac{\partial^2 \cdot}{\partial x^2} + \dots, \quad a = \sqrt{a_{22}}.$$

Если  $d < 0$ , то уравнение  $\mathbb{D}v = 0$  будет *эллиптического* типа. Оно не имеет реальных характеристик, оба выражения  $\xi_1, \xi_2$  мнимые и никак не влияют на решение  $v$ . Такие уравнения характерны для

установившихся процессов. У них нет особых областей влияния, распространения информации. На решение в любой точке  $t \in \Omega$  влияют краевые условия, заданные на всей границе  $\partial\Omega$ .

Если  $d = 0$ , то уравнение  $\mathbb{D}v = 0$  будет *параболического* типа с одной характеристикой  $\xi$ . При этом, характеристическое уравнение вырождается в линейное уравнение  $a_{11}dx_2 - a_{12}dx_1 = 0$ . Для выполнения равенства нулю при любых коэффициентах, обычно, принимают характеристику  $\xi$  равной  $x_1 = const$ . При этом в операторе (3.37) исчезает вторая производная по  $x_1$ , но может оставаться аналогичная первая производная в членах с многоточием. Такие уравнения характерны для нестационарных задач теплопроводности, диффузии, где  $x_1$  становится временем  $t$ , а  $x_2 \stackrel{\text{def}}{=} x$ . При этом оператор  $\mathbb{D}$  может принимать вид

$$\mathbb{D} \cdot = \frac{\partial \cdot}{\partial t} - a \frac{\partial^2 \cdot}{\partial x^2}, \quad a = a_{22},$$

а характеристическое направление  $\frac{dx}{dt} = \infty$  задаёт бесконечную скорость распространения возмущений тепла или концентрации вещества в каждый момент времени  $t$ . Такие задачи решаются при  $a > 0$  только в положительном направлении времени.

В реальных физических процессах скорость возмущений тепла или концентрации вещества безусловно конечна, но уравнения параболического типа получают при пренебрежении инерционностью молекулярных процессов. С удалением от своего источника величина таких возмущений асимптотически убывает. Поэтому возмущения на большом расстоянии от источника могут реально проявляться только спустя некоторое время, что может отражаться на условиях управляемости параболической системой.

Уравнение  $\mathbb{D}v = 0$  с непостоянными на  $\Omega$  коэффициентами  $a_{11}$ ,  $a_{12}$ ,  $a_{22}$  может иметь разный тип на разных частях  $\Omega$ . Такие уравнения называются уравнениями *смешанного* типа .

Оператор  $\mathbb{D}$  (3.37), в фиксированной точке  $\tau \in \Omega$ , посредством замены независимых переменных на  $\xi_{1,2}$ , т.е. замены  $v(x_1, x_2)$  на  $v(\xi_1, \xi_2)$ , может быть представлен в простом каноническом виде, наглядно показывающим тип исходного уравнения:

- $d > 0$  (гиперболический тип) —

$$\frac{\partial^2 v}{\partial \xi_1^2} - \frac{\partial^2 v}{\partial \xi_2^2} = \Phi_1 \text{ или } \frac{\partial^2 v}{\partial \xi_1 \partial \xi_2} = \Phi_2;$$

- $d < 0$  (эллиптический тип) —

$$\frac{\partial^2 v}{\partial \xi_1^2} + \frac{\partial^2 v}{\partial \xi_2^2} = \Phi_3;$$

- $d = 0$  (параболический тип) —

$$\frac{\partial^2 v}{\partial \xi^2} = \Phi_4.$$

Здесь  $\Phi_i$  — некоторые функции, зависящие от  $\xi_{1,2}$ ,  $v$ ,  $\partial v / \xi_{1,2}$ . Если  $\mathbb{D}$  — линейный оператор, то приведение его к каноническому виду в точке  $\tau$  автоматически приводит к каноническому виду и во всей области  $\Omega$ . После приведения исходного уравнения к каноническому виду оно иногда настолько упрощаются, что оказывается возможным его решить аналитически.

Остановимся подробнее на *нестационарных гиперболических уравнениях* в виде системы уравнений первого порядка. Такие системы характерны для задач газовой динамики. Здесь в решении  $v$  может появляться негладкость, и даже разрывы (ударные волны). Если же краевые условия обеспечивают гладкость решения, то систему дифференциальных уравнений первого порядка можно дополнительным

дифференцированием привести к предыдущему случаю в виде одного уравнения второго порядка.

Рассмотрением квазилинейную систему двух уравнений с оператором  $\mathbb{D}$  и состоянием  $v \in L_2^{m=2}(\bar{\Omega})$ :

$$\mathbb{D} \cdot = \frac{\partial \cdot}{\partial t} + A \frac{\partial \cdot}{\partial x} + F,$$

где матрица  $A$  размерности  $2 \times 2$  имеет два вещественных *собственных числа*  $\lambda_{1,2}$ . Они находятся через решение квадратного уравнения, полученного из определителя:

$$|A - \lambda I| = 0, \quad (3.38)$$

где  $I$  — единичная матрица. При этом, ранее рассмотренное уравнение характеристических направлений  $\frac{dx_2}{dx_1} = \frac{a_{12} \pm \sqrt{d}}{a_{11}}$ , принимает вид:

$$\frac{dx}{dt} = \lambda_{1,2},$$

задающее (после интегрирования) характеристики  $\xi_1, \xi_2$  в виде двух траекторий распространения возмущений в области  $\bar{\Omega}$ .

Характеристики для гиперболических уравнений очень важны. Вдоль характеристик система уравнений в частных производных может быть представлена как два уравнения в обыкновенных дифференциалах (в одномерном случае). Это делается посредством записи дифференциальных уравнений на *собственных векторах* матрицы  $A$ .

Собственные векторы — это такие векторы  $\Lambda_{1,2}$ , которые на собственных числах  $\lambda_{1,2}$  удовлетворяют условиям:

$$\Lambda_1^T A = \lambda_1 \Lambda_1^T, \quad \Lambda_2^T A = \lambda_2 \Lambda_2^T.$$

Находятся они из соответствующих уравнений:

$$\Lambda_1^T (A - \lambda_1 I) = 0, \quad \Lambda_2^T (A - \lambda_2 I) = 0, \quad (3.39)$$

где  $0$  — нулевой вектор.

Если теперь систему дифференциальных уравнений умножить слева на векторы  $\Lambda_{1,2}$ , то в системе исчезнет матрица  $A$ . Мы получим по одному уравнению вдоль каждого характеристического направления  $\lambda_1$ ,  $\lambda_2$ , т.е. вдоль характеристик  $\xi_1$  и  $\xi_2$ . Такой вид уравнений называют уравнениями в *характеристической форме*:

$$\Lambda_1^T \left( \frac{\partial v}{\partial t} + A \frac{\partial v}{\partial x} + F \right) = \Lambda_1^T \left( \frac{\partial v}{\partial t} + \lambda_1 \frac{\partial v}{\partial x} \right) + \Lambda_1^T F = 0,$$

$$\Lambda_2^T \left( \frac{\partial v}{\partial t} + A \frac{\partial v}{\partial x} + F \right) = \Lambda_2^T \left( \frac{\partial v}{\partial t} + \lambda_2 \frac{\partial v}{\partial x} \right) + \Lambda_2^T F = 0.$$

Если учесть что  $\lambda_{1,2} = \frac{dx}{dt}$ , то получим в данных уравнениях члены  $\lambda_{1,2} \frac{\partial v}{\partial x} = \frac{\partial v}{\partial x} \frac{dx}{dt} \Big|_{\xi_{1,2}}$ , а сумму членов  $\frac{\partial v}{\partial t} + \lambda_{1,2} \frac{\partial v}{\partial x} = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \frac{dx}{dt} \Big|_{\xi_{1,2}} = \frac{dv}{dt} \Big|_{\xi_{1,2}}$ , где полные производные берутся вдоль характеристик  $\xi_1$  и  $\xi_2$ . При этом характеристические уравнения в частных производных превращаются в два уравнения в обыкновенных дифференциалах:

$$\begin{aligned} \Lambda_1^T \frac{dv}{dt} + \Lambda_1^T F &= 0, & \text{на } \xi_1, \\ \Lambda_2^T \frac{dv}{dt} + \Lambda_2^T F &= 0, & \text{на } \xi_2. \end{aligned} \tag{3.40}$$

Данная особенность часто используется при построении алгоритмов численного решения гиперболических задач. Состояние гиперболической системы  $v(\tau)$  находят во всех точках  $\tau \in \Omega$  где пересекаются характеристики обеих семейств  $\xi_1$ ,  $\xi_2$ , выходящие из краевых условий.

В некоторых случаях возможно дальнейшее упрощение характеристической формы заменой переменных. Вместо вектора состояния  $v(\tau)$  в каждом уравнении (3.40) переходят к одной функции  $r(\tau, v)$ , которая называется *инвариантом Римана*, и система (3.40) принимает вид:

$$\frac{dr_{1,2}}{dt} = g_{1,2}, \quad \text{на } \xi_{1,2},$$

где  $g_{1,2}(\tau, r_{1,2})$  — некоторые функции. Если  $g_{1,2} = 0$ , что имеет место при  $F = 0$ , то инварианты постоянны вдоль своих характеристик (чем и объясняется их название) и решение  $v$  легко находится.

Инварианты, если они существуют, могут сохраняться только в случае *однородных* квазилинейных систем уравнений. Наличие свободного члена  $F$  приводит к изменению инвариантов, а следовательно, к изменению интенсивности (амплитуды) волн. Например, если  $F$  задаёт *течение в конфузоре* (в сопле), то при *дозвуковых течениях* амплитуда волн будет нарастать с отчётливо выраженным фронтом и дальнейшим возможным превращением в ударную волну. Если  $F$  задаёт течение с *диссипацией энергии* или *течение в диффузоре*, особенно если оно вязкое, то амплитуда волн будет убывать, волны будут затухать, а их фронт размазываться.

### 3.5.2. Тип уравнений сопряжённой задачи

Здесь мы исследуем *сопряжённый оператор*  $\mathbb{V}^*$ , действующий на сопряжённое состояние  $f$ . Сразу же отметим, что соответствующая сопряжённая задача (3.33) всегда линейна относительно  $f$ , независимо от линейности или нелинейности исходной задачи. То есть оператор  $\mathbb{D}$  исходной задачи может зависеть от  $v$ , а оператор  $\mathbb{V}^*$  сопряжённой задачи никогда не зависит от  $f$ .

В предыдущих примерах раздела 3.4.3, при определении градиента целевого функционала, мы уже получали различные выражения операторов  $\mathbb{V}^*$ , содержащих члены с производными  $A \frac{\partial}{\partial x}$ ,  $A \frac{\partial^2}{\partial x^2}$ ,  $A \frac{\partial^2}{\partial x_1 \partial x_2}$ . Здесь мы используем ранее полученные сопряжённые выражения для анализа получаемых сопряжённых уравнений.

Рассмотрим случай исходной задачи с одним *стационарным* двумерным уравнением ( $m = 1, N = 2$ ), который соответствует (3.37), где оператор уравнений

$$\mathbb{D} \cdot = a_{11} \frac{\partial^2 \cdot}{\partial x_1^2} + 2a_{12} \frac{\partial^2 \cdot}{\partial x_1 \partial x_2} + a_{22} \frac{\partial^2 \cdot}{\partial x_2^2} + \dots$$

Напомним, многоточие означает члены, содержащие производные менее второго порядка.

Слагаемые со вторыми производными  $a_{11} \frac{\partial^2}{\partial x_1^2}$ ,  $a_{22} \frac{\partial^2}{\partial x_2^2}$ ,  $2a_{12} \frac{\partial^2}{\partial x_1 \partial x_2}$  соответствуют ранее рассмотренным примерам для сопряжённых операторов с коэффициентами  $A$ . В нашем случае  $A = a_{11}$ , или  $A = a_{22}$ , или  $A = 2a_{12}$ , соответственно. Получаем сопряжённый оператор

$$\mathbb{V}^* \cdot = a_{11} \frac{\partial^2 \cdot}{\partial x_1^2} + 2a_{12} \frac{\partial^2 \cdot}{\partial x_1 \partial x_2} + a_{22} \frac{\partial^2 \cdot}{\partial x_2^2} + \dots$$

Мы видим, что

$$\mathbb{V}^* = \mathbb{D} + \dots, \quad \frac{dx_2}{dx_1} = \frac{a_{12} \pm \sqrt{d}}{a_{11}}.$$

Тип сопряжённого уравнения совпадает с типом исходного уравнения. Более того, поскольку коэффициенты уравнений при старших производных не изменились, то и характеристические направления  $dx_2/dx_1$ , а значит и характеристики  $\xi_{1,2}$  обеих задач будут одинаковыми.

Рассмотрим *нестационарные* одномерные процессы, где  $N = 1$ . Они описываются уравнениями параболического и гиперболического типов. При этом задают  $t \stackrel{\text{def}}{=} x_1$ , а  $x \stackrel{\text{def}}{=} x_2$ . Нестационарные процессы требуют наличия начального условия, которое мы будем задавать на  $\Gamma_0 = X \times t_0$ .

Начнём с *параболического уравнения*. Здесь оператор  $\mathbb{D}$  имеет вид:

$$\mathbb{D} \cdot = \frac{\partial \cdot}{\partial t} - a \frac{\partial^2 \cdot}{\partial x^2} + \dots$$

Опять же, согласно рассмотренным ранее примерам, следует принять

$A \frac{\partial}{\partial x} = \frac{\partial}{\partial t}$ ,  $A \frac{\partial^2}{\partial x^2} = a \frac{\partial^2}{\partial x^2}$ . В итоге получаем сопряжённый оператор

$$\mathbb{V}^* \cdot = -\frac{\partial \cdot}{\partial t} - a \frac{\partial^2 \cdot}{\partial x^2} + \dots$$

Отрицательный знак при производной по времени прекрасно согласуется с терминальным условием на  $\Gamma_1 = X \times t_1$ , которое мы ранее получили для сопряжённой задачи (3.33).

Таким образом, если сопряжённую задачу решать в обратном по времени направлении, т.е. переопределить  $t \leftarrow -t$ , то мы получим сопряжённый оператор  $\mathbb{V}^* = \mathbb{D} + \dots$  с характеристическим направлением  $\frac{dx}{dt} = \infty$ , где  $dt < 0$ . Мы видим, что тип сопряжённого оператора совпадёт с типом исходного. Характеристическое направление аналогично исходной задаче, но оно требует решать сопряжённую параболическую задачу в обратном времени.

Теперь обратимся к нестационарным *гиперболическим уравнениям* ( $m = 1, N = 1$ ). Рассмотрим оператор второго порядка

$$\mathbb{D} \cdot = \frac{\partial^2 \cdot}{\partial t^2} - a^2 \frac{\partial^2 \cdot}{\partial x^2} + \dots$$

Здесь, согласно рассмотренным ранее примерам, для первого слагаемого

принимаем  $A \frac{\partial^2}{\partial x^2} = \frac{\partial^2}{\partial t^2}$ , для второго  $A \frac{\partial^2}{\partial x^2} = a^2 \frac{\partial^2}{\partial x^2}$ . В итоге получаем

$$\mathbb{V}^* \cdot = \frac{\partial^2 \cdot}{\partial t^2} - a^2 \frac{\partial^2 \cdot}{\partial x^2} + \dots$$

Как мы видим

$$\mathbb{V}^* = \mathbb{D} + \dots, \quad \frac{dx}{dt} = \pm a.$$

Тип сопряжённого оператора совпадает с типом исходного, оба с двумя одинаковыми характеристическими направлениями  $\pm a$ .

Теперь рассмотрим гиперболическую систему двух уравнений ( $m \geq 2, N = 1$ ) с оператором первого порядка

$$\mathbb{D} \cdot = \frac{\partial \cdot}{\partial t} + A \frac{\partial \cdot}{\partial x} + \dots$$

Как и ранее, для первого слагаемого принимаем  $A \frac{\partial}{\partial x} = \frac{\partial}{\partial t}$ , для второго оставляем  $A \frac{\partial}{\partial x}$ . В итоге получаем оператор  $\mathbb{V}^*$  по-прежнему гиперболической системы:

$$\mathbb{V}^* \cdot = -\frac{\partial \cdot}{\partial t} - A^T \frac{\partial \cdot}{\partial x} + \dots$$

Поскольку собственные числа матриц  $A$  и транспонированной  $A^T$  одинаковы, то значит и характеристики в сопряжённой задаче будут теми же. Решается она в обратном по времени направлении при  $dt < 0$ , следовательно, и с обратным движением волновых возмущений по тем же характеристикам, что и в исходной задаче.

Полученные выражения сопряжённых операторов  $\mathbb{V}^*$  говорят о том, что обоснование корректности решения  $f \in V^*(\Omega)$  линейной сопряжённой задачи (3.33) следует делать на основе уже имеющихся условий корректности исходной задачи. Очевидно, что такое обоснование не будет сложнее чем в исходной задаче.

При этом надо учитывать следующие особенности сопряжённой задачи. Во-первых, «переворачивание» её краевых (начальных, граничных) условий. Во-вторых, появление слагаемого  $I'_v|_\omega$  в дифференциальных уравнениях или в краевых условиях. Именно эти особенности и их влияние на управляемость мы далее и обсудим.

### 3.5.3. Условия управляемости

На сегодняшний день не выработано единого понятия и подхода к анализу управляемости для произвольных распределённых систем. Преобладают попытки обобщения результатов управляемости, полученных для систем с сосредоточенными параметрами [21, 28, 43, 52, 57]. При этом *управляемость* трактуется как возможность перевода

системы из заданного начального состояния в заданное финальное (терминальное по времени) состояние при помощи допустимых управлений.

Из-за пространственной распределённости модели, управления и целевого функционала традиционное понятие управляемости зачастую оказывается бесполезным. В системах с распределёнными параметрами появляется огромное разнообразие задач оптимизации [14, 92, 91], не характерных для сосредоточенных систем. При этом управляемость по заданному финальному состоянию не гарантирует управляемость по условиям, накладываемым на какое-либо состояние в некотором пространственном месте, где задан целевой функционал.

Традиционное понятие и условия управляемости всегда связывают с видом оператора  $\mathbb{D}$  исходной задачи и не связывают с видом критерия оптимизации  $J$ , т.е. его не связывают ни с видом функции цели оптимизации  $I(v, u)$ , ни с пространственно-временным множеством задания цели  $\omega$ . По сути, традиционным понятием управляемости в некоторой степени подменяют понятие классической корректности исходной задачи с оператором  $\mathbb{D}$ . Но такая управляемость не является достаточной для обеспечения корректности обратной задачи оптимизации.

Представляется целесообразным, особенно для систем с распределёнными параметрами, расширить понятие управляемости от «перевода системы в финальное состояние» до более объемлемого и практичного понятия «возможности управления системой на основе заданной цели управления» [92, 93, 91]. Новое понятие управляемости основано на теории некорректных задач и связывает управляемость системы с целевым функционалом  $J$ .

Ранее в разделе 3.3 «Регуляризация» мы указывали на то, что задачи оптимального управления являются *обратными задачами*. В

выражении (3.14) был показан прямой характер отображения в исходной задаче

$$U(S) \rightarrow V(\bar{\Omega}), \quad S \subset \bar{\Omega},$$

а в (3.15) показано обратное отображение в задаче оптимизации:

$$V(\omega) \xrightarrow{\min J} u_*(\tau) \in U(S), \quad \omega \subset \bar{\Omega}.$$

Обратные задачи, в общем случае некорректны. Для их решения А. Н. Тихонов ввёл отличное от *классического* понятие корректности, заключающееся в требовании ограничения пространства решений  $U$  до *компактного множества*  $\mathcal{U}$  существования, единственности и устойчивости управлений. Для сужения множества (пространства)  $U$  классически некорректных решений до компакта корректности  $\mathcal{U}$  используют регуляризирующие алгоритмы. Такие алгоритмы играют роль *оператора сужения* решения из пространства  $U$  в компакт  $\mathcal{U}$ .

В связи с этим в работе [92] предлагается трактовка управляемости как корректности по А. Н. Тихонову обратной задачи условного отображения.

**Определение 3.3** Математическая модель распределённой системы  $\mathbb{D}(\tau, v, u)v = 0$ ,  $\tau \in \bar{\Omega}$  в задаче (1.1), характеризуемая прямым отображением

$$U(S) \rightarrow V(\bar{\Omega}), \quad S \subset \bar{\Omega},$$

*управляема* посредством  $u(\tau) \in U(S)$  относительно целевого функционала  $J$ , когда обратная задача

$$V(\omega) \xrightarrow{\min J} u_* \in \mathcal{U}(S), \quad \omega \subset \bar{\Omega}$$

отображения пространства  $V(\omega)$  состояний модели в элемент  $u_*$  при условии  $\min J$  является *корректной по Тихонову*.

Приведенное определение формализует подход к анализу управляемости для произвольных систем относительно любых целевых

функционалов. Анализ управляемости может быть относительно просто реализован, если использовать прямой экстремальный подход (1.2).

Прямой подход состоит из двух, практически самостоятельных, подзадач. Первая — это минимизация целевого функционала  $J$  на основе его градиента  $\nabla J$ , вторая — определение градиента целевого функционала. При этом проблема управляемости также разбивается на две подзадачи:

1. Регуляризация в алгоритмах прямой минимизации, использующих градиент  $\nabla J$ ;
2. Корректность нахождения градиента  $\nabla J$  (из производной Фреше).

Первое требование выполняется автоматически, если допустимое начальное приближение управления  $u^0$  принадлежит компакту корректности  $\mathcal{U}$ , а последующие его коррекции экстремальными алгоритмами, например, МРНС осуществляются с *удовлетворительными параметрами регуляризации*  $b^k \alpha^k$ . Уточним последнее понятие.

**Определение 3.4.** Параметры  $b^k$  (шаговый множитель) и  $\alpha^k$  (параметр регулирования направления спуска для МРНС),  $k = 0, 1, \dots$ , являются *удовлетворительными параметрами регуляризации* в экстремальных алгоритмах, если при  $u^k \in \mathcal{U}$  последующие управления  $u^{k+1} \in \mathcal{U}$ , т.е. не выходят из компакта корректности  $\mathcal{U}$ .

Таким образом, при  $u^0 \in \mathcal{U}$  и при удовлетворительных параметрах регуляризации, принципиальная возможность достижения оптимального управления  $u_*$ , будет полностью зависеть от корректности нахождения градиента.

Для анализа корректности градиента  $\nabla J(u)$  необходимо обратиться к символической форме сопряжённой задачи (3.33) и градиента (3.35):

$$\begin{aligned} \mathbb{V}^* f + I'_v|_\omega &= 0 \in V^*(\Omega), \\ \bar{\mathbb{V}}^* f + I'_v|_\omega &= 0 \in V^*(\Gamma^*), \quad \bar{\bar{\mathbb{V}}}^* f = 0 \in V^*(\Gamma^*); \\ \nabla J(u) &= \left\{ \begin{array}{l} \mathbb{U}_\emptyset^* f \in U^*(\Omega) \\ \bar{\mathbb{U}}^* f \in U^*(\bar{\Gamma}_{\partial 1}) \\ \bar{\bar{\mathbb{U}}}^* f \in U^*(\bar{\bar{\Gamma}}_{\partial 2}) \end{array} \right\} \in U^*(S). \end{aligned}$$

Мы видим, что корректность  $\nabla J$  зависит от классической корректности решения  $f$  сопряжённой задачи, от особенностей распространения возмущений  $I'_v$  из  $\omega$  на  $\bar{\Omega}$  и от свойств отображения соответствующего линейного оператора из набора  $\mathbb{U}_\emptyset^* \cdot = \mathbb{U}^* \cdot + I'_u|_\omega, \bar{\mathbb{U}}^*, \bar{\bar{\mathbb{U}}}^*$ .

Далее ограничимся рассмотрением сути проблемы с одним оператором  $\mathbb{U}_\emptyset^*$  на  $\Omega$ . Начнём рассмотрение со случая, когда целевая функция  $I$  не зависит явно от  $u$ , т.е. когда  $I'_u = 0$ ,  $\mathbb{U}_\emptyset^* = \mathbb{U}^*$  и градиент  $\nabla J(u) = \mathbb{U}^* f$ . Итак, для обеспечения корректности  $\nabla J$  необходимо:

1. Обосновать корректность решения сопряжённой задачи;
2. Выявить область  $V^*(\Omega)$ ,  $\Omega \subset \Omega$  определения оператора  $\mathbb{U}^*$ , из которой возможно однозначное отображение в область значений  $U^*(S)$ ;
3. Оценить корректность работы оператора  $\mathbb{U}^*: V^*(\Omega) \rightarrow U^*(S)$ .

Начнём с сопряжённой задачи. Её тип и алгоритм решения определяется оператором  $\mathbb{V}^*$ . В предыдущем подразделе мы уже отмечали, что сопряжённая задача является прямой, линейной, того же типа и с теми же характеристиками, что и исходная задача, которая определяется оператором  $\mathbb{D}$ . Анализ корректности решения сопряжённой задачи не будет сложнее аналогичного анализа исходной задачи. Поэтому далее будем считать, что она имеет корректное решение.

Возмущения сопряжённого состояния  $f$ , порождённые производной  $I'_v|_\omega$ , должны распространяться из области  $\omega$  на такое подмножество  $\Omega \subset \bar{\Omega}$ , где оператор  $\mathbb{U}^*$  может реализовать однозначное отображение из  $V^*(\Omega)$  в  $U^*(S)$ . Уточним необходимые далее понятия.

**Определение 3.5.** Множество сопряжённых состояний  $V^*(\Omega)$ ,  $\Omega \subset \bar{\Omega}$ , корректно решённой сопряжённой задачи, будем называть *областью определения* линейного оператора  $\mathbb{U}^*$ , если  $f \in V^*(\Omega)$  однозначно зависит от управления  $u$  на  $S$ , и при этом *областью значений* оператора  $\mathbb{U}^*$  должно быть множество  $U^*(S)$ , т.е.

$$U(S) \ni u \xrightarrow{\text{однозначно}} I'_v|_\omega \xrightarrow{\text{однозначно}} f|_\Omega \in V^*(\Omega) \xrightarrow{\mathbb{U}^*} U^*(S).$$

Можно проще описать обсуждаемую проблему. Во-первых, целевой функционал  $J(u)$ , заданный на  $\omega$ , должен быть *чувствителен к управлению*, заданному на  $S$  (чувствительность характеризуется производной  $I'_v|_\omega$ ). Это первая ветка в отображениях. Во-вторых, из множества сопряжённых решений на всём  $\bar{\Omega}$  надо выделить такое подмножество  $\Omega$ , где сопряжённые решения  $f|_\Omega$  будут однозначно зависеть от значений целевого функционала в виде его производной  $I'_v|_\omega$ . На всей области  $\bar{\Omega}$  такой зависимости может и не быть. Это вторая ветка в отображениях. В-третьих, множество  $\Omega$  должно обеспечивать оператору  $\mathbb{U}^*$  возможность отображения сопряжённых состояний  $f|_\Omega$  из  $\Omega$  на  $S$ . Такое отображение представлено последней веткой, где оператор  $\mathbb{U}^*$  из полученной области определения  $V^*(\Omega)$  может осуществить отображение в область значений  $U^*(S)$ , где существует градиент  $\nabla J = \mathbb{U}^* f|_\Omega$ .

Мы видим, что область  $\Omega$  играет очень важную роль в выделении из производной Фреше  $J'_u$  именно градиента  $\nabla J$ . В свою очередь область  $\Omega$  существенным образом зависит от множества  $\omega$ . Если  $\omega$  задано не в

том месте  $\bar{\Omega}$ , не тех размеров, то ожидать получения корректной области  $\Omega$  для оператора  $\mathbb{U}^*$  не стоит.

Потеря однозначности при правильном  $\omega$  также возможна из-за различных *помех*. Например, численное решение дифференциальных уравнений различными методами аппроксимаций может заметно влиять, во-первых, на конфигурацию области  $\Omega$ , во-вторых, — на значения  $f$  на  $\Omega$ . Кроме того, при управлении *неоднородными* уравнениями со свободными членами, отражающими нарушения законов сохранения, возможна существенная *диссипации энергии*. Наличие диссипации приводит к затуханию и искажению возмущений как в исходном состоянии  $v$ , при их распространении от  $S$  к  $\omega$ , так и в сопряжённом состоянии  $f$ , при распространении возмущений от  $\omega$  к  $\Omega$ . Во всех этих случаях сопряжённое состояние  $f$  на  $\Omega$  может представлять существенно искажённую область определения  $V^*(\Omega)$  для оператора  $\mathbb{U}^*$ , настолько, что его значение  $\mathbb{U}^*f$  не будет принадлежать пространству градиентов  $U^*(S)$ .

Теперь обсудим свойства линейного сопряжённого оператора  $\mathbb{U}^*$ . При оптимальном управлении  $u_*$  его значение  $\mathbb{U}^*f$  должно обращаться в ноль. Этого требует НУО. При этом, сопряжённое состояние  $f$  будет нулевым, поскольку линейный оператор  $\mathbb{U}^*$  — однородный. Такой оператор, с нулевым значением только при  $f = 0$ , называется *невыврожденным*<sup>1</sup>.

Если оператор  $\mathbb{U}^*$  оказывается *вырожденным*, то необходимо оценить его *ядро*<sup>2</sup>  $\text{Ker } \mathbb{U}^* \equiv \{f(\tau)\}_{\text{Ker}}$ , т.е. выделить множество особых решений  $\{f(\tau)\}_{\text{Ker}} \subset V^*(\Omega)$ , когда при  $u \neq u_*$  значение оператора  $\mathbb{U}^*f$  обращается в ноль на  $S$ , или хотя бы на некотором его подмножестве

<sup>1</sup> Линейный оператор  $\mathbb{U}^*$  *невыврожденный*, если его значение  $\mathbb{U}^*f = 0$  на  $S$  только при  $f = 0$  на  $\Omega$ .

<sup>2</sup> Ядро вырожденного оператора  $\mathbb{U}^*$  — это множество таких  $f$ , на которых  $\mathbb{U}^*$  вырожденный. Кратко ядро обозначается как множество  $\text{Ker } \mathbb{U}^* = \{f | \mathbb{U}^*f = 0\} \equiv \{f\}_{\text{Ker}}$ . Если  $\text{Ker } \mathbb{U}^* = \{0\}$ , т.е. ядро состоит из одного нулевого элемента  $f = 0$ , то оператор  $\mathbb{U}^*$  — невырожденный.

ненулевой меры. Наличие ядра, когда  $\mathbb{U}^*(f \neq 0) = 0$ , не позволит корректно находить градиент, а следовательно, и оптимальное управление  $u_*$ .

С практической точки зрения важно не знание непосредственно ядра оператора  $\mathbb{U}^*$ , в виде множества особых решений  $\{f\}_{Ker}$ , а знание конкретных условий, приводящих к появлению ядра. В зависимости от условий образования ядра необходимо изменить условия задачи, например, изменить целевую функцию  $I$  или множество  $\omega$  так, чтобы наблюдаемое ядро исчезло (обнулилось). Только тогда можно говорить о корректном определении градиента  $\nabla J$ .

Возможна и ситуация выбора неудачной модели распределённой системы  $\mathbb{D}(\tau, v, u)v = 0$  или условий её функционирования, когда состояние  $v$  практически не изменяется при изменении управления  $u \in U$ . Естественно, управлять такой системой бессмысленно, и мы такие ситуации обсуждать не будем. Выбор той или иной модели и её адекватность определяется решением задачи идентификации.

Если целевая функция  $I$  зависит явно от управления  $u$ , то соответствующий линейный оператор  $\mathbb{U}_\emptyset^* \cdot = \mathbb{U}^* \cdot + I'_u|_\omega$  уже *не будет однородным*. Очевидно, что если  $\mathbb{U}^*$  был невырожденным, то и оператор  $\mathbb{U}_\emptyset^*$  будет невырожденным. При этом, произойдёт всего лишь смещение ранее нулевого ядра  $\{f\}_{Ker} = f_{Ker} = 0$  к новому значению в новом оптимуме  $u_*$ , где градиент  $\nabla J = \mathbb{U}_\emptyset^* f$  обратится в ноль. Такой оптимум не будет сопровождаться нулевым сопряжённым состоянием  $f$ . Новое ядро  $f_{Ker} \neq 0$  неоднородного оператора  $\mathbb{U}_\emptyset^*$  не делает его вырожденным.

На рисунке 3.1 символически представлены обсуждаемые отображения. Шаг 1 — это прямое отображение управления  $u \in U(S)$  оператором  $\mathbb{D}$  в некоторое состояние  $v \in V(\bar{\Omega})$  и далее, с

использованием  $J$ , — в сопряжённое пространство функций  $I'_v(\tau) \in V^*(\omega)$ . Примеры таких функции изображены линиями  $I'_v$ .

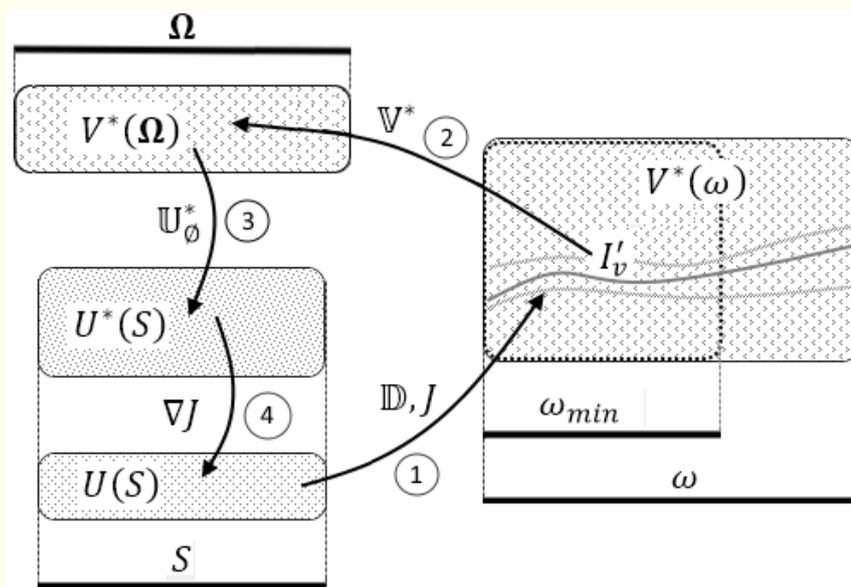


Рисунок 3.1 — Отображения пространств при получении градиента

Шаг 2 представляет собой отображение функций  $I'_v$  (через состояние  $f$ ) при помощи оператора  $V^*$  из множества  $\omega$  в  $\Omega$ . Заметим, что для *однозначного* отображения из  $\omega$  во все точки  $\Omega$  может потребоваться не всё множество  $\omega$ , а лишь его часть  $\omega_{min} \subset \omega$ . Эта ситуация, с наличием  $\omega_{min} \subset \omega$ , соответствует *избыточности*  $\omega$ . При этом, попытка учёта всего множества  $\omega$  может приводить к неоправданным *помехам* в вычислении градиента и, соответственно, — к помехам в управлении. Если же отображение  $I'_v$  из  $\omega$  попадает на такое  $\Omega$ , что область  $V^*(\Omega)$  значений оператора  $U^*$  оказывается определённой не на всём  $S$ , то это свидетельствует о *недостаточности*  $\omega$  для формирования полноценной области определения оператора  $U^*$  и, как следствие, — невозможности управления на  $S$ .

Шаг 3 на рисунке 3.1 — это отображение пространства сопряжённого состояния  $V^*(\Omega)$  в пространство градиентов  $U^*(S)$  при помощи оператора  $\mathbb{U}^*$ .

Шаг 4 — это отображение градиентов из  $U^*(S)$  в множество управлений  $U(S)$ . Последний шаг реализуется экстремальными алгоритмами, и мы его уже обсудили. Поскольку имеет место самосопряжённость  $U^* = U$ , то последний шаг можно назвать не отображением, а смещением на множестве  $U$  к улучшенному управлению на основе  $\nabla J$ .

Подытожим сказанное следующим утверждением для практического анализа управляемости в рамках прямого экстремального подхода.

**Теорема 3.4.** Математическая модель распределённой системы  $\mathbb{D}(\tau, v, u)v = 0, \tau \in \bar{\Omega}$  в задаче (1.1) управляема посредством  $u(\tau)$  на  $S \subset \bar{\Omega}$  по функционалу  $J$  если:

- существует область  $V^*(\Omega)$  определения оператора  $\mathbb{U}^*$  (или  $\bar{\mathbb{U}}^*, \bar{\bar{\mathbb{U}}}^*$ ) с его однозначными значениями в области  $U^*(S)$ ;
- оператор  $\mathbb{U}^*$  (или  $\bar{\mathbb{U}}^*, \bar{\bar{\mathbb{U}}}^*$ ) — невырожденный;
- прямой экстремальный подход (1.2) при  $u^0 \in \mathcal{U}$  использует удовлетворительные параметры регуляризации.

Выявление конкретных условий управляемости, обеспечивающих выполнение требований теоремы 3.4, осуществляется в процессе анализа на корректность решения исходной и сопряжённой задач, в процессе выявления области  $V^*(\Omega)$  определения оператора  $\mathbb{U}^*$  (или  $\bar{\mathbb{U}}^*, \bar{\bar{\mathbb{U}}}^*$ ), и при оценке ядра оператора  $\mathbb{U}^*$  (или  $\bar{\mathbb{U}}^*, \bar{\bar{\mathbb{U}}}^*$ ).

Задачу оптимального управления (1.1) можно считать корректно поставленной для применения экстремального подхода (1.2) только при выполнении условий управляемости. Если они не выполнены, то

производная Фреше  $J'_u$ , формально найденная из первой вариации целевого функционала  $\delta J = \langle J'_u, \delta u \rangle_{U^*(S)}$ , будет удовлетворять НУО в оптимуме, но не будет градиентом:  $J'_u \neq \nabla J$ . При этом невозможно применение экстремальных алгоритмов направленного поиска оптимума.

Мы получили относительно простую процедуру контроля управляемости распределённой системы по целевому функционалу. Введённое понятие управляемости в виде определения 3.3 и следующая из него теорема 3.4 делают эту процедуру наглядной и универсальной.

### 3.6 Управление с ограничениями

При решении задач оптимизации часто присутствуют какие-либо ограничения. Они выдвигаются организационными, техническими, экономическими требованиями, соображениями безопасности и т.д. Исключением являются задачи идентификации. Здесь ограничений не бывает, поэтому исследователями, разработчиками в области идентификации данный подраздел может быть пропущен.

Ограничения на управление могут присутствовать в виде явных ограничений на управление, или неявных ограничений на управление в виде некоторых условий на состояние системы. Эти два типа ограничений явно или неявно определяют множество допустимых управлений  $U$  и они реализуются существенно разными способами.

*Ограничения на состояние системы*, обычно, приводят к виду некоторого, например, штрафного функционала, который далее минимизируется вместе с целевым функционалом  $J$ . Это естественный и очевидный способ реализации такого ограничения, поскольку любое условие, содержащее состояние  $v$ , по сути является одной из целей оптимизации, также, как и главная цель  $J$ . Естественно их объединять.

Если же некоторое условие, содержит и состояние  $v$ , и управление  $u$ , и оно определено на  $\Omega$ , то такое условие не является ограничением. Это часть модели распределённой системы. При этом, возможно, размерность состояния  $v$  должна быть увеличена на единицу, чем ранее предполагалось. Если условие на  $v$  и  $u$  определено где-либо на границе  $\partial\Omega$ , то это тоже не ограничение. Это граничное условие распределённой системы.

Если некоторое условие, содержит только управление  $u$ , то условие является явным ограничением на управление. При использовании экстремального подхода такие ограничения реализуются различными методами в рамках экстремальных алгоритмов [20]. Например, это может быть широко известные методы проецирования управлений на допустимое множество  $U$ .

Итак, можно выделить *два типа* ограничений на управление. Первый тип относится к неявным ограничениям, которые следует учитывать посредством уточнения цели оптимизации  $J$  при определении градиента  $\nabla J$ . Второй тип относится к явным ограничениям, которые следует учитывать после определения градиента, в экстремальных алгоритмах.

В указанных типах ограничений можно выделить подтипы, влияющие на методы реализации ограничений в рамках своего типа:

## 1. Первый тип (неявные ограничения на управление).

1.1. Неявные ограничения на управление с **явно** (непосредственно) заданным *множеством допустимых состояний системы*  $V \subset L_2^m$ .

1.2. Неявные ограничения на управление с **неявно** заданным множеством  $V$ , в виде каких-либо условий для состояния системы.

## 2. Второй тип (явные ограничения на управление).

2.1. Явные ограничения на управление с **явно** (непосредственно) заданным *множеством допустимых управлений*  $U \subset L_2^n$ .

2.2. Явные ограничения на управление с **неявно** заданным множеством  $U$ . Это ограничения на некоторые выражения, содержащие управление.

Далее мы будем указывать к какому типу принадлежит то или иное ограничения для быстрого понимания проблемы и возможных способов, методов её решения.

Ограничения любого типа могут быть активными и неактивными. В задачах оптимизации с *активными ограничениями* оптимальное управление  $u_*$  будет находиться на границе  $\partial U$  допустимого множества, т.е. доставлять целевому функционалу  $J(u)$  *краевой минимум*. Заметим, что такая функция управления  $u_*(\tau)$  может либо полностью лежать на границе  $\partial U$ , либо лежать частично на  $\partial U$ , а частично — внутри  $U$ . В любом случае мы получаем управление  $u_* \in \partial U$  как точку (в пространстве  $L_2^n$ ) краевого минимума целевого функционала.

Также необходимо заметить, что в общем случае, даже для строго выпуклого функционала, не всякая точка  $u \in \partial U$  локального краевого минимума может быть оптимальным управлением  $u_*$ . То есть какой-либо локальный краевой минимум может и не быть глобальным минимумом целевого функционала  $J(u)$ . Напомним, что подобная ситуация может иметь место при минимизации на *невыпуклом* множестве  $U$ . Иллюстрация подобной ситуации была ранее приведена на рисунке 1.9 для примера двумерной оптимизации. Проверка, подтверждение выпуклости допустимого множества управлений  $U$  может оказаться нетривиальной задачей (как и подтверждение выпуклости  $J(u)$ ). Поэтому на практике всякую функцию управления,

доставляющую физически разумный краевой минимум функционалу  $J(u)$ , считают оптимальной.

Далее мы рассмотрим практические методы учёта указанных выше типов ограничений при численном решении задач оптимизации распределённых систем в рамках прямого экстремального подхода. Для большей наглядности примем  $n = m = 1$ .

### 3.6.1. Штрафные функции (тип 1.1)

Рассмотрим ограничения первого типа 1.1 на множество допустимых состояний системы, которые представляют собой явные выражения разрешённых состояний системы, обычно, в виде диапазона:

$$v_{\min}(\tau) \leq v(\tau) \leq v_{\max}(\tau), \quad \tau \in \bar{\Omega}. \quad (3.41)$$

То есть, множество допустимых состояний  $V = \{v \mid v_{\min} \leq v \leq v_{\max}\} \subset L_2^m(\bar{\Omega})$ . При этом множество допустимых управлений  $U$  в явном виде недоступно.

Ограничения (3.41) традиционно реализуются различными штрафными функциями, которые добавляются к целевому функционалу задачи (1.1) в виде штрафного функционала, например:

$$J(u) = \int_{\omega} I(v) d\omega + \kappa \int_{\Omega} I_F(v, v_{\min}, v_{\max})|_{\omega_F} d\Omega \rightarrow \min, \quad (3.42)$$

где  $\kappa \in E_+^1$  — положительный весовой коэффициент,  $I_F$  — штрафная функция с параметрами  $v, v_{\min}, v_{\max}$ . Здесь  $\cdot|_{\omega_F}$  — оператор сужения действия штрафной функции из всей области  $\bar{\Omega}$  на множество  $\omega_F \subset \bar{\Omega}$  возможного активного действия штрафного ограничения.

Штрафная функция может быть внутренней, внешней или смешанной. Смысл штрафных функций  $I_F$  заключается в штрафовании, т.е. в увеличении их значений в целевом функционале, при приближении к границам допустимого диапазона изнутри (внутренние штрафы,

барьерные функции) или при выходе за пределы допустимого диапазона (внешние штрафы). Такие функции и их смысл иллюстрируются рисунком 3.2 для одномерного случая  $v \in E$ .

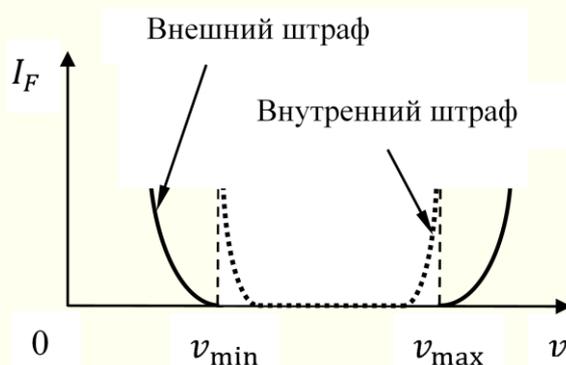


Рисунок 3.2 — Внешние и внутренние штрафные функции

По теории и методам штрафных функций имеется значительное количество публикаций, например, [95, 99, 29, 13, 41] и многие другие, включая обширную базу свободных Internet-ресурсов. Поэтому здесь мы подробно на этом останавливаться не будем. Конкретный вид штрафной функции  $I_F$ , её влияние на целевой функционал, т.е. коэффициент  $k$ , определяются желаниями, опытом разработчика и численным экспериментом оптимизации.

Легче всего реализуются алгоритмы внешних штрафов. Если незначительный выход за пределы заданного диапазона не допустим, то придётся приложить усилия по реализации внутренних штрафов с барьерными функциями.

Можно дать несколько рекомендаций для программирования задач с *барьерными функциями*. Прежде всего надо приложить усилия для выбора корректного начального приближения  $u^0(\tau)$ . Оно должно, во-первых, удовлетворять неявному ограничению типа (3.41), во-вторых, должно находиться достаточно далеко от допустимых границ, где внутренняя штрафная функция может иметь значение близкое к

машинной бесконечности. Кроме того, в процессе итераций возможен шаг, который выведет  $u^{k+1}$  за границы допустимой области, где барьерная функция будет неопределённой или бесконечной. Поэтому необходима явная проверка допустимости каждой последующей точки  $u^{k+1}$  на предмет удовлетворения условия (3.41).

При целевом функционале (3.42), со штрафной функцией, задача оптимизации (1.1) становится задачей оптимизации без ограничений. Легко видеть, что поиск градиента  $\nabla J$  в новой задаче не сложнее исходной. Там, где ранее при вариации  $\delta J$  присутствовала производная  $I'_v|_\omega$  теперь будет присутствовать сумма производных  $I'_v|_\omega + \kappa(I_F)'_v|_{\omega_F}$ .

### 3.6.2. Дополнительные функционалы (тип 1.2)

Ограничения типа 1.2 могут иметь вид каких-либо выражений, отличных от (3.41), включающих в себя состояние  $v$  (без управления  $u$ ). Например, это могут быть дополнительные  $i$ -тые функционалы  $J_i \rightarrow \min_{v_i}$ , также требующие минимизации. Пример такого ограничения рассмотрен в разделе 7, когда целевой функционал  $J(u)$  должен минимизироваться при условии минимума другого функционала.

Реализация рассматриваемых ограничений может осуществляться, как и в предыдущем подразделе, посредством совместной минимизации целевого функционала  $J$  с дополнительными функционалами-ограничениями  $J_i$ :

$$J(u) + \sum_i \kappa_i J_i(v) \rightarrow \min. \quad (3.43)$$

Естественно, объединение всех функционалов должно производиться с весовыми коэффициентами  $\kappa_i$ .

### 3.6.3. Проецирование на допустимое множество (тип 2.1)

Рассмотрим ограничения второго *типа* 2.1, когда допустимое множество управлений  $U$  задано явно в виде *диапазона* допустимых изменений управления:

$$u_{\min}(\tau) \leq u(\tau) \leq u_{\max}(\tau), \quad \tau \in S. \quad (3.44)$$

То есть,  $U = \{u \mid u_{\min} \leq u \leq u_{\max}\} \subset L_2(S)$ . Такой диапазон представляет собой *выпуклое* множество, в виде бесконечномерного параллелепипеда. Например, в конечномерном случае  $U \subset E^n$  — это будет «классический» параллелепипед:

$$u_{i,\min} \leq u_i \leq u_{i,\max}, \quad i = 1 \dots n.$$

Наличие ограничения (3.44) превращает алгоритмы минимизации, например, МРНСГ (3.4) в методы проекции спуска на допустимое множество:

$$u^{k+1} = P_U \left( u^k - b^k \alpha \nabla J(u^k) \right), \quad k = 0, 1 \dots,$$

где  $P_U(\cdot)$  — *оператор проецирования* на допустимое множество управлений  $U$ . При этом заметим, что шаговый множитель  $b^k$  нельзя определять методами линейного поиска.

На рисунке 3.3 показана суть работы оператора проецирования для двумерного параллелепипеда  $U \subset E^2$ . При выходе какой-либо компоненты вектора управления за допустимый диапазон (в примере — это компонента  $u_2$  выходит за предел  $u_{2\max}$ ) происходит скольжение управления по соответствующей границе параллелепипеда.

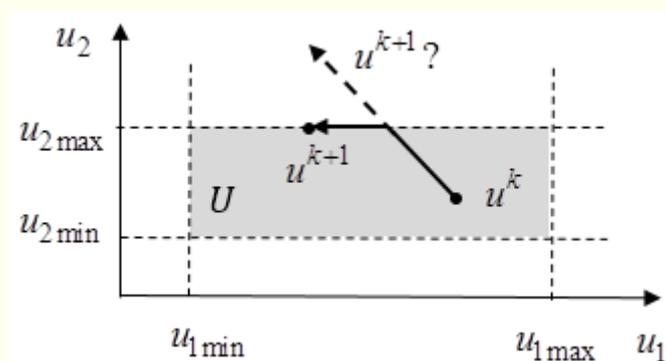


Рисунок 3.3 — Скользящее управление по границе

Если бесконечномерное допустимое множество — это параллелепипед, что мы и имеем в условии (3.44), то проектирование работает на множестве  $S$  следующим очень простым образом:

$$\begin{cases} \text{Если } u^{k+1}(\tau) < u_{\min}(\tau), & \text{тогда } u^{k+1}(\tau) \leftarrow u_{\min}(\tau). \\ \text{Если } u^{k+1}(\tau) > u_{\max}(\tau), & \text{тогда } u^{k+1}(\tau) \leftarrow u_{\max}(\tau). \end{cases} \quad (3.45)$$

То есть, если очередное приближение управления  $u^{k+1}$  в точке  $\tau$  выходит за какую-либо границу параллелепипеда  $U$ , то значению  $u^{k+1}(\tau)$  присваивается соответствующее граничное значение. Если очередное приближение  $u^{k+1}$  в точке  $\tau$  не выходит за границы параллелепипеда  $U$ , то оператор проектирования  $P_U$  в этой точке не включается в работу.

#### 3.6.4. Изопериметрическое условие (тип 2.2)

Рассмотрим пример ограничения второго типа 2.2 с неявно заданным множеством  $U$ . Это ограничение в виде *изопериметрического условия*<sup>1</sup>, накладываемого на управление:

$$\int_S u(\tau) dS = C. \quad (3.46)$$

<sup>1</sup> Здесь рассматривается одна из разновидностей изопериметрических задач. В классической постановке целью изопериметрической задачи является поиск фигуры наибольшей возможной площади, граница которой имеет заданную длину.

где  $C$  — заданная изопериметрическая константа. Такое ограничение может встречаться, например, в задачах оптимизации режимов подачи энергетических и других ресурсов, когда задана суточная, месячная и т.п. норма потребления ресурсов. Также возможна оптимизация формы какого-либо объекта с заданной его площадью или объёмом.

Давайте получим необходимую модификацию экстремальных алгоритмов для учёта изопериметрических ограничений. Очевидно любое начальное приближение  $u^0(\tau)$  в рассматриваемой задаче будет пересекать область значений точного решения  $u_*(\tau)$ , что будет сопровождаться знакопеременным значением начального градиента  $\nabla J(\tau; u^0)$ . В таких условиях нельзя применять МРНС с вычислением  $\alpha(\tau)$  по формулам (3.7)–(3.9). Поэтому далее возьмём за основу бесконечномерный градиентный алгоритм (1.7).

Изопериметрическое ограничение (3.46) можно реализовать следующим образом [92]. Запишем градиентный метод в интегральной форме:

$$\int_S u^{k+1} dS = \int_S u^k dS - \int_S b^k \nabla J^k dS, \quad k = 0, 1 \dots$$

Если  $k$ -тое приближение управления удовлетворяло изопериметрическому условию, т.е.  $\int_S u^k dS = C$  и новое  $k + 1$  приближение тоже удовлетворяет  $\int_S u^{k+1} dS = C^{k+1} = C$ , то из этого следует, что при изопериметрических коррекциях управления интегральный шаг должен быть нулевым:

$$\int_S b^k \nabla J^k dS = 0, \quad k = 0, 1 \dots$$

Для получения нулевого интегрального шага при изопериметрической переменной  $C^{k+1}$ , не равной изопериметрической

константе  $C$ , необходимо делать шаг с некоторой поправкой, которую можно выразить из интегральной формы метода коррекции:

$$\int_S \frac{C^{k+1}}{\text{mes } S} dS = \int_S \frac{C}{\text{mes } S} dS - \int_S b^k \nabla J^k dS, \quad k = 0, 1, \dots,$$

где  $\text{mes } S = \int_S dS$  — мера множества  $S$ . Тогда интегральный шаг с поправкой будет иметь вид:

$$\int_S \left( b^k \nabla J^k + \frac{C^{k+1} - C}{\text{mes } S} \right) dS = 0, \quad k = 0, 1, \dots$$

Откуда следует изопериметрическая коррекция управления

$$u^{k+1} = u^k - b^k \nabla J^k - \frac{C^{k+1} - C}{\text{mes } S} \quad \text{на } S, \quad k = 0, 1, \dots$$

Таким образом градиентный метод с изопериметрическим ограничением принимает вид следующего двухшагового метода:

$$u^{k+1} = u^k - b^k \nabla J^k, \\ u^{k+1} \leftarrow u^{k+1} - \frac{C^{k+1} - C}{\text{mes } S} \quad \text{на } S, \quad k = 0, 1, \dots \quad (3.47)$$

Множитель  $b^k$ , как и в предыдущей задаче с явным ограничением на управление, нельзя определять методами линейного поиска.

Алгоритм (3.47) реализуется следующим образом. Во-первых, задаётся начальное приближение  $u^0(\tau)$ , удовлетворяющее изопериметрическому условию. Во-вторых, на каждой последующей итерации сначала делается обычный шаг градиентного метода без учёта изопериметрического условия, находится управление  $u^{k+1}$  и для него рассчитывается значение изопериметрической переменной  $C^{k+1}$ . Если  $C^{k+1} \neq C$ , то полученное управление  $u^{k+1}$  дополнительно корректируется смещением на число  $\frac{C^{k+1} - C}{\text{mes } S}$ . Именно повторный шаг с указанной коррекцией реализует изопериметрическое условие.

Не рекомендуем находить значение изопериметрической константы аналитически, поскольку аналитическое  $C$  и численно проинтегрированное  $C^{k+1}$  могут заметно отличаться из-за вычислительных погрешностей интегрирования. В этом случае изопериметрическое ограничение (3.46) будет реализовываться существенно неточно.

### 3.6.5. Кусочно-постоянное управление (тип 2.2)

Рассмотрим следующий пример ограничения второго типа 2.2 с неявно заданным множеством  $U$ . Здесь ограничение задаётся в виде требования принадлежности управлений к классу *кусочно-постоянных* функций, определённых на заданных подмножествах  $\hat{S}_i$  из  $S$ :

$$u(\tau) = \{u_i \text{ на } \hat{S}_i \subseteq S\}_{i=1}^{\hat{n}}. \quad (3.48)$$

Здесь множество допустимых управлений  $U$  — это множество функций из  $L_2(S)$ , которые являются кусочно-постоянными. Такое ограничение встречается при использовании технических устройств с дискретным переключением управления.

По сути, ограничение (3.48) требует отыскивать управление как набор  $\hat{n}$  чисел  $u_i$ , которые можно трактовать как компоненты некоторого вектора управления  $u = (u_1, \dots, u_{\hat{n}}) \in E^{\hat{n}}$ . Из интегрального выражения первой вариации (3.34), определяющего градиент целевого функционала  $J$ , следует, что здесь вместо бесконечномерного градиента (3.36), мы получаем конечномерный градиент [92]:

$$\nabla J(u) = \left\{ \int_{\hat{S}_i} \nabla J(u; \tau) dS \right\}_{i=1}^{\hat{n}} \in E^{\hat{n}}.$$

Для применения, МРНСг (3.4) или МРНСсг (3.11) необходимо осуществлять переход в конечномерное пространство для всего отрегулированного направления спуска  $d^k$  относительно антиградиента

$-\nabla J^k$  или сопряжённых градиентов  $p^k$ , т.е. задавать направление минимизации вектором

$$d(u^k) = - \left\{ \int_{\hat{S}_i} \alpha(\tau) \nabla J^k(u^k; \tau) dS \right\}_{i=1}^{\hat{n}}$$

или

$$d(u^k) = \left\{ \int_{\hat{S}_i} \alpha(\tau) p(u^k; \tau) dS \right\}_{i=1}^{\hat{n}}.$$

При этом МРНС превращаются в конечномерные алгоритмы вида (2.1):

$$u_i^{k+1} = u_i^k + b^k d_i(u^k), \quad i = 1 \dots \hat{n}, \quad k = 0, 1 \dots \quad (3.49)$$

Поскольку исходное ограничение (3.48) не ограничивает пределы множества допустимых значений нового вектора управления  $u = (u_1, \dots, u_{\hat{n}})$ , то мы приходим к задаче неограниченной оптимизации конечномерной целевой функции  $J(u)$ ,  $u \in E^{\hat{n}}$ . Для выбора глубины каждого шага минимизации можно использовать методы линейного поиска  $b^k$ .

Таким образом, задача минимизации на неявно заданном множестве  $U \subset L_2(S)$  в виде кусочно-постоянных функций свелась к задаче неограниченной минимизации в конечномерном пространстве  $E^{\hat{n}}$ . Начальное приближение  $u^0$ , естественно, должно быть постоянным или кусочно-постоянным, согласно (3.48).

### 3.6.6. Ограничение на скорость изменения управления

(тип 2.2)

Рассмотрим ещё один пример ограничения второго типа 2.2 с неявно заданным множеством  $U$ . Это ограничение на *скорость изменения* функции управления. Такое ограничение встречается при использовании технических устройств, которые не могут реализовывать

произвольную интенсивность изменения управления. Для наглядности предположим, что управление зависит только от времени:

$$\vartheta_{\min} \leq \frac{du(t)}{dt} \leq \vartheta_{\max}, \quad t \in (t_0, t_1). \quad (3.50)$$

Здесь  $\vartheta_{\min}$  и  $\vartheta_{\max}$  — предельные значения скорости изменения функции управления  $u(t)$ .

Ограничение (3.50) — это *дифференциальное неравенство*, которому должно удовлетворять управление. Данное ограничение аналогично ограничению в виде (3.44), однако здесь границы допустимого параллелепипеда  $U$  оказываются плавающими.

Поясним сказанное. Ограничение (3.50) можно представить как

$$u(t) + \vartheta_{\min} dt \leq u(t + dt) \leq u(t) + \vartheta_{\max} dt.$$

Поскольку распределённые системы практически всегда решаются численно, то  $dt$  аппроксимируется шагом  $\Delta t$  конечно-разностной сетки дифференцирования. Обозначим индексом  $j$  номер узла шага сетки и введём обозначения:

$$\begin{aligned} t_{j+1} &= t_j + \Delta t, \\ u_{\min}^{j+1} &= u(t_j) + \vartheta_{\min} \Delta t, \\ u_{\max}^{j+1} &= u(t_j) + \vartheta_{\max} \Delta t. \end{aligned}$$

Тогда ограничение (3.50) на каждом  $j + 1$  шаге конечно-разностной сетки примет вид уже рассмотренного ранее ограничения (3.44):

$$u_{\min}^{j+1} \leq u(t_{j+1}) \leq u_{\max}^{j+1}.$$

Для реализации такого ограничения естественно использовать *проецирование* каждого нового приближения управления оператором  $P_U(u^{k+1})$  на плавающие границы  $u_{\min}^{j+1}$  и  $u_{\max}^{j+1}$ . При этом мы получим метод аналогичный (3.45):

$$\begin{cases} \text{Если } u^{k+1}(t_{j+1}) < u_{\min}^{j+1}, & \text{тогда } u^{k+1}(t_{j+1}) \leftarrow u_{\min}^{j+1}. \\ \text{Если } u^{k+1}(t_{j+1}) > u_{\max}^{j+1}, & \text{тогда } u^{k+1}(t_{j+1}) \leftarrow u_{\max}^{j+1}. \end{cases} \quad (3.51)$$

Рассмотренные шесть видов ограничений (3.41), (3.43), (3.44), (3.46), (3.48), (3.50) являются типичными и широко распространёнными при оптимизации распределённых систем, но, конечно, не являются исчерпывающими. Как мы показали, данные ограничения относительно просто реализуются в рамках прямого экстремального подхода.

## 4 Особенности задач идентификации

### 4.1 Общая постановка задачи

*Идентификация* — это процедура построения оптимальной в определённом смысле математической *модели* рассматриваемого *объекта* по данным наблюдений за объектом. Полученная модель далее может использоваться для управления этим объектом, его исследования, прогнозирования состояния объекта и т.п. Задачи идентификации укладываются в общее представление задач оптимизации (1.1).

При решении задач оптимального управления сначала формируется цель управления  $J$ , а потом подбираются уравнения  $\mathbb{D}v = 0$ , которые наилучшим образом могут минимизировать цель  $J$ . При идентификации — наоборот, сначала подбираются принципиальные уравнения модели  $\mathbb{D}v = 0$ , а потом формируется цель  $J$  как критерий качества идентичности состояний *модели* и *объекта*.

На рисунке 4.1 схематично представлена суть *задачи идентификации*. Одинаковые входные данные подаются на объект и его математическую модель, требующую идентификации. С объекта и модели снимают показания выходных данных и оценивают качество идентификации по расхождению этих данных. Если расхождение недопустимо велико, то включается процедура идентификации, редактирующая математическую модель.

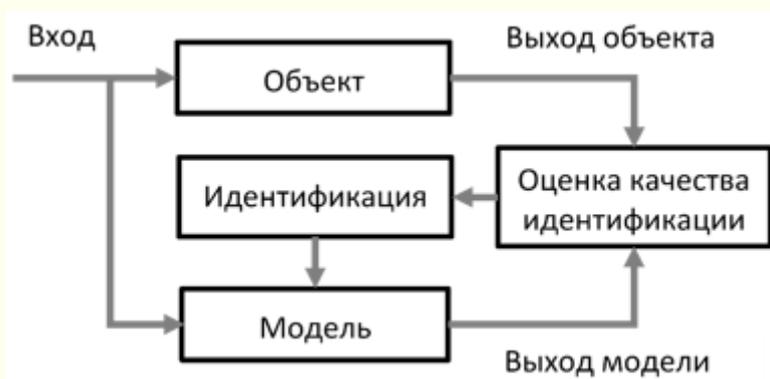


Рисунок 4.1 — Принципиальная схема идентификации

Если выходные данные объекта, при одних и тех же входных данных, с течением времени изменяются, то идентификация объекта должна периодически повторяться. Если периоды идентификаций оказываются очень короткими, то может потребоваться *on-line* идентификация. Здесь возникают особые, высокие требования к быстродействию вычислительных ресурсов, а главное, к методам идентификации. Использование традиционных экстремальных алгоритмов вида (1.5), формально обобщённых на бесконечномерные пространства, может потребовать тысячи и даже миллионы итераций для достижения хотя бы частично разумных решений. Понятно, что подобные методы для *on-line* идентификации неприемлемы.

Возможно, в подобных ситуациях даже придётся отказаться от математического моделирования в виде сложных распределённых систем уравнений и перейти к упрощённым имитационным моделям в виде *передаточных функций*. Естественно, такие модели не могут с высокой точностью описывать поведение объекта в широком диапазоне входных данных. Придётся постоянно подстраивать упрощённую модель под изменяющееся поведение объекта.

Далее будем рассматривать задачи идентификации для высоко реалистичных математических моделей, построенных на физических

законах функционирования распределённых объектов. При этом, в экстремальных алгоритмах идентификации будем использовать МРНС, что может радикально снизить количество итераций для достижения желаемой точности моделирования. Такой подход можно будет использовать не только в *off-line*, но и в умеренно-оперативной *on-line* идентификации.

Когда говорят об идентификации, то следует различать два разных этапа или типа идентификации математической модели: *структурная* идентификация; *параметрическая* идентификация.

На этапе структурной идентификации, которая, естественно, выполняется в *off-line* режиме, из анализа динамичности, нелинейности, пространственной распределённости и других априорных знаний о функционировании объекта строится структура математической модели в виде подходящих уравнений. В нашем случае — это дифференциальные уравнения в частных производных, принципиально описывающих поведение  $v$  объекта. Например, в задаче (1.1) структура математической модели описывается уравнением  $\mathbb{D}v = 0$ , где дифференциальный оператор  $\mathbb{D}$  настраивается параметром  $u$ . Задача параметрической идентификации заключается в определении (часто употребляют термин — оценка) подходящего значения параметра  $u$  ранее созданной структуры уравнений.

После структурной идентификации нередко возникает ситуация, когда искомый параметр остаётся зависимым от состояния объекта, т.е.  $u = u(v)$ . Это свидетельствует о незавершённости структурной идентификации. В таких случаях необходимо дополнительно структурировать искомый параметр, т.е. строить явную аналитическую зависимость  $u$  от  $v$ , содержащую новые параметры-коэффициенты  $u_{new}$ . Такая дополнительная структурная идентификация может проводиться на основе более глубокого исследования объекта, либо формальным

применением полиномиальных зависимостей  $u$  от  $v$  с коэффициентами полиномов  $u_{new}$ . При этом задача идентификации приводится к параметрической идентификации относительно набора коэффициентов полиномов.

Необходимо отметить, что параметры дифференциальных уравнений модели — это, как правило, *косвенно измеримые* физические величины, например, — коэффициент трения, коэффициент теплоотдачи и т.п. Достоверное значение таких параметров могут быть получены только из параметрической идентификации, даже если для этих параметров существуют справочные данные.

Дело в том, что справочные данные получают из неявного (не всегда осознаваемого экспериментаторами) решения задачи параметрической идентификации для других, существенно упрощённых объектов и моделей. Эксперименты проводят в стационарных, пространственно-осреднённых и других упрощённых условиях для моделей в виде простых алгебраических уравнений. Это означает, что если в реальную математическую модель распределённого объекта подставить полученные таким образом справочные значения параметров-коэффициентов, то может оказаться невозможным получить достаточно точное решение модели  $v(\tau)$  на  $\bar{\Omega}$ . Здесь нередко моделирование принимает «извращённую» форму ручной подгонки изначально не очень подходящих справочных значений параметров уравнений. Вряд ли это можно считать идентификацией математической модели и в дальнейшем доверять её поведению в относительно широком диапазоне входных данных.

Задачи параметрической идентификации, в отличие от задач оптимального управления, оптимального дизайна не содержат ограничений ни на состояние объекта, ни на множество параметров-управлений. Это всегда задачи без ограничений. Здесь множество

параметров-управлений  $U = E^n$  (конечномерная идентификация) или  $U = L_2^n(S)$  (бесконечномерная идентификация, возможно при  $S = S_\Delta$ ). То есть, управления принадлежат соответствующим неограниченным пространствам.

Методы оценки параметров распределённых систем могут быть весьма разнообразны. Так в обзорной работе [10] их выделяется одиннадцать: градиентные методы (наиболее популярны); стохастической аппроксимации; стохастические методы наименьших квадратов; калмановской фильтрации; нелинейной фильтрации; нелинейного программирования; максимального правдоподобия; методы теории статистических решений; модельного поиска; квазилинеаризации; методы решения алгебраических уравнений. Кроме выбора и реализации метода оценки параметров, существует и множество других проблем.

*Задача идентификации* распределённых систем традиционно требует решения следующих практически самостоятельных подзадач:

1. Выбор и обоснование математической *модели* объекта на этапе структурной идентификации, выявление параметров  $u$  для идентификации модели;
2. Планирование *измерений* (наблюдений) состояния системы для обеспечения её *наблюдаемости* посредством выбора типов датчиков и их местоположений;
3. Формирование критерия качества идентификации  $J$  на основе наблюдений состояния объекта и модели;
4. Анализ *чувствительности* критерия качества  $J$  к искомым параметрам  $u$ , и определение условий *идентифицируемости* модели;
5. Реализация физического эксперимента на *объекте*;

6. Выбор и реализация метода *оценки*<sup>1</sup> значений отыскиваемых параметров математической модели;
7. Выбор и реализация метода оптимизации<sup>2</sup> оценочных значений параметров.

После решения всех этих подзадач мы должны получить оптимальные значения параметров  $u_*$  математической модели. В теории идентификации, когда говорят об оптимальной оценке параметров (см., например, [98], с. 59), то формулируют задачу идентификации в виде экстремальной задачи:

$$u_* = \arg \min J(u).$$

При такой формулировке задачи идентификации она полностью совпадает с *прямым экстремальным подходом* (1.2).

Почти все перечисленные выше задачи, за очевидным исключением первой (уравнения модели) и пятой (эксперимент), могут успешно решаться именно прямым экстремальным подходом, что мы и будем далее делать на основе полученных ранее условий, методов, формул.

## 4.2 Критерий качества идентификации

Возможные критерии идентификации широко обсуждаются в литературе, например, [98, 39, 37]. Наибольшее распространение получил *критерий качества идентификации* квадратичный по *невязке* (часто употребляют термины — *рассогласование*, *ошибка*) наблюдений модельного и реального состояний объекта в виде функционала некоторой нормы:

$$J(u) = \|M(v) - v_e\|^2, \quad (4.1)$$

---

<sup>1</sup> Оценку параметров, в рамках прямого экстремального подхода, следует понимать как определение разумного начального приближения  $u^0$  недалеко отстоящего от оптимального  $u_*$ .

<sup>2</sup> Максимально возможное приближение к  $u_*$  из оценочного значения  $u^0$ .

где  $M(\cdot)$  — оператор наблюдений модельного состояния  $v$ , позволяющий сравнивать состояние модели  $v \in V = L_2^m(\bar{\Omega})$  с экспериментально наблюдаемым (измеряемым) состоянием объекта  $v_e \in V_e = L_2^{m_e}(\omega)$ . Здесь  $m_e \leq m$ , т.е. не все компоненты вектора состояния могут наблюдаться на объекте,  $\omega \subseteq \bar{\Omega}$  — пространственно-временное множество наблюдений состояния объекта.

Норма в целевом функционале (4.1) будет вычисляться в  $L_2^{m_e}(\omega)$ . При этом функция цели  $I(v) = (M(v) - v_e)^2$ , как квадрат невязки модели и объекта, должна определяться на  $\omega$ . У неё не будет производной  $I'_u$  и поэтому градиент  $\nabla J$  в задачах идентификации будет соответствовать общему выражению (3.36) с однородным оператором  $\mathbb{U}^*$  (или  $\bar{\mathbb{U}}^*$ ,  $\bar{\bar{\mathbb{U}}}^*$ ).

Конкретный вид области  $\omega$  и вид оператора  $M$ , а возможно и другая функция невязки  $I(v)$ , тесно связаны с характером наблюдений за объектом, т.е. с типом измерений и местоположением датчиков для сбора данных, достаточных для наблюдаемости состояния. *Наблюдаемость* — это возможность полного восстановления состояния системы по некоторым ограниченным наблюдениям за ней.

Наблюдения в модели реализуются оператором  $M$  — это выбор пространственно-временной области  $\omega$  из всей области  $\bar{\Omega}$ , а также выбор  $m_e$  компонент из всех  $m$  в векторе  $v \in L_2^m(\bar{\Omega})$ . Можно сказать короче, что  $M$  — это оператор отображения пространства состояний модели в пространство экспериментально наблюдаемых состояний объекта, т.е.

$$M : V(\bar{\Omega}) \rightarrow V_e(\omega).$$

Рассмотрим далее примеры конкретных типов критериев качества идентификации, акцентировав внимание на пространственно-временные области их определения. Для наглядности примем  $m_e = m = 1$ .

Если мы имеем дискретные наблюдения стационарного состояния объекта  $v_{e,p} \stackrel{\text{def}}{=} v_e(x_p)$  на множестве точек  $\{x_p\}_{p=1}^P \subset X$ , то норма в критерии (4.1) будет определяться как норма в пространстве  $E^P$  и мы получаем следующий критерий качества идентификации:

$$J(u) = \sum_{p=1}^P (v|_{x_p} - v_{e,p})^2 = \int_X \sum_{p=1}^P (v - v_{e,p})^2 \delta(x - x_p) dx. \quad (4.2)$$

Здесь оператор наблюдений  $M(\cdot) = \cdot|_{x_p}$  представляет собой оператор сужения действия функции  $v(x)$  на множество точек  $\{x_p\}_{p=1}^P$ . Данный оператор реализуется при помощи *дельта-функций Дирака*. Напомним, что  $\delta$ -функция — это специальная математическая функция, в одномерном случае удовлетворяющая условиям:

$$\delta(x - x_p) = \begin{cases} +\infty, & x = x_p \\ 0, & x \neq x_p \end{cases}, \quad \int_{-\infty}^{+\infty} \delta(x - x_p) dx = 1.$$

Если  $x$  — это вектор из  $E^N$ ,  $N > 1$ , то  $\delta$ -функцию необходимо понимать как произведение  $N$  покомпонентных  $\delta$ -функций:

$$\delta(x - x_p) = \delta(x_1 - x_{1,p}) \cdots \delta(x_N - x_{N,p}).$$

Если мы имеем пространственно-дискретные наблюдения динамического состояния объекта, то норма определяется как в пространстве  $L_2^P(t_0, t_1)$ , а критерий качества идентификации (4.1) принимает вид:

$$\begin{aligned} J(u) &= \sum_{p=1}^P \int_{t_0}^{t_1} (v|_{x_p} - v_{e,p})^2 dt \\ &= \int_{\Omega} \sum_{p=1}^P (v - v_{e,p})^2 \delta(x - x_p) dx dt. \end{aligned} \quad (4.3)$$

Здесь оператор наблюдений  $M(\cdot) = \cdot|_{x_p}$  сужает значения функции  $v(\tau)$  на множество  $P$  интервалов  $\{x_p \times (t_0, t_1)\}_{p=1}^P$ . Данный оператор

реализуется, как и в предыдущем случае, набором  $\delta$ -функций относительно точек  $x_p$ . Аналогичным образом может быть получен и функционал с дискретными измерениями по времени.

Часто, в случае непрерывных во времени наблюдений на границе одномерного объекта, например, на  $x_a$ , критерий (4.1) записывают в виде следующего функционала:

$$J(u) = \int_{t_0}^{t_1} (v|_{x_a} - v_e)^2 dt. \quad (4.4)$$

При непрерывных наблюдениях и во времени, и в пространстве на  $\Omega$  получаем из (4.1) следующий критерий качества идентификации:

$$J(u) = \int_{\Omega} (v - v_e)^2 d\Omega. \quad (4.5)$$

Реальные наблюдения состояния объекта всегда имеют пространственно-дискретный характер, но в большинстве случаев, применяя интерполяцию экспериментальных данных, можно перейти к непрерывному представлению состояния объекта, и использовать интегральный критерий (4.5) или (4.4).

Все приведённые критерии (4.2)–(4.5) можно рассматривать как частные случаи целевого функционала в задаче оптимизации (1.1).

Когда речь идёт об идентификации, то всегда необходимо обсуждать *помехи*, в виде *погрешностей* экспериментальных наблюдений состояния объекта  $v_e$ , погрешностей структурной идентификации модели и вычислительных алгоритмов. Все возможные погрешности будут содержаться в идентифицированных параметрах. Это совершенно нормальная ситуация. Конечно, желательно заранее снизить эти погрешности до минимума. Но полностью от них избавиться вряд ли возможно.

Поэтому надо понимать, что идентифицированные параметры могут быть хороши только для конкретных математических моделей,

численных методов решения, а также для принятых методик измерений и экспериментального оборудования.

Совсем неоднозначна ситуация, связанная с увеличением количества и типа датчиков наблюдения состояния объекта при формировании критерия качества идентификации  $J$ . Количество датчиков должно быть всего лишь достаточным для идентифицируемости модели [89]. Увеличение их числа может приводить к увеличению погрешностей в наблюдении состояния объекта. Выбор типа датчика особенно актуален для систем с многокомпонентным состоянием. В этом случае надо *измерять* ту компоненту вектора состояния, которая наиболее чувствительна к идентифицируемому параметру и имеет наименьшую погрешность измерений.

Несколько слов о чувствительности критерия качества  $J$  к искомым параметрам  $u$ . *Чувствительность*, как мы уже говорили, определяется производной  $\partial J/\partial u$ , т.е. градиентом  $\nabla J$ . В задачах бесконечномерной оптимизации чувствительность характеризуется значениями функции  $\nabla J(u; \tau)$  на  $S$  при каких-либо оценочных значениях  $u$ .

Функция чувствительности в виде  $\nabla J(u; \tau)$  может иметь существенно разные значения в разных точках  $\tau \in S$ . Это означает, что при идентификации параметра  $u(\tau)$  может оказаться невозможным равномерное достижение точного значения  $u_*(\tau)$  на  $S$ . При дальнейшем использовании такой модели, с произвольными входными данными, возможно её неадекватное поведение. Выходом в этой ситуации может быть минимизация  $J(u)$  на основе использования МРНС с равномерной сходимостью к  $u_*(\tau)$  на  $S$  (или на  $S_\Delta$ ).

Чувствительность, так же, как и наблюдаемость, представляют собой отдельные аспекты одной проблемы — идентифицируемости, которую мы далее и обсудим.

### 4.3 Идентифицируемость

Давайте сформулируем понятие идентифицируемости. Воспользуемся ранее полученными определениями и условиями управляемости, построенными на теории обратных задач с использованием экстремальных алгоритмов. Подобный взгляд на идентифицируемость распределённых систем впервые был предложен в работе автора [87].

Параметрическая идентификация является классическим представителем *обратных задач*. Такие задачи характеризуются *обратным отображением*

$$v_e \rightarrow u_*. \quad (4.6)$$

Однако, в прямом экстремальном подходе это отображение удобнее представлять в форме аналогичной (3.15):

$$V_e \xrightarrow{\min J} u_* \in U,$$

где  $U = L_2^n$  — пространство неограниченных управлений. Поиск оптимальных параметритов  $u_*$  осуществляется по какому-либо экспериментально наблюдаемому «оптимальному» состоянию  $v_e \in V_e \subseteq V$ . Подобные задачи в оптимизации, с заранее известным оптимальным состоянием системы, существуют и вне идентификации. Это задачи *синтеза оптимального управления*.

Как мы отмечали ранее, требование корректности обратных задач заключается в ограничении пространства управлений  $U$  до *компактного множества*  $\mathcal{U} \subseteq U$  существования, единственности и устойчивости решений. Для сужения множества классически некорректных решений

до компакта корректности  $\mathcal{U}$ , естественно, будем применять экстремальные алгоритмы, для которых регуляризирующие свойства сужения  $U \rightarrow \mathcal{U}$  нами ранее были рассмотрены и доказаны.

Итак, для определения идентифицируемости возьмём за основу определение 3.3 раздела 3.5.3 «Условия управляемости», перефразировав его в терминах задачи идентификации.

**Определение 4.1.** Математическая модель распределённой системы  $\mathbb{D}(\tau, v, u)v = 0$ ,  $\tau \in \bar{\Omega}$  в задаче (1.1), характеризуемая прямым отображением

$$U(S) \rightarrow V(\bar{\Omega}), \quad S \subset \bar{\Omega},$$

идентифицируема посредством параметра  $u(\tau) \in U(S)$  относительно целевого функционала  $J$ , когда обратная задача

$$V_e(\omega) \xrightarrow{\min J} u_*(\tau) \in \mathcal{U}(S), \quad \omega \subset \bar{\Omega}$$

отображения экспериментально наблюдаемых элементов пространства состояний объекта  $V_e(\omega)$  в элемент модели  $u_*$  при условии  $\min J$  является *корректной по Тихонову*.

Данное утверждение, с учётом теоремы 3.4 об управляемости, может быть сформулировано в виде следующего утверждения для практического анализа идентифицируемости в прямом экстремальном подходе.

**Теорема 4.1.** Математическая модель распределённой системы  $\mathbb{D}(\tau, v, u)v = 0$ ,  $\tau \in \bar{\Omega}$  в задаче (1.1) *идентифицируема* посредством параметра  $u(\tau)$  на  $S \subset \bar{\Omega}$  по функционалу  $J$ , если:

- существует область  $V^*(\Omega)$  определения оператора  $U^*$  (или  $\bar{U}^*$ ,  $\bar{\bar{U}}^*$ ) с его однозначными значениями в области  $U^*(S)$ ;
- оператор  $U^*$  (или  $\bar{U}^*$ ,  $\bar{\bar{U}}^*$ ) — невырожденный;
- прямой экстремальный подход (1.2) при  $u^0 \in \mathcal{U}$  использует удовлетворительные параметры регуляризации.

Данная теорема даёт нам простую процедуру контроля идентифицируемости. Давайте подробнее разберём утверждения теоремы в нашем случае. Ограничимся рассмотрением ситуации невырожденности операторов.

Пусть мы имеем подходящую оценку параметров модели  $u^0 \in \mathcal{U}$  и умеем для алгоритма МРНСг (3.5) находить удовлетворительные параметры регуляризации  $b^k$  и  $\alpha^k$ . Далее нам надо выделить область определения  $V^*(\Omega)$  необходимого оператора из набора  $\mathbb{U}^*$ ,  $\bar{\mathbb{U}}^*$ ,  $\overline{\mathbb{U}}^*$ . Остановимся на этом немного подробнее.

В задачах идентификации  $I'_v|_\omega$  — это производная невязки состояний объекта  $v_e$  и модели  $v$ , которая, например, для функционалов  $J(u)$  типа (4.3)–(4.5) принимает следующий вид:

$$(4.3) \text{ — } I'_v|_\omega = 2 \sum_{p=1}^P (v - v_{e,p}) \delta(x - x_p),$$

$$\omega = \{x_p\}_{p=1}^P \times (t_0, t_1) \subset \Omega;$$

$$(4.4) \text{ — } I'_v|_\omega = 2(v - v_e), \quad \omega = x_a \times (t_0, t_1) \subset \partial\Omega;$$

$$(4.5) \text{ — } I'_v|_\omega = 2(v - v_e), \quad \omega = (x_a, x_b) \times (t_0, t_1) = \Omega.$$

Величина  $I'_v$  и её область определения  $\omega$  представляют собой источники возмущений сопряжённого состояния  $f \in V^*(\bar{\Omega})$  в сопряжённой задаче (3.33).

В случае критерия (4.3) мы имеем множество источников возмущений в разных пространственных точках  $x_p$ . В критерии (4.4) мы имеем один граничный на  $x_a$  источник возмущений. В (4.5) область источников возмущений совпадает со всей областью  $\Omega$  (без границ).

Обычно, из априорных знаний в каждой конкретной задаче известно, что где-либо на  $\omega \subset \bar{\Omega}$  производная  $I'_v \neq 0$ , а вот какой

конкретно должна быть область наблюдений  $\omega$  — это вопрос, требующий тщательных исследований. Имеющаяся или планируемая область  $\omega$  может определять такие источники возмущений  $I'_v$ , которые не смогут однозначно изменять сопряжённое состояние  $f$  в  $V^*(\Omega)$  для нахождения градиента  $\nabla J \in U^*(S)$ . В этом случае исходная система (модель объекта) будет неидентифицируема. Для того, чтобы добиться идентифицируемости придётся менять методику *измерений*, возможно, менять датчики измерительной аппаратуры, их пространственное положение, и формировать новый вид критерия качества идентификации.

В целом, анализ идентифицируемости проводится по описанной ранее процедуре анализа управляемости. В разделах шесть и восемь такой анализ реализуется при решении прикладных задач параметрической идентификации распределённых систем, параболической и гиперболической.

Имея понятие идентифицируемости математической модели в прямом экстремальном подходе, отпадает необходимость в понятиях наблюдаемости состояния и чувствительности критерия качества. Они кажутся устаревшими.

## 5 Простейшая задача оптимального управления тепловыми процессами

Начиная с данного раздела и до конца книги, мы будем решать и обсуждать прикладные задачи оптимизации, оптимального управления, идентификации для систем с распределёнными параметрами с точки зрения теории прямого экстремального подхода. Некоторые целевые функционалы  $J(u)$  будут квадратичными (раздел 5), некоторые только выпуклыми и локально квадратичными (разделы 6–8) или даже выпукловогнутыми (раздел 9). Будет рассмотрено большое разнообразие помех и ограничений на управление.

Как показали расчёты, алгоритмы МРНС с постоянным параметром  $\alpha = \alpha^0$  дают хорошие результаты в рассмотренных задачах выпуклой оптимизации (разделы 5–8). Особо выделяется задача максимизации функционала из раздела 9 с несколькими локальными максимумами. Здесь применение МРНС невозможно, были использованы градиентные методы с адаптивной релаксацией шагов.

Во всех задачах проведён анализ управляемости (идентифицируемости), выявлены области определения сопряжённых состояний  $V^*(\Omega)$  операторов  $\mathcal{U}^*$ ,  $\bar{\mathcal{U}}^*$ ,  $\bar{\bar{\mathcal{U}}}^*$  для нахождения градиента  $\nabla J$ . В задаче раздела 7 оказалось, что требуемый оператор может быть вырожденным.

Описание постановок всех задач будем начинать с традиционной формы, привычной для той прикладной области, где они возникают. В то же время, мы будем проводить параллели с терминологией и обозначениями предыдущих разделов книги для того, чтобы использовать полученные ранее результаты, и чтобы теория прямого экстремального подхода наполнилась наглядной практикой.

## 5.1 Постановка задачи

Начнём рассмотрение с простого примера [92] в виде задачи *синтеза оптимального управления*. В таких задачах известно оптимальное состояние системы и необходимо найти соответствующее оптимальное управление. Задачи синтеза допускает постановку наглядных тестовых расчётов оптимизации. Именно поэтому первая задача является основной тестовой задачей для иллюстрации особенностей оптимизации в бесконечномерных пространствах и демонстрации работоспособности, эффективности алгоритмов, методов, описанных ранее. Эту задачу мы разберём особенно подробно.

Итак, необходимо найти оптимальное управление внешним потоком тепла в некотором гипотетическом химическом реакторе — рисунок 5.1.

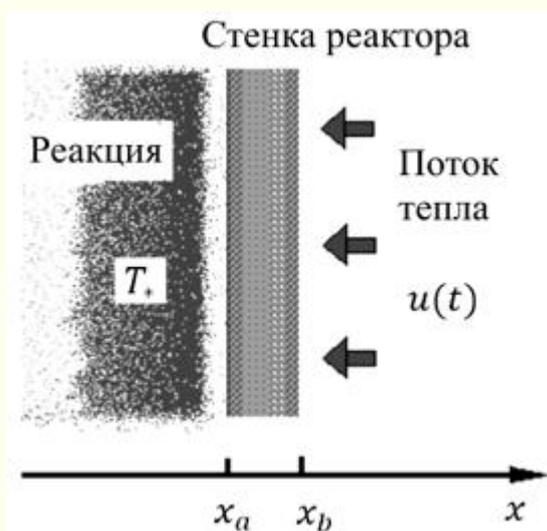


Рисунок 5.1 — К задаче оптимального управления потоком тепла в химическом реакторе

Предполагается, что химическая реакция в таком реакторе протекает с известной плотностью поглощаемого потока тепла  $q(t)$  и должна протекать при температуре  $T_*(t)$ . Значения  $q$  и  $T_*$  задаются с внутренней стороны стенки реактора на границе  $x_a$ , а управление плотностью потока вводимого тепла  $u(t)$  осуществляется на ее наружной поверхности при  $x_b$ . Целью управления является удержание реакции в заданном температурном режиме  $T_*(t)$ .

Сформулируем *исходную задачу* теплопереноса. Пространственно-распределённый нестационарный процесс теплопередачи описывается следующим *линейным, однородным параболическим* уравнением:

$$C\rho \frac{\partial T}{\partial t} - \lambda \frac{\partial^2 T}{\partial x^2} = 0, \quad (x, t) \stackrel{\text{def}}{=} \tau \in \Omega = (x_a, x_b) \times (t_0, t_1), \quad (5.1)$$

где  $T(x, t)$  — функция температуры в стенке реактора,  $C$ ,  $\rho$  и  $\lambda$  — положительные числа, соответственно, теплоемкость, плотность и теплопроводность.

Граничные условия *второго рода* для уравнения (5.1) имеют вид:

$$\lambda \frac{\partial T}{\partial x} = q \quad \text{на } \Gamma_a = x_a \times (t_0, t_1), \quad (5.2)$$

$$\lambda \frac{\partial T}{\partial x} = u \quad \text{на } \Gamma_b \equiv S = x_b \times (t_0, t_1). \quad (5.3)$$

Начальное условие:

$$T = T_0 \quad \text{на } \Gamma_0 = [x_a, x_b] \times t_0, \quad (5.4)$$

Отметим, что граничная область  $\Gamma_\partial = \Gamma_a \cup \Gamma_b$ , а также начальная  $\Gamma_0$  и терминальная  $\Gamma_1 = [x_a, x_b] \times t_1$  образуют границу  $\partial\Omega$  ограниченного прямоугольника  $\Omega$ , а замыкание  $\bar{\Omega} = \Omega \cup \partial\Omega$ .

Степень отклонения температуры  $T(x_a, t)$  химической реакции от заданной температуры  $T_*(t)$  на  $\Gamma_a$  будем оценивать *целевым функционалом*

$$J(u) = \int_{t_0}^{t_1} (T(x_a, t) - T_*(t))^2 dt. \quad (5.5)$$

Задача оптимального управления реактором формулируется следующим образом. Для распределённой системы (5.1) с краевыми условиями (5.2)–(5.4) необходимо найти плотность граничного потока тепла  $u(t) \in L_2(S)$ , доставляющую минимум функционалу (5.5).

Давайте уточним связь терминологий и обозначений поставленной задачи с предыдущими разделами теории прямого экстремального подхода.

Состояние системы (5.1) — это положительная температура  $T(x, t) \stackrel{\text{def}}{=} v(\tau) \in V = L_{2+}(\bar{\Omega})$ , определённая на  $\bar{\Omega}$ , которая регулируется управлением  $u(t) \in U \subseteq L_2(S)$ , определённым на  $S = \Gamma_b$ . Положительность состояния не является ограничением в задаче оптимизации, а лишь требованием разумности решения исходной задачи. При этом, управление может быть, как положительным, так и отрицательным. Тепло в реактор может и подаваться, и забираться. Размерности функций (вектор-функций) управления и состояния равны  $n = m = 1$ .

Уравнение (5.1) можно представить в операторном виде аналогично системе (3.20):

$$\mathbb{D}v = C\rho \frac{\partial v}{\partial t} - \lambda \frac{\partial^2 v}{\partial x^2} = 0 \quad \text{на } \Omega,$$

где дифференциальный оператор  $\mathbb{D} \cdot = C\rho \frac{\partial \cdot}{\partial t} - \lambda \frac{\partial^2 \cdot}{\partial x^2}$  не содержит коэффициентов, зависящих от  $\tau$ ,  $v$ ,  $u$ . Свободный член  $F$  отсутствует. Функционал (5.5) можно представить в виде второго граничного интеграла в общем выражении (3.19):

$$J(u) = \int_{\partial\Omega} I(v)|_{\omega} d\theta\Omega = \int_{\Gamma_a} I(v) dt,$$

где  $I(v; t) = (T(x_a, t) - T_*(t))^2$ , а область  $\omega = \Gamma_a$ .

Задача оптимального управления (5.1)–(5.5), может иметь и ограничения, когда  $U \subset L_2(S)$ , т.е. множество допустимых управлений  $U$  не совпадает со всем пространством  $L_2$ .

По техническим соображениям могут присутствовать ограничения на минимально допустимые значения потока тепла в реактор. Здесь множество допустимых управлений задаётся явно, например, в виде полуограниченного (ограничен только с одной стороны) диапазона, что относится к ограничениям типа 2.1:

$$U = \{u(t) \in L_2(S) \mid u(t) \geq u_{\min}\}. \quad (5.6)$$

Может присутствовать ограничение на множество состояний системы в виде неостывания реактора:

$$V = \{T(x, t) \in L_{2+}(\bar{\Omega}) \mid T(x, t) \geq T_{\min}\}. \quad (5.7)$$

Оно относится к типу 1.1, как неявное ограничение на управление, заданное явным выражением для допустимых состояний системы. Множество  $U$  в явном виде здесь неизвестно.

Кроме того, рассмотрим и возможное изопериметрическое ограничение в виде требования заданного суммарного расхода тепла для управления реактором в течение рассматриваемого времени. Это явное ограничение на управление  $u \in U \subset L_2$  типа 2.2 с неявно заданным множеством  $U$ :

$$\int_{t_0}^{t_1} u(t) dt = C, \quad (5.8)$$

где  $C$  — изопериметрическая постоянная расхода тепла.

По техническим причинам может оказаться, что подача тепла в реактор не может быть реализована плавным регулированием, а только ступенчатым, с шагом  $\hat{\Delta}t$ . Это означает, что решение задачи

оптимального управления мы должны искать в классе *кусочно-постоянных функций*. Это также ограничение *типа 2.2*:

$$u(t) = \{u_i \text{ на } \hat{\Delta}t_i\}_{i=1 \dots \hat{n}}, \quad (5.9)$$

где  $\hat{n}$  — заданное количество регулировок механизмов подачи тепла, а сумма  $\sum_{i=1}^{\hat{n}} \hat{\Delta}t_i = t_1 - t_0$ .

И последнее ограничение, которое также относится к *типу 2.2*. Рассмотрим ситуацию, когда механизмы регулировки подачи тепла не могут реализовывать подачу с любой интенсивностью. То есть мы имеем ограничение на *скорость изменения* плотности потока тепла в виде дифференциального неравенства:

$$\frac{du(t)}{dt} \leq \vartheta_{\max}, \quad t \in (t_0, t_1). \quad (5.10)$$

Мы привели пример пяти возможных ограничений (5.6)–(5.10), которые далее реализуем в рамках прямого экстремального подхода.

## 5.2 Градиент целевого функционала

Для прямой минимизации целевого функционала (5.5) нам необходимо найти его градиент. Для этого будем использовать технику, описанную в разделе 3.4 «Градиент неявно заданного функционала».

### 5.2.1. Выделение ограничений на управление

Ограничение первого типа (5.7), неявно заданное на управление, следует учитывать в виде *штрафного функционала*

$$\int_{\Omega} I_F(T, T_{\min})|_{\omega_F} dt dx.$$

Будем использовать внешнюю *штрафную функцию*

$$I_F|_{\omega_F} = \begin{cases} (T - T_{\min})^2, & \text{если } T(\tau) < T_{\min}, \tau \in \omega_F, \\ 0 & \text{иначе, } \tau \notin \omega_F. \end{cases}$$

Приведённое выражение определяет не только значение функции  $I_F$  в разных точках  $\tau$  области  $\bar{\Omega}$ , но и область  $\omega_F$  *активного* действия штрафа, где температура опускается ниже минимально допустимой  $T_{\min}$ .

Теперь, согласно (3.42), вместо критерия качества управления (5.5), получаем *целевой функционал*:

$$J(u) = \int_{\Gamma_a} I(v)dt + \kappa \int_{\Omega} I_F(T, T_{\min})|_{\omega_F} dt dx. \quad (5.11)$$

При этом ограничение на состояние системы в виде условия (5.7) исчезнет.

Обратимся к разнообразным ограничениям второго типа (5.6), (5.8)–(5.10), явно заданным на управление. Их следует учитывать после определения градиента  $\nabla J$ , в экстремальных алгоритмах, как это было показано в подразделах 3.6.3–3.6.6 «Управление с ограничениями».

Таким образом, получаем, что множества состояний и управлений  $V = L_{2+}(\bar{\Omega})$ ,  $U = L_2(S)$  и далее будем рассматривать задачу неограниченной минимизации  $J(u)$ .

Для определения градиента  $\nabla J$  необходимо выписать главную линейную часть приращения  $J$ . Это можно сделать посредством варьирования, что мы далее и будем делать.

### 5.2.2. Линеаризация задачи

Вариация плотности теплового потока  $\delta u$  приводит к вариации температуры  $\delta T \stackrel{\text{def}}{=} \delta v$ . При этом вариация функционала (5.11), по аналогии с (3.22), принимает вид:

$$\begin{aligned} \delta J(u) &= \langle I'_v, \delta v \rangle_{V^*(\Gamma_a)} + \langle \kappa I'_{Fv} |_{\omega_F}, \delta v \rangle_{V^*(\Omega)}, \\ I'_v &= 2(T - T_*), \quad I'_{Fv} = 2(T - T_{\min}). \end{aligned} \quad (5.12)$$

Здесь вариация  $\delta v$  присутствует на левой границе  $\Gamma_a$  стенки реактора.

Вариация уравнения теплопереноса в операторном виде приводит к выражению типа (3.23):

$$\delta \mathbb{D}v = \mathbb{V}\delta v = 0 \in V(\Omega),$$

или в явном виде:

$$\mathbb{V}\delta T = C\rho \frac{\partial \delta T}{\partial t} - \lambda \frac{\partial^2 \delta T}{\partial x^2} = 0 \quad \text{на } \Omega. \quad (5.13)$$

В сравнении с (3.23) здесь оператор  $\mathbb{V} = \mathbb{D}$ , а оператор  $\mathbb{U}$  на  $\Omega$  вообще отсутствует, т.е. имеем одно отображение

$$\mathbb{V} : V(\Omega) \rightarrow V(\Omega).$$

### 5.2.3. Отображение линеаризованных уравнений в сопряжённые пространства

Возьмём в сопряжённом пространстве произвольный линейный функционал  $\tilde{f}(\tau) \in V^*(\bar{\Omega})$  и согласно (3.24) запишем его значение на вариации системы (5.13):

$$\langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} = 0.$$

Это выражение следует преобразовать к значениям некоторых линейных функционалов, заданных непосредственно на элементах  $\delta v$ , т.е. представить в виде тождества Лагранжа типа (3.25). В нашем одномерном случае, с первой производной по времени и второй производной по пространству, оно будет иметь вид:

$$\begin{aligned} \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} &\equiv \langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\Omega)} \\ &+ \langle \bar{\mathbb{V}}^* \tilde{f}, \delta v \rangle_{V^*(\partial\Omega)} + \langle \bar{\bar{\mathbb{V}}}^* \tilde{f}, \frac{\partial \delta v}{\partial x} \rangle_{V^*(\Gamma_\partial)} = 0. \end{aligned} \quad (5.14)$$

Теперь запишем данное тождество в явном виде и найдём выражение для сопряжённых операторов  $\mathbb{V}^*$ ,  $\bar{\mathbb{V}}^*$ ,  $\bar{\bar{\mathbb{V}}}^*$ . Напомним, что одна черта над  $\bar{\mathbb{V}}^*$  соответствует граничным сопряжённым операторам при  $\delta v$ , а две черты над  $\bar{\bar{\mathbb{V}}}^*$  — при производных  $\frac{\partial \delta v}{\partial x}$ .

Сделаем подробно необходимые преобразования типа (3.27), используя правило дифференцирования произведений:

$$\begin{aligned}
\langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} &\equiv \iint_{\Omega} \tilde{f} C \rho \frac{\partial \delta T}{\partial t} dx dt - \iint_{\Omega} \tilde{f} \lambda \frac{\partial^2 \delta T}{\partial x^2} dx dt \\
&= \iint_{\Omega} \left[ \frac{\partial}{\partial t} (\tilde{f} C \rho \delta T) - \frac{\partial \tilde{f}}{\partial t} C \rho \delta T \right] dx dt - \iint_{\Omega} \left[ \frac{\partial}{\partial x} \left( \tilde{f} \lambda \frac{\partial \delta T}{\partial x} \right) - \frac{\partial \tilde{f}}{\partial x} \lambda \frac{\partial \delta T}{\partial x} \right] dx dt \\
&= \int_{x_a}^{x_b} \tilde{f} C \rho \delta T \Big|_{t_0}^{t_1} dx - \iint_{\Omega} C \rho \frac{\partial \tilde{f}}{\partial t} \delta T dx dt - \int_{t_0}^{t_1} \tilde{f} \lambda \frac{\partial \delta T}{\partial x} \Big|_{x_a}^{x_b} dt \\
&\quad + \iint_{\Omega} \lambda \frac{\partial \tilde{f}}{\partial x} \frac{\partial \delta T}{\partial x} dx dt,
\end{aligned}$$

то есть, тождество (5.14) на данном этапе преобразований принимает вид:

$$\begin{aligned}
\langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} &\equiv \langle -C \rho \frac{\partial \tilde{f}}{\partial t}, \delta T \rangle_{V^*(\Omega)} + \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Omega)} \\
&\quad + \langle C \rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_1)} - \langle C \rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_0)} - \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Gamma_b)} + \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Gamma_a)}.
\end{aligned}$$

В пространстве  $V^*(\Omega)$ , как мы видим, всё ещё осталась производная от вариации  $\frac{\partial \delta T}{\partial x}$ . Избавимся от неё в соответствующем слагаемом:

$$\begin{aligned}
\langle \lambda \frac{\partial \tilde{f}}{\partial x}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Omega)} &\equiv \iint_{\Omega} \lambda \frac{\partial \tilde{f}}{\partial x} \frac{\partial \delta T}{\partial x} dx dt \\
&= \iint_{\Omega} \left[ \frac{\partial}{\partial x} \left( \lambda \frac{\partial \tilde{f}}{\partial x} \delta T \right) - \lambda \frac{\partial^2 \tilde{f}}{\partial x^2} \delta T \right] dx dt \\
&= \int_{t_0}^{t_1} \lambda \frac{\partial \tilde{f}}{\partial x} \delta T \Big|_{x_a}^{x_b} dt - \iint_{\Omega} \lambda \frac{\partial^2 \tilde{f}}{\partial x^2} \delta T dx dt \\
&= \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_b)} - \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_a)} - \langle \lambda \frac{\partial^2 \tilde{f}}{\partial x^2}, \delta T \rangle_{V^*(\Omega)}
\end{aligned}$$

В итоге тождество (5.14) принимает вид:

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} \\
& \equiv \langle -C\rho \frac{\partial \tilde{f}}{\partial t} - \lambda \frac{\partial^2 \tilde{f}}{\partial x^2}, \delta T \rangle_{V^*(\Omega)} \\
& + \langle C\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_1)} - \langle C\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_0)} \\
& - \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Gamma_b)} + \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Gamma_a)} \\
& + \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_b)} - \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_a)} = 0.
\end{aligned} \tag{5.15}$$

Полученное выражение значительно проще описанных ранее (3.27), (3.29) поскольку в нашем случае матрица  $A(x, v, u) = -\lambda$ , т.е. — это всего лишь постоянное число.

Из (5.15) видно, что сопряжённый оператор  $\mathbb{V}^*$  имеет следующий вид с соответствующим отображением:

$$\mathbb{V}^* \cdot = -C\rho \frac{\partial \cdot}{\partial t} - \lambda \frac{\partial^2 \cdot}{\partial x^2} : V^*(\Omega) \rightarrow V^*(\Omega).$$

Граничные операторы  $\bar{\mathbb{V}}^*$ ,  $\bar{\bar{\mathbb{V}}}^*$  и их общий вид, как вид операторов, не имеет ценности с точки зрения их дальнейшего анализа. Всё же, для удовлетворения любопытства, приведём их:

$$\bar{\mathbb{V}}^* \cdot := -C\rho \cdot|_{\Gamma_0} + C\rho \cdot|_{\Gamma_1} - \lambda \frac{\partial \cdot}{\partial x} \Big|_{\Gamma_a} + \lambda \frac{\partial \cdot}{\partial x} \Big|_{\Gamma_b} \quad \text{на } \partial\Omega,$$

$$\bar{\bar{\mathbb{V}}}^* \cdot := \lambda \cdot|_{\Gamma_a} - \lambda \cdot|_{\Gamma_b} \quad \text{на } \Gamma_a \cup \Gamma_b.$$

В дальнейших выкладках будем всегда писать граничные выражения в явном, а не операторном виде. Это должно упростить понимание краевых условий задачи.

#### 5.2.4. Объединение задачи

Объединим линейные функционалы из (5.12) и (5.15) в одинаковых пространствах на элементах, к которым они относятся, и

получим выражение вариации целевого функционала (без учёта краевых условий), аналогичное (3.30):

$$\begin{aligned}
\delta J(u) = & \langle \mathbb{V}^* \tilde{f} + \kappa I_{F_v}'|_{\omega_F}, \delta T \rangle_{V^*(\Omega)} \\
& + \langle C\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_1)} - \langle C\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_0)} \\
& - \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Gamma_b)} + \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Gamma_a)} \\
& + \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_b)} + \langle -\lambda \frac{\partial \tilde{f}}{\partial x} + I_v', \delta T \rangle_{V^*(\Gamma_a)}.
\end{aligned} \tag{5.16}$$

Здесь производная от штрафной функции  $I_{F_v}'$  присутствует в первом слагаемом на  $\Omega$ , а производная от целевой функции  $I_v'$  присутствует в последнем слагаемом на границе  $\Gamma_a$ .

### 5.2.5. Учёт краевых условий

Варьируем граничные и начальные условия (5.2)–(5.4):

$$\lambda \frac{\partial \delta T}{\partial x} = 0 \quad \text{на } \Gamma_a, \tag{5.17}$$

$$\lambda \frac{\partial \delta T}{\partial x} = \delta u \quad \text{на } \Gamma_b = S, \tag{5.18}$$

$$\delta T = 0 \quad \text{на } \Gamma_0, \tag{5.19}$$

Теперь значения вариаций  $\frac{\partial \delta T}{\partial x}$ ,  $\delta T$  из (5.17)–(5.19) подставим в вариацию функционала  $\delta J$  (5.16) на соответствующих множествах  $\Gamma_a$ ,  $\Gamma_b$ ,  $\Gamma_0$ . Получим следующее выражение *первой вариации* целевого функционала:

$$\begin{aligned}
\delta J(u) = & \langle \mathbb{V}^* \tilde{f} + \kappa I_{F_v}'|_{\omega_F}, \delta T \rangle_{V^*(\Omega)} + \langle C\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_1)} \\
& + \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_b)} + \langle -\lambda \frac{\partial \tilde{f}}{\partial x} + I_v', \delta T \rangle_{V^*(\Gamma_a)} \\
& - \langle \tilde{f}, \delta u \rangle_{U^*(S)}.
\end{aligned} \tag{5.20}$$

В вариации целевого функционала в последнем слагаемом (при граничном условии второго рода на  $\Gamma_b = S$ ) появилась и вариация

управления  $\delta u$ , откуда получаем сопряжённый оператор  $\bar{\bar{U}}^*$  с соответствующим отображением:

$$\bar{\bar{U}}^* = -1 : V^*(S) \rightarrow U^*(S).$$

Напомним, что две черты над  $\bar{\bar{U}}^*$  означают граничный сопряжённый оператор на части  $\partial\Omega$ , где имело место граничное условие второго рода.

### 5.2.6. Выделение градиента

В выражении (5.20) зададим значение  $\tilde{f} = f$ , такое, чтобы все линейные функционалы при  $\delta T$  обратились в ноль:

$$\mathbb{V}^* f = -C\rho \frac{\partial f}{\partial t} - \lambda \frac{\partial^2 f}{\partial x^2} + \kappa I_{Fv}' \Big|_{\omega_F} = 0 \quad \text{на } \Omega, \quad (5.21)$$

$$-\lambda \frac{\partial f}{\partial x} + I_v' = 0 \quad \text{на } \Gamma_a, \quad (5.22)$$

$$\lambda \frac{\partial f}{\partial x} = 0 \quad \text{на } \Gamma_b, \quad (5.23)$$

$$f = 0 \quad \text{на } \Gamma_1. \quad (5.24)$$

Мы получили *сопряжённую задачу*, аналогичную (3.33), в виде дифференциального уравнения (5.21) с граничными условиями (5.22), (5.23) и начальным (терминальным) условием (5.24) для расчёта сопряжённого состояния  $f \in V^*(\bar{\Omega})$ .

При таком значении сопряжённой переменной  $f$  первая вариация целевого функционала (5.20) принимает вид аналогичный (3.34):

$$\delta J(u) = \langle \bar{\bar{U}}^* f, \delta u \rangle_{U^*(S)} = \langle \nabla J, \delta u \rangle_{U^*(S)},$$

следовательно, *градиент целевого функционала*  $\nabla J = \bar{\bar{U}}^* f$  и имеет вид:

$$\nabla J(u; t) = -f \in U^*(S). \quad (5.25)$$

Полученное выражение градиента соответствует третьей строке общей формы градиента (3.36), когда управление присутствует только в граничном условии второго рода.

В рассматриваемом случае операторы  $U^* = \bar{U}^* = 0$ . Сравним полученное выражение оператора  $\bar{U}^* = -1$  с «теоретическим» выражением для граничного условия второго рода с управлением, которое нами рассматривалось ранее в разделе 3.4.5. Там, через граничный оператор  $\bar{V}^*$ , мы имели  $\bar{U}^* = \lambda^{-1}\bar{V}^*$ . Если подставить найденное сейчас на  $\Gamma_b$  значение  $\bar{V}^* = -\lambda$ , то получим  $\bar{U}^* = \lambda^{-1}\bar{V}^* = -1$ , что мы и получили в градиенте (5.25), не используя явные выражения граничных операторов  $\bar{V}^*$ ,  $\bar{V}^*$ .

Напомним, что выражение (5.25) будет представлять градиент  $\nabla J$ , а не просто некоторую производную Фреше  $J'_u$ , только в случае выполнения условий управляемости, которые мы далее обсудим.

### 5.3 Управляемость

Рассмотрим условия корректной постановки задачи оптимального управления (5.1)–(5.4) для возможности её решения прямым экстремальным подходом. Выясним, когда производная Фреше  $J'_u$  в первой вариации  $\delta J$  будет градиентом  $\nabla J$  (5.25). То есть, нам предстоит ответить на вопрос: «При каких условиях распределённая система (5.1)–(5.3) будет управляема по целевому функционалу (5.4)?». Для этого мы воспользуемся теоремой 3.4 об управляемости.

Согласно теореме, для проведения анализа управляемости следует начинать с обоснования классической корректности исходной и сопряжённой задач. Далее необходимо выявить область определения  $V^*(\Omega)$  и область значений  $U^*(S)$  оператора  $\bar{U}^*$ , сделать оценку возможной вырожденности  $\bar{U}^*$ . Это обеспечит получение градиента  $\nabla J$ . В конечном итоге, на основе градиента  $\nabla J$ , необходимо корректно (при удовлетворительных параметрах) применить экстремальные методы.

Сразу же отметим очевидный факт: оператор  $\bar{\bar{U}}^* = -1$  невырожденный, значение  $\bar{\bar{U}}^* f$  может обращаться в ноль только при  $f = 0$  на  $S$ . У оператора  $\bar{\bar{U}}^*$  существует только нулевое ядро. Таким образом, выполняется второе условие теоремы 3.4 «оператор  $\bar{\bar{U}}^*$  — невырожденный».

Исходная задача (5.1)–(5.4) с положительными коэффициентами является параболической. При заданных краевых условиях она имеет классически корректное решение на  $\bar{\Omega}$  [83, 76, 31].

Посмотрим на линейную сопряжённую задачу (5.21)–(5.24). Уравнение (5.21) с полученным оператором  $\mathbb{V}^*$  будет параболическим, если его решать в обратном времени от  $t_1$  к  $t_0$ . Знак минус при производной по времени именно это и делает. Данное обстоятельство хорошо согласуется с терминальным условием (5.24) на  $\Gamma_1$ , которое является результатом переворота начального условия (5.4) с  $\Gamma_0$  на  $\Gamma_0^* \equiv \Gamma_1$ . Уравнение (5.21) описывает сопряжённые тепловые процессы с тем же самым коэффициентом теплопроводности  $\lambda$ , что и в исходном уравнении (5.1).

Очевидно, что сопряжённая задача (5.21)–(5.24), с приведёнными граничными и начальным (терминальным) условиями, аналогично исходной прямой задаче имеет классически корректное решение на всей области  $\bar{\Omega}$ .

Найдём область определения оператора  $\bar{\bar{U}}^*$  и область его значений, воспользовавшись определением 3.5. Будем рассматривать ситуацию без ограничения на температуру потока тепла, т.е. примем  $\kappa = 0$ . Ограничение не будет принципиально влиять на условия управляемости.

Ненулевое решение  $f$  на внешней стенке реактора  $\Gamma_b = S$ , может появляться только за счёт источника сопряжённого теплового возмущения  $I'_v|_\omega = 2\delta T$ , который находится на внутренней стенке

реактора, где  $\omega = \Gamma_a$ ,  $\delta T = T - T_*$ . Первопричиной сопряжённых возмущений является *вариация управления*  $\delta u = u - u_*$  на  $\Gamma_b$ .

Уравнения исходной и сопряжённой задач однородные и не содержат диссипативных членов, которые могли бы исказить возмущения при их распространении. Искажения возможны только из-за вычислительных помех.

Напомним, что возмущения в параболических уравнениях распространяются мгновенно и бесконечно далеко, с асимптотическим убыванием от своего источника. Поэтому, можно утверждать, что возмущения  $\delta u$  на правой границе  $\Gamma_b$  должны однозначно (в силу корректности исходной задачи) приводить к возмущениям  $\delta T = T - T_*$  на левой границе  $\Gamma_a$ , т.е. должна выполняться левая ветка определения 3.5:

$$U(S) \ni u \xrightarrow{\text{однозначно}} I'_v|_{\omega}.$$

В сопряжённой задаче источником возмущений изначально нулевого состояния  $f$  на  $\Gamma_1$  является производная  $I'_v|_{\omega} \neq 0$  на  $\Gamma_a$ . В силу параболичности сопряжённой задачи, возмущения  $f(x_a, t)$  на левой границе мгновенно приводят к появлению возмущений  $f(x_b, t)$  на правой границе. Однако, оно может быть очень малым, даже бесконечно малым. С учётом вычислительных погрешностей<sup>1</sup> такие изменения в  $f(x_b, t)$  могут быть нулевыми. Только через некоторый, относительно большой промежуток времени  $\Delta t$ , сопряжённые возмущения слева могут реально проявиться справа при  $t < t_1 - \Delta t$ . Величина интервала  $\Delta t$  зависит от величины возмущений  $\delta T$  слева, и от толщины стенки реактора  $[x_a, x_b]$ .

---

<sup>1</sup> Такие помехи могут быть и в исходной задаче при расчёте температуры на левой границе в начальные моменты времени, но они не «успевают повлиять» на решение сопряжённой задачи на правой границе.

Это означает, что в определении 3.5 отображение во второй (средней) ветке

$$I'_v|_\omega \xrightarrow{\text{однозначно}} f|_\Omega$$

может выполняться только при  $\Omega = S_\Delta$ , где  $S_\Delta = x_b \times (t_0, t_1 - \Delta t)$ . При этом область значений оператора  $\bar{U}^*$  будет не  $U^*(S)$ , а  $U^*(S_\Delta)$ . Последняя ветка упомянутого определения принимает вид:

$$V^*(\Omega) \xrightarrow{\bar{U}^*} U^*(S_\Delta).$$

Итак, мы нашли область определения  $V^*(\Omega)$  и область значений  $U^*(S_\Delta)$  оператора  $\bar{U}^*$ .

Мы видим, первое условие теоремы 3.4 «существует область  $V^*(\Omega)$  определения оператора  $\bar{U}^*$  с его значениями в области  $U^*(S)$ » выполняется только для  $U^*(S_\Delta)$ . Это означает, что модель распределённой системы (5.1)–(5.4) не может быть управляемой по целевому функционалу (5.5) на множестве  $S$ . Избавиться от этой неприятности можно сужением области управления до подмножества:

$$S_\Delta = x_b \times (t_0, t_1 - \Delta t) \subset S.$$

В принципе, возможно  $S_\Delta = S$ , но для этого необходимо либо  $x_a \approx x_b$ , либо осуществлять искусственное усиление возмущений  $I'_v$  какими-либо весовыми коэффициентами в виде функций времени в целевом функционале  $J$ . Однако, как выбирать такие весовые функции неизвестно.

На рисунке 5.2 в виде серой вертикальной линии справа показано множество

$$\Omega = S_\Delta$$

на котором существует область  $V^*(\Omega)$  определения оператора  $\bar{U}^*$ , принимающего значения  $\bar{U}^* f = \nabla J$  в области  $U^*(S_\Delta)$ . Серая линия слева — это область  $\omega$ , где задан целевой функционал  $J$ . Стрелка из этой

области указывает направление распространения сопряжённых возмущений.

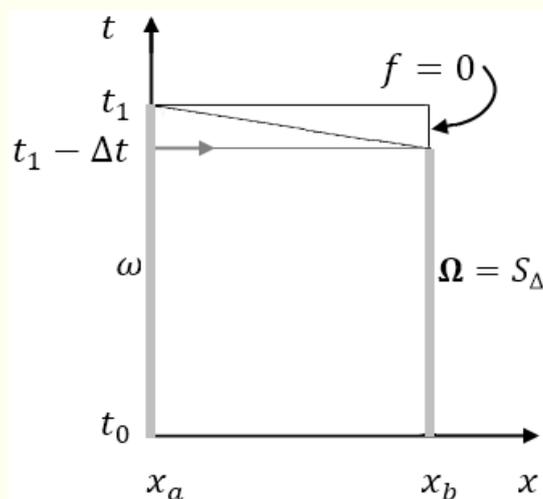


Рисунок 5.2 — Пространственно-временная диаграмма с областью  $\Omega$

Мы получили время принципиальной возможности управления системой в задаче (5.1)–(5.5):

$$t \in (t_0, t_1 - \Delta t), \quad (5.26)$$

которое является *условием управляемости* в рассматриваемой задаче. Управление  $u(t)$  вблизи конечного времени  $t_1$  не влияет на значение целевого функционала и там оно не может быть оптимизировано.

Таким образом, согласно Теореме 3.4, распределённая система (5.1) с краевыми условиями (5.2)–(5.4) управляема посредством  $u(t)$  на  $S_\Delta$  по функционалу  $J$  (5.5). При выполнении условия управляемости (5.26) производная Фреше  $J'_u = \nabla J$ , и становится возможным применение экстремальных алгоритмов для спуска к оптимальности  $u_*(t)$  на  $S_\Delta$ .

Кроме условий управляемости существует и ещё одна особенность рассматриваемой задачи бесконечномерной оптимизации, а именно, — вопрос поточечной сходимости управлений  $u^k(t)$  к  $u_*(t)$ . Местом

отсутствия такой сходимости, очевидно, является некоторая окрестность точки  $(x_b, t_1)$ . Рассматриваемая величина  $\Delta t$  должна исключать из  $S$  относительно большую<sup>1</sup> окрестность конечной меры, где может никогда не достигаться сходимость управлений к оптимуму. Поэтому, интервал  $\Delta t$  должен быть достаточно большим, чтобы можно было говорить не только об управляемости, но и о равномерной сходимости  $u^k(t) \rightarrow u_*(t)$  на  $S_\Delta$ . Именно такую сходимость мы будем далее реализовать алгоритмами МРНС.

## 5.4 Численное решение

Тестовая задача синтеза оптимального управления решалась со следующими исходными данными для гипотетического химического реактора. Толщина стенки реактора, т.е. отрезок  $[x_a, x_b] = [0.2 \text{ м}]$ , теплоёмкость стенки  $C = 500 \text{ Дж}/(\text{кг} \cdot \text{К})$ , плотность  $\rho = 7800 \text{ кг}/\text{м}^3$ , теплопроводность  $\lambda = 477 \text{ Вт}/(\text{м} \cdot \text{К})$ , плотность оттока тепла на внутренней стенке реактора  $q = 150 \text{ кДж}/(\text{м}^2 \cdot \text{с})$ , начальная температура стенки реактора  $T_0 = 300 \text{ К}$ .

### 5.4.1. Конечно-разностные аппроксимации

Исходная (5.1)–(5.4) и сопряжённая (5.21)–(5.24) параболические задачи решались численно по *неявной конечно-разностной схеме* Кранка-Николсона [23, 71, 55] с шеститочечным шаблоном (рисунок 5.3) второго порядка точности. Задавались сетка  $\{x_i = x_a + i\Delta x, i = 0 \dots m, \Delta x = (x_b - x_a)/m\}$  и  $\{t_j = t_0 + j\Delta t, j = 0 \dots n, \Delta t = (t_1 - t_0)/n\}$ . Решения задач, принадлежащие к слоям по времени, будем отмечать верхним индексом  $j$ , а по пространству — нижним индексом  $i$ .

<sup>1</sup> Смотрите обсуждение равномерной сходимости в разделе 1.2.

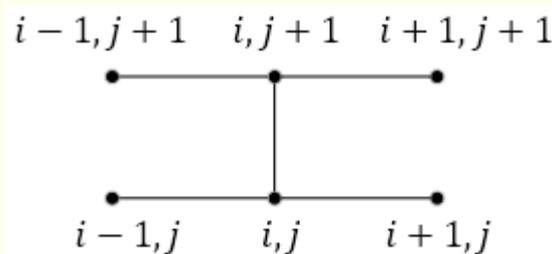


Рисунок 5.3 — Шеститочечный шаблон неявной конечно-разностной схемы Кранка-Николсона

Аппроксимация дифференциальных уравнений исходной линейной задачи по указанной схеме имеет вид:

$$c\rho \frac{T_i^{j+1} - T_i^j}{\Delta t} - \frac{\lambda}{2} \left( \frac{T_{i-1}^{j+1} - 2T_i^{j+1} + T_{i+1}^{j+1}}{\Delta x^2} + \frac{T_{i-1}^j - 2T_i^j + T_{i+1}^j}{\Delta x^2} \right) = 0.$$

Данная аппроксимация, после приведения подобных членов, на каждом временном слое  $j + 1$  приводит к системе  $m+1$  (с учётом граничных условий) линейных уравнений, характеризуемых трёхдиагональной матрицей:

$$T_{i-1}^{j+1} - cT_i^{j+1} + T_{i+1}^{j+1} = F_{i\pm 1}^j, \quad i = 1 \dots m - 1,$$

$$T_i^{j+1} = T_{i+1}^{j+1} - q_i^{j+1} \frac{\Delta x}{\lambda}, \quad i = 0,$$

$$T_i^{j+1} = T_{i-1}^{j+1} + u_i^{j+1} \frac{\Delta x}{\lambda}, \quad i = m,$$

где коэффициент  $c = 2(1 + \frac{\Delta x^2 c\rho}{\lambda\Delta t})$ , свободный член  $F_{i\pm 1}^j = -T_{i-1}^j + 2(1 - \frac{\Delta x^2 c\rho}{\lambda\Delta t})T_i^j - T_{i+1}^j$ . При  $j = 0$  мы имеем начальное условие  $T_i^0 = T_0 \forall i$ .

В результате прямой прогонки на каждом новом временном слое находились прогоночные коэффициенты:

$$\mu_{i+1} = \frac{1}{c - \mu_i}, \quad v_{i+1} = \frac{v_i - F_{i\pm 1}^j}{c - \mu_i}, \quad i = 0 \dots m - 1,$$

$$\mu_0 = 1, \quad v_0 = -q_0^{j+1} \frac{\Delta x}{\lambda}.$$

В обратной прогонке для  $j = 0 \dots n - 1$  находилась температура:

$$T_i^{j+1} = \mu_{i+1} T_{i+1}^{j+1} + v_{i+1}, \quad i = m - 1 \dots 0,$$

$$T_m^{j+1} = \frac{v_m + u_m^{j+1} \frac{\Delta x}{\lambda}}{1 - \mu_m}.$$

Сопряжённая задача (5.21)–(5.24) решается в обратном по времени направлении, и её аппроксимация имеет вид:

$$-C\rho \frac{f_i^j - f_i^{j+1}}{-\Delta t} - \frac{\lambda}{2} \left( \frac{f_{i-1}^{j+1} - 2f_i^{j+1} + f_{i+1}^{j+1}}{\Delta x^2} + \frac{f_{i-1}^j - 2f_i^j + f_{i+1}^j}{\Delta x^2} \right) + \kappa \left( I_{Fv}' \Big|_{\omega_F} \right)_i^{j+1/2} = 0.$$

При  $j = n$  мы имеем начальное (терминальное) условие для сопряжённого состояния  $f_i^n = 0 \forall i$ . На каждом следующем временном слое мы получаем систему линейных уравнений для метода прогонки:

$$f_{i-1}^j - cf_i^j + f_{i+1}^j = F_{i\pm 1}^{j+1}, \quad i = 1 \dots m - 1,$$

$$f_i^j = f_{i+1}^j - 2(T - T_*)_i^j \frac{\Delta x}{\lambda}, \quad i = 0,$$

$$f_i^j = f_{i-1}^j, \quad i = m,$$

свободный член

$$F_{i\pm 1}^{j+1} = -f_{i-1}^{j+1} + 2 \left( 1 - \frac{\Delta x^2 C\rho}{\lambda \Delta t} \right) f_i^{j+1} - f_{i+1}^{j+1} + \frac{4\Delta x^2}{\lambda} \kappa (T - T_{\min})_i^{j+1/2} \Pi \left( \tau_i^{j+1/2}; \omega_F \right).$$

Оператор сужения  $\cdot|_{\omega_F}$  из области  $\Omega$  на множество  $\omega_F$  для штрафа  $I_{Fv}'$  представлен многомерной прямоугольной функцией

$$\Pi(\tau; \omega_F) = \begin{cases} 1, & \tau \in \omega_F, \\ 0, & \tau \notin \omega_F, \end{cases}$$

т.е.  $(T - T_{\min})|_{\omega_F} = (T - T_{\min})\Pi(\tau; \omega_F)$ . Напомним, множество  $\omega_F$  активного действия штрафной функции определяется условием: если  $T(\tau) < T_{\min}$ , тогда  $\tau \in \omega_F$ .

Прогоночные коэффициенты для сопряжённой задачи имеют вид:

$$\mu_{i+1} = \frac{1}{c - \mu_i}, \quad \nu_{i+1} = \frac{\nu_i - F_{i\pm 1}^{j+1}}{c - \mu_i}, \quad i = 0 \dots m - 1,$$

$$\mu_0 = 1, \quad \nu_0 = -2(T - T_*)_0^j \frac{\Delta x}{\lambda}.$$

В обратной прогонке для  $j = n - 1 \dots 0$  находилось сопряжённое состояние:

$$f_i^j = \mu_{i+1}f_{i+1}^j + \nu_{i+1}, \quad i = m - 1 \dots 0,$$

$$f_m^j = \frac{\nu_m}{1 - \mu_m}.$$

Итак, получаем градиент

$$\nabla J(u; t_j) = -f(x_m, t_j), \quad j = 1 \dots n - 1,$$

который в этом выражении определён на дискретном выражении множества  $S$  в виде временной сетки  $\{t_j\}_{j=1}^{n-1}$  на правой границе в узлах  $x_m$ . Заметим, что множество  $S$  — это открытый интервал, не включающий граничные точки, где  $j = 0, n$ .

Расчёты состояний исходной и сопряжённой задач проводились на пространственно-временной сетке с  $m = 10$ ,  $n = 100$ . Общее время решения задачи составляло 150 с, т.е.  $[t_0, t_1] = [0, 150]$  с. Шаг по

пространству  $\Delta x = 0.02$  м, шаг по времени  $\Delta t = 1.5$  с. Такая сетка удовлетворяла условию:

$$\frac{\lambda}{c\rho} \frac{\Delta t}{(\Delta x)^2} \leq 0.5.$$

В частности, при  $\Delta t = 1.5$  с значение данного выражения составляло 0.45. Приведенное условие обеспечивает монотонность решения для неявной, устойчивой схемы Кранка-Николсона.

Как будет показано далее выбранный шаг  $\Delta t$  оказывается достаточно большим и вполне удовлетворяет условию управляемости (5.26). То есть, при решении исходной задачи на отрезке времени  $[t_0, t_1]$  мы будем оптимизировать управление  $u(t)$  на более узком «вычислительном» множестве

$$S_{\Delta} = x_b \times (t_0, t_1 - \Delta t) = x_b \times [t_0 + \Delta t, t_1 - \Delta t] = x_b \times [1.5, 148.5] \text{ с.}$$

Дискретное выражение множества  $S_{\Delta} = S$  для выбранного шага  $\Delta t$ .

#### 5.4.2. Оценка выпуклости целевого функционала

Оценка выпуклости целевого функционала  $J(u)$  позволит нам получить представление о компакте  $\mathcal{U}$ , что необходимо экстремальным алгоритмам оптимизации для выбора подходящего начального приближения  $u^0 \in \mathcal{U}$  и контроля дальнейших приближений  $u^k \in \mathcal{U}$ ,  $k = 1, 2, \dots$ . Для оценки выпуклости давайте посмотрим, насколько  $J(u)$  является квадратичным. Такую оценку будем выполнять графически. Строгое доказательство выпуклости, квадратичности требует аналитических исследований, которые для распределённых систем могут оказаться непреодолимой задачей.

Построим следующий вычислительный эксперимент. Будем задавать множество постоянных по времени управлений. При этом функционал  $J(u)$  станет одномерной функцией. Если  $J(u)$  —

квадратичный, то для любого множества из трёх значений управления (точек на оси постоянных управлений) можно будет построить множество парабол  $J(u)$  и все они должны будут совпасть друг с другом.

Зададим оптимальное управление  $u_* = 300$  кДж/(м<sup>2</sup>с) и два множества управлений  $\{u\}_1 = \{150, 200, 250\}$  и  $\{u\}_2 = \{350, 400, 450\}$ . Применим квадратичное интерполирование значений функции  $J_1(u)$  на множестве  $\{u\}_1$  и  $J_2(u)$  на множестве  $\{u\}_2$ . На рисунке 5.4 показаны полученные интерполяционные параболы. Обе параболы совпали, поэтому целевой функционал  $J(u)$  будем считать *квадратичным*.

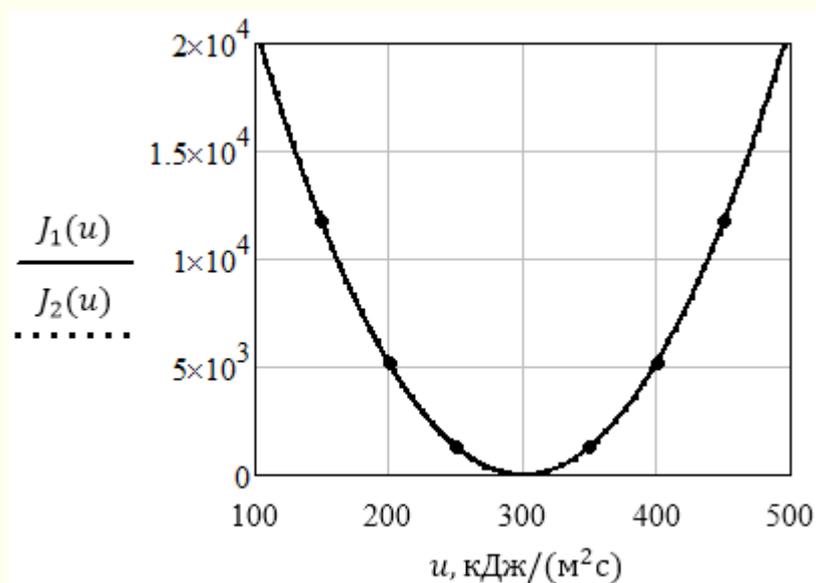


Рисунок 5.4 — Оценка квадратичности целевого функционала

Из рисунка следует, во-первых, *компактом корректности*  $\mathcal{U}$ , достаточным для решения поставленной задачи управления, можно считать, как минимум, множество

$$\mathcal{U} = [100, 500] \text{ кДж/(м}^2\text{с)}$$

с не слишком сильным колебанием функций управления в нём. В пределах этого множества может корректно определяться градиент  $\nabla J$ , указывающий направление спуска к оптимуму.

Во-вторых, квадратичность функционала говорит о возможности в дальнейшем применять МРНС с параметром-функцией  $\alpha(t) = \alpha^0(t)$ , определяемым на начальной итерации и неизменным на всех последующих итерациях.

### 5.4.3. Оптимизация без ограничений

Тестовые расчёты для задачи оптимизации (5.1)–(5.5) организовывались следующим образом. Задавалось управление

$$u_*(t) = 150 + 50 \sin \frac{2\pi t}{t_1 - t_0} \frac{\text{кДж}}{\text{м}^2\text{с}} \quad \text{на } S \quad (5.27)$$

и рассчитывалась температура на внутренней поверхности стенки реактор  $x_a$ . Данная температура принималась оптимальной температурой  $T_*$  и решалась обратная задача о восстановлении управления  $u(t)$ , доставляющего минимум целевому функционалу (5.5) с полученной температурой  $T_*$ .

Все тесты заканчивались при практическом прекращении сходимости по критерию (1.24):

$$\frac{\|u^k - u^{k-1}\|_{L_2(S_\Delta)}}{\|u^{k-1}\|_{L_2(S_\Delta)}} \leq \varepsilon_4, \quad \varepsilon_4 = 10^{-6}.$$

Глубина шагов  $b^k$  задавалась линейным поиском согласно *четвёртой стратегии* по методу Носедала-Райта с начальной глубиной шагов равной  $\|u^0 - u_*\|$ .

Зададим начальное приближение управления:

$$u^0(t) = 400 \text{ кДж}/(\text{м}^2\text{с}).$$

Для него невязка управлений  $\|u^0 - u_*\|_{L_2(S_\Delta)} = 3.05 \cdot 10^6$ , целевой функционал  $J^0 = 4.1 \cdot 10^4$ , норма градиента  $\|\nabla J^0\|_{L_2(S_\Delta)} = 3.6 \cdot 10^{-2}$ .

В первой строке (после заголовков) таблицы 5.1 приведено начальное значение функции градиента  $\nabla J(u^0; t)$ , рассчитанное на  $S_\Delta = x_b \times [1.5, 148.5]$  для некоторых моментов времени. Из таблицы видно, что  $\nabla J^0$  убывает на 6 порядков от начального момента времени  $t_0 + \Delta t = 1.5$  с (при  $i = 1$ ) и до последнего момента  $t_1 - \Delta t = 148.5$  с (при  $i = 99$ ). Это означает, что функционал  $J$  слабо чувствителен к управлению в последние моменты времени, как и было предсказано в анализе управляемости. Однако последнее значение  $\nabla J(u^0; t_1 - \Delta t) = 4 \cdot 10^{-9}$  не выходит за пределы вычислительной точности<sup>1</sup> и отражает реальный вклад возмущения  $I'_v = 2(T - T_*)$  в градиент. То есть заданное время  $[1.5, 148.5]$  удовлетворяет условию управляемости (5.26).

Таблица 5.1 — Значение функции градиента на начальной итерации и на конечной итерации для МНС

$t, c$	1.5	30	60	90	135	148.5
$i$	1	20	40	60	90	99
$\nabla J(u^0; t) \cdot 10^{-3}$	4.9	4.3	3.3	2.9	0.029	$4 \cdot 10^{-6}$
$\nabla J(u^{78}; t) \cdot 10^{-7}$	0.11	-9.1	-1.5	-4.5	1.8	$3.9 \cdot 10^{-4}$

Перейдём к минимизации целевого функционала традиционными экстремальными алгоритмами. Начнём с градиентного метода (1.8) наискорейшего спуска (МНС), формально обобщённого на бесконечномерное пространство при  $S_\Delta$ :

<sup>1</sup> Вычисления производились в среде Delphi для переменных Extended с мантисой в 20 цифр.

$$u^{k+1}(t) = u^k(t) - b^k \nabla J(u^k; t) \quad \text{на } S_{\Delta}.$$

На рисунке 5.5 показаны результаты минимизации. Наилучшие результаты МНС показал с шаговым множителем  $b^k$  по методу Носедала-Райта при параметре Вольфе  $c_2=0.5$ .

Как мы видим, оптимизация закончилась на итерации  $k = 78$  (пунктирная кривая) с плохой неравномерной даже на  $S_{\Delta}$  сходимостью. При этом на последней итерации МНС:

$$\|u^{78} - u_*\|_{L_2(S_{\Delta})} = 1.21 \cdot 10^6, \quad J^{78} = 0.9, \quad \|\nabla J^{78}\|_{L_2(S_{\Delta})} = 5.8 \cdot 10^{-6}.$$

То есть, невязка управлений уменьшилась всего лишь в 2.5 раза, а функционал — в 46 тысяч раз, градиент — в 6 тысяч раз. Здесь мы столкнулись с ярким примером ситуации, обсуждаемой на рисунке 1.4 и в формуле (1.25), о косвенных критериях близости к оптимуму. Значительное уменьшение целевой функции, её градиента ничего не говорят о близости управления к оптимальному значению.

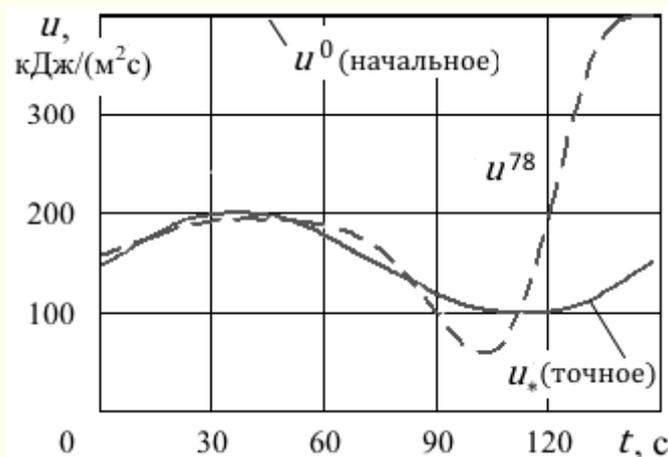


Рисунок 5.5 — Оптимизация посредством МНС

В последней строке таблицы 5.1 приведены значения функции градиента на последней итерации МНС. Как мы видим, значение

$\nabla J(u^{78}; t_i)$  в начальные моменты времени уменьшилось на 4—5 порядков, а в последние моменты времени — только на 2 порядка. Такое изменение градиента демонстрирует неравномерную на  $S_\Delta$  сходимость  $\nabla J(u^k; t) \rightarrow 0$  при  $k \rightarrow \infty$ . То есть, и управление, и градиент сходятся к своим предельным значениям неравномерно на  $S_\Delta$ .

Следующим реализовывался метод (3.10) сопряжённых градиентов (МСГ), формально обобщённый на бесконечномерное пространство:

$$u^{k+1}(t) = u^k(t) + b^k p(u^k; t) \quad \text{на } S_\Delta,$$

$$p^k = -\nabla J^k + \beta^k p^{k-1}, \quad \beta^k = \frac{\|\nabla J^k\|_{L_2(S_\Delta)}^2}{\|\nabla J^{k-1}\|_{L_2(S_\Delta)}^2}, \quad \beta^0 = 0,$$

где  $b^k$  вычислялся с параметром Вольфе  $c_2=0.1$ . Очистка сопряжённого направления ( $\beta^k=0$ ) устанавливалась на каждой 15-ой итераций.

По сути МСГ не улучшил ситуацию, что иллюстрирует пунктирная кривая на рисунке 5.6. Сходимость метода закончилась на итерации  $k=44$ . При этом на последней итерации СМГ:

$$\|u^{44} - u_*\|_{L_2(S_\Delta)} = 1.20 \cdot 10^6, \quad J^{44} = 0.5, \quad \|\nabla J^{44}\|_{L_2(S_\Delta)} = 4.1 \cdot 10^{-6}.$$

То есть, достигнутые значения невязки управлений, целевого функционала и нормы градиента практически такие же, как у МНС, но были получены за меньшее количество итераций. Мы видим, что решение МСГ не сильно отличается от неудачного решения предыдущим МНС.

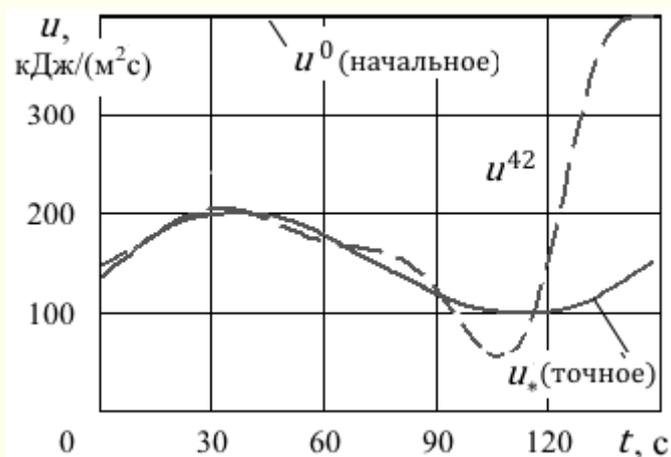


Рисунок 5.6 — Оптимизация посредством МСГ

Иногда встречаются утверждения о равноценности применения бесконечномерных и конечномерных методов оптимизации, если последние применить к дискретному представлению бесконечномерной задачи в узлах конечноразностной сетки решения дифференциальных уравнений. Для проверки данного утверждения мы построили следующий тест.

Исходная бесконечномерная задача оптимизации была преобразована аналогично [8] к конечномерной задаче посредством разложения функции  $u(t) = \sum_{i=1}^{n-2} u_i B_i(t)$  на  $S_\Delta$  через  $B$ -сплайны нулевого порядка (кусочно-постоянные функции) с носителями, равными шагу  $\Delta t$ . Для выбранной конечно-разностной сетки с  $n=100$ , вектор управления  $u \in E^{98}$ . Это управление отыскивалось квазиньютоновским методом с экономной памятью L-BFGS, где обратная матрица Гессе представлялась пятью парами векторов. Наилучшая минимизация была достигнута с шагом  $b^k$  при  $c_2=0.5$ . Результаты минимизации представлены на рисунке 5.7.

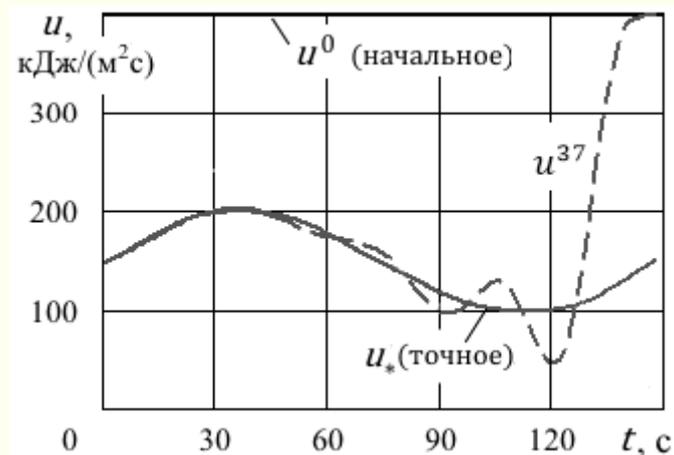


Рисунок 5.7 — Оптимизация методом L-BFGS

Пунктирная кривая — это решение, полученное методом L-BFGS на последней итерации  $k = 37$ . Мы видим, что попытка решения бесконечномерной задачи оптимизации традиционными конечномерными методами не привела к успеху. Равномерной сходимости на  $S_\Delta$  как не было, так и не появилось. При этом на последней итерации L-BFGS:

$$\|u^{37} - u_*\|_{L_2(S_\Delta)} = 0.85 \cdot 10^6, \quad J^{37} = 0.039, \quad \|\nabla J^{37}\|_{L_2(S_\Delta)} = 0.75 \cdot 10^{-6}.$$

Целевая функция и градиент, по сравнению с предыдущими методами, стали заметно меньше. Однако, это не существенно улучшило решение, что демонстрирует рисунок 5.7.

Во всех рассмотренных примерах реализации МНС, МСГ, L-BFGS хорошо выполняется классическое НУО (1.11). Нормы градиента в  $L_2(S_\Delta)$  стремятся к нулю (уменьшаются на 4—5 порядков), однако, управление при этом плохо сходится к оптимальному значению — неравномерно. Полученное интегральное стремление  $\|\nabla J^k\|_{L_2(S_\Delta)} \rightarrow 0$  не сопровождается равномерным стремлением, т.е. здесь  $|\nabla J(u^k; t)| \not\rightarrow 0$  на  $S_\Delta$ , как видно из всех предыдущих рисунков.

Таким образом, при решении простейшей задачи оптимального управления прямым экстремальным подходом с традиционными методами оптимизации как в бесконечномерном пространстве, так и в конечномерном представлении, получить оптимальное решение с поточечной близостью к  $u_*(t)$  на  $S_\Delta$  не удаётся. Традиционные методы не смогли реализовать равномерную сходимость как в пространстве управлений, так и в сопряжённом пространстве градиентов.

Перейдём к реализации алгоритмов МРНС. Начнём с МРНСг (3.4) в виде:

$$u^{k+1}(t) = u^k(t) - b^k \alpha(t) \nabla J(u^k; t) \quad \text{равномерно на } S_\Delta.$$

Все шаги, начиная с первого, будем реализовывать с шаговым множителем по методу Носедала-Райта, в котором примем параметр Вольфе  $c_2 = 0.5$ .

Определим для МРНСг параметр регулирования направления спуска  $\alpha(t)$ . Выбранное начальное приближение управления  $u^0 = 400$  кДж/(м<sup>2</sup>с), в виде постоянной функции, делает одинаковыми рассмотренные ранее *шаблоны*  $\tilde{u}^0$ : изменение управления (3.7) «под 45°»; пропорциональное изменение (3.8). Зададим, например,  $\delta = 0.2u^0$  или  $\lambda = 0.2$  и получим

$$\alpha(t) = \left| \frac{0.2u^0}{\nabla J(u^0; t)} \right|. \quad (5.28)$$

Отметим, что для принятого начального управления  $u^0$  градиент  $\nabla J(u^0; t) \neq 0$  на  $S_\Delta$ , что необходимо для реализации МРНСг (3.4).

Оценим поведение функции градиента  $\nabla J(u^k; t)$  после первых  $k = 50$  итераций МРНСг по данным полученной таблицы 5.2. Сравним эту таблицу с аналогичной таблицей 5.1 для МНС. Мы видим, что при МРНСг, в отличие от МНС, значение функции градиента уменьшилось на 3–4 порядка как в начальные, так и в конечные моменты времени.

Таким образом при МРНСГ с параметром  $\alpha$  из (5.28) мы наблюдаем в сопряжённом пространстве сходимость функции градиента  $\nabla J(u^k; t) \rightarrow 0$  на  $S_\Delta$  равномерно!

Таблица 5.2 — Значение градиента на начальной и 50-й итерации для МРНСГ

$t, c$	1.5	30	60	90	135	148.5
$i$	1	20	40	60	90	99
$\nabla J(u^0; t) \cdot 10^{-3}$	4.9	4.3	3.3	2.9	0.029	$4 \cdot 10^{-6}$
$\nabla J(u^{50}; t) \cdot 10^{-7}$	10	-7.7	-3.6	-2.2	0.28	$5.2 \cdot 10^{-5}$

На рисунке 5.8 показан ход процесса оптимизации для итераций  $k = 5$  (пунктирная кривая),  $k = 50$  (точечная кривая) и на конечной итерации  $k = 376$  (штрихпунктирный кончик справа). Видно, что уже и на 50-й итерации управление очень близко к оптимальному значению. Характер сходимости очевиден — это равномерная сходимость в пространстве управлений!

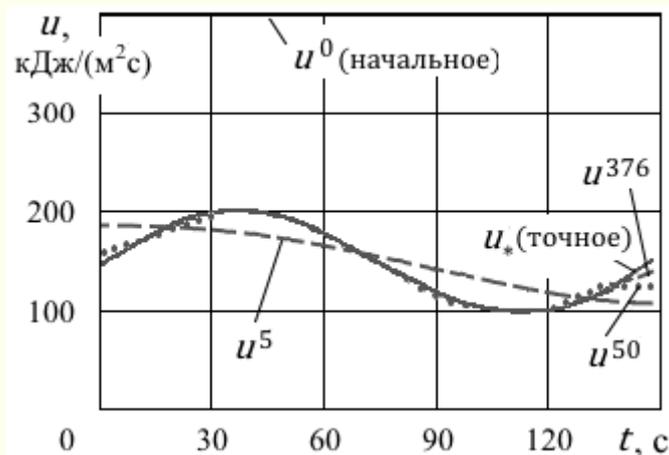


Рисунок 5.8 — Оптимизация посредством МРНСг

На последней итерации мы получили:

$$\begin{aligned} \|u^{376} - u_*\|_{L_2(S_\Delta)} &= 2.3 \cdot 10^4, \quad J^{376} = 1.4 \cdot 10^{-5}, \quad \|\nabla J^{376}\|_{L_2(S_\Delta)} \\ &= 8.4 \cdot 10^{-9}. \end{aligned}$$

По сравнению с традиционными методами невязка управлений уменьшилась на 2 порядка сильнее, значение целевой функции — на 4 порядка лучше, значение нормы градиента — на 3 порядка лучше. Достигнута очень высокая точность оптимизации.

Перейдём к реализации МРНСг (3.11):

$$u^{k+1}(t) = u^k(t) - b^k \alpha(t) p(u^k; t) \quad \text{равномерно на } S_\Delta.$$

Значение параметра  $\alpha$  возьмём, как и ранее, из выражения (5.27) и параметр Вольфе оставим  $c_2 = 0.5$ .

Оптимизация МРНСг закончилась на шаге  $k = 101$ . Полученное оптимальное управление  $u^{101}$  совпало с оптимальным управлением  $u^{376}$  МРНСг на рисунке 5.8. То есть, тот же результат равномерной сходимости был достигнут почти в четыре раза быстрее.

Все МРНС продемонстрировали выполнение НУО (3.2) сильной сходимости (смотрите теорему 3.2). При этом *равномерная сходимость*

градиентов в сопряжённом пространстве  $U^* = L_2(S_\Delta)$  сопровождалась и равномерной сходимостью в пространстве управлений  $U = L_2(S_\Delta)$ , что позволило решить поставленную задачу с высокой точностью.

Далее мы исследуем пять ранее поставленных задач ограниченной оптимизации, где оптимальные решения будут доставлять *краевой минимум* целевому функционалу.

#### 5.4.4. Оптимизация в заданном диапазоне управлений

Перейдём к реализации возможных ограничений, когда  $U \subset L_2$ . Все параметры будем задавать так, чтобы ограничения были активными, т.е. когда при движении из точки  $u^0 \in \text{int } U$  мы обязательно попадём на границу  $\partial U$ . Здесь нам придётся использовать только МРНСГ, поскольку при наличии ограничений сопряжённость градиентов может нарушаться и применение МРНСГ не обосновано.

Ограничение (5.6) типа 2.1, в виде полуограниченного диапазона, будем реализовывать посредством проецирования  $P_U(u^{k+1})$  очередного спуска МРНСГ на допустимое множество  $U = \{u(t) \in L_2(S_\Delta) \mid u(t) \geq u_{\min}\}$ . В нашем случае алгоритм проецирования (3.45) принимает вид:

$$\text{Если } u^{k+1}(t) < u_{\min}, \text{ тогда } u^{k+1}(t) \leftarrow u_{\min}. \quad (5.29)$$

Чтобы указанное ограничение было активным, зададим значение нижней границы  $u_{\min} = 150$  кДж/(м<sup>2</sup>с). Обратим внимание на то, что используемое начальное приближение  $u^0 = 400$  кДж/(м<sup>2</sup>с) находится внутри допустимого множества, т.е.  $u^0 \in \text{int } U$ .

На рисунке 5.9 представлены результаты оптимизации. Параметр  $\alpha$  задавался, как и ранее, по формуле (5.28). Шаговый множитель  $b^k$  вычислялся по *второй стратегии* адаптивной релаксации (1.14) с коэффициентами  $b_1 = 1.005$ ,  $b_2 = 0.8$ . Для МРНСГ, с известным первым шаблонным шагом, всегда будем задавать  $b^0 = 1$ . Коэффициент

усиления сходимости  $b_1$  можно было бы увеличить, но как показали расчёты, такое ускорение сходимости сопровождается ухудшением точности решения.

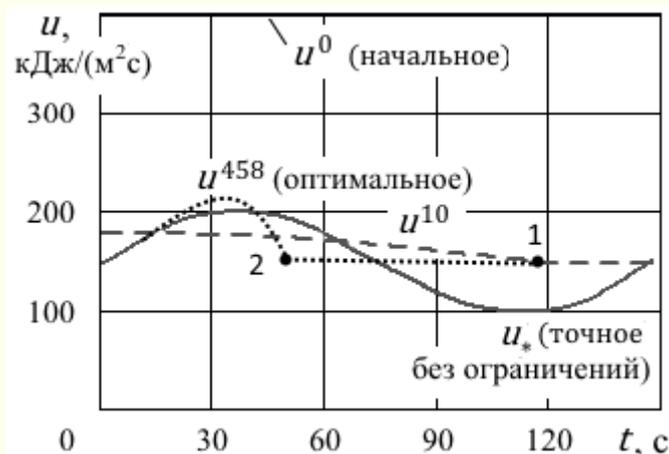


Рисунок 5.9 — Оптимальное управление тепловым потоком посредством МРНСг с ограничением на управление типа 2.1

При указанных условиях алгоритм минимизации завершился на итерации  $k = 458$ . Здесь оптимальное управление — точечная кривая доставляет целевому функционалу (5.5) *краевой минимум*. На итерации  $k=10$  функция управления  $u^k(t)$  в конце времени управления выходит за нижнюю границу допустимого множества  $U$ , где срабатывает проецирование (5.29) и возвращает вышедшее управление на границу, что демонстрирует пунктирная кривая. Начиная с 10-й итерации управление лежит на границе. Дальнейший спуск к минимуму целевого функционала (5.5) представляет собой скольжение вдоль нижней границы допустимого множества  $U$ , о чем свидетельствует смещение характерного излома управления  $u(t)$ , перемещаемого из точки 1 в точку 2.

Целевой функционал с начального значения  $J^0 = 4.1 \cdot 10^4$  уменьшился до  $J^{458} = 8.8$ , что, естественно, хуже, чем при оптимизации без ограничений, где на последней итерации  $J^{376} = 1.4 \cdot 10^{-5}$ . Тем не менее, полученные результаты наглядно иллюстрируют «разумный» процесс оптимизации с проецированием на явно заданное множество допустимых управлений.

#### 5.4.5. Оптимизация с ограничением на состояние

Теперь реализуем неявное ограничение на управление (5.7), связанное с состоянием (температурой) системы в виде требования не остывания реактора. Ограничения такого типа обычно реализуются штрафными функционалами, что мы и сделали в целевом функционале (5.11).

Прежде чем перейти к расчётам, давайте посмотрим на поведение температуры без ограничений при начальном управлении  $u^0$ . На рисунке 5.10 показана температура ( $k = 0$ ) в стенке реактора в начальный момент времени  $T^0(x, t_0)$  и конечный момент  $T^0(x, t_1)$ , а также при оптимальном управлении в конечный момент времени  $T^{376}(x, t_1)$ , найденном ранее посредством МРНСГ (рисунок 5.8). Точечная линия — это запланированное ограничение на управление в виде минимально допустимой температуры  $T_{\min} = 280$  К.

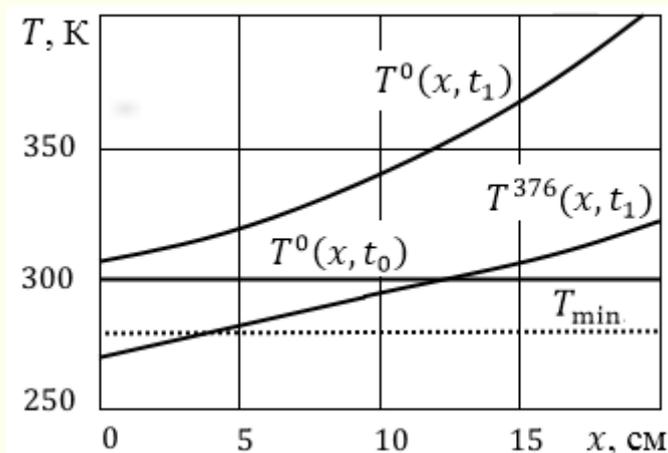


Рисунок 5.10 — Температура в стенке реактора при управлении без ограничений

Мы видим, что начальное приближение управления  $u^0$  выбрано удачно, и оно находится вне зоны активного действия ограничения на допустимую температуру. Здесь  $T^0(x, t) > T_{\min}$  для всех  $(x, t)$ . В тоже время оптимальная температура  $T^{376}(x, t_1)$  нарушает возможное ограничение с внутренней стороны реактора вблизи химической реакции.

Теперь включим ограничение на температуру. Оно реализовывалось с весовым коэффициентом  $\kappa = 10^3$ . Для минимизации целевого функционала со штрафом (5.11), как и ранее, применялся МРНСГ с параметром  $\alpha$  (5.28). Шаговый множитель  $b^k$  рассчитывался адаптивно по *второй стратегии* методом (1.14) с  $b_1 = 1.005$ ,  $b_2 = 0.8$ . Сходимость МРНСГ завершилась на итерации  $k = 15$ .

На рисунке 5.11 показана полученная температура в конце управления при  $t_1$ , а на рисунке 5.12 — оптимальное управление  $u^{15}$ . Исходный целевой функционал (5.5) с начального значения  $J^0 = 4.1 \cdot$

$10^4$  уменьшился до  $J^{15} = 4.1 \cdot 10^2$ , что, естественно, хуже чем при оптимизации без ограничений, где было получено  $J^{376} = 1.4 \cdot 10^{-5}$ .

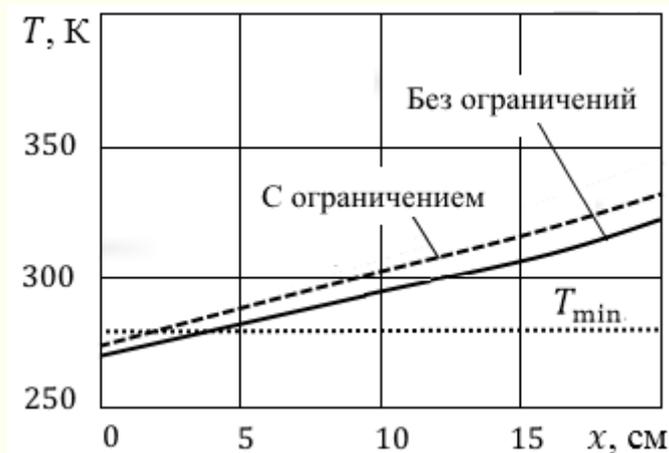


Рисунок 5.11 — Конечная температура в стенке реактора

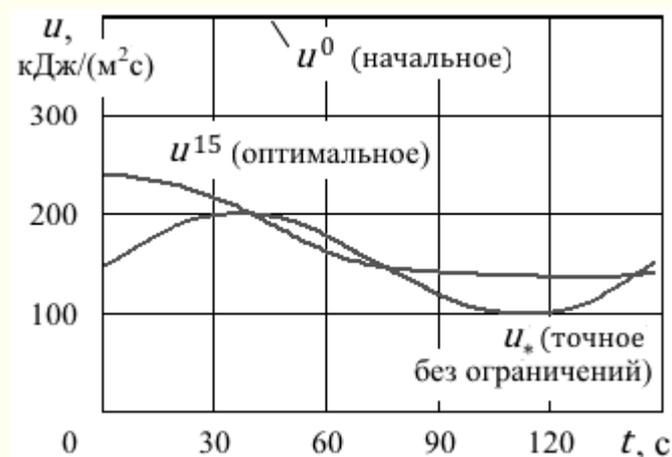


Рисунок 5.12 — Оптимальное управление тепловым потоком посредством МРНСГ с ограничением на температуру типа 1.1

Мы видим, для того чтобы температура в стенке реактора вблизи реакции не опускалась ниже ограничения (допускается незначительное нарушение ограничения, характерное для внешних штрафов), нам

необходимо в начальные моменты времени задавать повышенный поток тепла  $u$  на внешней стенке реактора. Полученное оптимальное управление  $u^{15}$  уже в меньшей степени нарушает ограничение на температуру чем без штрафных ограничений. Для точного предотвращения нарушения ограничений на температуру необходимо использовать не внешние, а внутренние штрафы.

Результаты демонстрируют удовлетворительный учёт ограничений типа 1.1 при использовании МРНСг со штрафными функционалами. Определённых усилий может потребовать выбор подходящей штрафной функции вместе со своим весовым коэффициентом и выбор удовлетворительного начального приближения управления. Кроме этого, могут потребоваться относительно медленные аккуратные приближения управлений к области активного действия ограничения, что регулируется коэффициентом  $b_1$  при адаптивном выборе длины шага, а также аккуратный выбор  $\delta$  или  $\lambda$  на начальном шаге шаблонного приближения. Последние рекомендации особенно актуальны при использовании внутренних штрафов, когда выход за ограничение принципиально недопустим.

#### 5.4.6. Оптимизация с изопериметрическим ограничением

Изопериметрическое ограничение (5.8) реализуется градиентным методом (3.47), который здесь принимает вид:

$$u^{k+1} = u^k - b^k \nabla J^k - \frac{C^{k+1} - C}{\text{mes } S_\Delta} \quad \text{на } S_\Delta, \quad k = 0, 1, \dots, \quad (5.30)$$

где  $\text{mes } S_\Delta = t_1 - 2\Delta t$ ,  $C^{k+1} = \int_{t_0+\Delta t}^{t_1-\Delta t} u^{k+1} dt$ . Значение изопериметрической константы  $C$  для тестовых расчётов мы определим через оптимальное значение управления (5.27):

$$C = \int_{t_0+\Delta t}^{t_1-\Delta t} u_* dt.$$

Все интегралы, на всех итерациях, брались численно. Шаговый множитель  $b^k$ , как и в предыдущих задачах с ограничениями, рассчитывался адаптивно по *второй стратегии* методом (1.14) с  $b_1=1.1$ ,  $b_2=0.8$ .

Зададим начальное приближение в виде линейной функции, удовлетворяющей изопериметрическому условию:

$$u^0(t) = 50 + \frac{200(t - t_0)}{t_1 - t_0} \text{ кДж}/(\text{м}^2\text{с}).$$

То есть функция  $u^0(t)$  будет линейно меняться от 50 кДж/(м<sup>2</sup>с) до 250 кДж/(м<sup>2</sup>с). Можно было бы поступить проще и задать начальное приближение постоянной величиной  $u^0(t) = 150$  кДж/(м<sup>2</sup>с) посередине синусоиды  $u_*(t)$ . Оба варианта  $u^0$  удовлетворяют изопериметрическому условию (5.8). Но мы остановимся на первом, чтобы посмотреть на характер сходимости к оптимуму в конечные моменты времени возле  $t_1$ , где ранее отсутствовала сходимость при оптимизации МНС и МСГ, без регулирования направления спуска и с оптимальным  $b^k$ . Напомним, что при изопериметрическом ограничении мы не можем регулировать направления спуска, а также выбирать  $b^k$  из условия полной релаксации.

На рисунке 5.13 представлены результаты оптимизации традиционным бесконечномерным градиентным методом с адаптивными шагами. Алгоритм минимизации завершился на итерации  $k=76$ . Найденное оптимальное управление  $u^{76}$  практически полностью совпало с точным решением  $u_*$ . Начальное значение целевого функционала  $J^0 = 3.4 \cdot 10^3$  уменьшилось существенно, до  $J^{76} = 5.5 \cdot 10^{-2}$ .

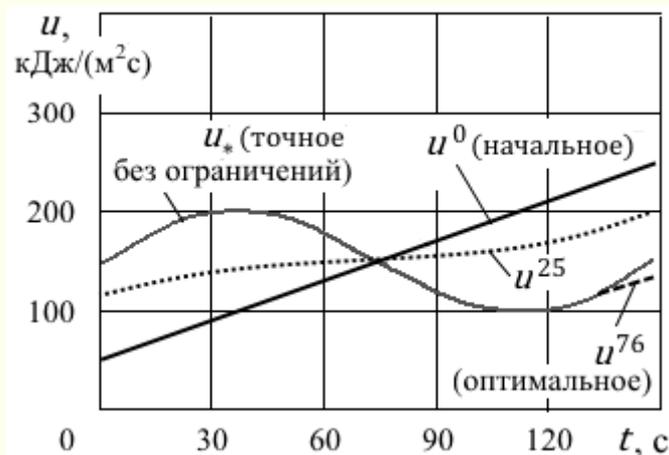


Рисунок 5.13 — Оптимальное управление тепловым потоком с изопериметрическим условием

Несмотря на то, что отсутствует регулирование направления спуска, мы всё-таки наблюдаем хорошую сходимость во все моменты времени управления, включая «неблагоприятную» область конца управления (сравните с рисунками 5.5–5.7). Это объясняется тем, что область изначально плохой сходимости заметно влияет на значение изопериметрической переменной  $C^{k+1} \neq C$ , что приводит к смещению всей функции управления  $u^{k+1}$  в алгоритме (5.30), в том числе и к смещению «неблагоприятной» правой области функции управления в сторону точного решения.

#### 5.4.7. Оптимизация с кусочно-постоянным управлением

Рассмотрим реализацию ограничения (5.9) в виде управления в классе кусочно-постоянных функций. Такое ограничение реализуется МРНС в конечномерной форме (3.49):

$$u_i^{k+1} = u_i^k + b^k d_i(u^k), \quad i = 1 \dots \hat{n}, \quad k = 0, 1 \dots,$$

где направление спуска для МРНСг и МРНСсг имеет вид:

$$d_i(u^k) = - \left\{ \int_{\hat{\Delta}t_i} \alpha(t) \nabla J(t; u^k) dt \right\}_{i=1}^{\hat{n}} \in U^* = E^{\hat{n}},$$

$$d_i(u^k) = \left\{ \int_{\hat{\Delta}t_i} \alpha(t) p(t; u^k) dt \right\}_{i=1}^{\hat{n}} \in U^* = E^{\hat{n}}.$$

Зададим шаг регулировок  $\hat{\Delta}t = 30$  с. При этом мы получим пять одинаковых областей регулирования плотности потока подачи тепла в реактор. Целевой функционал (5.5) становится 5-мерной *целевой функцией* вектора управления  $u = (u_1, \dots, u_5) \in E^5$ , и ограничения на управление исчезают. При этом мы можем использовать линейный поиск для выбора  $b^k$ , например, в *четвёртой стратегии* методом Носедала-Райта.

Зададим то же самое начальное управление  $u^0 = 400$  кДж/(м<sup>2</sup>с).

На рисунке 5.14 показаны результаты кусочно-постоянного управления. Как и ранее, параметр регулировки  $\alpha$  задавался по формуле (5.28), а в методе Носедала-Райта  $c_2 = 0.5$ .

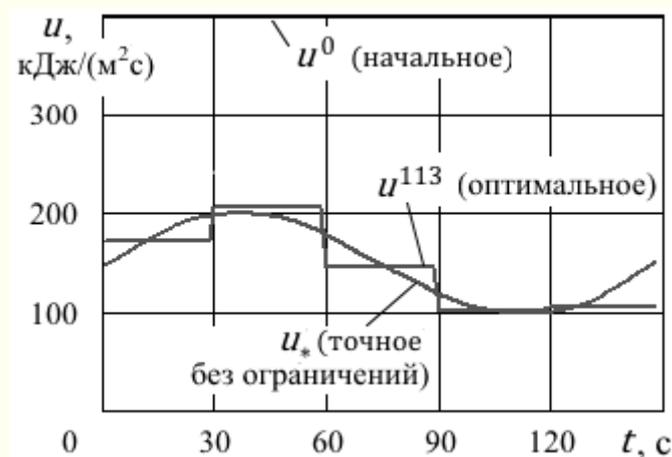


Рисунок 5.14 — Оптимальное управление посредством МРНСг при кусочно-постоянном тепловом потоке

Алгоритм минимизации завершился на итерации  $k = 113$ . Найденное управление  $u^{113}$  уменьшило целевой функционал с начального значения  $J^0 = 4.1 \cdot 10^4$  до конечного  $J^{113} = 4.8 \cdot 10^{-2}$ . Очевидно, что управление в классе кусочно-постоянных функций снижает качество управления по сравнению с произвольными функциями, где на последней итерации МРНСг значение целевого функционала  $J^{376} = 1.4 \cdot 10^{-5}$ .

Заметим, что если не регулировать спуск параметром  $\alpha$ , то на последних двух отрезках функции управления сходимость оказывалась существенно хуже. Так, например, значение управления на последнем отрезке без регулировки спуска составило  $350 \text{ кДж}/(\text{м}^2\text{с})$ , сходимость закончилась на итерации  $k = 63$ , а целевой функционал принимал значение  $J^{63} = 1.6$ , т.е. на два порядка хуже.

Реализация МРНСг привела к тем же визуальным результатам, что и на рисунке 5.14, но не за 113, а за  $k = 75$  итераций. При этом метод Носедала-Райта применялся с параметром  $c_2 = 0.1$ . Если не регулировать спуск параметром  $\alpha$ , то на последних двух отрезках функции управления сходимость по-прежнему отсутствовала.

Мы видим, что поиск оптимальных управлений типа 2.2 в виде кусочно-постоянных функций, даже при переходе к конечномерному управлению, для достижения положительных результатов на всей области управления  $S_\Delta$ , потребовал применения МРНС.

#### **5.4.8. Оптимизация с ограничением на скорость изменения управления**

Будем полагать, что химический реактор ранее работал в номинальном режиме с оптимальным управлением  $u_*$  (5.27). Но возникла нештатная ситуация и механизмы подачи тепла перешли в

режим (5.10) пониженной интенсивности, со скоростью ввода тепла  $\frac{du(t)}{dt} \leq \vartheta_{\max} = 0.3 \text{ кДж}/(\text{м} \cdot \text{с})^2$ .

Реализуем возникшее ограничение, используя проецирование  $P_U(u^{k+1})$  на множество  $U$  с подвижной границей максимально допустимой плотности потока тепла. Необходимый метод проецирования (3.51) в рассматриваемой задаче принимает вид:

$$\text{Если } u^{k+1}(t_{j+1}) > u_{\max}^{j+1}, \quad \text{тогда } u^{k+1}(t_{j+1}) \leftarrow u_{\max}^{j+1}, \quad (5.32)$$

где  $t_{j+1} = t_j + \Delta t$ ,  $u_{\max}^{j+1} = u(t_j) + \vartheta_{\max} \Delta t$ ,  $j = 1 \dots n - 2$  — временные слои конечно-разностной сетки в  $S_{\Delta}$ .

Будем искать оптимальное управление посредством МРНСг с параметром  $\alpha$ , который вычислялся ранее по формуле (5.28). Шаговый множитель  $b^k$  будем рассчитывать по *второй стратегии* методом (1.14) с коэффициентами  $b_1=1.005$ ,  $b_2=0.8$ .

На рисунке 5.15 пунктиром показано найденное оптимальное управление, которое было получено на итерации  $k=376$ . Мы видим, что в начальные (приблизительно 60 с) и конечные (приблизительно 40 с) промежутки времени управления срабатывало ограничение на скорость ввода тепла. За всё время оптимального управления  $u^{376}$  скорость плотности потока тепла  $du(t)/dt$  удерживалась в рамках заданной границы  $\vartheta_{\max}$ , что полностью реализовало нештатный режим работы регуляторов подачи тепла.

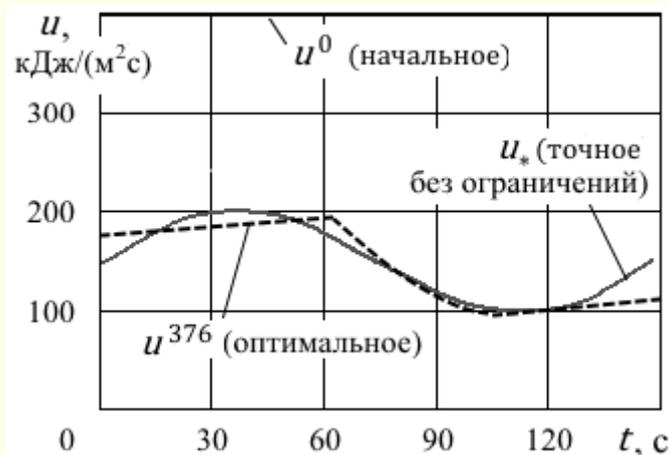


Рисунок 5.15 — Оптимальное управление потоком тепла посредством МРНСГ при ограничении на скорость ввода тепла

Целевой функционал (5.5) уменьшился от  $J^0 = 4.1 \cdot 10^4$  до  $J^{376} = 0.36$ , что, естественно, хуже, чем при управлении без ограничений, где целевой функционал уменьшался до  $1.4 \cdot 10^{-5}$ .

На этом мы закончили исследование типичных ограничений, приводящих к краевым минимумам исходного целевого функционала  $J$ .

#### 5.4.9. Оптимизация с помехами

Рассмотрим оптимальное управление плотностью теплового потока при наличии помех (1.18)–(1.20). В частности, будем искусственно вводить помехи в характерные места задачи, а именно, в температуру  $T_*$  целевого функционала  $J$ , в градиент  $\nabla J$  и наблюдать характер сходимости к оптимальному управлению  $u_*$ .

Начнём со случайных помех, которые могут появляться в температуре  $T_*$  в том случае, если температура будет задаваться не заранее, а из экспериментальных наблюдений за состоянием объекта, т.е. когда вместо  $T_*$  мы будем иметь  $T_e$ , что характерно для задач идентификации.

Рассмотрим *абсолютные случайные помехи* (1.20) в виде:

$$T_e(t) = T_*(t) + 2\Delta T(\xi(t) - 0.5), \quad t \in S,$$

где  $\xi(t) \in [0,1)$  — функция вещественных случайных чисел,  $\Delta T$  — максимально возможная погрешность температуры. Далее в «экспериментальной» температуре  $T_e$  примем  $\Delta T = 1$  К, что типично для термодатчиков.

Поскольку наличие случайных помех может приводить к осцилляциям функций управлений  $u^k(t)$  относительно точного значения  $u_*(t)$ , то для завершения итерационного процесса будем использовать не только критерий (1.24) интегральной малости изменений управления, но и критерий (1.25) малости изменения целевого функционала:

$$\frac{\|u^k - u^{k-1}\|_{L_2(S)}}{\|u^{k-1}\|_{L_2(S)}} \leq \varepsilon_4, \quad \frac{|J(u^k) - J(u^{k-1})|}{J(u^{k-1})} \leq \varepsilon_5.$$

Примем параметры малости  $\varepsilon_4 = \varepsilon_5 = 10^{-6}$ .

Полученные количественные результаты минимизации функционала (5.5) в виде критерия

$$J(u) = \int_{t_0}^{t_1} (T|_{x_a} - T_e)^2 dt$$

оказались практически одинаковыми для всех ранее рассмотренных экстремальных методов. Количество итераций и точность минимизации была приблизительно одинаковой.

Визуальные результаты минимизации были немного разными. Например, для МСГ и МРНСГ они представлены двумя рисунками 5.16 и 5.17 соответственно.

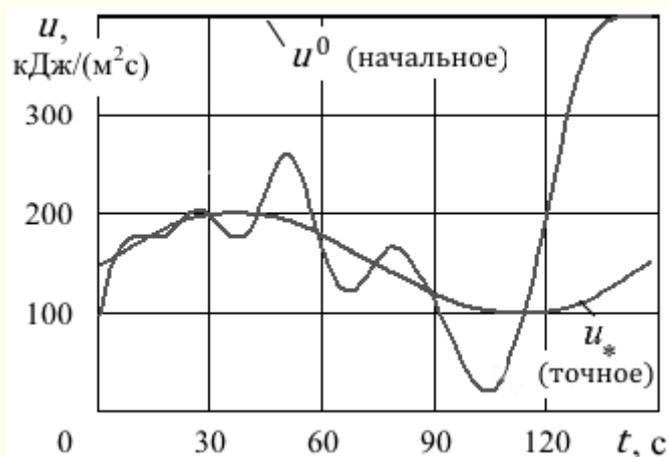


Рисунок 5.16 — Оптимальное управление потоком тепла посредством МСГ при наличии случайных погрешностей в функционале

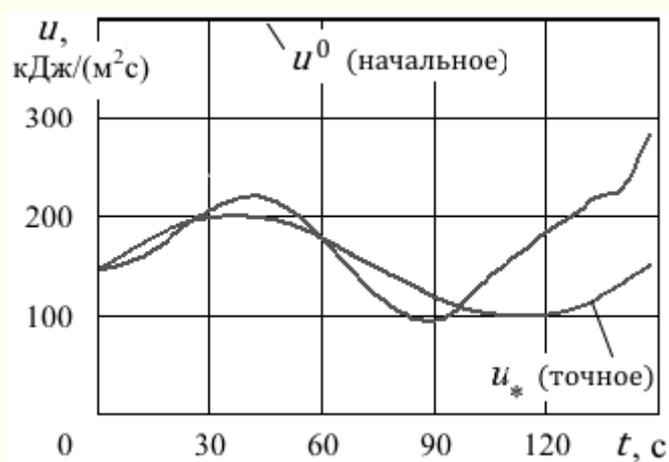


Рисунок 5.17 — Оптимальное управление потоком тепла посредством МРНСГ при наличии случайных погрешностей в функционале

Как и следовало ожидать, случайные погрешности в наблюдении температуры  $T_e$  привели к умеренным осцилляциям в решении задачи оптимизации. Все экстремальные методы продемонстрировали устойчивое приближение к оптимуму. Это объясняется

регуляризирующими свойствами экстремальных алгоритмов, которые мы подробно обсуждали в разделе 3.3 «Регуляризация».

Перейдём к детерминированным помехам, характерным для компьютерных вычислений и систематических неточностей в исходных данных. Начнём с *абсолютных детерминированных помех* (1.18) в градиенте, т.е. вместо градиента мы имеем  $\nabla J^k + r$ . Пусть помехи будут постоянные во времени и на всех итерациях. При этом направление спуска МРНСГ

$$d^k = -\alpha(\nabla J^k + r).$$

Рассмотрим следующие два случая значения помех  $r$ :

$$r_1 = 0.2 \max_t \nabla J(u^0; t);$$

$$r_2 = -0.002 \max_t \nabla J(u^0; t).$$

Результаты оптимизации посредством МСГ для  $r_1$  показаны сплошной линией  $u^3$  ( $J^3 = 525$ ) и для  $r_2$  — пунктирной линией  $u^{25}$  ( $J^{25} = 18.6$ ) на рисунке 5.18. Аналогичные расчёты посредством МРНСГ для  $r_1$  показаны на рисунке 5.19 сплошной линией  $u^5$  ( $J^5 = 24.3$ ) и для  $r_2$  — пунктирной линией  $u^4$  ( $J^4 = 42.6$ ).

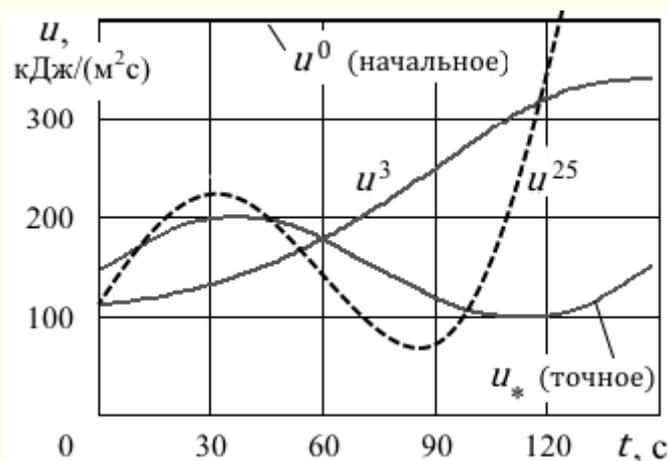


Рисунок 5.18 — Оптимальное управление потоком тепла посредством МСГ с погрешностью  $r_1$  (сплошная линия) и  $r_2$  (пунктирная линия)

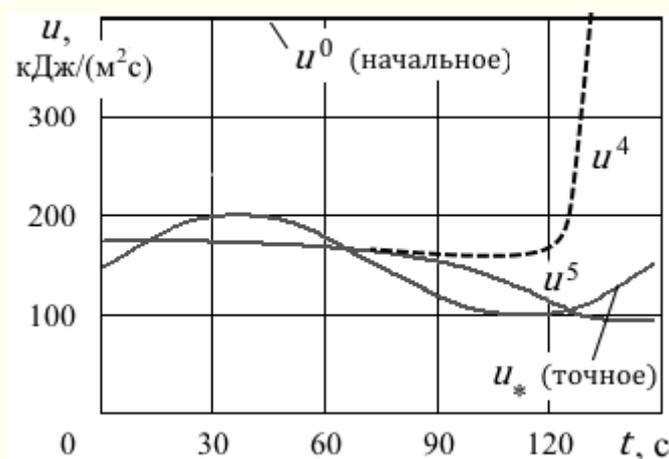


Рисунок 5.19 — Оптимальное управление потоком тепла посредством МРНСГ с погрешностью  $r_1$  (сплошная линия) и  $r_2$  (пунктирная линия)

Как мы видим, рассматриваемые абсолютные детерминированные помехи существенным образом сказываются на результатах оптимизации. Особенно следует отметить катастрофические результаты оптимизации при относительно малой отрицательной погрешности  $r_2$  на фоне изначально положительного значения градиента на  $S_\Delta$ . Это

объясняется тем, что абсолютное значение погрешности  $r_2$  к концу времени управления становится соизмеримой и даже превосходит значения градиента, что приводит к изменению положительного направление коррекции  $\nabla J(u^k; t)$  на ложное отрицательное  $\nabla J(u^k; t) + r_2$  при  $t \gtrsim 100$  с.

Перейдём к *относительным детерминированным помехам*  $r^k$  в (1.19), зависящими от значения градиента в каждый момент времени. Это погрешности приближённых, ошибочных вычислений, которые должны убывать по мере приближения к  $u_*$ . Зададим такие помехи коэффициентом  $\varepsilon(t)$  при градиенте, т.е.  $r^k(t) = \varepsilon(t)\nabla J(u^k; t)$ . При этом направлении спуска для МРНСГ будет

$$d^k = -\alpha(\nabla J^k + r^k) = -\alpha(1 + \varepsilon)\nabla J^k = -\alpha_\varepsilon \nabla J^k,$$

где  $\alpha_\varepsilon$  — это  $\alpha$ , выбранный описанными ранее способами, при наличии относительных детерминированных помех.

Рассмотрим три случая помех с зашумлением градиента, пока что не более 50%. При этом максимальное значение  $\varepsilon$  — это  $\varepsilon_0 = 0.5$ .

Варианты зашумления:

– Прямая линия:

$$\varepsilon(t) = \varepsilon_0;$$

– Убывающая наклонная линия, при  $t_0$  равная  $\varepsilon_0$ , а при  $t_1$  — ноль:

$$\varepsilon(t) = \varepsilon_0 \frac{t - t_1}{t_0 - t_1};$$

– Перевернутая парабола, принимающая на своих концах  $t_0, t_1$  нулевые значения, а в центре — максимальное  $\varepsilon_0$ ;

$$\varepsilon(t) = 4\varepsilon_0 \frac{t - t_0}{t_1 - t_0} \left(1 - \frac{t - t_0}{t_1 - t_0}\right).$$

Полученные результаты оптимизации посредством МСГ и МРНСГ визуально полностью совпали с результатами предыдущих

оптимизацией без помех на рисунках 5.6 (МСГ без равномерной сходимости) и 5.8 (МРНСГ с равномерной сходимостью). Несущественные различия содержались в общем количестве итераций и конечном значении целевого функционала.

Далее помехи были существенно усилены. Так, в частности, для наклонной линии помех было задано  $\varepsilon_0 = 100$ , что характерно для рассматриваемой задачи теплопроводности с асимптотическим убыванием градиента на несколько порядков при  $t \rightarrow t_1$  (смотрите таблицу 5.1). При таких помехах МСГ продемонстрировал существенно худшую сходимости (смотрите рисунок 5.20), которая закончилась уже на второй итерации.

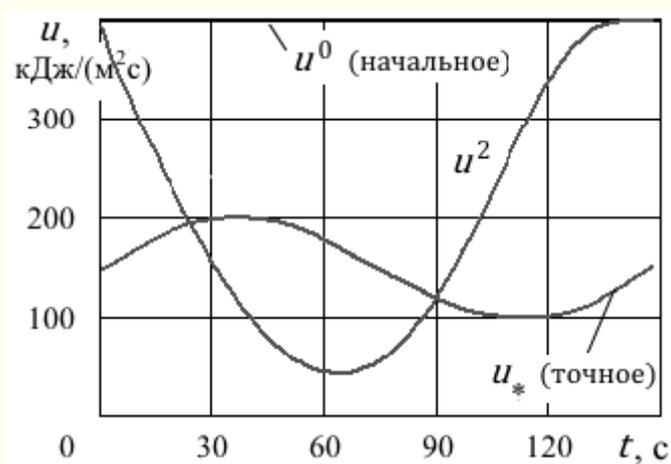


Рисунок 5.20 — Оптимальное управление потоком тепла посредством МСГ с линейной функцией помех при  $\varepsilon_0 = 100$

В то же время МРНСГ, с той же помехой, показал прекрасную сходимости, такую же как при управлении без помех (рисунок 5.8). Относительные помехи полностью компенсировались выбором

параметра  $\alpha(t)$  по формуле (5.28), обеспечив при этом равномерную сходимость на  $S_{\Delta}$ .

Подведём итоги.

Наличие умеренных *абсолютных случайных помех*, во-первых, не нарушило устойчивость алгоритмов оптимизации (задействуются регуляризирующие свойства экстремальных алгоритмов), во-вторых, получаемый случайный характер осцилляций оптимального управления не нарушил разумность и физичность решения.

*Абсолютные детерминированные помехи* в градиенте могут приводить к катастрофическому искажению оптимального управления, физически не реализуемого.

*Относительные детерминированные помехи* могут вообще не проявлять себя при использовании МРНС. Это тот самый случай, когда вычислительные *ошибки* (приближённые вычисления) могут не повлиять на решение. Алгоритмы МРНС способны очень эффективно устранять подобные ошибки. Относительно данных помех МРНС являются *робастными*.

## 6 Идентификация параметров при формировании отливок

### 6.1 Постановка задачи

Процессы затвердевания отливок существенным образом влияют на химические и механические свойства получаемого материала. Точность математического моделирования этих процессов, в основном, определяется точностью задания параметров-коэффициентов, входящих в уравнения конвекции расплава и тепломассопереноса в нём. Такие уравнения довольно громоздки, при численном решении требуют значительных компьютерных ресурсов и не гарантируют желаемой точности расчётов из-за многочисленных эмпирических, полуэмпирических параметров-коэффициентов. Качественное управление тепловыми процессами с такими моделями вряд ли возможно.

Значительное снижение вычислительных затрат может быть достигнуто введением эффективных коэффициентов теплопроводности и диффузии, что позволяет отказаться от расчёта уравнений конвекции и существенно снизить число экспериментально определяемых параметров. Естественно, что такие коэффициенты не могут быть измерены непосредственно. Их достоверные и точные значения могут быть получены только из решения соответствующих задач параметрической идентификации.

Здесь мы рассмотрим задачу идентификации эффективного коэффициента теплопроводности  $\lambda_{эф}(t)$  стальной отливки [33, 64, 26] — рисунок 6.1.

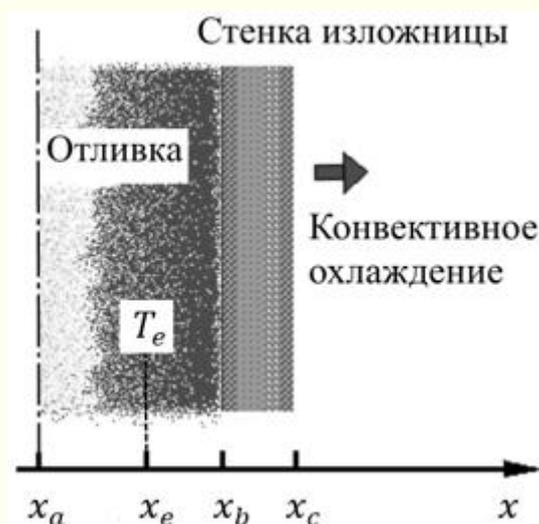


Рисунок 6.1 — К задаче идентификации эффективного коэффициента теплопроводности

В цилиндрическую изложницу (оснастку) заливается металл. Предполагается следующие: между отливкой и изложницей имеет место идеальный тепловой контакт; соотношения линейных размеров отливки позволяют рассматривать задачу в одномерной постановке вдоль радиуса отливки. Также полагаем задачу теплопереноса симметричной относительно центральной оси отливки.

В рамках принятых предположений *исходную задачу* процесса теплопередачи в системе отливка-изложница в стадии затвердевания можно описать следующей моделью в виде *линейного, однородного параболического* дифференциального уравнения:

$$C\rho \frac{\partial T}{\partial t} - \frac{\partial}{\partial x} \left( \lambda \frac{\partial T}{\partial x} \right) = 0, \quad (x, t) \stackrel{\text{def}}{=} \tau \in \Omega, \quad (6.1)$$

где  $T$  — температура в системе,  $C$ ,  $\rho$  — положительные константы теплоёмкости и плотности. Пространственно-временная область  $\Omega = (x_a, x_c) \times (t_0, t_1)$ , где  $x_a$  — центр отливки,  $x_c$  — внешняя граница

контакта изложницы с окружающей средой, а граница контакта отливки с изложницей — это  $x_b$ , при этом  $x_a < x_b < x_c$ .

Считаем, что эффективный коэффициент теплопроводности отливки  $\lambda_{\text{эф}}$  равномерно распределён на  $[x_a, x_b)$ , т.е. он однородный в расплаве, но в процессе затвердевания меняется со временем  $t \in (t_0, t_1)$ . Данный коэффициент для нас представляет собой управление, т.е.  $\lambda_{\text{эф}}(t) \stackrel{\text{def}}{=} u(t)$ . Теплопроводность в стенке изложницы считается известной и равной  $\lambda_{\text{ст}}$ .

Используя  $\Theta$ -функцию Хевисайда, можно записать коэффициент теплопроводности всей системы как

$$\lambda(x, t) = u(t)(1 - \Theta(x - x_b)) + \lambda_{\text{ст}}\Theta(x - x_b), \quad (6.2)$$

где  $\Theta$ -функция в зоне отливки, при  $x < x_b$ , удаляет из  $\lambda$  коэффициент теплопроводности стенки  $\lambda_{\text{ст}}$ , а в стенке изложницы, при  $x \geq x_b$ , удаляет из  $\lambda$  коэффициент теплопроводности отливки  $u$ . Здесь мы доопределили значение  $\Theta$ -функции в точке разрыва  $x_b$  величиной  $\Theta(0) = 1$ , т.е. точка  $x_b$  принадлежит стенке.

Граничные условия для модели (6.1) задаются на оси отливки и на внешней границе изложницы, в виде условий *второго рода* и *третьего рода*<sup>1</sup> соответственно:

$$\lambda \frac{\partial T}{\partial x} = 0 \quad \text{на } \Gamma_a = x_a \times (t_0, t_1), \quad (6.3)$$

$$\lambda \frac{\partial T}{\partial x} = -\gamma(T - T_c) \quad \text{на } \Gamma_c = x_c \times (t_0, t_1), \quad (6.4)$$

где  $\gamma$  — известный коэффициент теплоотдачи,  $T_c$  — температура окружающей среды на внешней границе изложницы.

<sup>1</sup> Граничное условие *третьего рода* (условие Роббина или Ньютона—Рихмана, это комбинация условий Неймана и Дирихле) описывает конвективный теплообмен поверхности тела с окружающей средой.

Надо отметить, что в точке  $x_b$ , по сути, имеет место граничное условие *четвёртого рода*<sup>1</sup> для тепловых потоков и температур:

$$\lambda \frac{\partial T}{\partial x} \Big|_{x_b^-} = \lambda \frac{\partial T}{\partial x} \Big|_{x_b^+}, \quad T|_{x_b^-} = T|_{x_b^+},$$

где  $x_b^-$ ,  $x_b^+$  — указывают на состояние процесса слева и справа от точки  $x_b$ . Разрывность коэффициента теплопроводности  $\lambda$  на границе двух сред, в нашем случае отливка-стенка, как известно [83], не приводит к разрыву температуры и потока тепла. Данное граничное условие мы явно использовать не будем, но оно позволит нам варьировать температуру системы на  $\Omega$  без выделения особенностей в точке  $x_b$ .

Начальное условие задавалось в виде:

$$T = \begin{cases} T_0 & \text{при } x < x_b \\ T_{\text{ст}} & \text{при } x \geq x_b \end{cases} \quad \text{на } \Gamma_0 = [x_a, x_c] \times t_0, \quad (6.5)$$

где  $T_0$  — температура залитого металла,  $T_{\text{ст}}$  — начальная температура стенки изложницы.

Отметим, что граничная область  $\Gamma_{\partial} = \Gamma_a \cup \Gamma_b$ , начальная  $\Gamma_0$  а также терминальная  $\Gamma_1 = [x_a, x_c] \times t_1$  образуют границу  $\partial\Omega$ . Замыкание  $\bar{\Omega} = \Omega \cup \partial\Omega$ .

Для идентификации модели (6.1) будем использовать экспериментальные наблюдения показаний термопары  $T_e$  в некоторой точке отливки  $x_e \in [x_a, x_b]$  в течение времени  $(t_0, t_1)$ . То есть, *множество наблюдений состояния объекта*  $\omega = x_e \times (t_0, t_1)$ .

Качество идентификации, согласно (4.1), будем оценивать по *невязке* наблюдаемого состояния модели  $T(x, t)$  и измеряемого состояния объекта  $T_e(t)$ :

<sup>1</sup> Граничное условие *четвёртого рода* описывает условия на границе контакта двух тел с различными теплофизическими характеристиками.

$$J(u) = \|M(T) - T_e\|_{L_2(\omega)}^2,$$

где  $M(\cdot) = \cdot|_{\omega}$  — оператор наблюдений модельного состояния, т.е. — это оператор сужения решения уравнений (6.1)–(6.5) из всей области  $\bar{\Omega}$  на область определения термопары  $\omega$ .

Критерий качества идентификации можно представить в виде целевого функционала типа (4.3) на  $\Omega$  с оператором  $M(\cdot)$  в виде  $\delta$ -функции:

$$J(u) = \int_{\Omega} (T - T_e)^2 \delta(x - x_e) d\Omega = \int_{t_0}^{t_1} (T|_{x_e} - T_e)^2 dt.$$

Теперь сравним терминологию и обозначения в поставленной задаче с предыдущими разделами теории прямого экстремального подхода.

Состояние системы — это положительная температура  $T(x, t) \stackrel{\text{def}}{=} v(\tau) \in V = L_{2+}(\bar{\Omega})$ , которая регулируется положительным управлением  $u(t) \in U = L_{2+}(S)$ ,  $S = (t_0, t_1)$ . Положительность управления не является ограничением в задаче оптимизации, а всего лишь условием параболичности системы (6.1). Размерности функций управления и состояния равны  $n = m = 1$ .

Заметим, что управление определено на  $S$ , и в то же время оно присутствует на  $\Omega$  в дифференциальном уравнении (6.1) и в граничном условии (6.3) на  $\Gamma_a$  через коэффициент теплопроводности  $\lambda$  в виде (6.2). Поэтому, на самом деле управление определено на пространственно-временной области  $\Omega \cup \Gamma_a$ , при этом оно однородно по пространству, но меняется со временем. При  $x \geq x_b$  управление закреплено значением  $\lambda_{\text{ст}}$ . Можно сказать, что  $S = P_t(\Omega|_{x < x_b})$ , т.е.,  $S$  является проекцией части (при  $x < x_b$ ) пространственно-временной области  $\Omega$  на ось времени  $t$ .

Уравнение (6.1) можно представить в операторном виде аналогично системе (3.20):

$$\mathbb{D}v = C\rho \frac{\partial v}{\partial t} - \frac{\partial}{\partial x} \left( \lambda \frac{\partial v}{\partial x} \right) = 0 \quad \text{на } \Omega, \quad (6.6)$$

где дифференциальный оператор  $\mathbb{D}(x, u) \cdot = C\rho \frac{\partial \cdot}{\partial t} - \frac{\partial}{\partial x} \left( \lambda \frac{\partial \cdot}{\partial x} \right)$  зависит от  $x$  через  $\Theta$ -функцию в  $\lambda$  и содержит управление  $u$ . Свободный член отсутствует, поэтому  $F = 0$ .

Целевой функционал можно записать в форме (3.19), когда граничный интеграл отсутствует:

$$J(u) = \int_{\Omega} I(v)|_{\omega} d\Omega, \quad (6.7)$$

$$I(v; x, t)|_{\omega} = (T(x, t) - T_e(t))^2 \delta(x - x_e),$$

где функция цели  $I$  отлична от нуля только в области  $\omega = x_e \times (t_0, t_1)$ .

Задача идентификации формулируется следующим образом. В модели (6.2)–(6.6) необходимо найти параметр (оптимальное управление)  $u(t)$ , который доставляет минимум функционалу (6.7). Ограничения на управление отсутствуют.

## 6.2 Градиент целевого функционала

Для нахождения градиента целевого функционала (6.7), согласно разделу 3.4, сразу перейдём ко второму шагу, поскольку ограничений в нашей задаче нет.

**2. Линеаризуем задачу.** Вариация коэффициента теплопроводности  $\delta u$  приводит к вариации температуры  $\delta T \stackrel{\text{def}}{=} \delta v$  во всей системе отливка-изложница. При этом вариация непосредственно целевого функционала (6.7), по аналогии с (3.22), принимает вид:

$$\delta J(u) = \langle I'_v|_{\omega}, \delta v \rangle_{V^*(\Omega)}, \quad (6.8)$$

где  $I'_v|_{\omega} = 2(T - T_e)\delta(x - x_e)$ .

Вариация уравнения теплопереноса в операторном виде (6.6) приводит к выражению типа (3.23):

$$\delta \mathbb{D}v = \mathbb{V}\delta v + \mathbb{U}\delta u = 0 \in V(\Omega),$$

или в явном виде:

$$C\rho \frac{\partial \delta T}{\partial t} - \frac{\partial}{\partial x} \left( \lambda \frac{\partial \delta T}{\partial x} \right) - \frac{\partial}{\partial x} \left( \frac{\partial \lambda}{\partial u} \frac{\partial T}{\partial x} \right) \delta u = 0 \quad \text{на } \Omega. \quad (6.9)$$

В сравнении с (3.23) здесь операторы

$$\begin{aligned} \mathbb{V} \cdot &= C\rho \frac{\partial \cdot}{\partial t} - \frac{\partial}{\partial x} \left( \lambda \frac{\partial \cdot}{\partial x} \right) : V(\Omega) \rightarrow V(\Omega), \\ \mathbb{U} \cdot &= -\frac{\partial}{\partial x} \left( \frac{\partial \lambda}{\partial u} \frac{\partial T}{\partial x} \right) \cdot : U(S) \rightarrow V(\Omega). \end{aligned}$$

Видно, что оператор  $\mathbb{U}$  отображает вариацию управления  $\delta u$  из области определения управления  $S$  в область  $\Omega$  функционирования распределённой системы.

**3. Отобразим линеаризованное уравнение (6.9) в сопряжённые пространства.** Возьмём произвольный линейный функционал  $\tilde{f}(\tau) \in V^*(\bar{\Omega})$  и, согласно (3.24), запишем его значение на вариации системы (6.9):

$$\langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} + \langle \tilde{f}, \mathbb{U}\delta u \rangle_{V^*(\Omega)} = 0.$$

Преобразуем данное выражение к значениям соответствующих линейных функционалов непосредственно на элементах  $\delta v$  и  $\delta u$ . В нашем одномерном случае, с первой производной по времени и второй производной по пространству, получим тождество Лагранжа типа (3.25) в виде:

$$\begin{aligned} \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} + \langle \tilde{f}, \mathbb{U}\delta u \rangle_{V^*(\Omega)} &\equiv \langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\Omega)} \\ &+ \langle \bar{\mathbb{V}}^* \tilde{f}, \delta v \rangle_{V^*(\partial\Omega)} + \langle \bar{\mathbb{V}}^* \tilde{f}, \frac{\partial \delta v}{\partial x} \rangle_{V^*(\Gamma_\partial)} \\ &+ \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)} = 0. \end{aligned} \quad (6.10)$$

Теперь запишем данное тождество в явном виде и найдём выражение для сопряжённых операторов  $V^*$  и  $U^*$ .

Начнём с преобразований при  $\delta v$ .

$$\begin{aligned}
\langle \tilde{f}, \nabla \delta v \rangle_{V^*(\Omega)} &= \iint_{\Omega} \tilde{f} C \rho \frac{\partial \delta T}{\partial t} dx dt - \iint_{\Omega} \tilde{f} \frac{\partial}{\partial x} \left( \lambda \frac{\partial \delta T}{\partial x} \right) dx dt \\
&= \iint_{\Omega} \left[ \frac{\partial}{\partial t} (\tilde{f} C \rho \delta T) - \frac{\partial \tilde{f}}{\partial t} C \rho \delta T \right] dx dt - \iint_{\Omega} \left[ \frac{\partial}{\partial x} \left( \tilde{f} \lambda \frac{\partial \delta T}{\partial x} \right) \right. \\
&\quad \left. - \frac{\partial \tilde{f}}{\partial x} \lambda \frac{\partial \delta T}{\partial x} \right] dx dt \\
&= \int_{x_a}^{x_b} \tilde{f} C \rho \delta T \Big|_{t_0}^{t_1} dx - \iint_{\Omega} C \rho \frac{\partial \tilde{f}}{\partial t} \delta T dx dt - \int_{t_0}^{t_1} \tilde{f} \lambda \frac{\partial \delta T}{\partial x} \Big|_{x_a}^{x_c} dt \\
&\quad + \iint_{\Omega} \frac{\partial \tilde{f}}{\partial x} \lambda \frac{\partial \delta T}{\partial x} dx dt.
\end{aligned}$$

У нас в последнем слагаемом осталась непреобразованная первая производная вариации состояния системы  $\frac{\partial \delta T}{\partial x}$ . Избавимся от неё.

$$\begin{aligned}
\left\langle \lambda \frac{\partial \tilde{f}}{\partial x}, \frac{\partial \delta T}{\partial x} \right\rangle_{V^*(\Omega)} &\equiv \iint_{\Omega} \lambda \frac{\partial \tilde{f}}{\partial x} \frac{\partial \delta T}{\partial x} dx dt \\
&= \iint_{\Omega} \left[ \frac{\partial}{\partial x} \left( \lambda \frac{\partial \tilde{f}}{\partial x} \delta T \right) - \frac{\partial \lambda}{\partial x} \frac{\partial \tilde{f}}{\partial x} \delta T - \lambda \frac{\partial^2 \tilde{f}}{\partial x^2} \delta T \right] dx dt \\
&= \int_{t_0}^{t_1} \lambda \frac{\partial \tilde{f}}{\partial x} \delta T \Big|_{x_a}^{x_c} dt - \iint_{\Omega} \frac{\partial}{\partial x} \left( \lambda \frac{\partial \tilde{f}}{\partial x} \right) \delta T dx dt.
\end{aligned}$$

В итоге тождество Лагранжа (6.10) при  $\delta v$  принимает вид:

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} \\
&= \langle -C\rho \frac{\partial \tilde{f}}{\partial t} - \frac{\partial}{\partial x} \left( \lambda \frac{\partial \tilde{f}}{\partial x} \right), \delta T \rangle_{V^*(\Omega)} \\
&+ \langle C\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_1)} - \langle C\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_0)} \\
&- \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Gamma_c)} + \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Gamma_a)} \\
&+ \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_c)} - \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_a)} = 0.
\end{aligned} \tag{6.11}$$

Из сравнения (6.10) с (6.11) видно, что сопряжённый оператор  $\mathbb{V}^*$  и его отображение — это

$$\mathbb{V}^* \cdot = -C\rho \frac{\partial \cdot}{\partial t} - \frac{\partial}{\partial x} \left( \lambda \frac{\partial \cdot}{\partial x} \right) : V^*(\Omega) \rightarrow V^*(\Omega).$$

Оставшаяся часть тождества (6.10) при вариации управления  $\delta u$  принимает вид:

$$\begin{aligned}
\langle \tilde{f}, \mathbb{U}\delta u \rangle_{V^*(\Omega)} &= \iint_{\Omega} -\tilde{f} \frac{\partial}{\partial x} \left( \frac{\partial \lambda}{\partial u} \frac{\partial T}{\partial x} \right) \delta u dx dt \\
&= \iint_{\Omega} \left[ -\frac{\partial}{\partial x} \left( \tilde{f} \frac{\partial \lambda}{\partial u} \frac{\partial T}{\partial x} \right) + \frac{\partial \lambda}{\partial u} \frac{\partial T}{\partial x} \frac{\partial \tilde{f}}{\partial x} \right] \delta u dx dt = \\
&= - \int_{t_0}^{t_1} \tilde{f} \frac{\partial T}{\partial x} \delta u \Big|_{x_a} dt + \int_{t_0}^{t_1} \left[ \int_{x_a}^{x_b} \frac{\partial T}{\partial x} \frac{\partial \tilde{f}}{\partial x} dx \right] \delta u dt.
\end{aligned}$$

Здесь было использовано то, что при  $x \geq x_b$  вариация  $\delta u = 0$  (закреплённость управления в стенке, вариация  $\delta u = \delta \lambda_{\text{ст}} = 0$ ), а при  $x < x_b$  производная  $\frac{\partial \lambda}{\partial u} = 1$  (следует из выражения (6.2)). Всё это отразилось на пределах интегрирования по  $x$ . Вместо интегрирования от  $x_a$  до  $x_c$  осталось интегрирование от  $x_a$  до  $x_b$ . Более того, значение первого интеграла в пределе  $|_{x_a}$  будет равно нулю, поскольку согласно граничному условию (6.3) при  $x_a$ , производная  $\frac{\partial T}{\partial x} = 0$ .

Таким образом, в тождестве (6.10) при  $\delta u$  получаем:

$$\langle \tilde{f}, \mathbb{U}\delta u \rangle_{V^*(\Omega)} \equiv \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)} = \int_{t_0}^{t_1} \left( \int_{x_a}^{x_b} \frac{\partial T}{\partial x} \frac{\partial \tilde{f}}{\partial x} dx \right) \delta u dt.$$

То есть сопряжённый оператор  $\mathbb{U}^*$  и его отображение — это

$$\mathbb{U}^* \cdot = \int_{x_a}^{x_b} \frac{\partial T}{\partial x} \frac{\partial \cdot}{\partial x} dx : V^*(\Omega|_{x < x_b}) \rightarrow U^*(S).$$

Оператор  $\mathbb{U}^*$  осуществляет отображение из пространства сопряжённых состояний, определённых на части  $\Omega$  где  $x < x_b$ , в пространство сопряжённых управлений, определённых на  $S = (t_0, t_1)$ . В своём отображении он осуществляет *проецирование*  $P_t(\Omega|_{x < x_b})$ , которое реализуется интегрированием по пространству на  $(x_a, x_b)$ .

**4. Объединим линейные функционалы** в соответствующих пространствах на элементах, к которым они относятся, и получим выражение вариации целевого функционала (без учёта граничных условий), аналогичное (3.30):

$$\begin{aligned} \delta J(u) &= \langle \mathbb{V}^* \tilde{f} + I'_v|_{\omega}, \delta T \rangle_{V^*(\Omega)} \\ &+ \langle C\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_1)} - \langle C\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_0)} \\ &- \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Gamma_c)} + \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial x} \rangle_{V^*(\Gamma_a)} \\ &+ \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_c)} - \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_a)} \\ &+ \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)}. \end{aligned} \quad (6.12)$$

Здесь в первом слагаемом была добавлена производная  $I'_v|_{\omega}$  от критерия качества идентификации.

**5. Учёт краевых условий.** Варьируем граничные и начальные условия (6.3)–(6.5):

$$\lambda \frac{\partial \delta T}{\partial x} = 0 \quad \text{на } \Gamma_a, \quad (6.13)$$

$$\lambda \frac{\partial \delta T}{\partial x} = -\gamma \delta T \quad \text{на } \Gamma_c, \quad (6.14)$$

$$\delta T = 0 \quad \text{на } \Gamma_0, \quad (6.15)$$

Теперь значения вариаций  $\frac{\partial \delta T}{\partial x}$  и  $\delta T$  из (6.13)–(6.15) подставим в вариацию функционала (6.12) на множествах  $\Gamma_a$ ,  $\Gamma_c$ ,  $\Gamma_0$ . Получим следующее выражение *первой вариации* целевого функционала:

$$\begin{aligned} \delta J(u) = & \langle \mathbb{V}^* \tilde{f} + I'_v|_\omega, \delta T \rangle_{V^*(\Omega)} + \langle C\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_1)} \\ & + \langle \gamma \tilde{f} + \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_c)} - \langle \lambda \frac{\partial \tilde{f}}{\partial x}, \delta T \rangle_{V^*(\Gamma_a)} \\ & + \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)}. \end{aligned} \quad (6.16)$$

**6. Выделение градиента.** В полученном выражении зададим  $\tilde{f} = f$ , такое, чтобы все линейные функционалы при  $\delta T$  обратились в ноль:

$$\mathbb{V}^* f + I'_v|_\omega = -C\rho \frac{\partial f}{\partial t} - \frac{\partial}{\partial x} \left( \lambda \frac{\partial f}{\partial x} \right) + I'_v|_\omega = 0 \quad \text{на } \Omega, \quad (6.17)$$

$$-\lambda \frac{\partial f}{\partial x} = 0 \quad \text{на } \Gamma_a, \quad (6.18)$$

$$\gamma f + \lambda \frac{\partial f}{\partial x} = 0 \quad \text{на } \Gamma_c, \quad (6.19)$$

$$f = 0 \quad \text{на } \Gamma_1. \quad (6.20)$$

Мы получили *сопряжённую задачу*, аналогичную (3.33).

При таком значении сопряжённой переменной  $f$  первая вариация целевого функционала принимает вид:

$$\delta J(u) = \langle \mathbb{U}^* f, \delta u \rangle_{U^*(S)} = \langle \nabla J, \delta u \rangle_{U^*(S)}.$$

Следовательно, *градиент целевого функционала*  $\nabla J = \mathbb{U}^* f$  имеет вид:

$$\nabla J(u; t) = \int_{x_a}^{x_b} \frac{\partial T}{\partial x} \frac{\partial f}{\partial x} dx \in U^*(S). \quad (6.21)$$

Полученное выражение соответствует первой строке общей формы градиента (3.36), когда искомый параметр (управление) присутствует в дифференциальном уравнении модели. Добавим, что (6.21) будет представлять градиент  $\nabla J$ , а не просто некоторую

производную Фреше  $J'_u$ , только при выполнении условий идентифицируемости.

### 6.3 Идентифицируемость

Для определения условий возможного корректного решения поставленной задачи идентификации в рамках прямого экстремального подхода воспользуемся теоремой 4.1 об идентифицируемости. Анализ идентифицируемости следует начинать с обоснования классической корректности исходной и сопряжённой задач. Далее, для определения градиента  $\nabla J$  на основе производной Фреше в выражении (6.21), необходимо выявить области определения и значений оператора  $\mathbb{U}^*$  и оценить возможность его вырожденности. В заключении мы должны корректно применить экстремальные алгоритмы.

Начнём с обоснования корректности обеих задач.

Уравнение (6.1) при всех положительных коэффициентах имеет параболический тип. Исходная задача с однородным уравнением (6.1), краевыми условиями второго и третьего рода (6.3), (6.4), с начальным условием (6.5) имеет корректное решение [83, 76, 31].

Сопряжённое уравнение (6.17) с дифференциальным оператором  $\mathbb{V}^*$  так же является параболическим с тем же самым коэффициентом теплопроводности  $\lambda$ , что и исходная задача. Знак минус при производной по времени указывает на необходимость решения сопряжённой задачи в обратном направлении с терминальным условием (6.20) на  $\Gamma_1$ . Граничные условия сопряжённой задачи (6.18), (6.19) аналогичны условиям (6.3), (6.4) исходной задачи (второго и третьего рода). Таким образом, сопряжённая задача (6.17)–(6.20) с учётом обратного времени, имеет классически корректное решение.

Найдём область определения и область значений оператора  $\mathbb{U}^*$ , воспользовавшись определением 3.5.

Изменение коэффициента теплопроводности посредством управления  $u(t)$  мгновенно (свойство параболических уравнений) меняет температуру  $T(x, t)$  во всей системе отливка-изложница. В силу корректности исходной задачи и отсутствия диссипаций, управление  $u(t)$  однозначно влияет на значение производной целевой функции  $I'_v|_\omega = 2(T - T_e)\delta(x - x_e)$ . То есть имеет место левая ветка отображений в определении 3.5:

$$U(S) \ni u \xrightarrow{\text{однозначно}} I'_v|_\omega.$$

В сопряжённой задаче свободный член  $I'_v|_\omega$  является источником возмущений изначально ненулевого сопряжённого состояния  $f$ . Возмущения  $f$  при  $t \rightarrow t_0$  мгновенно (опять же свойство параболических уравнений) распространяются в разные стороны от точки  $x_e$  по всей пространственной области  $[x_a, x_c]$ , как это показано на графике рисунка 6.4а (подробнее рисунок обсудим позже).

При численном решении дифференциальных уравнений, конечно же будут присутствовать вычислительные помехи. Поэтому, возмущения  $I'_v$  хоть и будут мгновенно распространяться в разные стороны, но заметными они станут на границах системы только после некоторого запаздывания  $\Delta t$ . Можно уверенно говорить о влиянии  $I'_v|_\omega$  на всё пространство отливка-изложница на  $[x_a, x_c]$  только с момента  $t_1 - \Delta t$  и далее — до момента  $t_0$ .

На рисунке 6.2 серым цветом показано множество  $\Omega = [x_a, x_b] \times [t_0, t_1 - \Delta t)$  области определения  $V^*(\Omega)$ , где оператор градиента  $\mathbb{U}^*$  принимает однозначные значения  $\mathbb{U}^*f$  в полуинтервале  $[t_0, t_1 - \Delta t)$ . Именно здесь в определении 3.5 могут выполняться обе правые ветки отображения

$$I'_v|_\omega \xrightarrow{\text{однозначно}} f|_\Omega \xrightarrow{\mathbb{U}^*} U^*([t_0, t_1 - \Delta t)).$$

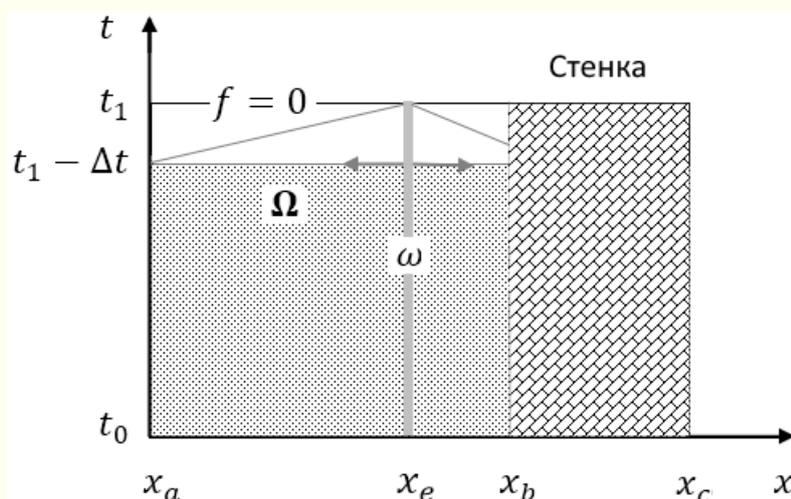


Рисунок 6.2 — Пространственно-временная диаграмма области  $\Omega$

Начальная линия  $[x_a, x_c] \times t_0$  в нашей задаче бесполезна, так как мы не проводим идентификацию начальных условий. Поэтому далее уберём эту линию из рассмотрения. Тогда, пространство сопряжённых состояний  $V^*(\Omega)$ , согласно определению 3.5, будет представлять собой область определения оператора  $\mathbb{U}^*$  с множеством

$$\Omega = [x_a, x_b] \times (t_0, t_1 - \Delta t).$$

При этом, в области значений оператора  $\mathbb{U}^*$ , т.е. в  $U^*(S)$ , вместо «старого» множества  $S = (t_0, t_1)$  получаем

$$S_\Delta = (t_0, t_1 - \Delta t).$$

Итак, мы нашли область определения  $V^*(\Omega)$  и область значений  $U^*(S_\Delta)$  оператора  $\mathbb{U}^*$ . Первое условие теоремы 3.4 «существует область  $V^*(\Omega)$  определения оператора  $\mathbb{U}^*$  с его значениями в области  $U^*(S)$ » выполняется только при  $U^*(S_\Delta)$ . Это означает, что эффективный коэффициент теплопроводности в модели (6.1), (6.3), (6.4) не идентифицируем на множестве  $S = (t_0, t_1)$  по функционалу (6.7). В то

же время, он может быть идентифицируем по этому же функционалу на множестве  $S_\Delta = (t_0, t_1 - \Delta t)$ .

Теперь оценим наличие возможной вырожденности оператора  $\mathbb{U}^*$ , т.е. выясним, имеет ли линейный оператор  $\mathbb{U}^*$  при каких-либо условиях ненулевое ядро  $\text{Ker } \mathbb{U}^* \neq 0$ . Иными словами, надо исследовать в каких случаях градиент  $\nabla J = \mathbb{U}^* f = \int_{x_a}^{x_b} \frac{\partial T}{\partial x} \frac{\partial f}{\partial x} dx$  может обращаться в ноль на  $S_\Delta$  или на значительной (ненулевой меры) части  $S_\Delta$  при  $f \neq 0$  всюду на  $\Omega$ , или п.в. $\Omega$ .

Наличие или отсутствие ненулевого ядра у оператора  $\mathbb{U}^*$  будет зависеть от точки  $x_e$  наблюдения температуры затвердевающего расплава. Для невырожденности  $\mathbb{U}^*$  необходимо, чтобы, во-первых, никакая из производных  $\frac{\partial T}{\partial x}$ ,  $\frac{\partial f}{\partial x}$  не обращалась в ноль на отрезке  $[x_a, x_b]$ , во-вторых, интегральные значения производных

$$\int_{x_a}^{x_e} \frac{\partial T}{\partial x} \frac{\partial f}{\partial x} dx \neq - \int_{x_e}^{x_b} \frac{\partial T}{\partial x} \frac{\partial f}{\partial x} dx, \quad \forall x_e \in (x_a, x_b),$$

т.е. они не сокращались для всевозможных мест размещения термопары.

Для наглядности дальнейших рассуждений обратимся к демонстрационным расчётам. Для некоторого  $u^0 > u_*$  на рисунке 6.3а показано характерное распределение температуры  $T$  в разные моменты времени от  $t_0$  до момента  $t_1$ , а на рисунке 6.3б — соответствующее распределение производной  $\frac{\partial T}{\partial x}$ . В начальный момент температура имеет разрыв на внутренней стенке  $x_b$  изложницы, поскольку температура стенки существенно ниже температуры заливаемого расплава. Разрыв со временем сглаживается, а температура всей системы отливка-изложница стремится выровняться. Направление наклона кривых температур, а, следовательно, и знак производной в окрестности разрыва, всегда одинаков (отрицательный). График 6.3б всплеска производной

температуры от максимального значения в точке разрыва  $x_b$  расширяется и стремится к нулевому значению равномерного распределения температуры в системе.

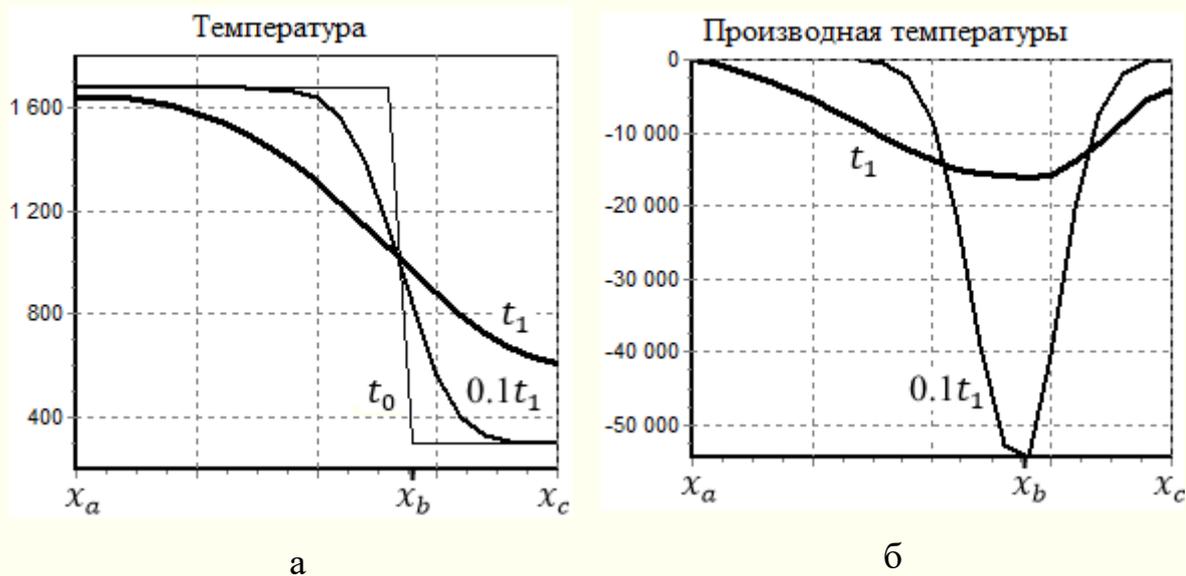


Рисунок 6.3 — Температура и её производная в системе отливка-изложница в разные моменты времени

На следующих рисунках 6.4 показано распределение сопряжённого состояния  $f$  и производной  $\frac{\partial f}{\partial x}$  для случая размещения термопары в центре рассматриваемой системы в точке  $x_e = 0.5(x_a + x_c)$ . Решение сопряжённой задачи  $f(x, t)$  (рисунок 6.4а) в каждый момент времени при движении от  $t_1$  к  $t_0$  имеет вид расширяющегося колокола  $f(x)$  с центром в точке  $x_e$ . Максимальный колокол образуется к моменту времени  $t_0$ . При этом значения производных (рисунок 6.4б) сопряжённого состояния являются знакопеременными относительно точки  $x_e$ .

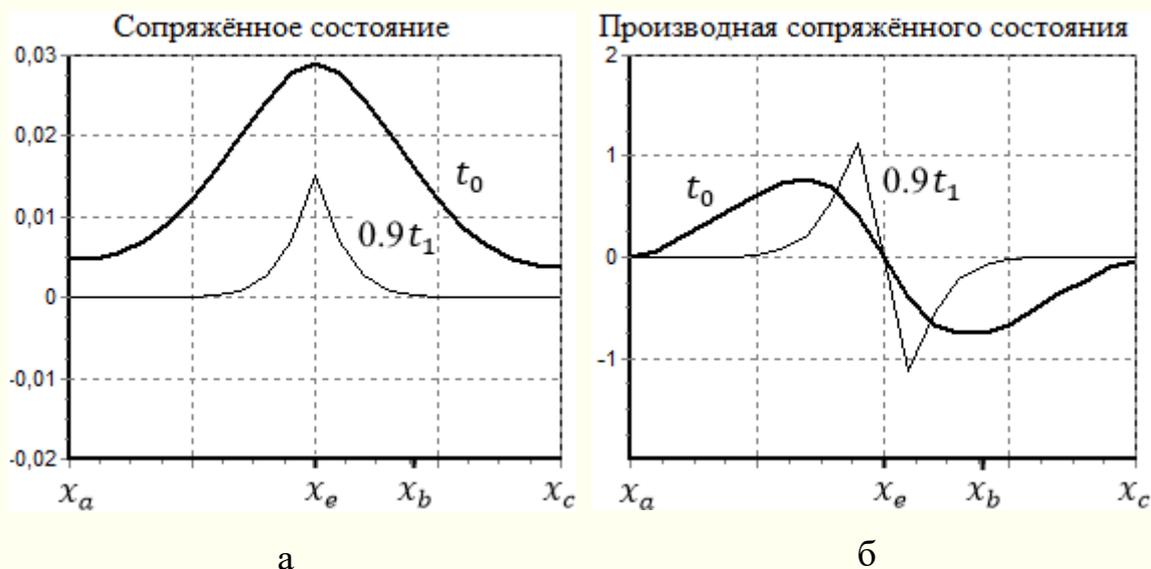


Рисунок 6.4 — Сопряжённое состояние и его производная в разные моменты времени

Оценивая визуально графики 6.3б и 6.4б для производных  $\frac{\partial T}{\partial x}$  и  $\frac{\partial f}{\partial x}$  слева и справа от точки  $x_e$ , можно утверждать, что возможно такое положение точки  $x_e$ , когда  $\int_{x_a}^{x_e} \frac{\partial T}{\partial x} \frac{\partial f}{\partial x} dx = - \int_{x_e}^{x_b} \frac{\partial T}{\partial x} \frac{\partial f}{\partial x} dx$  на всём множестве  $S_\Delta$ , или на его части конечной меры. То есть, будет  $\nabla J(u; t) = 0$  при  $f \neq 0$  и  $u \neq u_*$ . Это означает, что оператор  $\mathbb{U}^*$ , при некотором положении термодары внутри отливки, станет *вырожденным*.

Наличие вырожденности у оператора  $\mathbb{U}^*$  иллюстрируется рисунками 6.5 с тестовыми расчётами значения градиента  $\nabla J(x_e; u^0; t)$  и его нормы  $\|\nabla J(x_e; u^0; t)\|_{L_2(S_\Delta)}$ . Мы видим из рисунка 6.5а, что при  $x_e \approx x_b$  градиент на большей части интервала  $(t_0, t_1)$  обращается в ноль. Значение нормы градиента (рисунок 6.5б), вблизи внутренней стенки изложницы, приобретает существенный провал, приближающийся к нулевому значению.

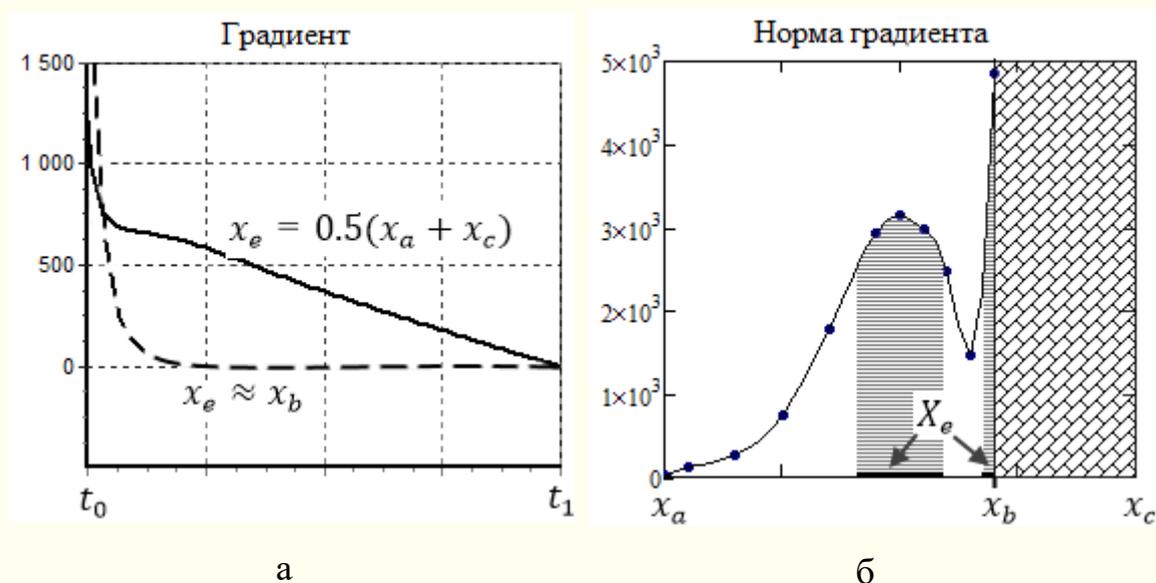


Рисунок 6.5 — Начальный градиент и его норма в зависимости от положения  $x_e$  термопары

Крайний левый всплеск градиента на рисунке 6.5а — это следствие разрыва начальной температуры  $T(x, t_0)$  на внутренней стенке изложницы  $x_b$ . С удалением точки наблюдений  $x_e$  от стенки  $x_b$  начальный всплеск в градиенте исчезает.

С учётом помех (приближённые вычисления состояний, производных...) можно выделить примерные две области размещения термопары, гарантирующие невырожденность оператора  $\mathbb{U}^*$  при существенно ненулевых нормах градиента. Это область  $X_e \subset (x_a, x_b]$  состоящая из двух частей, которые показаны на рисунке 6.5б. Первая левая часть  $X_e$  находится в расплаве примерно посередине всей системы, вторая — в малой зоне, прилегающей к стенке изложницы. Левее зоны  $X_e$ , как показали тестовые расчёты, помехи заметно ухудшают точность идентификации, поскольку  $\nabla J \approx 0$  в силу граничного условия  $\left. \frac{\partial f}{\partial x} \right|_{x_a} = 0$ . Вторая узкая часть  $X_e$  прилегает к внутренней стенке изложницы.

Термопару можно было бы встроить и в стенку изложницы. Но здесь мы эту ситуацию рассматривать не будем.

Таким образом, для преодоления вырожденности оператора  $\mathbb{U}^*$  необходимо задавать следующее положение термопары:

$$x_e \in X_e. \quad (6.22)$$

При этом будет выполняться второе условие теоремы 4.1 «оператор  $\mathbb{U}^*$  — невырожденный».

Условие (6.22), совместно с ранее полученным требованием

$$t \in S_\Delta = (t_0, t_1 - \Delta t), \quad (6.23)$$

представляют собой *условия идентифицируемости* в рассматриваемой задаче.

Согласно Теореме 4.1, математическая модель распределённой системы (6.1) с краевыми условиями (6.3)–(6.5) идентифицируема посредством параметра  $u(t)$  на  $S_\Delta$  (6.23) при условии (6.22) по функционалу  $J$  (6.7). При выполнении обоих условий идентифицируемости (6.22) и (6.23) производная Фреше  $J'_u = \nabla J$  в выражении (6.21), и становится возможным применение экстремальных алгоритмов для спуска к оптимуму  $u_*(t)$ . Остаётся корректно применить данные алгоритмы.

## 6.4 Численное решение

Было решено две задачи идентификация эффективного коэффициента теплопроводности: тестовая задача с гипотетической системой отливка-изложница и прикладная задача идентификация для отливки цилиндров поршней двигателей внутреннего сгорания.

Тестовая задача решалась при следующих параметрах: теплоёмкость  $C = 650$  Дж/(кг · К), плотность  $\rho = 6950$  кг/м<sup>3</sup>, теплопроводность изложницы  $\lambda_{ст} = 45$  Вт/(м · К), коэффициент

теплоотдачи  $\gamma = 580 \text{ Вт}/(\text{м}^2 \cdot \text{К})$ , начальные температуры отливки  $T_0 = 1673 \text{ К}$  и стенки изложницы  $T_{\text{ст}} = 300 \text{ К}$ , температура окружающей среды  $T_c = 300 \text{ К}$ . Радиус отливки составлял  $0.08 \text{ м}$ , толщина стенки изложницы  $0.03 \text{ м}$ .

Точка наблюдений температуры  $T_e$  термопары была задана на внутренней поверхности стенки изложницы:

$$x_e = x_b,$$

что удовлетворяет условию идентифицируемости (6.22).

Параметры прикладной задачи описаны непосредственно при её решении.

#### 6.4.1. Конечно-разностные аппроксимации

С учётом выражения (6.2) для  $\lambda(x, t)$  и значением производной тета-функции  $\Theta' = \delta$ -функция, представим уравнение (6.1) исходной задачи и (6.17) сопряжённой задачи в виде удобном для численной реализации:

$$C\rho \frac{\partial T}{\partial t} = \lambda \frac{\partial^2 T}{\partial x^2} + (\lambda_{\text{ст}} - u)\delta(x - x_b) \frac{\partial T}{\partial x}, \quad (6.24)$$

$$-C\rho \frac{\partial f}{\partial t} = \lambda \frac{\partial^2 f}{\partial x^2} + (\lambda_{\text{ст}} - u)\delta(x - x_b) \frac{\partial f}{\partial x} - 2(T - T_e)\delta(x - x_e). \quad (6.25)$$

Исходная и сопряжённая параболические задачи решались по явной конечно-разностной схеме с шаблонами аппроксимаций показанными на рисунке 6.6. Такая схема имеет первый порядок точности по  $\Delta t$  (далее верхний индекс  $j$  для слоёв по времени) и второй порядок по  $\Delta x$  (нижний индекс для узлов по пространству  $i$ ).

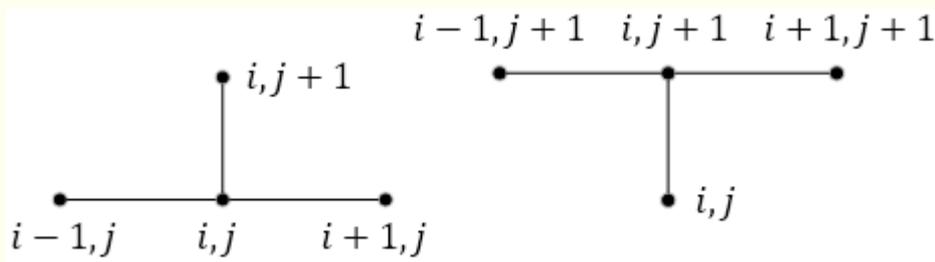


Рисунок 6.6 — Четырёхточечные шаблоны явной конечно-разностной схемы для решения исходной (слева) и сопряжённой (справа) задач

Аппроксимируем уравнение (6.24):

$$C\rho \frac{T_i^{j+1} - T_i^j}{\Delta t} - \lambda_i^j \frac{T_{i-1}^j - 2T_i^j + T_{i+1}^j}{\Delta x^2} - \Lambda_i^j \frac{T_{i+1}^j - T_{i-1}^j}{2\Delta x} = 0,$$

где  $\Lambda_i^j = (\lambda_{ст} - u^j)\delta(x_i - x_b)$ . Получаем

$$T_i^{j+1} = T_i^j + \frac{\Delta t}{2C\rho\Delta x^2} [(2\lambda_i^j - \Lambda_i^j\Delta x)T_{i-1}^j - 4\lambda_i^j T_i^j + (2\lambda_i^j + \Lambda_i^j\Delta x)T_{i+1}^j],$$

$$i = 1 \dots m - 1, \quad j = 0 \dots n - 1,$$

а на границах системы при  $i = 0$  (на оси отливки при  $x_a$ )

$$T_0^{j+1} = T_1^{j+1} \quad \text{на } \Gamma_a$$

и при  $i = m$  (на внешней поверхности изложницы при  $x_c$ ), из правой конечно-разностной производной температуры по  $x$  получаем

$$T_m^{j+1} = \frac{T_{m-1}^{j+1} + eT_c}{1 + e}, \quad e = \frac{\gamma\Delta x}{\lambda_{ст}} \quad \text{на } \Gamma_c.$$

В начальный момент времени при  $j = 0$  температура в системе

$$T_i^{j=0} = \begin{cases} T_0 & \text{при } x_i < x_b, \\ T_{ст} & \text{при } x_i \geq x_b, \end{cases} \quad i = 0 \dots m.$$

Теперь аппроксимируем сопряжённое уравнение (6.25) с учётом обратного времени:

$$-C\rho \frac{f_i^j - f_i^{j+1}}{-\Delta t} - \frac{f_{i-1}^{j+1} - 2f_i^{j+1} + f_{i+1}^{j+1}}{\Delta x^2} - \Lambda_i^{j+1} \frac{f_{i+1}^{j+1} - f_{i-1}^{j+1}}{2\Delta x} - 2(T - T_e)^j \delta(x_i - x_e) = 0.$$

Находим

$$f_i^j = f_i^{j+1} + \frac{\Delta t}{C\rho\Delta x^2} [(2\lambda_i^{j+1} - \Lambda_i^{j+1}\Delta x)f_{i-1}^{j+1} - 4\lambda_i^{j+1}f_i^{j+1} + (2\lambda_i^{j+1} + \Lambda_i^{j+1}\Delta x)f_{i+1}^{j+1}] - \frac{2\Delta t}{C\rho}(T - T_e)^j \delta(x_i - x_e),$$

$$i = 1 \dots m - 1, \quad j = n - 1 \dots 0.$$

На границах при  $i = 0$  и  $m$ :

$$f_0^j = f_1^j; \quad f_m^j = \frac{f_{m-1}^j}{1 + e}.$$

Начальное условие сопряжённой задачи (терминальное при  $j = n$ ):

$$f_i^n = 0, \quad i = 0 \dots m.$$

Градиент целевого функционала находится численным дифференцированием и интегрированием по пространству в соответствии с выражением (6.21):

$$\nabla J(u; t^j) = \int_{x_a}^{x_b} \frac{\partial T(x, t^j)}{\partial x} \frac{\partial f(x, t^j)}{\partial x} dx, \quad j = 1 \dots n - 1.$$

Присутствующие в нескольких местах *дельта* и *тета* функции аппроксимировались следующим образом:

$$\delta(x_i - y) = \begin{cases} \frac{1}{\Delta x} & \text{если } x_i \in \left(y - \frac{\Delta x}{2}, y + \frac{\Delta x}{2}\right), \text{ где } y \in \{x_e, x_b\}, \\ 0 & \text{иначе;} \end{cases}$$

$$\Theta(x_i - x_b) = \begin{cases} 0 & \text{если } x_i < x_b, \\ 1 & \text{иначе.} \end{cases}$$

Здесь дельта-функция представлена прямоугольным столбиком шириной  $\Delta x$  и высотой  $1/\Delta x$ , которая получена из условия нормировки (единичной площади) этого столбика.

Была выбрана пространственно-временная сетка:  $m = 20$  для индекса  $i$  по пространству;  $n = 100$  для индекса  $j$  по времени. При этом координаты системы отливка-изложница  $x_a = 0$  ( $i = 0$ ),  $x_e = x_b = 0.08$  м ( $i = 14$ ),  $x_c = 0.11$  м ( $i = 20$ ), начальное время  $t_0 = 0$  ( $j = 0$ ), конечное время  $t_1 = 45$  с ( $j = 100$ ). Напомним, значение координаты  $x_e$  удовлетворяет первому условию идентифицируемости (6.22).

Критерий устойчивости конечно-разностных схем, при  $\Delta x = 0.0055$  м и  $\Delta t = 0.45$  с, удовлетворял условию  $\frac{\lambda}{c\rho} \frac{\Delta t}{\Delta x^2} < 0.5$ , т.е. описанные вычислительные схемы были устойчивы.

Как показали расчёты, выбранное  $\Delta t$  оказалось достаточно большим для удовлетворения второго условия идентифицируемости (6.23). То есть, при решении исходной задачи на отрезке времени  $[t_0, t_1]$  мы будем оптимизировать значение  $u(t)$  на «вычислительном» множестве

$$S_{\Delta} = (t_0, t_1 - \Delta t] = [t_0 + \Delta t, t_1 - \Delta t] = [0.45, 44.55] \text{ с.}$$

#### 6.4.2. Оценка выпуклости целевого функционала

Для удовлетворения условий идентифицируемости теоремы 4.1, у нас остался открытым вопрос корректного применения экстремальных алгоритмов. Для этого необходимо оценить размеры компакта корректности  $\mathcal{U}$ , чтобы контролировать шаги  $u^k \in \mathcal{U}$ ,  $k = 0, 1, \dots$

Давайте оценим выпуклость целевого функционала (6.7) аналогично тому, как мы это делали в предыдущей задаче оптимизации. Зададим «оптимальное» управление  $u_* = 30$  Вт/(м · К) и два множества управлений  $\{u\}_1 = \{5, 15, 25\}$  и  $\{u\}_2 = \{50, 70, 90\}$ . Применим квадратичное интерполирование значений функции  $J_1(u)$  на множестве  $\{u\}_1$  и  $J_2(u)$  на множестве  $\{u\}_2$ .

На рисунке 6.7 показаны полученные интерполяционные параболы. Они не совпадают друг с другом, следовательно целевой функционал  $J(u)$  является *неквадратичным*. С разных сторон от оптимального управления он имеет разную выпуклость. Однако, с обеих сторон он проявляет себя как *выпуклый* в рассматриваемой области управлений.

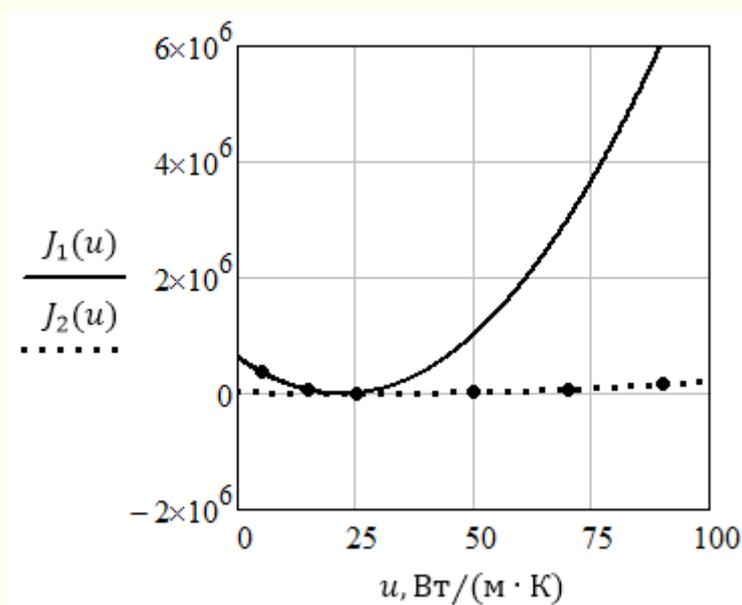


Рисунок 6.7 — Оценка выпуклости целевого функционала

Примем далее компакт корректности, достаточный для решения поставленной задачи оптимизации, в виде отрезка

$$\mathcal{U} = [5, 90] \frac{\text{Вт}}{\text{м} \cdot \text{К}}.$$

В этой области должно задаваться начальное управление  $u^0$  и корректно определяться все значения градиентов  $\nabla J(u^k)$ ,  $k = 1, 2, \dots$ . Можно сказать, что любое  $u^0 \in \mathcal{U}$  является удовлетворительной оценкой

значения идентифицируемого параметра  $\lambda_{\text{эф}}$ . Далее оценочное значение эффективной теплопроводности надо оптимизировать.

Существенно заметная неквадратичность целевого функционала может накладывать некоторые ограничения на применение алгоритмов МРНС с постоянным параметром  $\alpha^k = \alpha^0$ . Степень влияния данного фактора на качество работы МРНС можно будет оценить только численным экспериментом. Забегая вперёд, отметим, что применение МРНС для любых  $u^0 \in \mathcal{U}$  с параметром  $\alpha^0$  оказалось вполне приемлемым и позволило достичь оптимального значения  $u_*$  с высокой точностью. Целевой функционал  $J(u)$  проявил себя как *локально квадратичный* на полученных траекториях спуска к  $u_*$ .

В заключении отметим, что выпуклость целевого функционала  $J(u)$ , ядро сопряжённого оператора  $\mathcal{U}^*$ , границы компакта  $\mathcal{U}$ , очевидно, будут изменяться при перемещении координаты  $x_e$  термопары.

### 6.4.3. Решение тестовой задачи

Тесты организовывались следующим образом. Задавалось значение эффективного коэффициента теплопроводности  $\lambda_{\text{эф}}(t)$ , которое далее считалось оптимальным  $u_*(t)$ :

$$u_*(t) = 25 + 5 \sin \frac{2\pi t}{t_1 - t_0} \frac{\text{Вт}}{\text{м} \cdot \text{К}} \quad \text{на } S.$$

После чего решалась прямая задача (6.1)–(6.5). Найденная температура считалась экспериментальной  $T_e$ . Потом решалась обратная задача нахождения функции  $u(t)$ , доставляющей минимум целевому функционалу (6.7) при начальном приближении, заданном внутри компакта  $\mathcal{U} = [5, 90]$ :

$$u^0(t) = 50 \frac{\text{Вт}}{\text{м} \cdot \text{К}}.$$

Такому  $u^0$  соответствовало максимальное расхождение температур

$$\max_{t \in S} |T|_{x_e}(t) - T_e(t) = 121 \text{ К.}$$

Точность идентификации задавалась по критерию (1.24) практического прекращения сходимости

$$\frac{\|u^k - u^{k-1}\|_{L_2(S_\Delta)}}{\|u^{k-1}\|_{L_2(S_\Delta)}} \leq \varepsilon_4, \quad \varepsilon_4 = 10^{-5}.$$

Глубина шагов  $b^k$  во всех экстремальных алгоритмах задавалась линейным поиском *четвёртой стратегии* по методу Носедала-Райта с глубиной шагов не более  $\|u^0 - u_*\|$ .

Начнём тестирование с классических методов. На рисунке 6.8 показаны результаты идентификации коэффициента теплопроводности методом (1.8) наискорейшего спуска (МНС), обобщённого на бесконечномерные пространства:

$$u^{k+1}(t) = u^k(t) - b^k \nabla J(u^k; t) \quad \text{на } S_\Delta, \quad k = 0, 1 \dots$$

Значение параметра Вольфе в методе Носедала-Райта  $c_2 = 0.9$ .

Идентификация закончилась на итерации  $k = 32$  с заметными осцилляциями в начальные моменты времени и неравномерной сходимостью в последние моменты. Метод сопряжённых градиентов (МСГ) показал примерно такую же сходимость, но за меньшее число итераций  $k = 22$ . Для МСГ были заданы моменты обновления направлений спуска на каждой 10-й итерации и параметр Вольфе  $c_2 = 0.1$ .

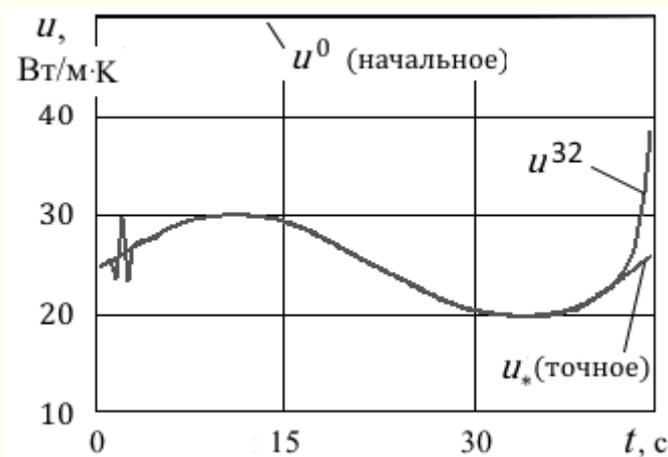


Рисунок 6.8 — Идентификация динамического коэффициента теплопроводности посредством МНС

Осцилляции управления в начальные моменты времени — это проявление особенностей используемой конечно-разностной схемы при расчёте разрывных состояний. Напомним, разрыв начальный температуры имеется в точке  $x_b$ . Через пять секунд начальный разрыв и осцилляции от него, как мы видим, полностью сглаживаются.

Обратимся к МРНСг (3.4), который здесь принимает вид:

$$u^{k+1}(t) = u^k(t) - b^k \alpha(t) \nabla J(u^k; t) \quad \text{равномерно на } S_\Delta, \quad k = 0, 1 \dots$$

Параметр регулирования направления спуска  $\alpha(t)$  задавался *шаблонным шагом* согласно (3.7) и (3.8), здесь они одинаковы:

$$\alpha(t) = \left| \frac{0.2u^0}{\nabla J(u^0; t)} \right|.$$

Параметр Вольфе  $c_2 = 0.9$ .

На рисунке 6.9 приведены полученные результаты. Идентификация закончилась на итерации  $k = 17$  при визуальном совпадении управления с точным значением  $u_*$ .

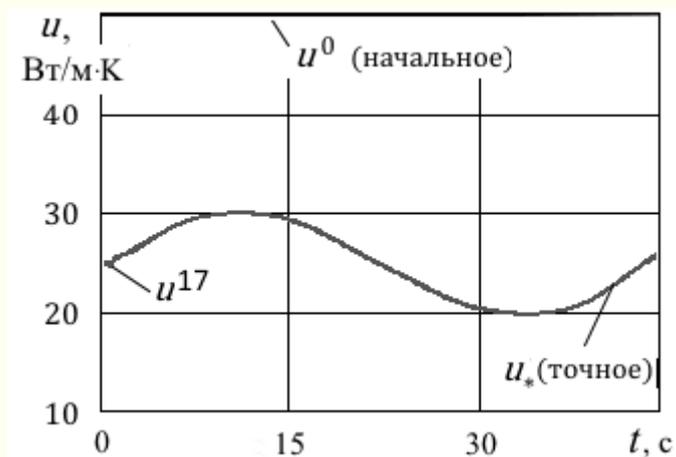


Рисунок 6.9 — Идентификация динамического коэффициента теплопроводности посредством МРНСг

Использование МРНСг (с базовым направлением сопряжённых градиентов) показало те же результаты, что и МРНСг на рисунке 6.9, но при меньшем количестве вычислений. Интересно отметить, что вычислительные *ошибки* (приближённые вычисления разрывных состояний) в виде осцилляций управления на предыдущем рисунке 6.8, оказались *относительными детерминированными помехами* и поэтому здесь, при использовании параметра  $\alpha(t)$  в МРНС, были полностью устранены.

Количественные результаты для всех методов сведены в таблицу 6.1.

Таблица 6.1 — Результаты идентификации динамического коэффициента теплопроводности

Метод	$k/\text{вычисления}$	$J^k$	$\ u^k - u_*\ $	$\max_t  T _{x_e} - T_e $
МНС	32/106	39	9.1	7.7
МСГ	22/488	73	11	9.2
МРНСГ	17/172	0.70	0.40	0.93
МРНСсг	13/55	1.4	0.66	0.92

Очевидна высокая *эффективность* МРНС. Здесь точность идентификации по невязке температур, при относительно незначительных вычислительных затратах, оказалась на порядок лучше традиционных алгоритмов. Сходимость к оптимальному значению — *равномерная*, без осциллирующих вычислительных помех.

#### 6.4.4. Решение прикладной задачи

Перейдём к решению реальной прикладной задачи идентификации эффективного коэффициента теплопроводности применительно к отливкам цилиндров поршней двигателей внутреннего сгорания, получаемым литьем чугуна в кокиль<sup>1</sup>.

По сравнению с предыдущей тестовой задачей здесь использовались следующие значения параметров: для чугунной отливки  $C = 540 \text{ Дж}/(\text{кг} \cdot \text{К})$ ,  $\rho = 6800 \text{ кг}/\text{м}^3$ ; для песчано-глинистой оснастки  $C = 920 \text{ Дж}/(\text{кг} \cdot \text{К})$ ,  $\rho = 1700 \text{ кг}/\text{м}^3$ ;  $\lambda_{\text{ст}} = 2 \text{ Вт}/(\text{м} \cdot \text{К})$ ,  $\gamma = 20 \text{ Вт}/(\text{м}^2 \cdot \text{К})$ ; начальные условия (6.5) для чугуна  $T_0 = 1613 \text{ К}$  и разогретого

<sup>1</sup> Кокиль, это металлическая или песчано-глинистая форма для получения отливок в литейном производстве повышенной точности и чистоты поверхности.

кокиля  $T_{ст} = 523$  К, температура окружающей среды  $T_c = 300$  К. При этих параметрах выполнялось условие устойчивости  $\frac{\lambda}{c\rho} \frac{\Delta t}{\Delta x^2} \leq 0.5$ .

Для данной задачи экспериментальная температура  $T_e$  была взята из натуральных наблюдений [27]. На рисунке 6.10 показаны дискретные измерения температуры (точки) и полученная посредством аппроксимации методом наименьших квадратов функция  $T_e(t)$ :

$$T_e(t) = 1366.6 - 40.2t + 3.6t^2 - 0.14t^3 + 0.00229597t^4. \quad (6.25)$$

При этом максимальное расхождение точек экспериментальных данных с аппроксимированной зависимостью (6.25) составило 4 К.

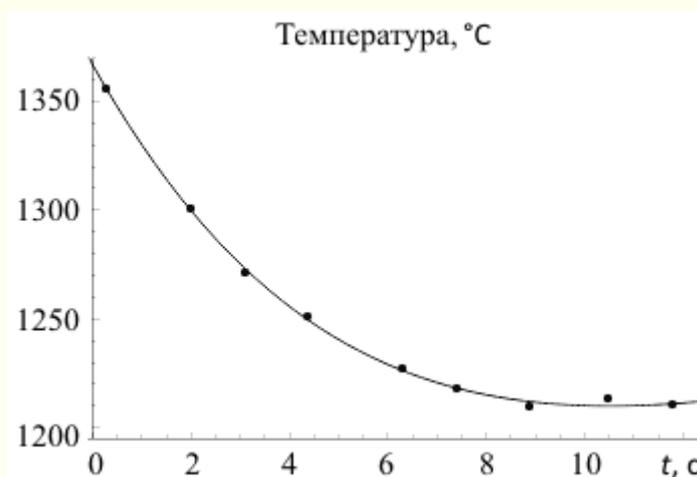


Рисунок 6.10 — Экспериментальная температура  $T_e(t)$

На основании температуры (6.25) было идентифицировано значение эффективного коэффициента теплопроводности  $\lambda_{эф} \stackrel{\text{def}}{=} u(t)$  посредством МРНСг. Результаты идентификации представлены на рисунке 6.11. Максимальное расхождение температур составило всего лишь  $\max_{t \in S} |T(x_e, t) - T_e(t)| = 6$  К. Это высокая точность моделирования реальных высокотемпературных процессов затвердевания на основе

экспериментальных данных с заметными (около 4 К) погрешностями наблюдений.

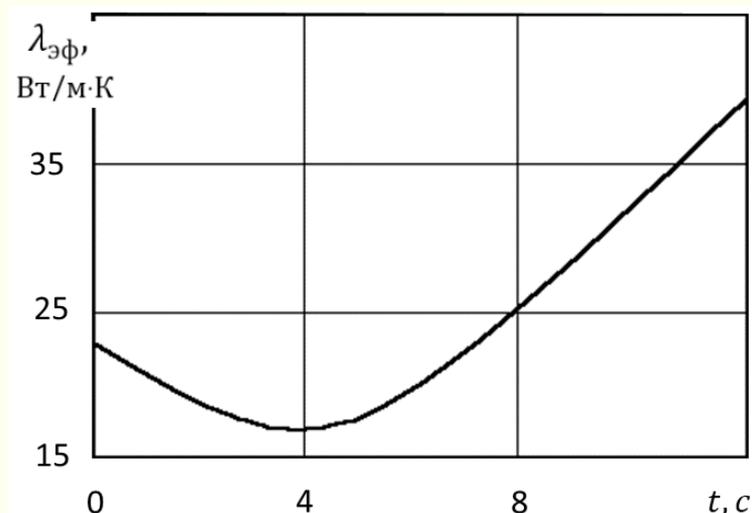


Рисунок 6.11 — Эффективный коэффициент теплопроводности, идентифицированный по экспериментальным данным

Как мы видим, полученное значение эффективного коэффициента теплопроводности не только не постоянно, но и существенно нелинейно. Именно такое динамичное поведение  $\lambda_{эф}(t)$  в упрощённой модели (6.1), не учитывающей конвекцию и другие физико-химические особенности затвердевающей отливки, позволило нам достичь высокой степени совпадения состояния модели с состоянием реального объекта.

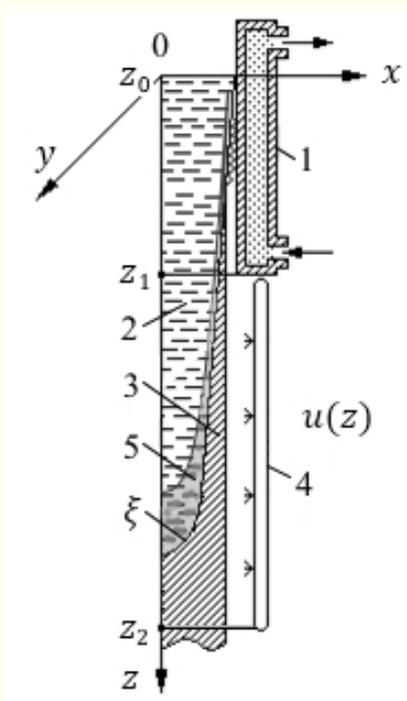
## 7 Оптимальное управление охлаждением непрерывного слитка

### 7.1 Постановка задачи

В данном разделе мы рассмотрим задачу определения наилучших условий затвердевания непрерывного слитка с целью формирования его структуры и свойств, связанных с различными физическими и химическими явлениями [64, 33, 16].

На рисунке 7.1 приведена принципиальная схема установки для непрерывного вертикального литья металла. В кристаллизатор 1 (первичный охладитель, формирующий твёрдый поперечный периметр слитка) заливается сверху жидкий металл 2. Вокруг слитка формируется твёрдая корка 3. Разу же за кристаллизатором следует внешний вторичный охладитель 4, в котором слиток затвердевает по всему объёму и на выходе из зоны вторичного охлаждения приобретает требуемую температуру. При внешнем охлаждении слитка в нём образуется жидкая лунка. По мере продвижения лунки вниз жидкий металл 2 остывает и перед полным затвердением имеет двухфазное жидкокристаллическое состояние 5, которое считается продолжением жидкой лунки.

Основным физическим параметром, в значительной мере предопределяющим структуру и свойства затвердевающего металла, является температура  $T$ . От ее распределения в расплаве зависят поля концентраций различных примесей, размеры переходной зоны твердожидкого состояния, распределение термических напряжений.



1 — кристаллизатор, 2 — жидкий металл, 3 — твёрдая корка, 4 —  
вторичный охладитель, 5 — жидкокристаллическое состояние  
Рисунок 7.1 — Принципиальная схема установки непрерывного  
вертикального литья металла.

Рассмотрим задачу снижения термонапряжений для предотвращения образования трещин в слитке посредством управления оттоком тепла  $u(z)$ , во вторичном охладителе. Для этого необходимо уменьшать горизонтальные градиенты температур  $T_x$ ,  $T_y$  по объёму корки слитка в зоне вторичного охлаждения. В работе [77] данное условие записывается как требование минимизации следующего функционала:

$$J(u) = \int_{z_1}^{z_2} \int_{\xi}^{y_b} \int_{\xi}^{x_b} (T_x^2 + T_y^2)^{\frac{1}{2}} dx dy dz, \quad (7.1)$$

где  $z_1$  и  $z_2$  — вертикальные координаты начала и конца зоны вторичного охлаждения слитка,  $x \in (0, x_b)$ ,  $y \in (0, y_b)$  — поперечные координаты слитка,  $\xi(x, y, z)$  — внутренняя граница затвердевшей части слитка.

Будем исследовать установившиеся тепловые процессы в непрерывном слитке посредством уравнения теплопроводности. В этом уравнении можно пренебречь кондуктивным членом с производной  $\frac{\partial^2 T}{\partial z^2}$ , поскольку конвективный теплоперенос вдоль оси слитка на два порядка превышает кондуктивный. Также полагаю, что влияние термоконвекции вынужденных циркуляционных потоков от воздействия струи и диффузия в твёрдой фазе пренебрежимо малы.

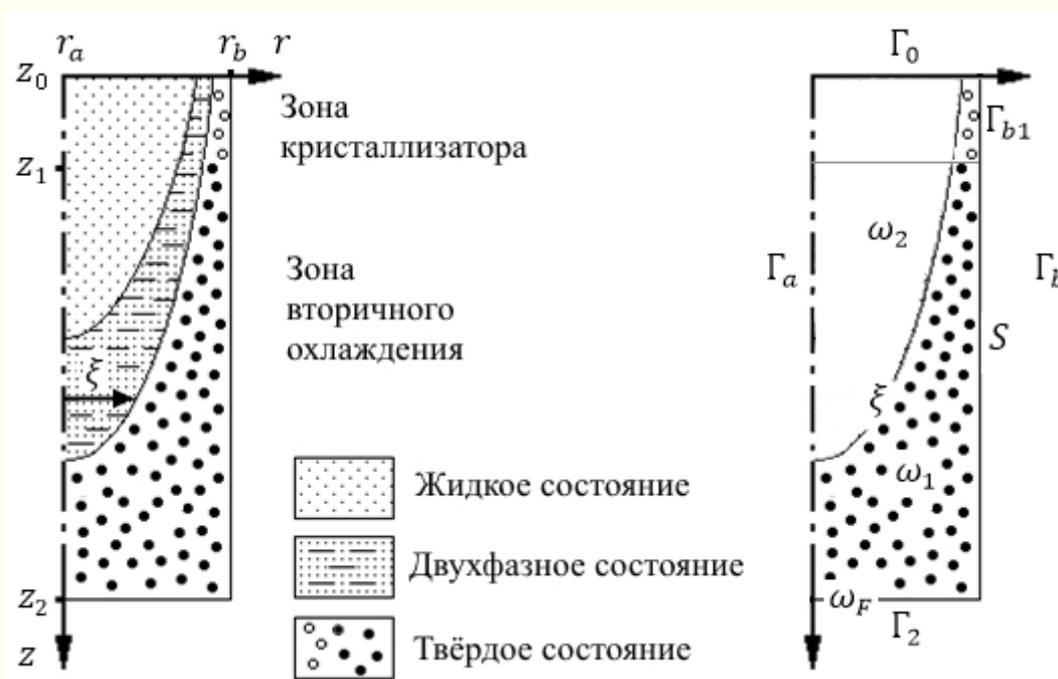


Рисунок 7.2 — Схема непрерывного цилиндрического слитка

Кроме того, будем рассматривать стальные слитки круглого сечения, используемые в качестве заготовок для трубо- и колесопрокатного производства. Принципиальная схема тепловой

обработки и фазовых состояний цилиндрического затвердевающего слитка приведена в левой части рисунка 7.2. Уравнение теплопереноса такой распределённой системы *исходной задачи*, при вышеуказанных допущениях, принимает вид *нелинейного, однородного* дифференциального уравнения [64, 78, 63] в цилиндрической системе координат:

$$WC\rho \frac{\partial T}{\partial z} - \frac{1}{r} \frac{\partial}{\partial r} \left( r\lambda \frac{\partial T}{\partial r} \right) = 0, \quad (7.2)$$

$$(r, z) \stackrel{\text{def}}{=} \tau \in \Omega = (r_a, r_b) \times (z_0, z_2)$$

где  $r$  — радиальная координата,  $r_b$  — радиус слитка, а на оси слитка естественно полагать, что  $r_a = 0$ ,  $T(r, z)$  — температура в слитке,  $W$  — постоянная скорость литья (скорость вытягивания слитка),  $C(T)$ ,  $\rho(T)$  и  $\lambda(T)$  — теплоёмкость, плотность и теплопроводность слитка, которые в жидкой, двухфазной и твёрдой зонах определяются следующим образом:

$$C = \begin{cases} C_l, & T \geq T_l, \\ C_l(1 - \zeta) + C_s\zeta - w \frac{\partial \zeta}{\partial T}, & T_s < T < T_l, \\ C_s, & T \geq T_s, \end{cases}$$

$$\rho = \rho_l(1 - \zeta) + \rho_s\zeta,$$

$$\lambda = \lambda_l(1 - \zeta) + \lambda_s\zeta.$$

Здесь нижние индексы  $l$  и  $s$  обозначают известные табличные значения параметров, соответственно, для жидкого (liquid) и твёрдого (solid) состояния слитка. Температуры  $T_s < T_l$  — температуры солидуса (начала плавления) и ликвидуса (конца плавления). Величина  $w$  — скрытая теплота кристаллизации. Переменная  $\zeta$  — доля твёрдой фазы в двухфазном состоянии, определяемая по уравнению для неравновесного рычага [60, 78]:

$$\zeta = 1 - \left( \frac{T_0 - T}{T_0 - T_l} \right)^{-\frac{1}{1-K}},$$

где  $K$  — равновесный коэффициент распределения примеси,  $T_0$  — температура заливки металла в кристаллизатор. Параметр  $\zeta$  используется для определения фазового состояния стали в заданной точке слитка. При  $\zeta \leq 0$  металл находится в жидком состоянии, при  $0 < \zeta < 0.95$  — в двухфазном, при  $\zeta \geq 0.95$  — в твёрдом.

Граничные условия для исходной задачи имеют вид условий *первого* (последнее по порядку), *второго* и *третьего* рода:

$$\begin{aligned} \frac{\partial T}{\partial r} \Big|_{\Gamma_a} &= 0, & \lambda_s \frac{\partial T}{\partial r} \Big|_{\Gamma_{b1}} &= -\gamma(T - T_c), \\ 2\pi r_b \lambda_s \frac{\partial T}{\partial r} \Big|_S &= -u, & T|_{\Gamma_0} &= T_0, \end{aligned} \quad (7.3)$$

где  $\Gamma_a = r_a \times [z_0, z_2]$  — ось слитка,  $\Gamma_0 = [r_a, r_b] \times z_0$  — верхняя граница слитка, где осуществляется заливка жидкого металла в кристаллизатор,  $\Gamma_{b1} = r_b \times (z_0, z_1)$  — зона кристаллизатора,  $T_c$  — температура кристаллизатора,  $\gamma$  — коэффициент теплоотдачи в зоне кристаллизатора,  $S = r_b \times (z_1, z_2)$  — зона вторичного охлаждения, где определено управление теплоотводом  $u(z)$ ,  $\Gamma_b = \Gamma_{b1} \cup S = r_b \times (z_0, z_2)$  — зона боковой поверхности слитка, взаимодействующая и с кристаллизатором, и со вторичным охладителем. Приведенные обозначения границ процесса иллюстрируются в правой части рисунка 7.2.

В соответствии с цилиндрической системой координат управление  $u(z)$  — это поток тепла по всей длине окружности слитка  $2\pi r_b$  в сечениях  $z$  вторичного охлаждения. Он измеряется в  $\text{Дж}/(\text{м} \cdot \text{с}) = \text{Вт}/\text{м}$ . Отвод тепла в зоне кристаллизатора можно оценить как

$$u_c(z) = 2\pi r_b \gamma (T - T_c).$$

В поставленной задаче, вместо функционала (7.1), будем минимизировать следующий *целевой функционал* термонапряжений:

$$J(u) = 2\pi \int_{z_0}^{z_2} \int_{r_a}^{r_b} \left( \frac{\partial T}{\partial r} \right)^2 \Theta(r - \xi) \Theta(z - z_1) r dr dz, \quad (7.4)$$

или компактнее

$$J(u) = 2\pi \int_{\Omega} \left( \frac{\partial T}{\partial r} \right)^2 \Big|_{\omega_1} r dr dz,$$

где  $\xi(r, z)$  — радиус границы твёрдой фазы и двухфазной зоны жидкой лунки слитка (смотрите рисунок 7.2). Множество точек, заданных радиусом  $\xi$ , образуют твёрдую границу  $\xi$ , которая определяется как первое появления доли твёрдой фазы  $\zeta \geq 0.95$  при  $r \rightarrow r_b$ . Величина  $r dr dz$  — элементарный объём цилиндра с точностью до константы  $2\pi$ , которая находится перед интегралом. *Тета-функции* Хевисайда представляют собой *оператор сужения*  $\cdot|_{\omega_1} = \cdot \Theta(r - \xi) \Theta(z - z_1)$  и выделяют из всего слитка затвердевшую область  $\omega_1$  (включая её границы) в зоне вторичного охлаждения, где мы будем управлять остаточными термонапряжениями. Значение  $\Theta$ -функции, связанное с границей  $\xi$ , легко определяется из выражения:

$$\Theta(r - \xi) = \begin{cases} 1 & \text{если } \zeta \geq 0.95, \\ 0 & \text{если } \zeta < 0.95. \end{cases}$$

Очевидно, что минимум функционала (7.4) достигается при  $\frac{\partial T}{\partial r} = 0$ , что означает отсутствие вторичного охлаждения и исчезновение твёрдой фазы. Весь слиток на выходе из машины непрерывного литья будет в двухфазном состоянии. Это бессмысленное охлаждение. Следовательно, целевой функционал (7.4) должен минимизироваться с некоторыми дополнительными ограничениями.

Зададим ограничение первого *типа* 1.1 в виде условий на температуру слитка при его выходе из зоны вторичного охлаждения, т.е. в виде множества допустимых состояний системы:

$$T|_{\Gamma_2} \leq T_{\max}, \quad (7.5)$$

где  $\Gamma_2 = [r_a, r_b] \times z_2$  — нижняя граница рассматриваемого слитка (смотрите рисунок 7.2).

Кроме термонапряжений целесообразно учитывать и возможность образования парогазовых полостей в виде пузырьков в двухфазной зоне. Данная особенность приводит к возникновению пористой структуры непрерывного слитка [78]. Наличие пористости, как и высоких остаточных термонапряжений, ослабляет сопротивление изделий механическим воздействиям.

Пористость слитка возрастает при снижении интенсивности отвода тепла в зоне вторичного охлаждения, т.е. при увеличении объёма жидкой лунки. Поэтому, в задачах оптимального охлаждения непрерывного слитка также целесообразно использовать и ограничение в виде минимального объёма жидкой лунки. Это будет ограничение опять-таки первого *типа*, но с подтипом 1.2. Его можно записать в форме функционала:

$$2\pi \int_{z_0}^{z_2} \int_{r_a}^{r_b} \Theta(T - T_s) \Theta(z - z_1) r dr dz \rightarrow \min. \quad (7.6)$$

Интеграл (7.6) включает в себя область жидкой лунки, когда  $T > T_s$ , внутри зоны вторичного охлаждения, когда  $z > z_1$ .

Теперь сравним терминологию и обозначения в поставленной задаче с предыдущими разделами теории прямого экстремального подхода.

Пространственно-распределённая система (7.2) представляет собой одно двумерное, стационарное дифференциальное уравнение в

частных производных с положительным состоянием  $v(\tau) = T(r, z) \in V = L_{2+}(\bar{\Omega})$ , которое регулируется распределённым управлением  $u(z) \in U \subset L_2(S)$  в виде отвода (теоретически допускается и ввод) тепла с поверхности цилиндрического слитка в зоне вторичного охладителя. Размерности функций управления и состояния равны  $n = m = 1$ .

Уравнение (7.2) можно представить в операторном виде аналогично системе (3.20):

$$\mathbb{D}v = WC\rho \frac{\partial v}{\partial z} - \frac{1}{r} \frac{\partial}{\partial r} \left( r\lambda \frac{\partial v}{\partial r} \right) = 0 \quad \text{на } \Omega, \quad (7.7)$$

где дифференциальный оператор  $\mathbb{D}(v) \cdot = WC\rho \frac{\partial \cdot}{\partial z} - \frac{1}{r} \frac{\partial}{\partial r} \left( r\lambda \frac{\partial \cdot}{\partial r} \right)$  зависит от  $v$  через свои положительные параметры  $C(v)$ ,  $\rho(v)$ ,  $\lambda(v)$ . Положительность параметров важна для определения типа уравнения (7.7). Свободный член отсутствует. Граница  $\Gamma_{\partial} = \Gamma_a \cup \Gamma_0 \cup \Gamma_b \cup \Gamma_2$  образует замыкание  $\bar{\Omega}$ .

Ограничения (7.5), (7.6) неявно задают допустимую область управлений  $U$  и приводят к формулировке трёх возможных задач оптимизации со следующими целевыми функционалами.

Первая задача — это минимизация термонапряжений  $J(u)$  в функционале (7.4) с обязательным ограничением (7.5) на выходную температуру слитка [16]. Здесь *целевой функционал* вместе с соответствующим *штрафным* функционалом для реализации явного ограничения на состояние системы типа 1.1 принимает вид интеграла в общем выражении (3.19):

$$J_1(u) = 2\pi \int_{\Omega} I(v)|_{\omega_1} r dr dz + \kappa_1 2\pi \int_{\Gamma_2} I_F(v, v_{\max})|_{\omega_F} r dr, \quad (7.8)$$

где  $I(v; r, z) = \left( \frac{\partial v(r, z)}{\partial r} \right)^2$  — функция термонапряжений в область  $\omega_1$  затвердевшей части слитка вторичного охлаждения,  $\kappa_1$  — весовой

коэффициент важности ограничения (7.5) в виде некоторой штрафной функции  $I_F(v, v_{\max}; r, z_2)$ . Напомним,  $r dr$  — это элементарная площадь поперечного сечения цилиндра с точностью до константы  $2\pi$ , которая, вообще говоря, не существенна при наличии  $\kappa_1$ .

Зададим внешнюю *штрафную функцию* в виде:

$$I_F = \begin{cases} (v - v_{\max})^2, & \text{если } v(r, z_2) > v_{\max}, \quad r \in \omega_F \subset \Gamma_2, \\ 0 & \text{иначе.} \end{cases}$$

Данное выражение определяет не только значение штрафной функции  $I_F$  в точках  $r$  на нижней границе слитка  $\Gamma_2$ , но и область  $\omega_F$  активного действия штрафа на этой границе. Оператор сужения  $\cdot|_{\omega_F}$  удобно представить *прямоугольной функцией*:

$$\Pi(r; \omega_F) = \begin{cases} 1, & r \in \omega_F, \\ 0, & r \notin \omega_F, \end{cases}$$

т.е. оператор  $\cdot|_{\omega_F} = \cdot \Pi(r; \omega_F)$  обращается в ноль если  $T(r, z_2) \leq T_{\max}$ .

Вторая задача — это минимизация термонапряжений и объёма жидкой лунки [64, 33]. Здесь мы получаем *целевой функционал* с учётом неявного ограничения на состояние системы типа 1.2 в виде:

$$J_2(u) = 2\pi \int_{\Omega} I(v)|_{\omega_1} r dr dz + \kappa_2 2\pi \int_{\Omega} 1|_{\omega_2} r dr dz, \quad (7.9)$$

где  $\omega_2$  — область жидкой лунки в зоне вторичного охлаждения, которая определяется *оператором сужения*  $\cdot|_{\omega_2} = \cdot \Theta(T - T_s)\Theta(z - z_1)$ . Влияние размеров лунки определяются весовым коэффициентом  $\kappa_2$ .

Третья задача — это минимизация термонапряжений с объёмом жидкой лунки и одновременно с ограничением на выходную температуру слитка. Здесь *целевой функционал*

$$\begin{aligned}
J_3(u) = & 2\pi \int_{\Omega} I(v)|_{\omega_1} r dr dz + \kappa_1 2\pi \int_{\Gamma_2} I_F(v, v_{\max})|_{\omega_F} r dr \\
& + \kappa_2 2\pi \int_{\Omega} 1|_{\omega_2} r dr dz.
\end{aligned} \tag{7.10}$$

Данный функционал можно рассматривать как

$$J_3 = \begin{cases} J & \text{если } \kappa_1 = 0, \kappa_2 = 0, \\ J_1 & \text{если } \kappa_2 = 0, \\ J_2 & \text{если } \kappa_1 = 0, \\ J_3 & \text{если } \kappa_1 \neq 0, \kappa_2 \neq 0. \end{cases}$$

То есть для минимизации  $J_3$  можно использовать суперпозицию предыдущих задач с  $J_1$  и  $J_2$  при ненулевых значениях штрафных коэффициентов  $\kappa_1$  и  $\kappa_2$ .

Заметим, что если  $\kappa_1$  и  $\kappa_2$  будут слишком малы, то все функционалы  $J_1, J_2, J_3$  будут близки к функционалу  $J$  (без ограничений) и жидкая лунка не закроется при выходе слитка из вторичного охладителя.

Применим прямой экстремальный подход для решения поставленных задач.

## 7.2 Градиент целевого функционала

Нам потребуется воспользоваться новым определением скалярного произведения, отличным от того, что было ранее. Введём для цилиндрической системы координат следующие *скалярные произведения*.

Для функций  $a, b$  из  $L_2(\Omega)$ , т.е. во всей области  $\Omega$  цилиндрического слитка скалярное произведение

$$\langle a, b \rangle_{L_2(\Omega)} = 2\pi \int_{\Omega} ab r dr dz.$$

Для функций, взятых в поперечном сечении слитка, например, на нижней границе  $\Gamma_2$  скалярное произведение

$$\langle a, b \rangle_{L_2(\Gamma_2)} = 2\pi \int_{\Gamma_2} ab r dr.$$

На коаксиальной цилиндрической поверхности слитка, при заданном радиусе  $r$ , например на боковой поверхности вторичного охлаждения при  $r = r_b$  скалярное произведение

$$\langle a, b \rangle_{L_2(S=\Gamma_b)} = 2\pi \int_{\Gamma_b} ab r dz.$$

Все эти произведения соответственно отражают объём цилиндрического слитка, площадь его поперечного сечения и площадь боковой поверхности слитка.

**1. Выделение ограничений на управление.** Ограничения (7.5) и (7.6) на состояние системы, по сути, уже учтены в целевых функционалах (7.8)–(7.10) и в явном виде теперь отсутствуют. Поэтому целевые функционалы  $J_1$ ,  $J_2$  и их суперпозиция  $J_3$  должны минимизироваться без ограничений, т.е. далее принимаем  $U = L_2(S)$  во всех задачах оптимизации.

Для получения градиента любого целевого функционала, достаточно рассмотреть задачи с функционалами  $J_1$  и  $J_2$ .

**2. Линеаризация задачи.** Вариация теплового потока  $\delta u$  во вторичном охладителе приводит к вариации температуры  $\delta T \stackrel{\text{def}}{=} \delta v$  в слитке. При этом вариации непосредственно целевых функционалов (7.8) и (7.9), по аналогии с (3.22), принимают вид:

$$\delta J_1(u) = \langle I'_v|_{\omega_1}, \delta v \rangle_{V^*(\Omega)} + \langle I'_v|_S, \delta v \rangle_{V^*(S)} + \langle \kappa_1 I'_{Fv}|_{\omega_F}, \delta v \rangle_{V^*(\Gamma_2)}, \quad (7.11)$$

$$\delta J_2(u) = \langle I'_v|_{\omega_1} + \kappa_2 (1|_{\omega_2})'_v, \delta v \rangle_{V^*(\Omega)} + \langle I'_v|_S, \delta v \rangle_{V^*(S)}, \quad (7.12)$$

где

$$I'_v|_{\omega_1} = -2 \left[ \left( \frac{\partial^2 T}{\partial r^2} + \frac{1}{r} \frac{\partial T}{\partial r} \right) \Theta(r - \xi) + \frac{\partial T}{\partial r} \delta(r - \xi) \right] \Theta(z - z_1),$$

$$I'_v|_S = 2 \frac{\partial T}{\partial r} \Big|_S,$$

$$I'_{Fv}|_{\omega_F} = 2(T - T_{\max}) \Pi(r; \omega_F),$$

$$(1|_{\omega_2})'_v = \delta(r - \xi) \Theta(z - z_1).$$

Здесь необходимо подробно написать, как получаются выражения указанных производных.

Первая и вторая по порядку производные  $I'_v|_{\omega_1}$ ,  $I'_v|_S$  получаются в результате следующего варьирования изначального функционала  $J(u)$  (7.4) с избавлением от производной вариации  $\frac{\partial \delta T}{\partial r}$ . С учётом производной тета-функции  $\Theta'_r = \delta(r - \xi)$ , приводящей к появлению  $\delta$ -функции, получаем:

$$\begin{aligned} \delta J(u) &= 2\pi \int_{z_0}^{z_2} \int_{r_a}^{r_b} 2 \left( \frac{\partial T}{\partial r} \right) \left( \frac{\partial \delta T}{\partial r} \right) \Theta(r - \xi) \Theta(z - z_1) r dr dz \\ &= 4\pi \int_{z_0}^{z_2} \int_{r_a}^{r_b} \left[ \frac{\partial}{\partial r} \left( r \frac{\partial T}{\partial r} \Theta(r - \xi) \delta T \right) - \frac{\partial T}{\partial r} \Theta(r - \xi) \delta T \right. \\ &\quad \left. - r \frac{\partial^2 T}{\partial r^2} \Theta(r - \xi) \delta T - r \frac{\partial T}{\partial r} \delta(r - \xi) \delta T \right] \Theta(z - z_1) dr dz. \end{aligned}$$

Интегрируем по  $r$  производную  $\frac{\partial}{\partial r}(\dots)$  с учётом  $\frac{\partial T}{\partial r}\Big|_{r_a} = 0$ ,  $\Theta(r - \xi)\Big|_{r_b} = 1$ , и в зоне вторичного охлаждения  $\Theta(z - z_1) = 1$ . Эта производная принимает вид:

$$\begin{aligned} & 4\pi \int_{z_0}^{z_2} \int_{r_a}^{r_b} \frac{\partial}{\partial r} \left( r \frac{\partial T}{\partial r} \Theta(r - \xi) \delta T \right) \Theta(z - z_1) dr dz \\ &= 4\pi \int_{z_0}^{z_2} \left( r \frac{\partial T}{\partial r} \Theta(r - \xi) \right) \Big|_{r_a}^{r_b} \Theta(z - z_1) \delta T dz \\ &= 4\pi \int_{z_1}^{z_2} r_b \frac{\partial T}{\partial r} \Big|_S \delta T dz = 2\pi r_b \int_{z_1}^{z_2} 2 \frac{\partial T}{\partial r} \Big|_S \delta T dz. \end{aligned}$$

Получаем

$$\begin{aligned} \delta J(u) &= 2\pi \int_{z_0}^{z_2} \int_{r_a}^{r_b} -2 \left[ \left( \frac{\partial^2 T}{\partial r^2} + \frac{1}{r} \frac{\partial T}{\partial r} \right) \Theta(r - \xi) + \frac{\partial T}{\partial r} \delta(r - \xi) \right] \\ &\quad \times \Theta(z - z_1) \delta T r dr dz + 2\pi r_b \int_{z_1}^{z_2} I'_v \Big|_S \delta T dz \\ &= \langle I'_v \Big|_{\omega_1}, \delta v \rangle_{V^*(\Omega)} + \langle I'_v \Big|_S, \delta v \rangle_{V^*(S)}. \end{aligned}$$

Такая относительно сложная вариация  $\delta J(u)$  с граничным членом на  $S$  появляется из-за наличия в функционале  $J(u)$  производной состояния  $\frac{\partial T}{\partial r}$ .

Выражение предпоследней в (7.12) производной  $I'_v \Big|_{\omega_F}$  представляется очевидным.

Остановимся на последней производной в выражении (7.12), где область  $\omega_2$  зависит от  $v = T$ :

$$(1|_{\omega_2})'_v = [\Theta(T - T_s) \Theta(z - z_1)]'_T = \delta(r - \xi) \Theta(z - z_1).$$

Здесь было учтено, что  $\Theta'_T = \delta(T - T_s)$  и понимание того, что  $T = T_s$  при  $r = \xi$ , т.е.  $\delta(T - T_s) = \delta(r - \xi)$ .

Вариация уравнения теплопереноса в операторном виде (7.7) приводит к выражению типа (3.23):

$$\delta \mathbb{D}v = \mathbb{V} \delta v = 0 \in V(\Omega),$$

где оператор  $\mathbb{V}$  в явном виде — это

$$\mathbb{V} \cdot = WC\rho \frac{\partial \cdot}{\partial z} + W \frac{\partial C\rho}{\partial z} \cdot - \frac{1}{r} \frac{\partial}{\partial r} \left( r\lambda \frac{\partial \cdot}{\partial r} + r \frac{\partial \lambda}{\partial r} \cdot \right), \quad (r, z) \in \Omega. \quad (7.13)$$

Здесь было учтено, что  $\frac{\partial C(T)\rho(T)}{\partial T} \frac{\partial T}{\partial z} = \frac{\partial C\rho}{\partial z}$  и  $\frac{\partial \lambda(T)}{\partial T} \frac{\partial T}{\partial r} = \frac{\partial \lambda}{\partial r}$ . Оператор  $\mathbb{U}$  на  $\Omega$  отсутствует.

**3. Отображение линеаризованных уравнений в сопряжённое пространства.** Возьмём произвольный линейный функционал  $\tilde{f}(\tau)$  из сопряжённого пространства  $V^*(\bar{\Omega})$  и по аналогии с (3.24) запишем его значение на вариации системы (7.13):

$$\langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} = 0.$$

Это выражение следует преобразовать к значениям линейных функционалов непосредственно на элементе  $\delta v$ , т.е. представить в виде тождества Лагранжа типа (3.25). В нашем случае, с первой производной по  $z$  и второй производной по  $r$ , оно будет иметь вид:

$$\begin{aligned} \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} &\equiv \langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\Omega)} \\ &+ \langle \bar{\mathbb{V}}^* \tilde{f}, \delta v \rangle_{V^*(\partial\Omega)} + \langle \bar{\bar{\mathbb{V}}}^* \tilde{f}, \frac{\partial \delta v}{\partial r} \rangle_{V^*(\Gamma_a \cup \Gamma_b)} = 0. \end{aligned} \quad (7.14)$$

Теперь запишем данное тождество в явном виде и найдём выражение для сопряжённого оператора  $\mathbb{V}^*$ . Граничные операторы  $\bar{\mathbb{V}}^*$  и  $\bar{\bar{\mathbb{V}}}^*$  в общем виде нам не понадобятся, поэтому их явно выписывать не будем.

Сделаем подробно необходимые преобразования типа (3.27):

$$\begin{aligned} &\langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} \\ &\equiv 2\pi \iint_{\Omega} \left[ \tilde{f}WC\rho \frac{\partial \delta T}{\partial z} + \tilde{f}W \frac{\partial C\rho}{\partial z} \delta T \right] r dr dz \\ &- 2\pi \iint_{\Omega} \tilde{f} \frac{\partial}{\partial r} \left( r\lambda \frac{\partial \delta T}{\partial r} + r \frac{\partial \lambda}{\partial r} \delta T \right) dr dz, \end{aligned}$$

где было учтено, что  $\frac{\partial C\rho}{\partial T} \frac{\partial T}{\partial z} \delta T = \frac{\partial C\rho}{\partial z} \delta T$ . Продолжим преобразования, используя правило дифференцирования произведений:

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} \\
& \equiv 2\pi \iint_{\Omega} \left[ \frac{\partial}{\partial z} (\tilde{f} W C \rho \delta T) - \frac{\partial \tilde{f}}{\partial z} W C \rho \delta T \right] r dr dz \\
& - 2\pi \iint_{\Omega} \left[ \frac{\partial}{\partial r} \left( \tilde{f} r \lambda \frac{\partial \delta T}{\partial r} + \tilde{f} r \frac{\partial \lambda}{\partial r} \delta T \right) - \frac{\partial \tilde{f}}{\partial r} r \lambda \frac{\partial \delta T}{\partial r} \right. \\
& \left. - \frac{\partial \tilde{f}}{\partial r} r \frac{\partial \lambda}{\partial r} \delta T \right] dr dz.
\end{aligned}$$

Здесь в самом начале сократились два слагаемых  $\tilde{f} W \frac{\partial C \rho}{\partial z} \delta T$  с разными знаками. Теперь проинтегрируем слагаемые  $\frac{\partial}{\partial z} (\tilde{f} W C \rho \delta T)$ ,  $\frac{\partial}{\partial r} \left( \tilde{f} r \lambda \frac{\partial \delta T}{\partial r} + \tilde{f} r \frac{\partial \lambda}{\partial r} \delta T \right)$  и получим:

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} \\
& \equiv 2\pi \int_{r_a}^{r_b} \tilde{f} W C \rho \delta T \Big|_{z_0}^{z_2} r dr - 2\pi \iint_{\Omega} \frac{\partial \tilde{f}}{\partial z} W C \rho \delta T r dr dz \\
& - 2\pi \int_{z_0}^{z_2} \left( \tilde{f} \lambda \frac{\partial \delta T}{\partial r} + \tilde{f} \frac{\partial \lambda}{\partial r} \delta T \right) \Big|_{r_a}^{r_b} r dz \\
& + 2\pi \iint_{\Omega} \left( \frac{\partial \tilde{f}}{\partial r} \lambda \frac{\partial \delta T}{\partial r} + \frac{\partial \tilde{f}}{\partial r} \frac{\partial \lambda}{\partial r} \delta T \right) r dr dz.
\end{aligned}$$

В пространстве  $V^*(\Omega)$ , как мы видим, в последней строке, всё ещё осталась производная от вариации  $\frac{\partial \delta T}{\partial r}$ . Избавимся от неё в соответствующем слагаемом:

$$\begin{aligned}
2\pi \iint_{\Omega} \frac{\partial \tilde{f}}{\partial r} \lambda \frac{\partial \delta T}{\partial r} r dr dz & = 2\pi \iint_{\Omega} \left[ \frac{\partial}{\partial r} \left( \frac{\partial \tilde{f}}{\partial r} r \lambda \delta T \right) - \frac{\partial}{\partial r} \left( \frac{\partial \tilde{f}}{\partial r} r \lambda \right) \delta T \right] dr dz \\
& = 2\pi \int_{z_0}^{z_2} \frac{\partial \tilde{f}}{\partial r} \lambda \delta T \Big|_{r_a}^{r_b} r dz - 2\pi \iint_{\Omega} \frac{1}{r} \frac{\partial}{\partial r} \left( r \lambda \frac{\partial \tilde{f}}{\partial r} \right) \delta T r dr dz.
\end{aligned}$$

В итоге тождество Лагранжа (7.14) принимает вид:

$$\begin{aligned}
\langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} &\equiv \langle -WC\rho \frac{\partial \tilde{f}}{\partial z} - \frac{1}{r} \frac{\partial}{\partial r} \left( r\lambda \frac{\partial \tilde{f}}{\partial r} \right) + \frac{\partial \lambda}{\partial r} \frac{\partial \tilde{f}}{\partial r}, \delta T \rangle_{V^*(\Omega)} \\
&+ \langle WC\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_2)} - \langle \tilde{f} C\rho W, \delta T \rangle_{V^*(\Gamma_0)} \\
&+ \langle \lambda \frac{\partial \tilde{f}}{\partial r} - \tilde{f} \frac{\partial \lambda}{\partial r}, \delta T \rangle_{V^*(\Gamma_b)} - \langle \lambda \frac{\partial \tilde{f}}{\partial r} - \tilde{f} \frac{\partial \lambda}{\partial r}, \delta T \rangle_{V^*(\Gamma_a)} \\
&- \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial r} \rangle_{V^*(\Gamma_b)} + \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial r} \rangle_{V^*(\Gamma_a)}.
\end{aligned} \tag{7.15}$$

Из полученного тождества видно, что сопряжённый оператор  $\mathbb{V}^*$  имеет вид:

$$\mathbb{V}^* \cdot = -WC\rho \frac{\partial \cdot}{\partial z} - \frac{1}{r} \frac{\partial}{\partial r} \left( r\lambda \frac{\partial \cdot}{\partial r} \right) + \frac{\partial \lambda}{\partial r} \frac{\partial \cdot}{\partial r}.$$

Если учесть, что  $-\frac{1}{r} \frac{\partial}{\partial r} \left( r\lambda \frac{\partial \cdot}{\partial r} \right) + \frac{\partial \lambda}{\partial r} \frac{\partial \cdot}{\partial r} = -\lambda \frac{\partial^2 \cdot}{\partial r^2} - \frac{\lambda}{r} \frac{\partial \cdot}{\partial r}$ , то получим:

$$\mathbb{V}^* \cdot = -WC\rho \frac{\partial \cdot}{\partial z} - \lambda \frac{\partial^2 \cdot}{\partial r^2} - \frac{\lambda}{r} \frac{\partial \cdot}{\partial r} \quad : \quad V^*(\Omega) \rightarrow V^*(\Omega). \tag{7.16}$$

Оператор в данной форме удобен для дальнейшего анализа типа уравнений получаемой сопряжённой задачи.

**4. Объединение задачи.** Объединим линейные функционалы в соответствующих пространствах на элементах, к которым они относятся и получим выражение первой вариации (без учёта краевых условий) целевых функционалов (7.8) и (7.9), аналогичное (3.30).

В первом случае объединяем (7.11) с (7.15):

$$\begin{aligned}
\delta J_1(u) &= \langle \mathbb{V}^* \tilde{f} + I'_v|_{\omega_1}, \delta T \rangle_{V^*(\Omega)} + \langle I'_v|_S, \delta T \rangle_{V^*(S)} \\
&+ \langle WC\rho \tilde{f} + \kappa_1 I'_{Fv}|_{\omega_F}, \delta T \rangle_{V^*(\Gamma_2)} - \langle WC\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_0)} \\
&+ \langle \lambda \frac{\partial \tilde{f}}{\partial r} - \tilde{f} \frac{\partial \lambda}{\partial r}, \delta T \rangle_{V^*(\Gamma_b)} - \langle \lambda \frac{\partial \tilde{f}}{\partial r} - \tilde{f} \frac{\partial \lambda}{\partial r}, \delta T \rangle_{V^*(\Gamma_a)} \\
&- \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial r} \rangle_{V^*(\Gamma_b)} + \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial r} \rangle_{V^*(\Gamma_a)}.
\end{aligned} \tag{7.17}$$

Во втором случае объединяем (7.12) с (7.15):

$$\begin{aligned}
\delta J_2(u) = & \langle \mathbb{V}^* \tilde{f} + I'_v |_{\omega_1} + \kappa_2 (1 |_{\omega_2})'_v, \delta T \rangle_{V^*(\Omega)} + \langle I'_v |_S, \delta T \rangle_{V^*(S)} \\
& + \langle WC\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_2)} - \langle WC\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_0)} \\
& + \langle \lambda \frac{\partial \tilde{f}}{\partial r} - \tilde{f} \frac{\partial \lambda}{\partial r}, \delta T \rangle_{V^*(\Gamma_b)} - \langle \lambda \frac{\partial \tilde{f}}{\partial r} - \tilde{f} \frac{\partial \lambda}{\partial r}, \delta T \rangle_{V^*(\Gamma_a)} \\
& - \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial r} \rangle_{V^*(\Gamma_b)} + \langle \lambda \tilde{f}, \frac{\partial \delta T}{\partial r} \rangle_{V^*(\Gamma_a)}.
\end{aligned} \tag{7.18}$$

**5. Учёт краевых условий.** Во-первых, сразу оценим значение коэффициента  $\lambda$  и его производной  $\frac{\partial \lambda}{\partial r}$  на границах  $\Gamma_a$  и  $\Gamma_b$ . На оси симметрии слитка  $\frac{\partial \lambda}{\partial r} \Big|_{\Gamma_a} = 0$ . На боковой поверхности слитка всегда имеется твёрдая корочка с постоянной теплопроводностью  $\lambda|_{\Gamma_b} = \lambda_s$  и поэтому  $\frac{\partial \lambda}{\partial r} \Big|_{\Gamma_b} = 0$ .

Теперь давайте варьируем краевые условия (7.3):

$$\begin{aligned}
\frac{\partial \delta T}{\partial r} \Big|_{\Gamma_a} = 0, \quad \lambda_s \frac{\partial \delta T}{\partial r} \Big|_{\Gamma_{b1}} = -\gamma \delta T, \\
2\pi r_b \lambda_s \frac{\partial \delta T}{\partial r} \Big|_S = -\delta u, \quad \delta T|_{\Gamma_0} = 0.
\end{aligned} \tag{7.19}$$

Значения вариаций  $\frac{\partial \delta T}{\partial r}$  и  $\delta T$  из (7.19) подставим в вариации функционалов (7.17), (7.18) на соответствующих граничных множествах. Множество  $\Gamma_b$  следует разделить на две части  $\Gamma_{b1}$  (область кристаллизатора) и  $S$  (область управления). Начнём с первого функционала (7.17):

$$\begin{aligned}
\delta J_1(u) = & \langle \mathbb{V}^* \tilde{f} + I'_v |_{\omega_1}, \delta T \rangle_{V^*(\Omega)} + \langle WC\rho \tilde{f} + \kappa_1 I'_{Fv} |_{\omega_F}, \delta T \rangle_{V^*(\Gamma_2)} \\
& + \langle \lambda_s \frac{\partial \tilde{f}}{\partial r}, \delta T \rangle_{V^*(\Gamma_{b1})} + \langle \lambda_s \frac{\partial \tilde{f}}{\partial r} + I'_v |_S, \delta T \rangle_{V^*(S)} \\
& - \langle \lambda \frac{\partial \tilde{f}}{\partial r}, \delta T \rangle_{V^*(\Gamma_a)} + \langle \gamma \tilde{f}, \delta T \rangle_{V^*(\Gamma_{b1})} + \langle \tilde{f}, \delta u \rangle_{U^*(S)}.
\end{aligned}$$

Объединив слагаемые на  $\Gamma_{b1}$ , получим:

$$\begin{aligned}
\delta J_1(u) = & \langle \mathbb{V}^* \tilde{f} + I'_v |_{\omega_1}, \delta T \rangle_{V^*(\Omega)} + \langle WC\rho \tilde{f} + \kappa_1 I'_{Fv} |_{\omega_F}, \delta T \rangle_{V^*(\Gamma_2)} \\
& + \langle \lambda_s \frac{\partial \tilde{f}}{\partial r} + \gamma \tilde{f}, \delta T \rangle_{V^*(\Gamma_{b1})} + \langle \lambda_s \frac{\partial \tilde{f}}{\partial r} + I'_v |_S, \delta T \rangle_{V^*(S)} \quad (7.20) \\
& + \langle \lambda \frac{\partial \tilde{f}}{\partial r}, \delta T \rangle_{V^*(\Gamma_a)} + \langle \tilde{f}, \delta u \rangle_{U^*(S)}.
\end{aligned}$$

Аналогично рассуждая, преобразуем второй функционал (7.18):

$$\begin{aligned}
\delta J_2(u) = & \langle \mathbb{V}^* \tilde{f} + I'_v |_{\omega_1} + \kappa_2 (1|_{\omega_2})'_v, \delta T \rangle_{V^*(\Omega)} + \langle WC\rho \tilde{f}, \delta T \rangle_{V^*(\Gamma_2)} \\
& + \langle \lambda_s \frac{\partial \tilde{f}}{\partial r} + \gamma \tilde{f}, \delta T \rangle_{V^*(\Gamma_{b1})} + \langle \lambda_s \frac{\partial \tilde{f}}{\partial r} + I'_v |_S, \delta T \rangle_{V^*(S)} \quad (7.21) \\
& + \langle \lambda \frac{\partial \tilde{f}}{\partial r}, \delta T \rangle_{V^*(\Gamma_a)} + \langle \tilde{f}, \delta u \rangle_{U^*(S)}.
\end{aligned}$$

Мы получили выражения *первых вариаций* для целевых функционалов  $J_1$  и  $J_2$ . При этом сопряжённый оператор  $\bar{\mathbb{U}}^*$  при граничном условии второго рода с соответствующим отображением имеет вид:

$$\bar{\mathbb{U}}^* = 1 : V^*(S) \rightarrow U^*(S),$$

что следует из последнего слагаемого в обеих выражениях первой вариации.

**6. Выделение градиента.** В выражениях (7.20) и (7.21) зададим значение  $\tilde{f} = f$ , такое, чтобы все линейные функционалы при  $\delta T$  обратились в ноль. Получим следующие *сопряжённые задачи*.

Сопряжённая задача для  $J_1$  первого функционала (7.8) с термонапряжениями и ограниченной температурой выхода слитка:

$$\mathbb{V}^* f + I'_v |_{\omega_1} = -WC\rho \frac{\partial f}{\partial z} - \lambda \frac{\partial^2 f}{\partial r^2} - \frac{\lambda}{r} \frac{\partial f}{\partial r} + I'_v |_{\omega_1} = 0 \quad \text{на } \Omega. \quad (7.22)$$

Краевые условия:

$$\left. \frac{\partial f}{\partial r} \right|_{\Gamma_a} = 0, \quad \lambda_s \left. \frac{\partial f}{\partial r} \right|_{\Gamma_{b1}} = -\gamma f, \quad (7.23)$$

$$\lambda_s \frac{\partial f}{\partial r} \Big|_S = -I'_v|_S, \quad f|_{\Gamma_2} = -\frac{\kappa_1}{WC\rho} I'_{Fv}|_{\omega_F}.$$

Сопряжённая задача для  $J_2$  второго функционала (7.9) с термонапряжениями и объёмом жидкой лунки:

$$\begin{aligned} \mathbb{V}^* f + I'_v|_{\omega_1} + \kappa_2 (1|_{\omega_2})'_v \\ = -WC\rho \frac{\partial f}{\partial z} - \lambda \frac{\partial^2 f}{\partial r^2} - \frac{\lambda}{r} \frac{\partial f}{\partial r} + I'_v|_{\omega_1} + \kappa_2 (1|_{\omega_2})'_v \\ = 0 \text{ на } \Omega. \end{aligned} \quad (7.24)$$

Краевые условия:

$$\begin{aligned} \frac{\partial f}{\partial r} \Big|_{\Gamma_a} = 0, \quad \lambda_s \frac{\partial f}{\partial r} \Big|_{\Gamma_{b1}} = -\gamma f, \\ \lambda_s \frac{\partial f}{\partial r} \Big|_S = -I'_v|_S, \quad f|_{\Gamma_2} = 0. \end{aligned} \quad (7.25)$$

Обе задачи незначительно отличаются друг от друга. Их различия определяются соответствующими целевыми функционалами  $J_1$  и  $J_2$ . В первой задаче имеется граничное условие в (7.23) с коэффициентом  $\kappa_1$  на нижней границе  $\Gamma_2$ . Во второй задаче имеется свободный член для жидкой лунки в уравнении (7.24) с коэффициентом  $\kappa_2$ .

При таких значениях сопряжённой переменной  $f$  первая вариация функционалов (7.8) и (7.9) принимает один и тот же вид:

$$\delta J_{1,2}(u) = \langle \bar{\mathbb{U}}^* f, \delta u \rangle_{U^*(S)} = \langle \nabla J, \delta u \rangle_{U^*(S)}.$$

Откуда следует, что *градиент целевых функционалов*  $\nabla J = \bar{\mathbb{U}}^* f$  и имеет вид:

$$\nabla J(u; z) = f \in U^*(S). \quad (7.26)$$

Полученное выражение градиента соответствует третьей строке общей формы градиента (3.36), когда управление присутствует только в граничном условии второго рода. Для того, чтобы выражение (7.26) действительно было градиентом, а не просто производной Фреше  $J'_u$ , необходимо получить условия управляемости.

### 7.3 Управляемость

Для корректного решения поставленной задачи оптимального управления посредством прямого экстремального подхода воспользуемся теоремой 3.4 об управляемости. Для реализации данной теоремы необходимо выявить область определения  $V^*(\Omega)$  и область значений  $U^*(S)$  оператора  $\bar{U}^*$ , а также сделать оценку возможной вырожденности  $\bar{U}^*$ . Это позволит найти градиент  $\nabla J$  из производной Фреше в выражении (7.26). Далее, на основе градиента  $\nabla J$ , мы сможем применить нужные нам экстремальные алгоритмы.

Но сначала следует обоснования классическую корректности исходной и сопряжённой задач. Уточним тип исходного дифференциального уравнения (7.2), или в форме (7.7). При быстром взгляде он кажется эллиптическим.

Согласно разделу 3.5.1 тип уравнения в частных производных определяется характеристическим уравнением через дискриминант  $d$ , в который входят коэффициенты при высших производных:

$$d = a_{12}^2 - a_{11}a_{22}.$$

Мы имеем оператор  $\mathbb{D} \cdot = W C \rho \frac{\partial \cdot}{\partial z} - \frac{1}{r} \frac{\partial}{\partial r} \left( r \lambda \frac{\partial \cdot}{\partial r} \right) = W C \rho \frac{\partial \cdot}{\partial z} - \lambda \frac{\partial^2 \cdot}{\partial r^2} + \Lambda \frac{\partial \cdot}{\partial r}$ ,

где  $\Lambda = \left( \frac{\lambda}{r} + \frac{\partial \lambda}{\partial r} \right)$ . Оператор имеет единственную вторую производную — по  $r$ . Можно записать коэффициенты при высших производных, как  $a_{11} = 0$ ,  $a_{12} = 0$ ,  $a_{22} = \lambda > 0$ . Получаем дискриминант  $d = 0$ .

В то же время, оператор  $\mathbb{D}$  имеет первую производную по  $z$  с положительными коэффициентами  $W C \rho$ . Значит, рассматриваемая задача имеет *параболический* тип с характеристическим направлением  $\frac{dr}{dz} = \infty$ , где  $z$  — аналог времени  $t$  для нестационарных задач. Поэтому, во-первых, данную задачу надо решать в сторону роста  $z$  от верхнего сечения слитка к нижнему (от  $\Gamma_0$  до  $\Gamma_2$ ). Во-вторых, в каждом

поперечном сечении слитка должны привлекаться два граничных условия (имеется вторая производная по  $r$ ), в нашем случае — это условия второго рода на оси симметрии слитка  $\Gamma_a$  и на внешней поверхности слитка — на  $\Gamma_b$ . При данных условиях однородная исходная параболическая задача (7.2), (7.3) имеет корректное решение [83, 76, 31].

Из корректности исходной задачи и отсутствия диссипативных потерь (исходная задача однородна) следует, что управление  $u(z)$  вторичного охлаждения однозначно влияет на производную целевой функции  $I'_v$  на  $\omega_1, \omega_F, S$ . То есть имеет место левая ветка отображений в определении 3.5 для всех задач оптимизации с функционалами  $J_1, J_2$ :

$$U(S) \ni u \xrightarrow{\text{однозначно}} I'_v|_{\omega_1}.$$

Напоминаем, что затвердевшей части слитка, т.е. множеству  $\omega_1$ , принадлежат все его границы, включая  $S$  и  $\omega_F$ . Поэтому здесь и далее мы не будем указывать дополнительно множества  $S$  и  $\omega_F$  при наличии  $\omega_1$ .

Теперь рассмотрим сопряжённые задачи: первая (7.22)–(7.23), вторая (7.24)–(7.25). Сразу же отметим, что неоднородность всех сопряжённых задач представлена только источниками возмущений от их целевых функционалов, следовательно, диссипация в рассматриваемых задачах отсутствует, и все возмущения сопряжённого состояния будут распространяться без искажений. Из вида сопряжённого оператора  $V^*$  (7.16) в обеих задачах, понятно, что его дискриминант  $d = 0$ .

Если направление  $z$  поменять на противоположное (что хорошо согласуется с наличием краевых условий в конце слитка на  $\Gamma_2$ ), то рассматриваемые уравнения станут параболическими с тем же самым коэффициентом теплопроводности  $\lambda$ , что и в исходной задаче. Граничные условия на  $\Gamma_a$  и  $\Gamma_b$  — это условия второго рода, аналогичные исходной задаче. Таким образом полученные сопряжённые

параболические задачи имеет классически корректное решение (с учётом обратного  $z$ ).

Обсудим область определения и область значений сопряжённого оператора  $\bar{U}^*$  в градиенте  $\nabla J = \bar{U}^* f$ . Воспользуемся для этого определением 3.5.

В первой задаче оптимального управления с функционалом  $J_1$  источниками возмущений, порождающими «сопряжённое тепло», являются функции  $I'_v|_{\omega_1}$  и  $I'_{Fv}|_{\omega_F}$ . Напомним, что тепловые возмущения в параболических системах распространяются мгновенно, с убыванием от своего источника. Для нас это означает следующее.

Во-первых, поскольку область источника термонапряжений вплотную соприкасается с боковой поверхностью слитка, то возмущения  $I'_v|_{\omega_1}$  от термонапряжений всегда будут достигать область управления  $S$ .

Во-вторых, никакие возмущения из верхней области между  $z_0$  и  $z_1$  не попадают на  $S$ , т.е. решать сопряжённую задачу в этой области бессмысленно.

Во-третьих, штрафные возмущения  $I'_{Fv}$  являются «начальным условием» на  $\Gamma_2 = [r_a, r_b] \times z_2$  для сопряжённой задачи, поэтому влияют на всю область  $\bar{\Omega}$ , включая  $S$ . Интенсивность влияния «начальных условий» определяется штрафным коэффициентом  $\kappa_1$ .

Таким образом, в первой задаче сопряжённое состояние  $f$  из  $V^*(\Omega)$  может отображаться в  $U^*(S)$  оператором  $\bar{U}^*$  для

$$\Omega = \omega_1,$$

т.е. для  $\Omega$ , равной всей затвердевшей части слитка, включая границы  $S$  и  $\omega_F$ . Если при управлении охлаждением слитка внешние твёрдые границы слитка не исчезают (естественное технологичное управление), то будут выполняться обе правые ветки отображений в определении 3.5:

$$I'_v|_{\omega_1} \xrightarrow{\text{однозначно}} f|_{\Omega} \xrightarrow{\bar{U}^*} U^*(S).$$

Рассмотрим вторую сопряжённую задачу (7.24)–(7.25) с функционалом  $J_2$ . Влияние сопряжённых тепловых возмущений  $I'_v|_{\omega_1}$  мы уже обсудили. Теперь рассмотрим, как влияют сопряжённые возмущения от жидкой лунки  $\kappa_2(1|_{\omega_2})'_v = \kappa_2\delta(r - \xi)\Theta(z - z_1)$ , которые представляют собой  $\delta$ -всплески на границе  $\xi$  начала твёрдой фазы при вторичном охлаждении. Очевидно, чем ближе граница  $\xi$  к внешней границе слитка  $r_b$  и чем больше значение<sup>1</sup>  $\kappa_2$ , тем существеннее будет влияние жидкой лунки на решение  $f$  на  $S$ . Поскольку область границы твёрдой фазы  $\xi$  входит в  $\omega_1$ , то полученное ранее  $\Omega = \omega_1$  не изменяется. С границы  $\xi$  сопряжённые возмущения переносятся через область  $\omega_1$  до границы  $S$ , где и формируется градиент  $\nabla J$ .

Очевидно, что в третьей задаче, в результате суперпозиции двух предыдущих, будем иметь ту же самую область  $\Omega$  для отображений оператором  $\bar{U}^*$  в  $U^*(S)$ .

Мы нашли для всех задач область определения  $V^*(\Omega)$ ,  $\Omega = \omega_1$  и область значений  $U^*(S)$  оператора  $\bar{U}^*$ . Таким образом, во всех задачах, выполняется первое условие теоремы 3.4 «существует область  $V^*(\Omega)$  определения оператора  $\bar{U}^*$  с его значениями в области  $U^*(S)$ ».

Теперь обратим внимание на сопряжённый оператор  $\bar{U}^* = 1$ . Очевидно, что выполняется и второе условие теоремы 3.4 «оператор  $\bar{U}^*$  — невырожденный».

Таким образом, полученные оценки для  $\Omega$  и  $\bar{U}^*$  означают, что распределённая система (7.2) с краевыми условиями (7.3) управляема посредством  $u(z)$  на  $S$  по функционалам  $J_1$ ,  $J_2$  и  $J_3$ . Особые,

<sup>1</sup> Значение  $\kappa_2$  имеет смысл только при вычислительных аппроксимациях  $\delta$ -функции.

дополнительные условия управляемости в данной задаче не требуются. Изначальная постановка всех трёх задач оптимального управления оказалась корректной. Производная Фреше  $J'_u$  безусловно совпадает с градиентом  $\nabla J$ . Остаётся корректно применить экстремальные алгоритмы. Следует ожидать равномерную на  $S$  сходимость, поскольку управляемость имеет место на всём множестве  $S$ .

## 7.4 Численное решение

Было решено две задачи. Первая тестовая задача — это синтез оптимального охлаждения непрерывного слитка по функционалу  $J = \left\| \frac{\partial T}{\partial r} - T'_* \right\|^2$ , где  $T'_*$  — заданное поле радиальных термонапряжений. Вторая задача — это реальная прикладная задача минимизации термонапряжений согласно функционалам  $J_1, J_2, J_3$ . Обе задачи применялись к слиткам с одинаковыми техническими характеристиками.

Для моделирования процессов затвердевания непрерывного цилиндрического слитка были приняты следующие значения параметров. Радиус слитка  $r_b - r_a = 0.1$  м, длина всего слитка  $z_2 - z_0 = 10$  м, длина кристаллизатора  $z_1 - z_0 = 1$  м, длина зоны вторичного охлаждения  $z_2 - z_1 = 9$  м. Температура кристаллизатора  $T_c = 293$  К, температура заливки металла  $T_0 = 1811$  К, коэффициент теплопередачи в зоне кристаллизатора  $\gamma = 300$  Вт/(м<sup>2</sup>К). Была выбрана сталь СТЗ, для которой температура ликвидуса  $T_l = 1796$  К, температура солидуса  $T_s = 1733$  К, скрытая теплота кристаллизации  $w = 272$  кДж/кг, равновесный коэффициент распределения примесей  $K = 0.4$ ,  $C_l = 838$  Дж/(кг · К),  $C_s = 661$  Дж/(кг · К),  $\rho_l = 7000$  кг/м<sup>3</sup>,  $\rho_s = 7620$  кг/м<sup>3</sup>,  $\lambda_l = 26.5$  Вт/(м · К),  $\lambda_s = 30.3$  Вт/(м · К). Скорость литья  $W = 0.02$  м/с.

### 7.4.1. Постановка тестовой задачи синтеза термонапряжений

Сформулируем тестовую задачу синтеза оптимального охлаждения непрерывного слитка по заданному полю радиальных термонапряжений  $\frac{\partial T}{\partial r} = T'_*$ . Необходимо найти оптимальный отток тепла  $u(z)$ , который обеспечивает заданные термонапряжения  $T'_*$  в затвердевшей части слитка  $\omega_1$  в зоне вторичного охлаждения. То есть необходимо минимизировать функционал

$$\begin{aligned} J(u) &= 2\pi \int_{z_0}^{z_2} \int_{r_a}^{r_b} \left( \frac{\partial T}{\partial r} - T'_* \right)^2 \Theta(r - \xi) \Theta(z - z_1) r dr dz \\ &= 2\pi \int_{\Omega} I(v)|_{\omega_1} r dr dz, \quad I(v) = \left( \frac{\partial T}{\partial r} - T'_* \right)^2. \end{aligned} \quad (7.27)$$

Такой функционал приведёт к небольшим изменениям в сопряжённой задаче (7.22) и (7.24). Здесь  $\kappa_1 = 0$  и  $\kappa_2 = 0$ .

Вариация тестового функционала (7.27) — это

$$\delta J(u) = \langle I'_v|_{\omega_1}, \delta v \rangle_{V^*(\Omega)} + \langle I'_v|_S, \delta v \rangle_{V^*(S)},$$

где

$$\begin{aligned} I'_v|_{\omega_1} &= -2 \left[ \left( \frac{\partial^2 T}{\partial r^2} - T''_* + \frac{1}{r} \left( \frac{\partial T}{\partial r} - T'_* \right) \right) \Theta(r - \xi) \right. \\ &\quad \left. + \left( \frac{\partial T}{\partial r} - T'_* \right) \delta(r - \xi) \right] \Theta(z - z_1), \\ I'_v|_S &= 2 \left( \frac{\partial T}{\partial r} - T'_* \right) \Big|_S. \end{aligned}$$

Производная  $T''_* = \frac{\partial T'_*}{\partial r}$ .

При этом сопряжённая задача примет вид:

$$-WC\rho \frac{\partial f}{\partial z} - \lambda \frac{\partial^2 f}{\partial r^2} - \frac{\lambda}{r} \frac{\partial f}{\partial r} + I'_v|_{\omega_1} = 0 \quad \text{на } \Omega. \quad (7.28)$$

Краевые условия:

$$\frac{\partial f}{\partial r}\Big|_{\Gamma_{a2}} = 0, \quad \lambda_s \frac{\partial f}{\partial r}\Big|_S = -2 \left( \frac{\partial T}{\partial r} - T'_* \right), \quad f|_{\Gamma_2} = 0,$$

где  $\Gamma_{a2} = r_a \times (z_1, z_2)$  — часть оси слитка ниже зоны кристаллизатора. Согласно анализу управляемости, решать сопряжённую задачу в зоне кристаллизатора бессмысленно.

Выражение для градиента останется тем же самым (7.26):

$$\nabla J(u; z) = f \quad \text{на } S.$$

### 7.4.2. Конечно-разностные аппроксимации

Для тестовой задачи исходное нелинейное и сопряженное линейное параболические уравнения решались численно по *явной конечно-разностной схеме* с четырёхточечным шаблоном (рисунок 7.3 с учётом направления оси  $z$  вниз). Такая схема имеет первый порядок точности по  $\Delta z$  и второй порядок по  $\Delta r$ . Задавалась пространственная сетка  $\{r_i = r_a + i\Delta r, i = 0 \dots m, \Delta r = (r_b - r_a)/m\}$ , и  $\{z_j = z_0 + j\Delta z, j = 0 \dots n, \Delta z = (z_{2(j=n)} - z_0)/n\}$ . Конец зоны кристаллизатора — это индекс  $j_c$ . Зона вторичного охлаждения охватывает индексы  $j \in \{j_c + 1 \dots n\}$ .

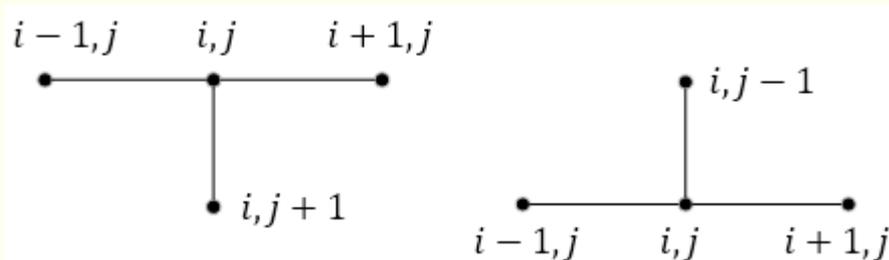


Рисунок 7.3 — Шаблоны явной конечно-разностной схемы для решения исходной (слева) и сопряжённой (справа) задач

Запишем нелинейное уравнение (7.2) в виде:

$$WC\rho \frac{\partial T}{\partial z} - \lambda \frac{\partial^2 T}{\partial r^2} - \Lambda \frac{\partial T}{\partial r} = 0,$$

где  $\Lambda = \left(\frac{\lambda}{r} + \frac{\partial \lambda}{\partial r}\right)$ . Аппроксимируем данное уравнение конечными разностями. Вертикальную сетку по  $z$ , будем отмечать верхним индексом  $j$ , а по радиусу, — нижним индексом  $i$ :

$$WC_i^j \rho_i^j \frac{T_i^{j+1} - T_i^j}{\Delta z} - \lambda_i^j \frac{T_{i-1}^j - 2T_i^j + T_{i+1}^j}{\Delta r^2} - \Lambda_i^j \frac{T_{i+1}^j - T_{i-1}^j}{2\Delta r} = 0.$$

Получаем

$$\begin{aligned} T_i^{j+1} = T_i^j + \frac{\Delta z}{2WC_i^j \rho_i^j \Delta r^2} [(2\lambda_i^j - \Lambda_i^j \Delta r)T_{i-1}^j - 4\lambda_i^j T_i^j \\ + (2\lambda_i^j + \Lambda_i^j \Delta r)T_{i+1}^j], \end{aligned} \quad (7.29)$$

$$i = 1 \dots m - 1, \quad j = 0 \dots n - 1.$$

На оси слитка при  $i = 0$  производная  $\frac{\partial T}{\partial r} = 0$  и мы имеем

$$T_0^{j+1} = T_1^{j+1},$$

а при  $i = m$ , на боковой поверхности, из правой конечно-разностной производной температуры по  $r$ , получаем

$$T_m^{j+1} = \frac{T_{m-1}^{j+1} + eT_c}{1 + e}, \quad e = \frac{\gamma \Delta r}{\lambda_s}, \quad \text{если } j + 1 \in \{1 \dots j_c\},$$

$$T_m^{j+1} = T_{m-1}^{j+1} - \frac{u_m^{j+1}}{2\pi\lambda_s m}, \quad \text{если } j + 1 \in \{j_c + 1 \dots n\}.$$

В начальном верхнем сечении слитка при  $j = 0$  температура заливки металла задана:

$$T_i^{j=0} = T_0, \quad i = 0 \dots m.$$

Целевой функционал (7.27) интегрировался численно методом трапеций по  $r$  и  $z$ . Для поставленной тестовой задачи при интегрировании по  $z$  в зоне вторичного охлаждения получаем:

$$J(u) \approx 2\pi\Delta z \left( \frac{J^{j_c+1} + J^n}{2} + \sum_{j=j_c+2}^{n-1} J^j \right), \quad (7.30)$$

где  $J^j$  — это результат интегрирования по  $r$ :

$$J^j = \Delta r \left( \frac{I_0^j r_0 + I_m^j r_m}{2} + \sum_{i=1}^{m-1} I_i^j r_i \right), \quad I_i^j = \left[ \left( \frac{\partial T}{\partial r} - T'_* \right)^2 \Theta \right]_i^j.$$

Производная  $\frac{\partial T}{\partial r}$  внутри слитка вычислялась через центральную конечно-разностную производную, а возле вертикальных границ — через соответствующие односторонние производные. *Функция Хевисайда*, для выделения твёрдой фазы, определяется как

$$\Theta_i^j = \begin{cases} 1 & \text{если } \zeta_i^j \geq 0.95, \\ 0 & \text{иначе.} \end{cases}$$

Теперь аппроксимируем сопряжённую задачу (7.28) аналогично исходной, но с учётом обратного движения по  $z$ , снизу вверх. Внутри слитка получаем:

$$\begin{aligned} -WC_i^j \rho_i^j \frac{f_i^{j-1} - f_i^j}{-\Delta z} - \lambda_i^j \frac{f_{i-1}^j - 2f_i^j + f_{i+1}^j}{\Delta r^2} - \frac{\lambda_i^j f_{i+1}^j - f_{i-1}^j}{r_i} \frac{1}{2\Delta r} \\ + (I'_v|_{\omega_1})_i^{j-1} = 0. \end{aligned}$$

Находим

$$\begin{aligned} f_i^{j-1} = f_i^j + \frac{\lambda_i^j \Delta z}{WC_i^j \rho_i^j \Delta r^2} \left[ \left( 1 - \frac{\Delta r}{2r_i} \right) f_{i-1}^j - 2f_i^j + \left( 1 + \frac{\Delta r}{2r_i} \right) f_{i+1}^j \right] \\ - \frac{(I'_v|_{\omega_1})_i^{j-1} \Delta z}{WC_i^j \rho_i^j}, \quad i = 1 \dots m-1, \quad j = n \dots j_c + 2. \end{aligned} \quad (7.31)$$

На вертикальных границах при  $i = 0$  и  $m$

$$f_0^{j-1} = f_1^{j-1}, \quad f_m^{j-1} = f_{m-1}^{j-1} - \frac{2}{\lambda_s} \left( T_m - T_{m-1} - T'_{*m-\frac{1}{2}} \Delta r \right)^{j-1}.$$

На нижней границе при  $j = n$

$$f_i^n = 0, \quad i = 0 \dots m.$$

Здесь значение  $(I'_v|_{\omega_1})_i^{j-1}$  находится численным дифференцированием температуры в виде центральной производной (первой и второй) относительно точки  $i$  на слое  $j - 1$ :

$$(I'_v|_{\omega_1})_i^{j-1} = -2 \left[ \left( \frac{T_{i-1} - 2T_i + T_i}{\Delta r^2} - \frac{T'_{*i+1} - T'_{*i-1}}{2\Delta r} + \frac{1}{r_i} \left( \frac{T_{i+1} - T_{i-1}}{2\Delta r} - T'_* \right) \right) \Theta_i + \left( \frac{T_{i+1} - T_{i-1}}{2\Delta r} - T'_* \right) \delta_i \right]^{j-1}.$$

Присутствующая в конце выражения  $\delta$ -функция аппроксимировалась прямоугольной функцией с единичной площадью на отрезке длиной  $\Delta r$ , где появляется граница  $\xi$  твёрдой фазы:

$$\delta_i^j = \begin{cases} 1/\Delta r & \text{если } \zeta_i^j \geq 0.95 \text{ впервые при } i \rightarrow m, \\ 0 & \text{иначе.} \end{cases}$$

Градиент целевого функционала в конечно-разностном представлении на  $S$  принимает вид:

$$\nabla J(u; z^j) = f_m^j, \quad j = j_c + 1 \dots n - 1.$$

Для вычислительной реализации полученных конечно-разностных алгоритмов задавался шаг по радиусу слитка  $\Delta r = 0.01$  м, по длине слитка  $\Delta z = 0.1$  м, что удовлетворяло условию устойчивости  $\frac{\lambda}{WC\rho} \frac{\Delta z}{\Delta r^2} < 0.5$ . Сетка определялась количеством узлов  $m=10$  по радиусу слитка, и  $n=100$  по длине слитка. Зона кристаллизатора заканчивалась через один метр от начала слитка, при  $j_c = 10$ .

### 7.4.3. Оценка выпуклости целевых функционалов

Здесь мы, на основе выпуклости исследуемых целевых функционалов, оценим размеры компакта корректности  $\mathcal{U}$ , чтобы в экстремальных алгоритмах правильно выбрать начальное приближение  $u^0 \in \mathcal{U}$  и контролировать дальнейшие приближения  $u^k \in \mathcal{U}$ ,  $k = 1, 2, \dots$

Начнём с тестовой задачи синтеза оптимального охлаждения. Построим две параболы вокруг оптимума, как мы это делали в предыдущих задачах оптимизации. Пусть  $u_* = 350$  кВт/м, а два множества управлений  $\{u\}_1 = \{275, 300, 325\}$  и  $\{u\}_2 = \{375, 400, 425\}$ . Сделаем квадратичное интерполирование значений функции  $J_{\{1\}}(u)$  на множестве  $\{u\}_1$  и  $J_{\{2\}}(u)$  на множестве  $\{u\}_2$ . На рисунке 7.4 показаны полученные интерполяционные параболы. Обе параболы совпадают, поэтому целевой функционал  $J(u)$  в тестовой задаче синтеза можно считать *квадратичным*.

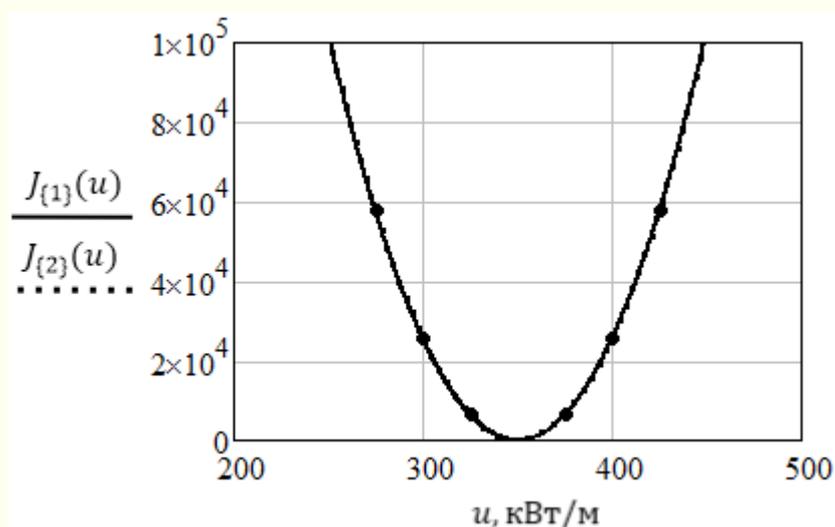
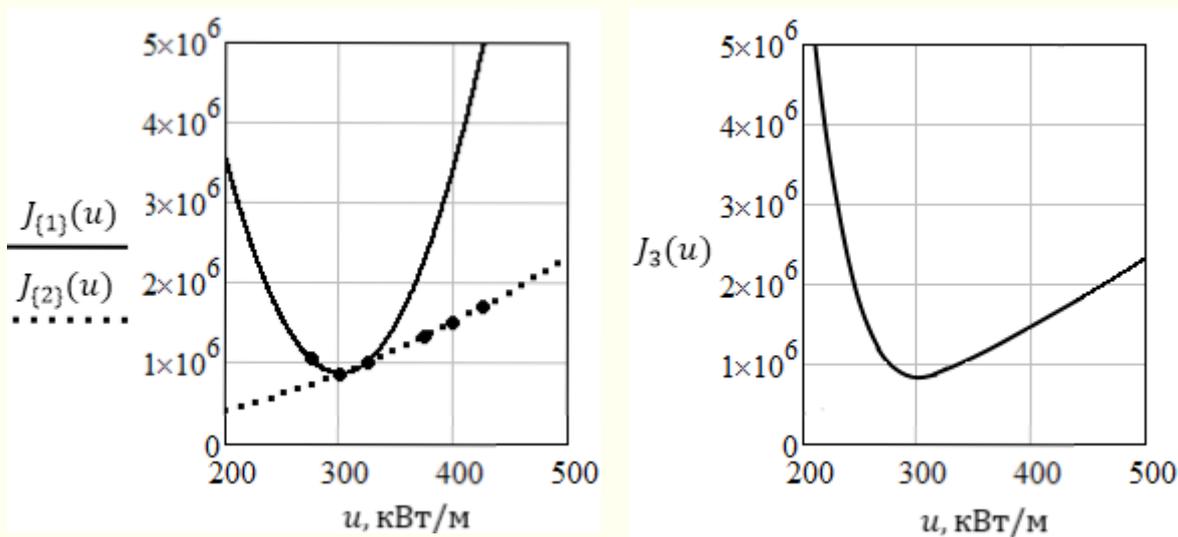


Рисунок 7.4 — Поведение целевого функционала  $J$  в тестовой задаче синтеза оптимального управления

Иная ситуация складывается с функционалом  $J_1$  (7.8) в реальной задаче оптимизации. Он отражает не только термонапряжения в слитке, но и добавляет штраф  $I_F$  при неприемлемо высокой температуре выхода слитка, в частности, более 1223 К. На рисунке 7.5а показаны полученные интерполяционные параболы, построенные по предыдущим множествам. Как и следовало ожидать, они далеки от совпадения друг с другом. Заметим, что точечная кривая  $J_{\{2\}}(u)$  — это поведение функционала  $J_1$  без действия штрафа, а сплошная кривая  $J_{\{1\}}(u)$  — это поведение функционала при активном действии штрафа.

На правом рисунке 7.5б показано значение всего целевого функционала  $J_1$ . Видно, что значение  $J_1$  резко возрастёт при снижении теплоотовода  $u < 300$ , где с появлением неприемлемо высокой температуры выхода слитка начинает действовать внешняя штрафная функция  $I_F$ .



а

б

Рисунок 7.5 — Поведение целевого функционала  $J_1$

Функционалы  $J_2$  и  $J_3$  ведут себя аналогично. Все они — *неквадратичные*. В то же время они всё же — *выпуклые*, как минимум, в рассматриваемой области. Будем далее считать компактом корректности, достаточным для решения поставленных задач оптимизации, диапазон управлений

$$\mathcal{U} = [100, 500] \text{ кВт/м.}$$

В этой области должно задаваться начальное управление  $u^0$  и корректно определяться все значения градиентов  $\nabla J(u^k)$ ,  $k = 1, 2 \dots$

Неквадратичность целевого функционала может накладывать ограничения на применение алгоритмов МРНС с постоянным параметром  $\alpha^k = \alpha^0$ . Но, забегая вперёд отметим, что дальнейшее применение МРНС показало допустимость использования  $\alpha^k = \alpha^0$  в полученных траекториях  $\{u^k\}$  спуска к  $u_*$ . Здесь целевой функционал  $J(u)$  на  $\mathcal{U}$  проявил себя как *локально квадратичный*.

#### 7.4.4. Решение тестовой задачи синтеза термонапряжений

Теперь перейдём к решению тестовой задачи. Зададим, например, синусоидальное тестовое оптимальное управление

$$u_*(z) = 150 + 50 \sin \frac{2\pi z}{z_2 - z_0} \frac{\text{кВт}}{\text{м}}, \quad z_1 < z < z_2. \quad (7.32)$$

После этого рассчитаем термонапряжения в слитке  $\frac{\partial T}{\partial r}$  и установим  $T'_* = \frac{\partial T}{\partial r}$ . Далее будем решать обратную задачу о восстановлении управления (7.32), доставляющего минимум целевому функционалу  $J$  (7.27).

Эту тестовую задачу и последующие прикладные задачи для функционалов  $J_1, J_2, J_3$  будем решать в одинаковых условиях. Рассмотрим бесконечномерные экстремальные алгоритмы МНС (1.8),

МСГ (3.10), МРНСг (3.4), МРНСсг (3.11) с начальным приближением теплоотвода

$$u^0(z) = 400 \frac{\text{кВт}}{\text{м}} \in \mathcal{U}.$$

Заканчивать минимизацию будем по критерию практического прекращения сходимости (1.24):

$$\frac{\|u^k - u^{k-1}\|}{\|u^{k-1}\|} \leq \varepsilon_4, \quad \varepsilon_4 = 10^{-6}.$$

Во всех случаях будем использовать *четвёртую стратегию* выбора шагового множителя  $b^k$  по методу Носедала-Райта с коэффициентом Вольфе:  $c_2 = 0.9$  для МНС;  $c_2 = 0.1$  для МСГ;  $c_2 = 0.5$  для МРНСг и МРНСсг. Максимальный шаг ограничим относительно малой величиной, равной  $0.1\|u^0 - u_*\|_{L_2(S)}$ .

Малый шаг особенно актуален для задач минимизации термонапряжений со штрафной функцией, поскольку здесь неквадратичный  $J(u)$  может резко менять выпуклость (рисунок 7.5б). Приближаться к оптимуму в подобных задачах необходимо очень медленно, иначе может существенно пострадать точность полученного управления из-за раннего прекращения сходимости. Здесь шаговый множитель  $b^k$  раньше времени превращается практически в нулевой, и поэтому сходимость завершается.

Для реализации МРНСг и МРНСсг выберем *шаблон*  $\tilde{u}^0$  в виде изменение управления «под 45°» (3.7) или пропорциональное изменение (3.8). Для выбранного начального приближения  $u^0$  они дают одинаковый результат. Зададим  $\delta = 0.05u^0$  или  $\lambda = 0.05$  и получим параметр регулирования направления спуска

$$\alpha(z) = \left| \frac{0.05u^0}{\nabla J(u^0; z)} \right|.$$

Найденные функции оптимального управления в тестовой задаче всеми методами минимизации оказались практически одинаковыми. На рисунке 7.6 полученные управления отмечены одной пунктирной линией. Точечная линия в начале — это отвод тепла  $u_c$  кристаллизатором, зона которого завершается при  $z_1 = 1$  м.

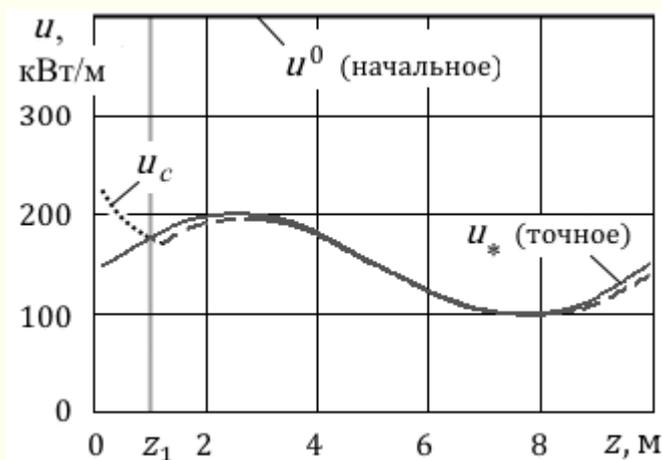


Рисунок 7.6 — Оптимальное управление (пунктирная линия) охлаждением непрерывного слитка для тестового функционала (7.27)

Начальное значение тестового функционала  $J^0 = 4.5 \cdot 10^5$  было уменьшено на четыре порядка. Все расчёты незначительно отличались количеством итераций и находились в диапазоне  $12 \leq k \leq 15$ .

Столь хорошее решение для всех рассмотренных бесконечномерных экстремальных алгоритмов в столь сложной задаче оптимального управления кажется неожиданным. Во всех случаях мы наблюдаем прекрасную *равномерную сходимость*. Этому есть наглядное объяснение.

Для заданного начального приближения управления можно построить *направление на оптимум*

$$d^0(z) = b \cdot (u_*(z) - u^0(z)), \text{ на } S, \quad b > 0 .$$

Такое направление  $d^0$  практически совпадает с направлением градиента, точнее антиградиента, как показано в таблице 7.1 для  $b = 15 \cdot 10^3$ . Можно сказать, что с достаточно хорошей точностью направление  $-\nabla J(u^0; z)$  проходит через оптимум  $u_*$ , и даже самый «слабый» из рассматриваемых методов, такой как МНС, здесь обеспечивает быструю равномерную сходимость к оптимуму. То есть, для удачно выбранного начального приближения (7.32), антиградиент  $-\nabla J(u^0; z)$  с достаточно хорошей точностью представляет собой направление кратчайшего пути к оптимуму.

Таблица 7.1 — Начальное значение градиента  $\nabla J(u^0; z)$  и направление на оптимум  $d^0(z)$

$z, \text{ м}$	1.1	3	5	7	9	9.9
$d^0(z)$	-11.2	-10.1	-11.9	-12.7	-14.6	-13.0,2
$-\nabla J(u^0; z)$	-10.4	-14.4	-17.6	-19.4	-14.9	-6.6

Можно считать, что в этой тестовой задаче нам повезло. Функционал (7.27) оказался практически одинаково чувствителен к управлению во всех точках множества  $S$ . Перейдём к реальной задаче минимизации термонапряжений.

Далее мы исследуем три ранее поставленные задачи ограниченной оптимизации с функционалами  $J_1, J_2, J_3$ , где оптимальные решения будут доставлять *краевой минимум* исходному целевому функционалу  $J$ .

### 7.4.5. Минимизация термонапряжений по функционалу $J_1$

Все задачи минимизации термонапряжений будем решать с теми же исходными данными и на той же пространственной сетке, что и рассмотренная выше тестовая задача. Состояние исходной прямой задачи будем рассчитывать по уже полученному конечноразностному выражению (7.29). Аналогичные выражения сопряжённых задач немного изменятся.

Начнём с минимизации первого функционала  $J_1$  (7.8) при ограничении (7.5) на выходную температуру слитка. Расчёт значения функционала  $J_1$  немного изменится по сравнению с выражением (7.30) тестовой задачи. Здесь  $T'_* = 0$ , а на нижней границе  $\Gamma_2$  появляется член  $I_F|_{\omega_F}$ , поэтому в конечноразностном представлении

$$J_1(u) \approx 2\pi\Delta z \left( \frac{J^{j_c+1} + J^n}{2} + \sum_{j=j_c+2}^{n-1} J^j \right) + \kappa_1 2\pi\Delta r \left( \frac{I_{F0}|_{\omega_F} r_0 + I_{Fm}|_{\omega_F} r_m}{2} + \sum_{i=1}^{m-1} I_{Fi}|_{\omega_F} r_i \right)^{j=n},$$

где штрафная функция  $I_{Fi}^n|_{\omega_F} = (T_i^n - T_{max})^2 \Pi_i^n$  вычисляется с использованием *прямоугольной П-функции*:

$$\Pi_i^n = \begin{cases} 1, & \text{если } T_i^n \geq T_{max}, \\ 0 & \text{иначе.} \end{cases}$$

Выражение  $J^j$  имеет вид:

$$J^j = \Delta r \left( \frac{I_0 r_0 + I_m r_m}{2} + \sum_{i=1}^{m-1} I_i r_i \right)^j, \quad I_i^j = \left[ \left( \frac{\partial T}{\partial r} \right)^2 \Theta \right]_i^j.$$

Сопряжённая задача при функционале  $J_1$  — это уравнение (7.22) с краевыми условиями (7.23). По сравнению с тестовой задачей (7.31) конечно-разностные выражения для расчёта  $f$  также немного изменятся.

Так, внутри расчётной области отличие проявляется в свободном члене при расчёте  $I'_v$ , который мы отдельно выпишем чуть ниже:

$$f_i^{j-1} = f_i^j + \frac{\lambda_i^j \Delta z}{WC_i^j \rho_i^j \Delta r^2} \left[ \left(1 - \frac{\Delta r}{2r_i}\right) f_{i-1}^j - 2f_i^j + \left(1 + \frac{\Delta r}{2r_i}\right) f_{i+1}^j \right] - \frac{(I'_v|_{\omega_1})_i^{j-1} \Delta z}{WC_i^j \rho_i^j} = 0, \quad i = 1 \dots m-1, \quad j = n \dots j_c + 2.$$

На вертикальных границах при  $i = 0$  и  $m$  (справа исчезает  $T'_*$ ) имеем:

$$f_0^{j-1} = f_1^{j-1}; \quad f_m^{j-1} = f_{m-1}^{j-1} - \frac{2}{\lambda_s} (T_m - T_{m-1})^{j-1}.$$

На нижней границе при  $j = n$  получаем новое выражение:

$$f_i^n = -\frac{\kappa_1}{WC_i^n \rho_i^n} (I'_{Fv}|_{\omega_F})_i^n, \quad i = 0 \dots m.$$

Здесь производные

$$(I'_v|_{\omega_1})_i^{j-1} = -2 \left[ \left( \frac{T_{i-1} - 2T_i + T_i}{\Delta r^2} + \frac{1}{r_i} \frac{T_{i+1} - T_{i-1}}{2\Delta r} \right) \Theta_i + \frac{T_{i+1} - T_{i-1}}{2\Delta r} \delta_i \right]^{j-1},$$

$$(I'_{Fv}|_{\omega_F})_i^n = 2(T_i^n - T_{max}) \Pi_i^n.$$

Перейдём к расчётам. Применим все ранее упомянутые методы МНС, МСГ, МРНСГ, МРНСсг. Зададим точность минимизации, глубину шагов по методу Носедала-Райта и регулирование направления спуска (для МРНСГ, МРНСсг) как в предыдущей тестовой задаче.

Оценим состояние слитка при  $\kappa_1 = 0$ , т.е. когда термонапряжения в функционале (7.4) минимизируются без ограничения (7.5) на максимально допустимую температуру  $T_{max}$  выхода слитка из зоны вторичного охлаждения.

На рисунке 7.7 показано полученное оптимальное управление и соответствующая температура на поверхности и в центре слитка. Графики теплоотвода  $u$  и температуры  $T$  на отрезке  $z \in [0, z_1]$  соответствуют зоне кристаллизатора.

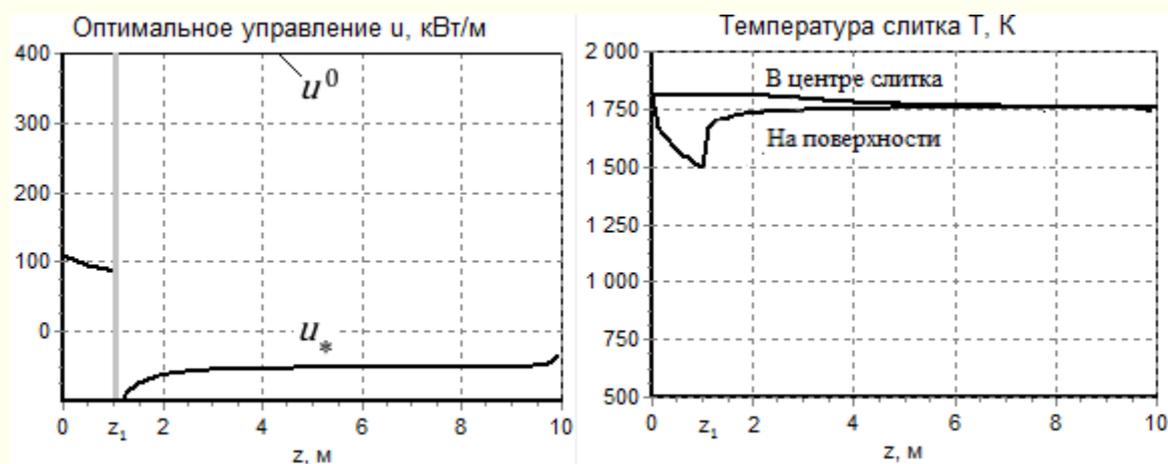


Рисунок 7.7 — МНС, МСГ, МРНСГ, МРНСсг: Оптимальное управление и температура слитка при минимизации  $J_1$  с  $\kappa_1 = 0$

Все методы минимизации потребовали около сотни итераций и уменьшили начальное значение функционала термонапряжений  $J^0 = 15 \cdot 10^5$  на 19 порядков! Мы видим, что в зоне кристаллизатора (первый метр) поверхность слитка заметно охлаждается от температуры заливки  $T_0 = 1811$  К примерно до  $T = 1500$  К. Далее, при  $z > 1$  м, в затвердевшей части слитка обнаруживаются высокие термонапряжения и оптимальное управление вторичного охлаждения начинает вместо охлаждения нагревать слиток, доводя температуру его поверхности до температуры центра слитка. Это правильное оптимальное управление, целью которого является только минимизация термонапряжений без

каких-либо ограничений. Действительно, значения термонапряжений снизились от  $\max_{\omega_1} \left| \frac{\partial T}{\partial r} \right| = 21011 \text{ К/м}$  до  $7396 \text{ К/м}$ , т.е. почти в три раза.

Формально задача минимизации термонапряжений решена: после кристаллизатора надо делать нагревание вместо охлаждения. Однако прикладной ценности такое решение не имеет. Температура выхода слитка (при  $z = 10 \text{ м}$ ) слишком высокая, чуть ниже температуры плавления  $T_s = 1733 \text{ К}$ . При этом глубина жидкой лунки доходила почти до конца слитка и находилась в пределах  $9.6\text{--}9.8 \text{ м}$ . Такое вторичное «охлаждение» не выполняет своё предназначение, слиток не охлаждается.

Включим ограничение (7.5) на температуру выхода слитка и зададим

$$T_{\max} = 1223 \text{ К.}$$

Рассмотрим оптимальное охлаждение слитка со штрафным коэффициентом  $\kappa_1 = 100$  для всех указанных ранее методов. На рисунках 7.8, 7.9 показаны полученные оптимальные управления  $u_*$  и распределения температур для разных методов оптимизации. Полученное решение доставляет *краевой минимум* функционалу  $J_1$ .

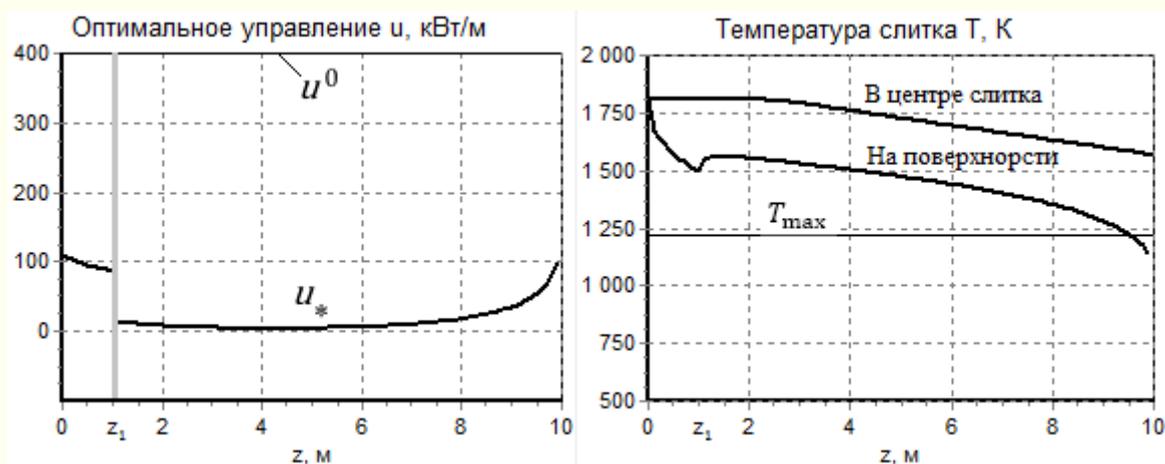


Рисунок 7.8 — МНС, МСГ: Оптимальное управление и температура в слитке при минимизации  $J_1$  с  $\kappa_1 = 100$

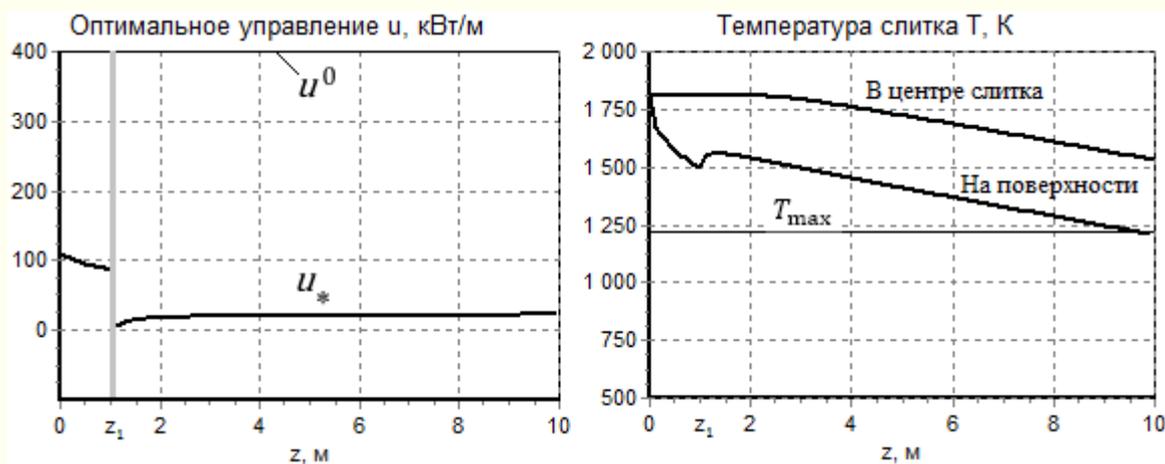


Рисунок 7.9 — МРНСг, МРНСсг: Оптимальное управление и температура для в слитке при минимизации  $J_1$  с  $\kappa_1 = 100$

В таблице 7.2 приведены значения некоторых характеристик результатов оптимизации, а именно: количества итераций  $k$  и вычислений функционала с градиентом до достижения заданной точности; конечное значение критерия (7.4) термонапряжений; максимальное значение полученных термонапряжений; глубина жидкой лунки.

Таблица 7.2 — Результаты оптимизации при  $\kappa_1 = 100$ 

Метод	Итераций $k$ / вычислений $J$ и $\nabla J$	Конечное $J^k \cdot 10^5$	$\max_{\omega_1} \left  \frac{\partial T}{\partial r} \right $	Глубина лунки, м
МНС	75/788	2.67	12979	4.9
МСГ	75/788	2.67	12979	4.9
МРНСГ	71/420	1.45	7396	4.7
МРНСсГ	72/508	1.45	7396	4.7

Мы видим, что МРНСГ и МРНСсГ продемонстрировали более высокую эффективность по сравнению с традиционными МНС и МСГ. МРНС потребовали в два раза меньше вычислений, лучше минимизировали критерий термонапряжений  $J$ , снизили максимальное значение локальных термонапряжений и, как оказалось, уменьшили глубину жидкой лунки.

Обсудим особенность, связанную с выбором шагового множителя в методе МРНСГ. Сменим четвертую стратегию выбора шагового множителя  $b^k$  по методу Носедала-Райта на *вторую стратегию* с адаптивной релаксацией по методу (1.14). Данная стратегия реализуется на много проще четвертой. Поэтому, если у вас нет готовых программных кодов (библиотеки методов линейного поиска), то, не сильно напрягаясь, вы можете реализовать метод (1.14).

Зададим параметры метода  $b_1 = 1.01$ ,  $b_2 = 0.1$ , которые могут обеспечить осторожное приближение к экстремуму. То есть не сильно ускорять сходимость при монотонном убывании функционала и быстро возвращаться назад для дальнейших малых шагов в случае нарушения монотонности. В таблице 7.3 приведены результаты оптимизации для МРНСГ.

Таблица 7.3 — Результаты оптимизации при  $\kappa_1 = 100$  с адаптивной релаксацией шагов

Метод	Итераций $k$ / вычислений $J$ и $\nabla J$	Конечное $J^k \cdot 10^5$	$\max_{\omega_1} \left  \frac{\partial T}{\partial r} \right $	Глубина лунки
МРНСГ	22/47	1.35	7396	4.8

Сравните результаты таблицы 7.3 с соответствующей строкой МРНСГ таблицы 7.2. Мы получили практически то же самое оптимальное управление, но с существенно меньшим количеством вычислений, на порядок меньшим! Адаптивный метод (1.14) продемонстрировал здесь высокую эффективность. С учётом всех результатов таблицы 7.2, мы далее будем минимизировать функционалы только посредством МРНСГ с адаптивной релаксацией шагов.

Обратите внимание на рисунки 7.8 и 7.9, во всех случаях ограничение на температуру выхода слитка удовлетворялось только на его поверхности. Внутри слитка ограничение  $T_{\max} = 1223$  К заметно нарушалось. Следовательно, необходимо увеличить штрафной коэффициент  $\kappa_1$ .

Зададим  $\kappa_1 = 10^6$ . Результаты такой оптимизации представлены на рисунке 7.10 и в таблице 7.4.

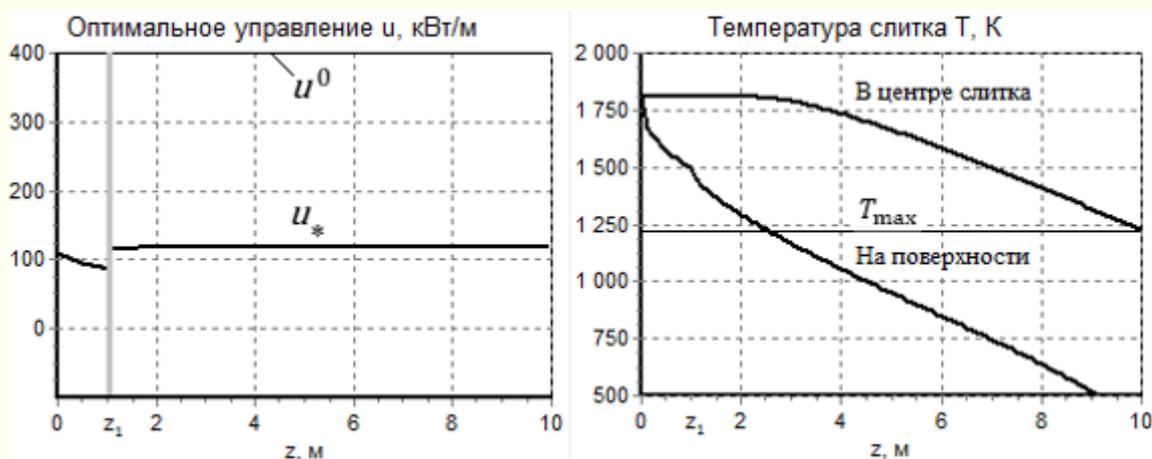


Рисунок 7.10 — МРНСГ: Оптимальное управление и температура в слитке при минимизации  $J_1$  с  $\kappa_1 = 10^6$

Мы видим, теперь температура выхода слитка из вторичного охладителя (при  $z = 10$  м) удовлетворяет желаемому ограничению (7.5):  $T|_{\Gamma_2} \leq T_{\max} = 1223$  К. При этом, максимальные термонапряжения в слитке значительно увеличились по сравнению со «слабым» штрафом при  $\kappa_1 = 100$  (сравните таблицы 7.2 и 7.4). Заметно уменьшилась глубина жидкой лунки, что и следовало ожидать.

Таблица 7.4 — Результаты оптимизации при  $\kappa_1 = 10^6$

Метод	Итераций $k$ / вычислений $J$ и $\nabla J$	Конечное $J^k \cdot 10^5$	$\max_{\omega_1} \left  \frac{\partial T}{\partial r} \right $	Глубина лунки
МРНСГ	7/17	7.16	14563	4.1

Полученные функции оптимального управления являются физически обоснованными и удовлетворяют требованиям поставленной задачи минимизации термонапряжений с ограничением на выходную температуру слитка.

### 7.4.6. Минимизация термонапряжений по функционалу $J_2$

Теперь рассмотрим минимизацию второго функционала (7.9) с ограничением на объём жидкой лунки. Здесь выражение  $J_2$  в зоне вторичного охлаждения, при интегрировании методом трапеций по  $z$ , имеет вид:

$$J_2(u) \approx 2\pi\Delta z \left( \frac{j_{c+1} + j^n}{2} + \sum_{j=j_c+2}^{n-1} j^j \right),$$

Слагаемые  $j^j$ , при интегрировании по  $r$ , приобретают вид:

$$j^j = \Delta r \left( \frac{I_0^j r_0 + I_m^j r_m}{2} + \sum_{i=1}^{m-1} I_i^j r_i \right),$$

где

$$I_i^j = \left[ \left( \frac{\partial T}{\partial r} \right)^2 \Theta \right]_i^j + \kappa_2 \Theta (T_i^j - T_s) = \begin{cases} \kappa_2, & \text{если } \zeta_i^j < 0.95, \\ \left[ \left( \frac{\partial T}{\partial r} \right)^2 \right]_i^j, & \text{если } \zeta_i^j \geq 0.95. \end{cases}$$

Сопряжённая задача при функционале  $J_2$  в конечно-разностном представлении имеет вид:

$$f_i^{j-1} = f_i^j + \frac{\lambda_i^j \Delta z}{WC_i^j \rho_i^j \Delta r^2} \left[ \left( 1 - \frac{\Delta r}{2r_i} \right) f_{i-1}^j - 2f_i^j + \left( 1 + \frac{\Delta r}{2r_i} \right) f_{i+1}^j \right] - \frac{\left[ I'_v |_{\omega_1} + \kappa_2 (1 |_{\omega_2})'_v \right]_i^{j-1} \Delta z}{WC_i^j \rho_i^j} = 0,$$

$$i = 1 \dots m - 1, \quad j = n \dots j_c + 2.$$

На вертикальных границах при  $i = 0$  и  $m$  имеем:

$$f_0^{j-1} = f_1^{j-1}; \quad f_m^{j-1} = f_{m-1}^{j-1} - \frac{2}{\lambda_s} (T_m - T_{m-1})^{j-1}.$$

На нижней границе при  $j = n$ :

$$f_i^n = 0, \quad i = 0 \dots m.$$

Здесь производные

$$\begin{aligned} (I'_v|_{\omega_1})_i^{j-1} &= -2 \left[ \left( \frac{T_{i-1} - 2T_i + T_i}{\Delta r^2} + \frac{1}{r_i} \frac{T_{i+1} - T_{i-1}}{2\Delta r} \right) \Theta_i \right. \\ &\quad \left. + \frac{T_{i+1} - T_{i-1}}{2\Delta r} \delta_i \right]^{j-1}, \\ [(1|_{\omega_2})'_v]_i^{j-1} &= \delta_i^{j-1}. \end{aligned}$$

Результаты минимизации термонапряжений и объёма жидкой лунки со штрафным коэффициентом  $\kappa_2 = 2 \cdot 10^6$  представлены на рисунке 7.11 и в таблице 7.5. Полученное решение доставляет *краевой минимум* функционалу  $J_2$ .

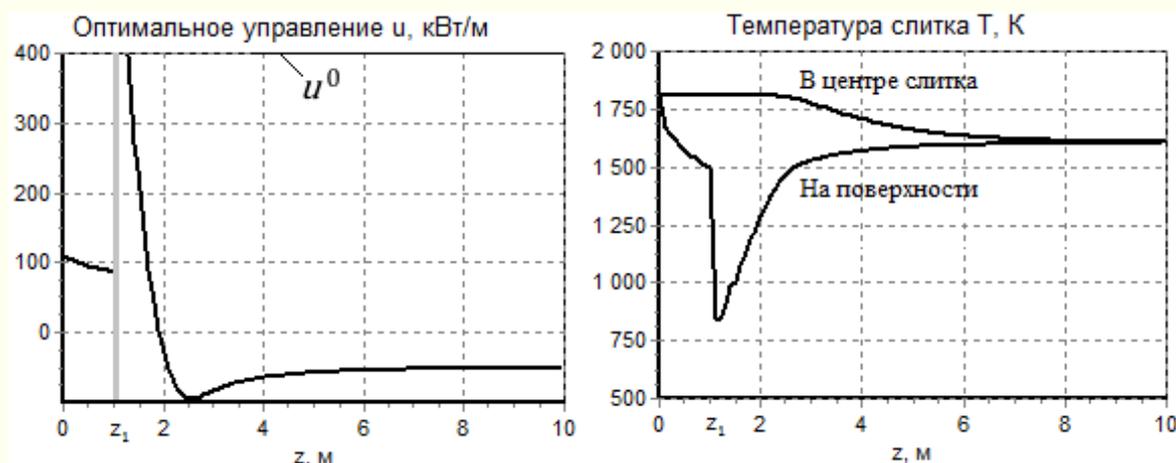


Рисунок 7.11 — МРНСГ: Оптимальное управление и температура в слитке при минимизации  $J_2$  с  $\kappa_2 = 2 \cdot 10^6$

Из левого рисунка мы видим, что для минимизации лунки необходимо сразу за кристаллизатором значительно увеличивать теплоотвод во вторичном охладителе. Высокий всплеск управления теплоотводом сопровождается на правом рисунке резким падением температуры поверхности слитка. Такое охлаждение слитка приводит к

существенному росту термонапряжений ( $\max_{\omega_1} \left| \frac{\partial T}{\partial r} \right| = 75290$  К/м по сравнению с начальным значением 21011 К/м). Далее по длине слитка начинает существенно сказываться требование минимизации термонапряжений, и вторичное охлаждение превращается в нагрев.

Таблица 7.5 — Результаты оптимизации при  $\kappa_2 = 2 \cdot 10^6$

Метод	Итераций $k$ / вычислений $J$ и $\nabla J$	Конечное $J^k \cdot$ $10^5$	$\max_{\omega_1} \left  \frac{\partial T}{\partial r} \right $	Глубина лунки
МРНСГ	92/187	0.00029	75290	3.7

Из таблицы видно, что глубина жидкой лунки действительно уменьшилась примерно на 1 м по сравнению с результатами минимизации предыдущего функционала  $J_1$ . Однако этот «успех» сопровождается значительными термонапряжениями. Полученное оптимальное управление и режим охлаждения имеют физический смысл, но вряд ли имеют технологическую ценность.

#### 7.4.7. Минимизация термонапряжений по функционалу $J_3$

В третьей задаче минимизации термонапряжений с ограничением выходной температуры и с минимизацией объёма жидкой лунки будем использовать выражения обеих предыдущих сопряжённых задач при ненулевых значениях  $\kappa_1$  и  $\kappa_2$ . В частности, зададим, как и ранее,  $\kappa_1 = 10^6$ ,  $\kappa_2 = 2 \cdot 10^6$ .

Результаты минимизации термонапряжений в этих условиях для  $J_3$  посредством МРНСГ представлены на рисунке 7.12 и в таблице 7.6. Для наглядности в таблице указаны результаты и первой задачи минимизации  $J_1$  с ограничением только на температуру.

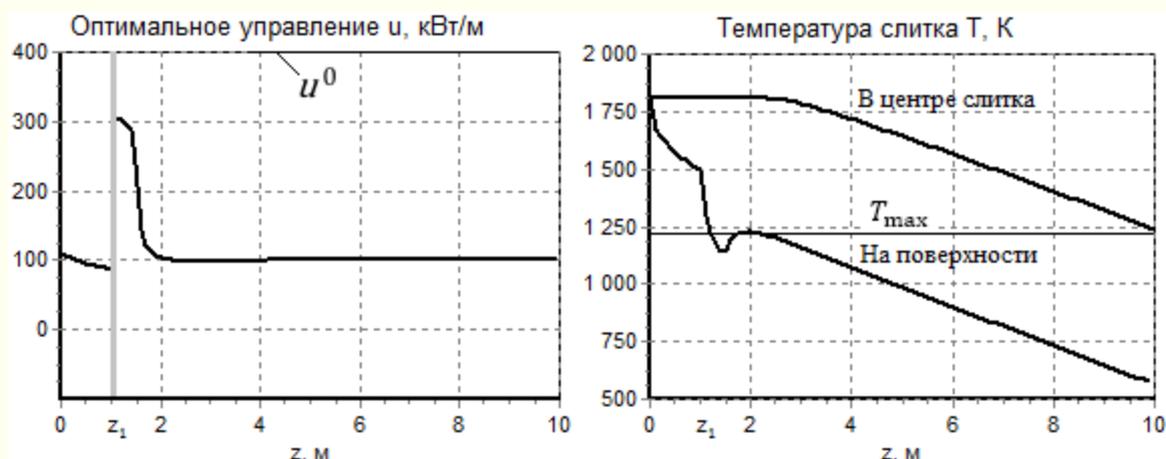


Рисунок 7.12 — МРНСГ: Оптимальное управление и температура в слитке при минимизации  $J_3$  с  $\kappa_1 = 10^6$ ,  $\kappa_2 = 2 \cdot 10^6$

Таблица 7.6 — Минимизация критерия термонапряжений  $J$  (7.4) при разных целевых функционалах посредством МРНСГ

Целевой функционал	Итераций $k$ / вычислений $J$ и $\nabla J$	Конечное $J \cdot 10^5$	$\max_{\omega_1} \left  \frac{\partial T}{\partial r} \right $	Глубина лунки
$J_1$	7/17	7.1	14453	4.1
$J_3$	9/21	5.81	30846	3.8

Из рисунка и таблицы мы видим, что, по сравнению с первой задачей  $J_1$ , в последней задаче с  $J_3$  выполняется ограничение на температуру выхода слитка, минимизируется жидкая лунка, минимизируется интегральное значение термонапряжений  $J$ . Однако возникающие здесь локальные значения термонапряжений примерно в два раза выше, чем в первой задаче. Полученный всплеск управления теплоотвода вблизи кристаллизатора можно считать вполне технологичным по сравнению с аналогичным всплеском оптимального управления во второй задаче с  $J_2$  (рисунок 7.11).

Окончательное решение о выборе необходимой задачи оптимизации и практической реализации режимов охлаждения непрерывного слитка могут принимать только специалисты по проектированию и эксплуатации машин непрерывного литья заготовок.

## 8 Идентификация параметров открытых русел

### 8.1 Постановка задачи

Математическое моделирование в гидродинамике, гидравлике открытых потоков применяется для прогнозирования речных течений и паводков, для проектирования различных водохозяйственных объектов, для автоматического управления и оптимизации режимов работы гидротехнических сооружений.

В настоящее время общепринятой математической моделью, используемой при исследовании нестационарных открытых потоков, являются системы уравнений Буссинеска [54, 96] и Сен-Венана [38, 80]. Формально они следуют из общих уравнений движения гидродинамики Навье-Стокса в предположении, что глубина потока мала по сравнению с длиной волны (приближение мелкой воды). Более общими являются уравнения Буссинеска, они учитывают внутренние волны. Если рассматривать только поверхностные волны, т.е. пренебречь вертикальным ускорением жидкости, то получаются уравнения Сен-Венана.

При решении задач водного хозяйства, применительно к рекам и каналам, целесообразно использовать одномерную (квазиодномерную) нестационарную модель потока в виде *квазилинейной, гиперболической, неоднородной* системы уравнений Сен-Венана [68, 54].

Соответствующая *исходная задача* будет иметь вид:

$$\begin{aligned} \frac{\partial Q}{\partial t} + 2w \frac{\partial Q}{\partial x} + B(c^2 - w^2) \frac{\partial Z}{\partial x} + F_{\text{тр}} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z &= 0, \\ \frac{\partial Z}{\partial t} + \frac{1}{B} \frac{\partial Q}{\partial x} - \frac{q}{B} &= 0, \quad (x, t) \stackrel{\text{def}}{=} \tau \in \Omega = (x_a, x_b) \times (t_0, t_1), \end{aligned} \quad (8.1)$$

где  $t$  — время,  $x$  — пространственная координата вдоль потока воды в русле (рисунок 8.1),  $Q(x, t)$  — расход воды в потоке,  $Z(x, t)$  — уровень

свободной поверхности потока относительно заданного горизонта ( $Z_{\text{дно}}$  — уровень дна русла),  $w = Q/\omega$  — скорость потока, осреднённая по живому поперечному сечению,  $\omega(x, Z)$  — площадь живого сечения,  $B(x, Z) = \frac{\partial \omega}{\partial Z}$  — ширина потока поверху,  $c = \sqrt{g\omega/B}$  — скорость распространения малых возмущений в потоке, равная скорости звука в воде,  $g$  — ускорение свободного падения,  $q(x, t)$  — боковой распределённый приток-отток.

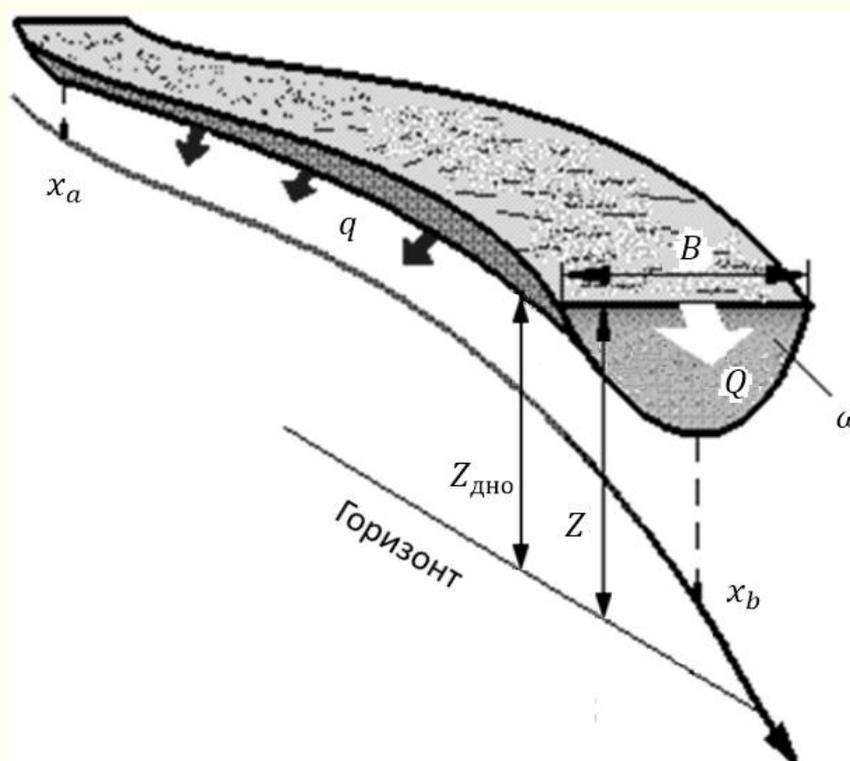


Рисунок 8.1 — Участок русла

Член  $\frac{\partial \omega}{\partial x} \Big|_Z$  определяется при фиксированном  $Z$  и характеризует уклон дна русла и неоднородность потока. Дело в том, что в речном русле поток далеко не всегда является строго одномерным. В этом случае течение считают *квазиодномерными*, что в рамках одномерной модели (8.1) реализуется появлением указанного члена.

Член трения

$$F_{\text{тр}} = \frac{gQ|Q|}{\omega RC^2},$$

где  $R(x, Z) = \omega/\chi$  — гидравлический радиус потока,  $\chi(x, Z)$  — смоченный поперечный периметр русла. Эмпирический коэффициент Шези принят по формуле Маннинга:

$$C(x, Z) = \frac{1}{n} R^{1/6},$$

где  $n(x)$  — коэффициент шероховатости, характеризующий вязкое трение воды о ложе русла, т.е. коэффициент, определяющий *диссипацию энергии* в пристеночной зоне потока.

Мы видим, что уравнения (8.1) неоднородные. Они содержат свободные члены в виде вязкого трения  $F_{\text{тр}}$  и члены, влияющие на одномерность потока  $\left. \frac{\partial \omega}{\partial x} \right|_Z$ ,  $\frac{q}{B}$ . Последние могут превращать квазиодномерный поток в *течение в диффузоре* или в *конфузоре*. В первом случае, для *дозвукового течения* воды (докритическое течение), т.е.  $w < c$ , будет происходить усиленная диссипация энергии (в модели — за счёт увеличения смоченного периметра  $\chi$  в  $F_{\text{тр}}$ , а на объекте — дополнительно за счёт увеличения турбулентности потока), приводящая к затуханиям волновых процессов и размытию их фронтов. Во втором случае скорость потока может стать *сверхзвуковой*, т.е.  $w > c$ , что является недопустимым для рассматриваемой задачи.

Возможности использования математической модели (8.1) на практике определяются точностью задания входящих в неё параметров, отражающих индивидуальные свойства каждого конкретного стока, к которым относятся: форма русла  $\omega$ ,  $B$ ,  $\chi$ ; шероховатость  $n$ ; боковой приток  $q$ . Значения всех этих параметров меняются со временем, что связано с зарастанием русел, заилением, размывом и т.п. Если форму

русла и боковой приток ещё можно периодически измерять, то шероховатость прямым измерениям не поддаётся. Поэтому далее мы рассмотрим задачу идентификации распределённого коэффициента шероховатости [92, 90, 34, 3].

Уточним терминологию и обозначения поставленной задачи согласно предыдущим разделам теории прямого экстремального подхода.

Состояние потока в системе (8.1) — это *вектор-функция*  $v = (Q, Z) \in V = L_{2+}^m(\bar{\Omega})$ ,  $m = 2$ , т.е., расход  $Q$  и уровень потока  $Z$  не могут быть отрицательными. Шероховатость  $n$  — это управление  $u \in U \subset L_{2+}(S)$ , где множество  $S = (x_a, x_b)$ . Можно сказать, что  $S = P_x(\Omega)$ , т.е.  $S$  является проекцией прямоугольной области  $\Omega$  на пространственную ось  $x$ .

Принадлежность состояния и управления положительным полупространствам не являются ограничениями в задаче идентификации, а всего лишь условием разумности решения исходной системы (8.1). Множество разумных (допустимых) управлений  $U$  мы обсудим позже при анализе идентифицируемости и оценке выпуклости целевого функционала, что поможет нам выявить компакт корректности  $\mathcal{U}$  для решения обратной задачи идентификации.

Запишем исходную задачу в операторном виде, аналогичном (3.20):

$$\mathbb{D}v + F = \frac{\partial v}{\partial t} + A \frac{\partial v}{\partial x} + F = 0 \quad \text{на } \Omega, \quad (8.2)$$

где дифференциальный оператор  $\mathbb{D} \cdot = \frac{\partial \cdot}{\partial t} + A \frac{\partial \cdot}{\partial x}$ , его матрица  $A$  и свободный член  $F$  (вектор-столбец) имеют вид:

$$A(x, v) = \begin{pmatrix} 2w & B(c^2 - w^2) \\ 1/B & 0 \end{pmatrix},$$

$$F(x, v, u) = \left( F_{\text{тр}} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z, -q/B \right).$$

Оператор  $\mathbb{D}$  зависит от  $x$  (через параметры  $B$ ,  $\omega$  в матрице  $A$ ) и от состояния  $v$ . Управление  $u$  присутствует в свободном члене  $F$ .

На границе  $\Gamma_{\partial} = \Gamma_a \cup \Gamma_b$ , т.е. концах стока, имеются граничные условия *первого рода*:

$$\begin{aligned} Z(x_a, t) &= Z_a(t) \text{ на } \Gamma_a = x_a \times (t_0, t_1), \\ Q(x_b, t) &= Q_b(t) \text{ на } \Gamma_b = x_b \times (t_0, t_1). \end{aligned} \quad (8.3)$$

Начальное условие:

$$v(x, t_0) = (Q_0(x), Z_0(x)) \text{ на } \Gamma_0 = [x_a, x_b] \times t_0. \quad (8.4)$$

Граничные области  $\Gamma_a$ ,  $\Gamma_b$ , а также начальная  $\Gamma_0$  и терминальная  $\Gamma_1 = [x_a, x_b] \times t_1$  образуют границу  $\partial\Omega$  прямоугольника  $\Omega$ .

Идентификацию модели (8.1), или в форме (8.2), будем проводить по экспериментальным наблюдениям за уровнем воды  $Z_e$  в нижнем створе стока при  $x_b$ . То есть, *множество наблюдений состояния объекта*  $\omega = \Gamma_b$ . *Измерение* именно уровня, а не расхода воды, объясняется, во-первых, апостериорно известной высокой чувствительностью  $Z$  к шероховатости  $n$ , во-вторых, достаточно малой погрешностью наблюдений уровня, в-третьих, простотой и не дороговизной эксперимента.

Расхождение модельного и наблюдаемого состояния объекта, согласно (4.1), будем оценивать критерием качества идентификации по *невязке уровней*:

$$J(u) = \|M(v) - Z_e\|_{L_{2+}^{me}(\omega)}^2 = \|Z|_{\omega} - Z_e\|_{L_{2+}^{me}(\omega)}^2,$$

где  $M$  — это *оператор наблюдений* модельного состояния  $v$ , сужающий  $v$  до одной его компоненты — уровня  $Z$  на множестве  $\omega = \Gamma_b$ .

Экспериментально наблюдаемое состояние объекта  $Z_e \in V_e = L_{2+}^{m_e}(\omega)$ , где  $m_e = 1$  — количество наблюдаемых компонент состояния потока.

Запишем критерий качества  $J(u)$  в виде целевого функционала (4.4):

$$J(u) = \int_{t_0}^{t_1} I(v)|_{\omega} dt, \quad I(v; x, t)|_{\omega=\Gamma_b} = (Z(x, t)|_{\Gamma_b} - Z_e(t))^2. \quad (8.5)$$

Задача идентификации формулируется следующим образом. Для распределённой системы (8.2) с краевыми условиями (8.3)–(8.4) необходимо найти параметр (оптимальное управление)  $u(x)$  в уравнениях (8.2), который доставляет минимум функционалу (8.5).

### 8.1.1. Характеристическая форма исходных уравнений

Как и всякая гиперболическая система, система (8.2) имеет характеристические направления  $\lambda_{1,2}$  описывающие волновые процессы в русле. Поскольку они имеют принципиальное значение для анализа состояния системы Сен-Венана и идентифицируемости, давайте их найдём.

Характеристические направления  $\lambda_{1,2}$  являются *собственными числами* матрицы  $A$ . Найдём их, воспользовавшись уравнением (3.6):

$$\begin{vmatrix} 2w - \lambda & B(c^2 - w^2) \\ 1/B & -\lambda \end{vmatrix} = \lambda^2 - 2w\lambda - (c^2 - w^2) = 0.$$

Откуда следует

$$\lambda_{1,2} = \frac{w \pm \sqrt{w^2 + (c^2 - w^2)}}{1} = w \pm c.$$

Это означает, что *характеристики*  $\xi_1$  и  $\xi_2$  в прямоугольнике  $\Omega = (x_a, x_b) \times (t_0, t_1)$  представляют собой семейство двух типов линий  $\xi_1 = x_1(t)$  и  $\xi_2 = x_2(t)$  вдоль которых распространяются волны в русле со скоростью

$$\lambda_1 = \frac{dx_1}{dt} = w + c \text{ вдоль } \xi_1, \quad \lambda_2 = \frac{dx_2}{dt} = w - c \text{ вдоль } \xi_2.$$

Вдоль характеристических направлений  $\lambda_{1,2}$  гиперболическая система (8.2) принимает характеристическую форму вида (3.40). Эта форма уравнений нам в дальнейшем понадобится. Для её получения надо исходную систему записать на *собственных векторах*  $\Lambda_{1,2}$  матрицы  $A$ .

Найдём первый собственный вектор  $\Lambda_1$  для  $\lambda_1 = w + c$ . В соответствии с (3.39)

$$\begin{aligned} \Lambda_1^T \begin{pmatrix} 2w - \lambda_1 & B(c^2 - w^2) \\ 1/B & -\lambda_1 \end{pmatrix} &= (\Lambda_{11}, \Lambda_{12})^T \begin{pmatrix} w - c & B(c^2 - w^2) \\ 1/B & -(w + c) \end{pmatrix} \\ &= (0, 0). \end{aligned}$$

Получаем

$$\Lambda_{11}(w - c) + \Lambda_{12}/B = 0, \quad \Lambda_{11}B(c^2 - w^2) - \Lambda_{12}(w + c) = 0.$$

Пусть  $\Lambda_{11} = 1$ , тогда  $\Lambda_{12} = -B(w - c)$ .

Найдём второй собственный вектор  $\Lambda_2$  для  $\lambda_2 = w - c$ :

$$\begin{aligned} \Lambda_2^T \begin{pmatrix} 2w - \lambda_2 & B(c^2 - w^2) \\ 1/B & -\lambda_2 \end{pmatrix} &= (\Lambda_{21}, \Lambda_{22})^T \begin{pmatrix} w + c & B(c^2 - w^2) \\ 1/B & -(w - c) \end{pmatrix} \\ &= (0, 0). \end{aligned}$$

Получаем

$$\Lambda_{21}(w + c) + \Lambda_{22}/B = 0, \quad \Lambda_{21}B(c^2 - w^2) - \Lambda_{22}(w - c) = 0.$$

Пусть  $\Lambda_{21} = 1$ , тогда  $\Lambda_{22} = -B(w + c)$ .

Теперь умножим исходную систему (8.2) сначала на первый вектор  $\Lambda_1^T$ , а потом повторим умножение системы на второй вектор  $\Lambda_2^T$ . При этом вдоль двух семейств характеристик  $\xi_1$  и  $\xi_2$  получим два уравнения типа (3.18) в виде:

$$\begin{aligned} \Lambda_1^T \left( \frac{\partial v}{\partial t} + \lambda_1 \frac{\partial v}{\partial x} \right) + \Lambda_1^T F \\ = \frac{\partial Q}{\partial t} - B(w - c) \frac{\partial Z}{\partial t} + (w + c) \frac{\partial Q}{\partial x} \\ - B(w - c)(w + c) \frac{\partial Z}{\partial x} + F_{fr} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z + (w - c)q = 0, \end{aligned}$$

$$\begin{aligned} \Lambda_2^T \left( \frac{\partial v}{\partial t} + \lambda_2 \frac{\partial v}{\partial x} \right) + \Lambda_2^T F \\ = \frac{\partial Q}{\partial t} - B(w + c) \frac{\partial Z}{\partial t} + (w - c) \frac{\partial Q}{\partial x} \\ - B(w + c)(w - c) \frac{\partial Z}{\partial x} + F_{fr} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z + (w + c)q = 0. \end{aligned}$$

В итоге получаем *характеристическую форму* уравнений Сен-Венана:

$$\begin{aligned} \frac{\partial Q}{\partial t} + (w \pm c) \frac{\partial Q}{\partial x} - B(w \mp c) \frac{\partial Z}{\partial t} + B(c^2 - w^2) \frac{\partial Z}{\partial x} + F_{тр} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z \\ + (w \mp c)q = 0, \end{aligned}$$

где верхний знак соответствует характеристикам первого семейства  $\xi_1$ , а нижний — характеристикам второго семейства  $\xi_2$ .

Если учесть, что вдоль  $\xi_1$  и  $\xi_2$  оператор  $\frac{\partial \cdot}{\partial t} + (w \pm c) \frac{\partial \cdot}{\partial x} = \frac{d \cdot}{dt} \Big|_{\xi_{1,2}}$ , то полученные уравнения в частных производных можно представить на характеристиках как уравнения в обыкновенных дифференциалах:

$$\frac{dQ}{dt} - B(w \mp c) \frac{dZ}{dt} + F_{тр} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z + (w \mp c)q = 0. \quad (8.6)$$

Именно в этой форме исходные уравнения мы далее будем решать численно с использованием конечно-разностных аппроксимаций производных вдоль характеристик.

## 8.2 Градиент целевого функционала

Поскольку ограничения на оптимизацию в поставленной задаче отсутствуют, то перейдём сразу ко второму шагу преобразований раздела 3.4.

**2. Линеаризуем задачу.** Вариация коэффициента шероховатости  $\delta u$  приводит к вариации состояния  $\delta v = (\delta Q, \delta Z)$ . При этом вариация непосредственно целевого функционала (8.5), по аналогии с (3.22), принимает вид:

$$\delta J(u) = \langle I'_v, \delta v \rangle_{V^*(\omega)} = \int_{t_0}^{t_1} I'_Z|_{\Gamma_b} \delta Z dt, \quad (8.7)$$

где сопряжённое пространство  $V^* = L_2^m$ , вектор-строка  $I'_v = (I'_Q, I'_Z)^T|_{\Gamma_b}$  имеет две компоненты, в которых первая при вариации  $\delta Q$  равна нулю, а вторая  $I'_Z|_{\Gamma_b} = 2(Z|_{\Gamma_b} - Z_e)$ . То есть  $I'_v = (0, 2(Z|_{\Gamma_b} - Z_e))^T$ .

Вариация уравнения (8.2) в операторном виде приводит к выражению типа (3.23):

$$\delta \mathbb{D}v = \mathbb{V}\delta v + \mathbb{U}\delta u = 0 \in V(\Omega),$$

или в явном виде:

$$\frac{\partial \delta v}{\partial t} + A \frac{\partial \delta v}{\partial x} + \frac{\partial A}{\partial v} \delta v \frac{\partial v}{\partial x} + F'_v \delta v + F'_u \delta u = 0 \quad \text{на } \Omega. \quad (8.8)$$

Получаем следующие операторы:

$$\mathbb{V} \cdot = \frac{\partial \cdot}{\partial t} + A \frac{\partial \cdot}{\partial x} + \mathbb{C}_1 \cdot + F'_v \cdot : V(\Omega) \rightarrow V(\Omega),$$

$$\mathbb{C}_1 \cdot = \frac{\partial A}{\partial v} \cdot \frac{\partial v}{\partial x},$$

$$\mathbb{U} \cdot = F'_u \cdot : U(S) \rightarrow V(\Omega).$$

Заметим, что оператор  $\mathbb{U}$  отображает вариацию управления  $\delta u$  из области определения управления  $S$  в область  $\Omega$ .

Появился оператор  $\mathbb{C}_1$ , как следствие зависимости матрицы  $A$  от состояния системы  $v$ . Такая зависимость рассматривалась нами ранее в теоретическом разделе 3.4.2, там фигурировал оператор  $\mathbb{C}_r$ . В нашем случае  $\mathbb{C}_r = \mathbb{C}_1$  поскольку содержит первую ( $r = 1$ ) производную  $\frac{\partial v}{\partial x}$ . Далее мы используем полученные ранее результаты для  $\mathbb{C}_1$ .

**3. Отобразим** линеаризованную систему уравнений (8.8) в сопряжённое пространства. Возьмём произвольный линейный функционал в виде двухкомпонентного вектора  $\tilde{f} = \{\tilde{f}_1, \tilde{f}_2\}$  из пространства  $V^*(\bar{\Omega}) = L_2^2(\bar{\Omega})$  сопряжённого с пространством  $V(\bar{\Omega})$ . Согласно выражению (3.24), запишем его значение на вариации системы (8.8):

$$\langle \tilde{f}, \delta(\mathbb{D}v + F) \rangle_{V^*(\Omega)} = \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} + \langle \tilde{f}, \mathbb{U}\delta u \rangle_{V^*(\Omega)} = 0.$$

Преобразуем данное выражение к значениям соответствующих линейных функционалов непосредственно на элементах  $\delta v$  и  $\delta u$ . Для системы с первыми производными и по времени, и по пространству получим тождество Лагранжа типа (3.25) в виде:

$$\begin{aligned} \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} + \langle \tilde{f}, \mathbb{U}\delta u \rangle_{V^*(\Omega)} &\equiv \langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\Omega)} + \langle \bar{\mathbb{V}}^* \tilde{f}, \delta v \rangle_{V^*(\partial\Omega)} \\ &+ \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)} = 0. \end{aligned} \quad (8.9)$$

Теперь запишем данное тождество в явном виде и найдём выражение для сопряжённых операторов  $\mathbb{V}^*$ ,  $\mathbb{U}^*$ .

Начнём с преобразований при  $\delta v$ .

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{V} \delta v \rangle_{V^*(\Omega)} \\
&= \iint_{\Omega} \left[ \frac{\partial \tilde{f}^T \delta v}{\partial t} - \frac{\partial \tilde{f}^T}{\partial t} \delta v \right] dx dt \\
&+ \iint_{\Omega} \left[ \frac{\partial \tilde{f}^T A \delta v}{\partial x} - \frac{\partial \tilde{f}^T}{\partial x} A \delta v - \tilde{f}^T \frac{\partial A}{\partial x} \delta v \right] dx dt \\
&+ \iint_{\Omega} \tilde{f}^T (\mathbb{C}_1 + F'_v) \delta v dx dt \\
&= \int_{x_a}^{x_b} \tilde{f}^T \delta v \Big|_{t_0}^{t_1} dx - \iint_{\Omega} \frac{\partial \tilde{f}^T}{\partial t} \delta v dx dt \\
&+ \int_{t_0}^{t_1} \tilde{f}^T A \delta v \Big|_{x_a}^{x_b} dt - \iint_{\Omega} \frac{\partial \tilde{f}^T}{\partial x} A \delta v dx dt \\
&+ \iint_{\Omega} \tilde{f}^T \left( -\frac{\partial A}{\partial x} + \mathbb{C}_1 + F'_v \right) \delta v dx dt.
\end{aligned}$$

В итоге тождество Лагранжа при  $\delta v$ , с учётом необходимости транспонирования векторов и матриц при их перестановке, принимает вид:

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{V} \delta v \rangle_{V^*(\Omega)} \\
&= \left\langle -\frac{\partial \tilde{f}}{\partial t} - A^T \frac{\partial \tilde{f}}{\partial x} + \left( -\frac{\partial A^T}{\partial x} + \mathbb{C}_1^* + F_v'^T \right) \tilde{f}, \delta v \right\rangle_{V^*(\Omega)} \quad (8.10) \\
&+ \langle \tilde{f}, \delta v \rangle_{V^*(\Gamma_1)} - \langle \tilde{f}, \delta v \rangle_{V^*(\Gamma_0)} \\
&+ \langle A^T \tilde{f}, \delta v \rangle_{V^*(\Gamma_b)} - \langle A^T \tilde{f}, \delta v \rangle_{V^*(\Gamma_a)} = 0.
\end{aligned}$$

Давайте найдём конкретный вид сопряжённого оператора  $\mathbb{C}_1^*$ , используя ранее полученную формулу (3.28). Поскольку, в нашей задаче матрица  $A = A(x, v(x))$ , при этом  $\frac{\partial A^T}{\partial x} = \sum_{i=1}^m \frac{\partial A^T}{\partial v_i} \frac{\partial v_i}{\partial x} + \frac{\partial A^T}{\partial x} \Big|_v$ , то согласно рекомендациям к выражению (3.28), целесообразно все свободные недифференциальные (без производных от  $\delta v$ ) члены находить одним выражением. Давайте такое выражение обозначим матричным оператором  $\mathbb{F}^*$ :

$$\mathbb{F}^* \cdot = \mathbb{C}_1^* \cdot - \frac{\partial A^T}{\partial x} \cdot + F_v'^T \cdot .$$

Согласно (3.28) первая матрица

$$\mathbb{C}_1^* = \begin{pmatrix} \frac{\partial a_{11}}{\partial Q} \frac{\partial Q}{\partial x} + \frac{\partial a_{12}}{\partial Q} \frac{\partial Z}{\partial x} & \frac{\partial a_{21}}{\partial Q} \frac{\partial Q}{\partial x} + \frac{\partial a_{22}}{\partial Q} \frac{\partial Z}{\partial x} \\ \frac{\partial a_{11}}{\partial Z} \frac{\partial Q}{\partial x} + \frac{\partial a_{12}}{\partial Z} \frac{\partial Z}{\partial x} & \frac{\partial a_{21}}{\partial Z} \frac{\partial Q}{\partial x} + \frac{\partial a_{22}}{\partial Z} \frac{\partial Z}{\partial x} \end{pmatrix}.$$

Вторая матрица

$$\begin{aligned} -\frac{\partial A^T}{\partial x} &= -\left( \frac{\partial A^T}{\partial Q} \frac{\partial Q}{\partial x} + \frac{\partial A^T}{\partial Z} \frac{\partial Z}{\partial x} \right) - \frac{\partial A^T}{\partial x} \Big|_v \\ &= -\begin{pmatrix} \frac{\partial a_{11}}{\partial Q} \frac{\partial Q}{\partial x} + \frac{\partial a_{11}}{\partial Z} \frac{\partial Z}{\partial x} & \frac{\partial a_{21}}{\partial Q} \frac{\partial Q}{\partial x} + \frac{\partial a_{21}}{\partial Z} \frac{\partial Z}{\partial x} \\ \frac{\partial a_{12}}{\partial Q} \frac{\partial Q}{\partial x} + \frac{\partial a_{12}}{\partial Z} \frac{\partial Z}{\partial x} & \frac{\partial a_{22}}{\partial Q} \frac{\partial Q}{\partial x} + \frac{\partial a_{22}}{\partial Z} \frac{\partial Z}{\partial x} \end{pmatrix} - \frac{\partial A^T}{\partial x} \Big|_v. \end{aligned}$$

После взаимного сокращения почти половины слагаемых, получаем значение свободного члена в виде суммы трёх матриц, умноженных на  $\tilde{f}$ :

$$\mathbb{F}^* \tilde{f} = \left[ \begin{pmatrix} \frac{\partial a_{12}}{\partial Q} \frac{\partial Z}{\partial x} - \frac{\partial a_{11}}{\partial Z} \frac{\partial Z}{\partial x} & \frac{\partial a_{22}}{\partial Q} \frac{\partial Z}{\partial x} - \frac{\partial a_{21}}{\partial Z} \frac{\partial Z}{\partial x} \\ \frac{\partial a_{11}}{\partial Z} \frac{\partial Q}{\partial x} - \frac{\partial a_{12}}{\partial Q} \frac{\partial Q}{\partial x} & \frac{\partial a_{21}}{\partial Z} \frac{\partial Q}{\partial x} - \frac{\partial a_{22}}{\partial Q} \frac{\partial Q}{\partial x} \end{pmatrix} - \frac{\partial A^T}{\partial x} \Big|_v + F_v'^T \right] \tilde{f}.$$

Найдём конкретное значение  $\mathbb{F}^*$  для имеющихся матриц  $A$  и  $F_v'$ .

Сначала подставим конкретные выражения элементов матриц:

$$\mathbb{F}^* \tilde{f} = \begin{pmatrix} \frac{\partial B(c^2 - w^2)}{\partial Q} \frac{\partial Z}{\partial x} - \frac{\partial 2w}{\partial Z} \frac{\partial Z}{\partial x} & -\frac{\partial 1/B}{\partial Z} \frac{\partial Z}{\partial x} \\ \frac{\partial 2w}{\partial Z} \frac{\partial Q}{\partial x} - \frac{\partial B(c^2 - w^2)}{\partial Q} \frac{\partial Q}{\partial x} & \frac{\partial 1/B}{\partial Z} \frac{\partial Q}{\partial x} \end{pmatrix} \tilde{f} - \begin{pmatrix} \frac{\partial 2w}{\partial x} \Big|_v & \frac{\partial 1/B}{\partial x} \Big|_v \\ \frac{\partial B(c^2 - w^2)}{\partial x} \Big|_v & 0 \end{pmatrix} \tilde{f} + \begin{pmatrix} \frac{\partial F_{\text{тр}} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z}{\partial Q} & -\frac{\partial q/B}{\partial Q} \\ \frac{\partial F_{\text{тр}} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z}{\partial Z} & -\frac{\partial q/B}{\partial Z} \end{pmatrix} \tilde{f}.$$

Возьмём указанные производные с учётом того,  $w(x, Q, Z) = Q/\omega$ ,  $\omega = \omega(x, Z)$  и  $\frac{\partial \omega}{\partial Z} = B(x, Z)$ ,  $c^2(x, Z) = g\omega/B$ ,  $F_{fr}(x, Q, Z) = \frac{gQ|Q|}{\omega RC^2}$ . Производные по  $x$  при  $v = \text{const}$  берутся от параметров, которые прямо<sup>1</sup> зависят от  $x$ , а не косвенно через  $Q$  и  $Z$ , последние при этом считаются константами.

$$\mathbb{F}^* \tilde{f} = \begin{pmatrix} 0 & -\frac{\partial 1/B}{\partial x} + \frac{\partial 1/B}{\partial x} \Big|_Z \\ 0 & -\frac{1}{B^2} \frac{\partial B}{\partial Z} \frac{\partial Q}{\partial x} \end{pmatrix} \tilde{f} - \begin{pmatrix} -\frac{2Q}{\omega^2} \frac{\partial \omega}{\partial x} \Big|_Z & \frac{\partial 1/B}{\partial x} \Big|_Z \\ \frac{\partial B(c^2 - w^2)}{\partial x} \Big|_v & 0 \end{pmatrix} \tilde{f} + \begin{pmatrix} \frac{\partial F_{\text{тр}}}{\partial Q} - \frac{2Q}{\omega^2} \frac{\partial \omega}{\partial x} \Big|_Z & 0 \\ \frac{\partial F_{fr}}{\partial Z} - \frac{\partial w^2}{\partial Z} \frac{\partial \omega}{\partial x} \Big|_Z & \frac{q}{B^2} \frac{\partial B}{\partial Z} \end{pmatrix} \tilde{f}$$

В итоге получаем свободный член с оператором  $\mathbb{F}^*$  в виде матрицы:

<sup>1</sup> Например,  $\frac{\partial w}{\partial x} \Big|_v = \frac{\partial Q/\omega}{\partial x} \Big|_v = -\frac{Q}{\omega^2} \frac{\partial \omega}{\partial x} \Big|_Z$

$$\mathbb{F}^* = \begin{pmatrix} \mathbb{f}_{11}^* & \mathbb{f}_{12}^* \\ \mathbb{f}_{21}^* & \mathbb{f}_{22}^* \end{pmatrix},$$

где элементы матрицы

$$\mathbb{f}_{11}^* = \frac{\partial F_{fr}}{\partial Q}, \quad \mathbb{f}_{12}^* = \frac{1}{B^2} \frac{\partial B}{\partial x},$$

$$\mathbb{f}_{21}^* = \frac{\partial F_{\text{тр}}}{\partial Z} + w^2 \frac{\partial B}{\partial x} \Big|_Z - g \frac{\partial \omega}{\partial x} \Big|_Z, \quad \mathbb{f}_{22}^* = \frac{1}{B^2} \left( q - \frac{\partial Q}{\partial x} \right) \frac{\partial B}{\partial Z}$$

и члены

$$\frac{\partial F_{\text{тр}}}{\partial Q} = \frac{2F_{\text{тр}}}{Q}, \quad \frac{\partial F_{\text{тр}}}{\partial Z} = -F_{\text{тр}} \left( \frac{B}{\omega} + \frac{4}{3} \frac{1}{R} \frac{\partial R}{\partial Z} \right).$$

Из сравнения тождества (8.9) с полученным выражением (8.10) мы видим, что сопряжённый оператор  $\mathbb{V}^*$  и его отображение — это

$$\mathbb{V}^* \cdot = -\frac{\partial \cdot}{\partial t} - A^T \frac{\partial \cdot}{\partial x} + \mathbb{F}^* \cdot : V^*(\Omega) \rightarrow V^*(\Omega).$$

Приведём и граничные операторы:

$$\bar{\mathbb{V}}^* \cdot = \cdot|_{\Gamma_1} - \cdot|_{\Gamma_0} + A^T \cdot|_{\Gamma_b} - A^T \cdot|_{\Gamma_a}.$$

Теперь обратимся к тождеству (8.9) при вариации управления  $\delta u$ :

$$\begin{aligned} \langle \tilde{f}, \mathbb{U} \delta u \rangle_{V^*(\Omega)} &\equiv \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)} \\ &= \iint_{\Omega} F_u'^T \tilde{f} \delta u dx dt = \int_{x_a}^{x_b} \left( \int_{t_0}^{t_1} F_u'^T \tilde{f} dt \right) \delta u dx. \end{aligned}$$

То есть, сопряжённый оператор  $\mathbb{U}^*$  и его отображение — это

$$\mathbb{U}^* \cdot = \int_{t_0}^{t_1} F_u'^T \cdot dt : V^*(\Omega) \rightarrow U^*(S),$$

где вектор-строка

$$F_u'^T = \left( \frac{\partial F_{\text{тр}}}{\partial u}, 0 \right)^T = \left( \frac{2F_{\text{тр}}}{u}, 0 \right)^T.$$

Оператор  $\mathbb{U}^*$  реализует отображение из пространства сопряжённых состояний, определённых на  $\Omega$ , в пространство сопряжённых управлений, определённых на  $S$ . В своём отображении он

осуществляет *проецирование*  $P_x(\Omega)$ , которое реализуется интегрированием по времени.

**4. Объединим** линейные функционалы в соответствующих пространствах на элементах, к которым они относятся, и получим выражение первой вариации целевого функционала (без учёта краевых условий задачи), аналогичное (3.30):

$$\begin{aligned} \delta J(u) = & \langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\Omega)} + \langle \tilde{f}, \delta v \rangle_{V^*(\Gamma_1)} - \langle \tilde{f}, \delta v \rangle_{V^*(\Gamma_0)} \\ & + \langle A^T \tilde{f} + I'_v, \delta v \rangle_{V^*(\Gamma_b)} - \langle A^T \tilde{f}, \delta v \rangle_{V^*(\Gamma_a)} \\ & + \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)}. \end{aligned} \quad (8.11)$$

Здесь на границе  $\Gamma_b$  добавлена производная  $I'_v$  от критерия качества идентификации.

**5. Варьируем краевые условия** (8.3), (8.4):

$$\delta Z = 0 \text{ на } \Gamma_a, \quad \delta Q = 0 \text{ на } \Gamma_b, \quad \delta v = 0 \text{ на } \Gamma_0.$$

Теперь учтём значения этих вариаций в (8.11) на соответствующих множествах  $\Gamma_a$ ,  $\Gamma_b$ ,  $\Gamma_0$ . Получим окончательное выражение *первой вариации* целевого функционала:

$$\begin{aligned} \delta J(u) = & \langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\Omega)} \\ & + \langle \tilde{f}, \delta v \rangle_{V^*(\Gamma_1)} + \langle A^T \tilde{f} + I'_v, (0, \delta Z) \rangle_{V^*(\Gamma_b)} \\ & - \langle A^T \tilde{f}, (\delta Q, 0) \rangle_{V^*(\Gamma_a)} + \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)}. \end{aligned} \quad (8.12)$$

**6. Выделение градиента.** В полученном выражении зададим  $\tilde{f} = f$ , такое, чтобы все линейные функционалы при  $\delta v = (\delta Q, \delta Z)$  обратились в ноль. В области  $\Omega$  имеем:

$$\mathbb{V}^* f = -\frac{\partial f}{\partial t} - A^T \frac{\partial f}{\partial x} + \mathbb{F}^* f = 0 \text{ на } \Omega. \quad (8.13)$$

На границе  $\Gamma_a$  необходимо  $-\langle A^T f, (\delta Q, 0) \rangle_{V^*(\Gamma_a)} = 0$ , т.е.

$$\left\langle \left( 2wf_1 + \frac{1}{B}f_2, B(c^2 - w^2)f_1 \right), (\delta Q, 0) \right\rangle_{L_2^m} = \left\langle 2wf_1 + \frac{1}{B}f_2, \delta Q \right\rangle_{L_2} = 0.$$

Следовательно

$$2wf_1 + \frac{1}{B}f_2 = 0 \quad \text{на } \Gamma_a. \quad (8.14)$$

На границе  $\Gamma_b$  необходимо  $\langle A^T f + I'_v, (0, \delta Z) \rangle_{V^*(\Gamma_b)} = 0$ , т.е.

$$\begin{aligned} & \left\langle \left( 2wf_1 + \frac{1}{B}f_2, B(c^2 - w^2)f_1 + I'_Z \right), (0, \delta Z) \right\rangle_{L_2^m} \\ & = \langle B(c^2 - w^2)f_1 + 2(Z - Z_e), \delta Z \rangle_{L_2} = 0. \end{aligned}$$

Следовательно

$$B(c^2 - w^2)f_1 + 2(Z - Z_e) = 0 \quad \text{на } \Gamma_b. \quad (8.15)$$

На терминальной линии  $\Gamma_1$  имеем:

$$f_1 = f_2 = 0 \quad \text{на } \Gamma_1. \quad (8.16)$$

Мы получили сопряжённую задачу (8.13)–(8.16), аналогичную (3.33), которая решается в обратном направлении по времени, начиная с нулевого терминального состояния.

Первая вариация целевого функционала принимает вид:

$$\delta J(u) = \langle \mathbb{U}^* f, \delta u \rangle_{U^*(S)} = \langle \nabla J, \delta u \rangle_{U^*(S)}$$

Следовательно, градиент целевого функционала  $\nabla J = \mathbb{U}^* f$  имеет вид:

$$\nabla J(u; x) = \int_{t_0}^{t_1} \frac{2F_{fr}}{u} f_1 dt \in U^*(S). \quad (8.17)$$

Полученное выражение соответствует первой строке общей формы градиента (3.36), когда искомым параметр (управление) присутствует в дифференциальном уравнении модели.

### 8.2.1. Характеристическая форма сопряжённых уравнений

Далее потребуется, как и в исходной задаче, характеристическая форма уравнений (8.13). Давайте её запишем.

Собственные числа матриц  $A$  и  $A^T$  одинаковые. Значит сопряжённая задача имеет те же самые характеристики  $\xi_{1,2} = w \pm c$ , что и исходная. А вот собственные векторы у транспонированной матрицы будут другими. Здесь

$$(\Lambda_{11}, \Lambda_{12}) = \left(1, \frac{1}{B(w+c)}\right), \quad (\Lambda_{21}, \Lambda_{22}) = \left(1, \frac{1}{B(w-c)}\right).$$

Умножим сопряжённую систему (8.13) сначала на  $\Lambda_1^T$ , потом на  $\Lambda_2^T$ . Получим вдоль характеристик  $\xi_1$  и  $\xi_2$  два характеристических уравнения (3.18) в виде:

$$\begin{aligned} \Lambda_1^T \left( -\frac{\partial f}{\partial t} - \lambda_1 \frac{\partial f}{\partial x} \right) + \Lambda_1^T \mathbb{F}^* f \\ = -\frac{\partial f_1}{\partial t} - \frac{1}{B(w+c)} \frac{\partial f_2}{\partial t} - (w+c) \frac{\partial f_1}{\partial x} - \frac{1}{B} \frac{\partial f_2}{\partial x} + \mathbb{f}_{11}^* f_1 \\ + \mathbb{f}_{12}^* f_2 + \frac{1}{B(w+c)} (\mathbb{f}_{21}^* f_1 + \mathbb{f}_{22}^* f_2) = 0, \end{aligned}$$

$$\begin{aligned} \Lambda_2^T \left( -\frac{\partial f}{\partial t} - \lambda_2 \frac{\partial f}{\partial x} \right) + \Lambda_2^T \mathbb{F}^* f \\ = -\frac{\partial f_1}{\partial t} - \frac{1}{B(w-c)} \frac{\partial f_2}{\partial t} - (w-c) \frac{\partial f_1}{\partial x} - \frac{1}{B} \frac{\partial f_2}{\partial x} + \mathbb{f}_{11}^* f_1 \\ + \mathbb{f}_{12}^* f_2 + \frac{1}{B(w-c)} (\mathbb{f}_{21}^* f_1 + \mathbb{f}_{22}^* f_2) = 0 \end{aligned}$$

Можно записать короче:

$$\begin{aligned} -B(w \pm c) \left( \frac{\partial f_1}{\partial t} + (w \pm c) \frac{\partial f_1}{\partial x} \right) - \frac{\partial f_2}{\partial t} - (w \pm c) \frac{\partial f_2}{\partial x} \\ + B(w \pm c) (\mathbb{f}_{11}^* f_1 + \mathbb{f}_{12}^* f_2) + \mathbb{f}_{21}^* f_1 + \mathbb{f}_{22}^* f_2 = 0. \end{aligned}$$

Если учесть, что  $\frac{\partial \cdot}{\partial t} + (w \pm c) \frac{\partial \cdot}{\partial x} = \frac{d \cdot}{dt} \Big|_{\xi_{1,2}}$  вдоль  $\xi_1$  и  $\xi_2$ , то полученные

уравнения можно представить в обыкновенных дифференциалах:

$$\begin{aligned} -B(w \pm c) \frac{df_1}{dt} - \frac{df_2}{dt} + B(w \pm c) (\mathbb{f}_{11}^* f_1 + \mathbb{f}_{12}^* f_2) \\ + \mathbb{f}_{21}^* f_1 + \mathbb{f}_{22}^* f_2 = 0, \end{aligned} \tag{8.18}$$

где верхний знак, как и ранее, соответствует характеристикам первого семейства  $\xi_1$ , а нижний — характеристикам второго семейства  $\xi_2$ .

Давайте запишем уравнения (8.18) с явным выражением элементов матрицы  $\mathbb{F}^*$ :

$$\begin{aligned}
& -B(w \pm c) \frac{df_1}{dt} - \frac{df_2}{dt} + B(w \pm c) \left( f_1 \frac{\partial F_{\text{тр}}}{\partial Q} + f_2 \frac{1}{B^2} \frac{\partial B}{\partial x} \right) \\
& + f_1 \left( \frac{\partial F_{\text{тр}}}{\partial Z} + w^2 \frac{\partial B}{\partial x} \Big|_Z - g \frac{\partial \omega}{\partial x} \Big|_Z \right) + f_2 \frac{1}{B^2} \left( q - \frac{\partial Q}{\partial x} \right) \frac{\partial B}{\partial Z} = 0.
\end{aligned}$$

Напомним, что  $\frac{\partial F_{\text{тр}}}{\partial Q} = \frac{2F_{\text{тр}}}{Q}$ ,  $\frac{\partial F_{\text{тр}}}{\partial Z} = -F_{\text{тр}} \left( \frac{B}{\omega} + \frac{4}{3} \frac{1}{R} \frac{\partial R}{\partial Z} \right)$ .

### 8.3 Идентифицируемость

Рассмотрим условия возможного корректного решения обсуждаемой задачи идентификации, т.е. условия идентифицируемости.

Для анализа идентифицируемости коэффициента шероховатости воспользуемся теоремой 4.1. Сначала обоснуем корректность решения исходной и сопряжённой задач. Далее уточним области определения и значений оператора  $U^*$ , сделаем оценку его возможной вырожденности и, на основе производной Фреше  $J'_u$  в первой вариации  $\delta J$ , получим градиент  $\nabla J$ .

О корректности обеих задач. При получении характеристической формы исходных уравнений мы нашли характеристики двух семейств  $\xi_{1,2}$  вдоль которых распространяются волны малых возмущений в русле со скоростью

$$\lambda_{1,2} = \frac{d\xi_{1,2}}{dt} = w \pm c.$$

Поскольку мы рассматриваем течения без ударных волн, то волны возмущений первого семейства, не пересекая друг друга, перемещаются вдоль потока воды к нижнему правому створу русла, а волны второго семейства — против течения к верхнему левому створу, также не пересекая друг друга.

При преобразованиях к характеристической форме сопряжённых уравнений мы получили те же характеристики  $\xi_{1,2}$ , что и в исходной задаче. Но здесь, сопряжённые волны, двигаясь по тем же

характеристикам, идут в обратном по времени направлении. От терминального времени  $t_1$  волны по характеристикам  $\xi_1$  движутся к верхнему створу русла, а по  $\xi_2$  — к нижнему створу.

Поскольку в каждую точку  $\tau \in \Omega$  попадают две характеристики разного семейства  $\xi_1, \xi_2$ , выходящие из участков границы  $\partial\Omega$  с известными решениями (заданными краевыми условиями), то исходная задача (8.2)–(8.4) и сопряжённая (8.13)–(8.16) обеспечивают корректное решение [70, 44] на  $\bar{\Omega}$  для двумерных вектор-функций  $v(\tau)$  и  $f(\tau)$ , соответственно.

К вопросу корректности следует сделать некоторые дополнения. В прямой задаче (8.2)–(8.4) множества существования состояния потока  $V$  и управления  $U$  определяются физически разумными, приемлемыми значениями состояния  $v \in L_{2+}^2(\bar{\Omega})$  и управления  $u \in L_{2+}(S)$ . Например, при достаточно больших значениях  $u$  в расчётах может быть получен уровень потока  $Z$  идущий ниже линии дна  $Z_{\text{дно}}$ . Поэтому, для разумного решения исходной задачи, ограничим  $U$  некоторым замкнутым множеством из условия:

$$\begin{aligned} U &= \{u: u(x) \geq 0, \forall x \in S, v \in V\}, \\ V &= \{v: Z(\tau) > Z_{\text{дно}}(\tau), \forall \tau \in \bar{\Omega}\}. \end{aligned} \quad (8.19)$$

Заметим, что последнее условие для  $V$  является более жёстким, чем ранее указанная принадлежность  $v \in L_{2+}^2(\bar{\Omega})$ . Здесь требуется принадлежность положительному полупространству глубины потока  $Z - Z_{\text{дно}} \in L_{2+}^2(\bar{\Omega})$ .

Приведенные условия не являются ограничениями для оптимизации и не требуют их реализации в экстремальных алгоритмах. Требования (8.19) означает, что прямая задача физически условно корректна. Её решение следует искать при ограниченном изменении коэффициента шероховатости. Множество таких допустимых

управлений  $u$ , при не слишком больших колебаниях функций  $u(x)$ , представляет собой *компакт корректности*  $\mathcal{U}$  для решения обратной задачи идентификации:

$$\mathcal{U} = U.$$

Все управления  $u^k$ , выходящие за пределы компакта  $\mathcal{U}$ , не позволят корректно найти ни градиент  $\nabla J^k$ , ни дальнейшее управление  $u^{k+1}$ .

Найдём область определения и значений оператора  $\mathbb{U}^*$ , воспользовавшись определением 3.5.

Из корректности исходной задачи следует, что изменение распределённого коэффициента шероховатости  $u(x)$  меняет уровень воды во всём потоке. Любые функции  $u(x) \in \mathcal{U}$ , при условии, что хотя бы одна характеристика  $\xi_1$  пройдёт без искажений (несущественная диссипация энергии волн) вдоль всего русла, будут однозначно влиять на значение производной целевой функции  $I'_Z|_\omega = 2(Z|_{\Gamma_b} - Z_e)$ . То есть имеет место левая ветка отображений в определении 3.5:

$$U(S) \ni u \xrightarrow{\text{однозначно}} I'_Z|_\omega.$$

Производная  $I'_Z|_\omega$  является источником возмущений сопряжённой задачи. Возмущения  $f$  при  $t \rightarrow t_0$  распространяются в виде сопряжённых волн по характеристикам первого семейства  $\xi_1$  от правого нижнего створа в сторону левого верхнего створа. В левом створе сопряжённые волны отражаются и переносятся обратно вправо по характеристикам второго семейства  $\xi_2$ . Распространение сопряжённых волн и пример множества  $\Omega$  (вся заштрихованная с разной плотностью область под верхней характеристикой  $\xi_1$ ) для области определения  $V^*(\Omega)$  оператора  $\mathbb{U}^*$  показана на рисунке 8.2.

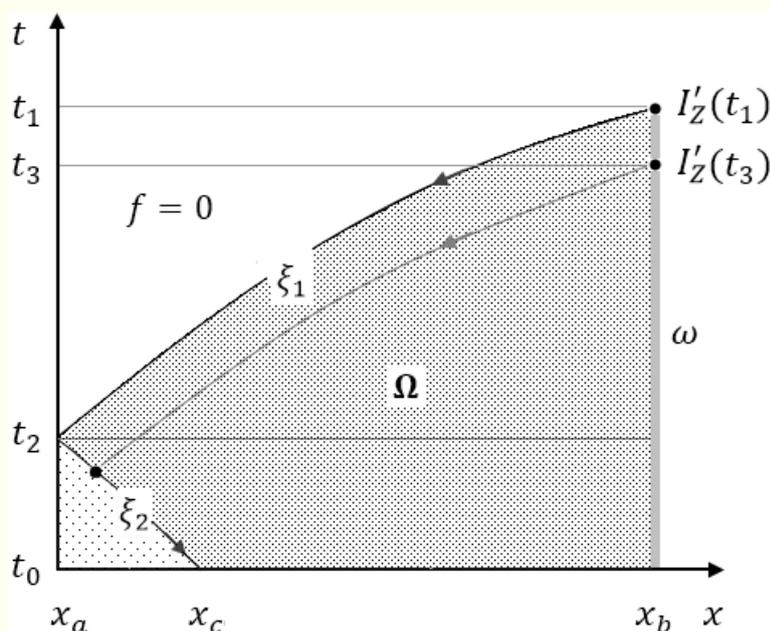


Рисунок 8.2 — Пространственно-временная диаграмма области  $\Omega$  с избыточной областью наблюдений  $\omega$

Можно выделить время  $t_2$  прихода верхней (замыкающей) характеристики  $\xi_1$  к левому створу русла. Ниже этого времени в области  $\Omega$ , под характеристикой  $\xi_2$ , начинает образовываться светлый треугольник области неоднозначного влияния значений функции  $I'_Z|_\omega$  на сопряжённое состояние  $f$ . В каждую точку данной области будет приходить две сопряжённые волны, вышедшие из разных точек  $\omega$  с разными значениями  $I'_Z$ , как это показано в качестве примера для возмущений  $I'_Z(t_1)$  и  $I'_Z(t_3)$ . Кроме того, характеристика  $\xi_2$  может нести дополнительную ненужную информацию об условиях течения в левом створе русла.

Всё это означает, что при  $t_0 < t_2$  применение оператора  $\mathbb{U}^*$  будет реализовываться в области его неоднозначного определения и может существенно исказить значение градиента  $\nabla J = \mathbb{U}^* f = \int_{t_0}^{t_1} F_u'^T f dt$ .

Чтобы такого не возникло, следует устранить *избыточность* области  $\omega$  задания целевого функционала до минимально достаточной  $\omega_{min}$ , как это показано на рисунке 8.3.

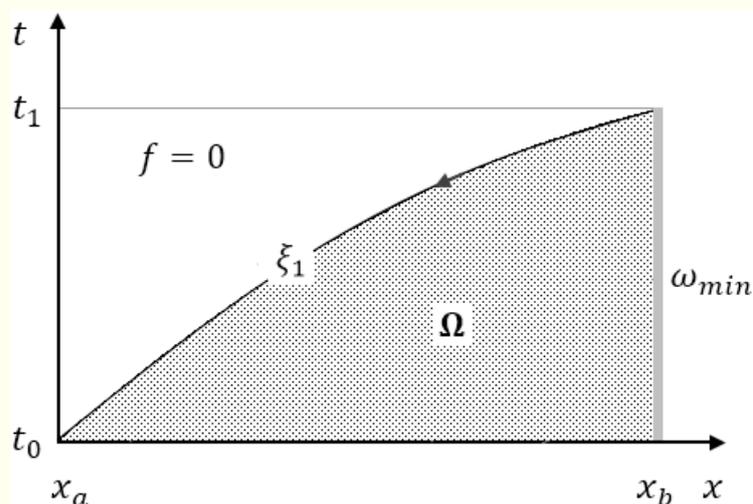


Рисунок 8.3 — Пространственно-временная диаграмма области  $\Omega$  с минимально достаточной областью наблюдений  $\omega_{min}$

Если диссипативные помехи в исходной и сопряжённой задачах не существенны, то однозначность области определения  $\mathbb{U}^*$  обеспечивается при

$$t_1 = t_0 + \int_{x_a}^{x_b} \frac{dx}{\lambda_1} \equiv t_0 + \int_{x_a}^{x_b} \frac{dx}{w + c}. \quad (8.20)$$

Данное условие представляет собой *условие идентифицируемости* в рассматриваемой задаче.

Наличие диссипативных членов в сочетании со значительной неоднородностью потока в русле, может приводить к существенному затуханию волн как исходной, так и сопряжённой задач. Затухание волн в виде уровня воды  $Z$  при движении слева-направо в исходной задаче будет искажать невязку  $(Z|_{\Gamma_b} - Z_e)$  в функции цели  $I$ , на основании которой решается сопряжённая задача, добавляя своё искажение

сопряжённых волн  $f$  от источника  $I'_v$  при движении обратно справа-налево.

Это означает, что возле верхнего створа русла точность идентификации шероховатости может ухудшаться, а при очень длинном русле, или сильно выраженной диффузорной неоднородности потока (большой уклон русла при низкой скорости течения) вообще может стать невозможной. Поэтому требование (8.20) следует дополнить фразой «с диссипацией, не существенно искажающей волновые процессы в обеих задачах».

В этих условиях выполняются оставшиеся ветки отображения в определении 3.5:

$$I'_Z|_\omega \xrightarrow{\text{однозначно}} f|_\Omega \xrightarrow{\mathbb{U}^*} U^*(S).$$

При полученном  $\Omega$  (серый треугольник на рисунке 8.3) оператор  $\mathbb{U}^*$  имеет область определения  $V^*(\Omega)$  с областью значений в  $U^*(S)$ .

Итак, мы нашли область определения и область значений оператора  $\mathbb{U}^*$ . Первое условие теоремы 4.1 «существует область  $V^*(\Omega)$  определения оператора  $\mathbb{U}^*$  с его значениями в области  $U^*(S)$ » выполняется.

Следует отметить, что нельзя полностью игнорировать явление диссипации в потоке воды и его неоднородность. Эти факторы оказывают значительное влияние на волновые процессы. Замыкающая характеристика  $\xi_1$ , изображённая на теоретическом рисунке 8.3, демонстрирует следующие условности. Во-первых, она указывает всего лишь на начало возникновения волны (малого возмущения) в потоке воды в момент  $t_0$  в левом створе и начало прихода этой волны в момент  $t_1$  к правому створу. Во-вторых, она указывает аналогичное начало возникновения и начало прихода сопряжённой волны от правого створа к левому. Понятно, что начало волн в обеих случаях будет

сопровождаться, с одной стороны, очень малым изменением уровня  $Z$  для формирования значимой величины  $I'_Z$ , и с другой стороны, — очень малым изменением сопряжённого состояния  $f_1$  для формирования градиента  $\nabla J$  особенно в левой части русла.

Если принять во внимание ещё и вычислительные погрешности, то на практике, чтобы волны достаточно заметно проявлялись в целевом функционале и далее — в градиенте, волнам необходимо дать дополнительное время  $t_\varepsilon$  — время ощутимого проявления фронта волны после появления впереди идущих малых возмущений.

Таким образом, практическое условие *идентифицируемости* будет выглядеть следующим образом:

$$t_1 = t_0 + \int_{x_a}^{x_b} \frac{dx}{\lambda_1} + t_\varepsilon. \quad (8.21)$$

Время  $t_\varepsilon$  не должно быть слишком большим, чтобы область неоднозначного влияния значений функции  $I'_Z|_\omega$  (светлый треугольник на рисунке 8.2) не стала преобладающей. Конкретное значение времени  $t_\varepsilon$  определяется в ходе вычислительного эксперимента.

Теперь оценим наличие возможной вырожденности оператора  $\mathbb{U}^*$  в градиенте (8.17):

$$\nabla J = \mathbb{U}^* f = \int_{t_1}^{t_0} \frac{2F_{\text{тр}}}{u} f_1 dt \in U^*(S).$$

Необходимо выяснить, имеет ли оператор  $\mathbb{U}^*$  при каких-либо условиях ядро  $\text{Ker } \mathbb{U}^* \neq 0$ , т.е. может ли градиент  $\nabla J$  при  $f_1 \neq 0$ , и соответственно при  $u \neq u_*$ , иметь нулевое значение на  $S$  или на значительной (ненулевой меры) части  $S$ .

Поскольку мы рассматриваем движущиеся потоки воды и только в одном направлении — вдоль русла к нижнему створу, то при  $u \in \mathcal{U}$  член, трения  $F_{\text{тр}} = \frac{gQ|Q|}{\omega RC^2} > 0$  и соответственно коэффициент в градиенте  $\frac{2F_{\text{тр}}}{u}$

никогда не обращается в ноль. В то же время, в разные моменты времени волны от невязки  $(Z|_{\Gamma_b} - Z_e)$ , идущие вдоль характеристик первого семейства, могут быть как волнами повышения ( $f_1$  положительное), так и волнами понижения ( $f_1$  отрицательное). Как при всяком волновом движении суммарное по времени (интегральное в  $\nabla J$ ) значение таких волн в некоторых точках  $x \in S$  (только в некоторых точках, а не на всём  $S$ ) может обращать в ноль значение  $\nabla J(u; x)$ . Однако, такие волны не являются стоячими, и при изменении управления от  $u^k$  на  $u^{k+1}$  ранее точечные нулевые значения градиента будут исчезать. Устойчивое нулевое значение интегралов по времени в градиенте  $\nabla J(u; x)$  на всём  $S$  возможно только когда  $(Z|_{\Gamma_b} - Z_e) = 0$ , что соответствует  $f_1 = 0$  при оптимальном решении  $u = u_*$ . Это означает, что оператор  $\mathbb{U}^*$  имеет только нулевое ядро. Выполняется второе условие теоремы 4.1 «оператор  $\mathbb{U}^*$  — невырожденный».

Таким образом, согласно теореме 4.1, математическая модель распределённой системы (8.1), или в форме (8.2), с краевыми условиями (8.3), (8.4) идентифицируема посредством параметра  $u(x)$  на  $S$  по функционалу  $J$  (8.5) при условии (8.20) с диссипацией, не существенно искажающей волновые процессы в обеих задачах.

При выполнении условия идентифицируемости (8.20) производная Фреше в выражении (8.17)  $J'_u = \nabla J$ , т.е. является градиентом. Остаётся выбрать разумное начальное приближение  $u^0 \in \mathcal{U}$  и корректно применить экстремальные алгоритмы. Следует ожидать равномерную на  $S$  сходимость, поскольку идентифицируемость имеет место на всём множестве  $S$ .

## 8.4 Численное решение

Для тестирования полученных условий и алгоритмов был рассмотрен гипотетический канал. Параметры канала и условия моделирования течения воды в нём описаны ниже при оценке выпуклости целевого функционала тестовой задачи.

Кроме тестовой задачи была решена задача идентификации шероховатости для реального участка Каракумского канала. Параметры Каракумского канала, исходные данные этой прикладной задачи описаны непосредственно перед её решением.

### 8.4.1. Конечно-разностные аппроксимации

Исходная (8.2)–(8.4) и сопряжённая (8.13)–(8.16) гиперболические задачи решались на основе характеристик по *явной конечно-разностной схеме* с шаблоном аппроксимаций показанном на рисунке 8.4. Пространственная сетка задавалась равномерной  $\{x_i = x_a + i\Delta x, i = 0 \dots n, \Delta x = (x_b - x_a)/n\}$ , сетка по времени была неравномерной  $\{t_j = t_{j-1} + \Delta t_j, j = 1 \dots m\}$ . Для исходной задачи новый временной слой — это  $j + 1$ , а для сопряжённой — слой  $j - 1$ .

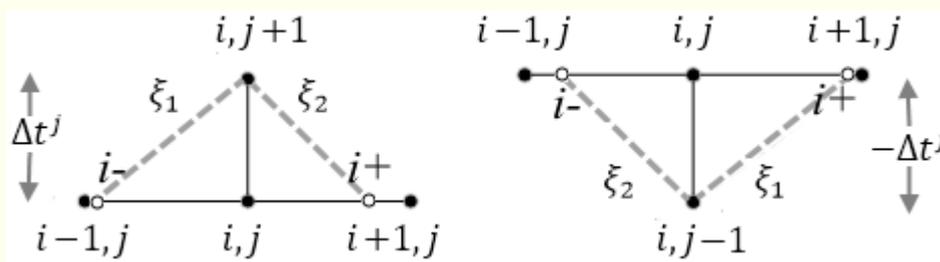


Рисунок 8.4 — Шаблоны конечно-разностных аппроксимаций для решения исходной (слева) и сопряжённой (справа) задач

Точки  $i \pm$  (белые точки на рисунке) соответствуют месту выхода характеристик  $\xi_1, \xi_2$  с предыдущего временного слоя слева или справа для попадания в точку  $i$  на новом временном слое. Рассматриваемая схема предполагает докритические течения.

Шаги по времени определяются условием устойчивости, которое имеет вид:  $\Delta t_j \leq \min_{\forall i} \frac{\Delta x}{|(w \pm c)_i^j|}$ , т.е. все самые «быстрые» характеристики со скоростью  $\lambda_{1,2} = (w \pm c)_i^j$ , пересекающие на новом временном слое узел  $i$ , не должны выходить со старого временного слоя из-за пределов отрезка  $[x_i - \Delta x, x_i + \Delta x]$ . В противном случае на новом временном слое узел  $i$  не попадёт в область влияния характеристик и решение задачи будет невозможным.

Поскольку гиперболическая система Сен-Венана не является однородной (присутствует вектор свободных членов  $F$ ), не является полуплинейной (имеются непостоянные коэффициенты в матрице  $A = A(x, v)$ ), то её инварианты  $r_{1,2}$  будут определяться сложным образом и не будут постоянными. Это затрудняет применение классического метода характеристик [70], который основан на инвариантах. Поэтому, мы реализуем метод характеристик для исходной и сопряжённой задач без инвариантов.

Начнём с исходной задачи. Аппроксимируем характеристические уравнения в обыкновенных дифференциалах (8.6) во внутренних точках  $0 < i < n$  для шага  $\Delta t_j$  по методу Эйлера (первый порядок точности), когда коэффициенты и свободные члены берутся с предыдущей точки. На характеристиках первого семейства получим:

$$\frac{Q_i^{j+1} - Q_{i-}^j}{\Delta t_j} - e_1 \frac{Z_i^{j+1} - Z_{i-}^j}{\Delta t_j} + \mathcal{F}_{i-}^j = 0,$$

где  $e_1 = B_{i-}^j (w - c)_{i-}^j$ ,  $\mathcal{F}_{i-}^j = \left( F_{\text{тр}} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z + (w - c)q \right)_{i-}^j$

На характеристиках второго семейства:

$$\frac{Q_i^{j+1} - Q_{i+}^j}{\Delta t_j} - e_2 \frac{Z_i^{j+1} - Z_{i+}^j}{\Delta t_j} + \mathcal{F}_{i+}^j = 0,$$

где  $e_2 = B_{i+}^j (w + c)_{i+}^j$ ,  $\mathcal{F}_{i+}^j = \left( F_{\text{тр}} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z + (w + c)q \right)_{i+}^j$ .

Для определения местоположения точек  $i \pm$  надо решать дифференциальные уравнения  $\frac{dx}{dt} = w \pm c$ :

$$x_{i+} = x_i - (w - c)_i^j \Delta t_j, \quad x_{i-} = x_i - (w + c)_i^j \Delta t_j.$$

Для определения параметров потока в точках  $x_{i\pm}$  необходимо делать интерполяцию значений параметров, например, квадратичную по трём точкам  $x_{i-1}^j$ ,  $x_i^j$ ,  $x_{i+1}^j$ . Для этого можно использовать многочлен Лагранжа, который для некоторой функции  $y$  заданной в точках  $x_i$  имеет вид:

$$\begin{aligned} L_2(x) = & y_{i-1} \frac{(x - x_i)(x - x_{i+1})}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} \\ & + y_i \frac{(x - x_{i-1})(x - x_{i+1})}{(x_i - x_{i-1})(x_i - x_{i+1})} \\ & + y_{i+1} \frac{(x - x_{i-1})(x - x_i)}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)}, \\ & x \in [x_{i-1}, x_{i+1}], \quad 0 < i < n. \end{aligned} \quad (8.22)$$

Мы получили систему двух уравнений с двумя неизвестными  $Q_i^{j+1}$  и  $Z_i^{j+1}$ . Откуда находим:

$$Z_i^{j+1} = (e_1 Z_{i-}^j - e_2 Z_{i+}^j - Q_{i-}^j + Q_{i+}^j + (\mathcal{F}_{i-}^j - \mathcal{F}_{i+}^j) \Delta t_j) \frac{1}{e_1 - e_2},$$

$$Q_i^{j+1} = Q_{i-}^j + e_1 (Z_i^{j+1} - Z_{i-}^j) - \mathcal{F}_{i-}^j \Delta t_j,$$

$$i = 1 \dots n - 1, \quad j = 0 \dots m - 1.$$

На левой границе русла, при  $i = 0$ , будем использовать одно уравнение из (8.6), соответствующее приходящий туда характеристике  $\xi_2$ , а на правой границе, при  $i = n$ , второе уравнение, соответствующее характеристике  $\xi_1$ . С учётом граничных условий (8.3), где известен уровень потока  $Z_{i=0}^{j+1} = Z_a^{j+1}$  слева на  $\Gamma_a$ , и расход воды  $Q_{i=n}^{j+1} = Q_b^{j+1}$  справа на  $\Gamma_b$ , получаем уравнения для расчёта неизвестных граничных компонент потока:

$$\begin{aligned} Q_i^{j+1} &= Q_{i+}^j + e_2(Z_a^{j+1} - Z_{i+}^j) - \mathcal{F}_{i+}^j \Delta t_j, & i = 0, \\ Z_i^{j+1} &= Z_{i-}^j + (Q_b^{j+1} - Q_{i-}^j + \mathcal{F}_{i-}^j \Delta t_j) \frac{1}{e_1}, & i = n. \end{aligned}$$

Теперь обсудим реализацию начальных условий. Естественно полагать, что измерений уровня и расхода воды вдоль всего русла нет. Поэтому будем считать, что начальное течение воды стационарное. При этом в системе уравнений (8.1) исчезнут производные по времени, и они превратятся, по сути, в одно уравнение в обыкновенных дифференциалах для расчёта начального уровня  $Z_0(x)$  воды в русле.

$$2wq + B(c^2 - w^2) \frac{dZ_0}{dx} + F_{\text{тр}} - w^2 \frac{\partial \omega}{\partial x} \Big|_Z, \quad Z_0(x_a) = Z_a(t_0) \quad (8.23)$$

Расход стационарного течения не связан с уровнем воды и элементарно определяется боковым притоком:

$$\frac{dQ_0}{dx} = q, \quad Q_0(x_b) = Q_b(t_0).$$

После численного решения задачи Коши (8.23), например, каким-либо методом Рунге-Кутты, получаем начальные условия (8.4) в узлах сетки:

$$Q_i^{j=0} = Q_0(x_i), \quad Z_i^{j=0} = Z_0(x_i), \quad i = 0 \dots n.$$

Мы получили необходимые конечно-разностные аппроксимации исходной задачи.

Теперь обратимся к сопряжённой задаче (8.13)–(8.16). Аппроксимируем её уравнения в характеристической форме (8.18), аналогично предыдущей задаче. На характеристиках первого семейства, с учётом движения в обратном направлении по времени, имеем:

$$-e_1 \frac{f_{1i}^{j-1} - f_{1i+}^j}{-\Delta t_j} - \frac{f_{2i}^{j-1} - f_{2i+}^j}{-\Delta t_j} + \mathcal{F}_{i+}^j = 0,$$

где  $e_1 = (B(w + c))_{i+}^j$ ,  $\mathcal{F}_{i+}^j = e_1(\mathbb{f}_{11}^* f_1 + \mathbb{f}_{12}^* f_2)_{i+}^j + (\mathbb{f}_{21}^* f_1 + \mathbb{f}_{22}^* f_2)_{i+}^j$ . На характеристиках второго семейства:

$$-e_2 \frac{f_{1i}^{j-1} - f_{1i-}^j}{-\Delta t_j} - \frac{f_{2i}^{j-1} - f_{2i-}^j}{-\Delta t_j} + \mathcal{F}_{i-}^j = 0,$$

где  $e_2 = (B(w - c))_{i-}^j$ ,  $\mathcal{F}_{i-}^j = e_2(\mathbb{f}_{11}^* f_1 + \mathbb{f}_{12}^* f_2)_{i-}^j + (\mathbb{f}_{21}^* f_1 + \mathbb{f}_{22}^* f_2)_{i-}^j$ .

Откуда получаем:

$$f_{1i}^{j-1} = (e_1 f_{1i+}^j - e_2 f_{1i-}^j + f_{2i+}^j - f_{2i-}^j - (\mathcal{F}_{i+}^j - \mathcal{F}_{i-}^j) \Delta t_j) \frac{1}{e_1 - e_2},$$

$$f_{2i}^{j-1} = f_{2i-}^j - e_2 (f_{1i}^{j-1} - f_{1i-}^j) - \mathcal{F}_{i-}^j \Delta t_j, \quad i = 1 \dots n - 1, \quad j = m - 1 \dots 0.$$

На левую границу, при  $i = 0$ , приходят характеристики первого семейства  $\xi_1$ , а на правую, при  $i = n$ , — второго семейства  $\xi_2$ . С учётом граничных условий (8.14), (8.15), которые берём на новом временном слое  $j - 1$ , получим на левой границе:

$$f_{1i}^{j-1} = (e_1 f_{1i+}^j + f_{2i+}^j - \mathcal{F}_{i+}^j \Delta t_j) \frac{1}{e_1 - 2(wB)_i^{j-1}},$$

$$f_{2i}^{j-1} = -2(wB)_i^{j-1} f_{1i}^{j-1}, \quad i = 0.$$

На правой границе:

$$f_{1i}^{j-1} = - \left( 2 \frac{Z - Z_e}{B(c^2 - w^2)} \right)_i^{j-1},$$

$$f_{2i}^{j-1} = f_{2i-}^j - e_2 (f_{1i}^{j-1} - f_{1i-}^j) - \mathcal{F}_{i-}^j \Delta t_j, \quad i = n.$$

Начальные, т.е. здесь терминальные, условия (8.16) известны:

$$f_{1i}^m = f_{2i}^m = 0, \quad i = 0 \dots n.$$

Осталось реализовать численное интегрирование на неравномерной сетке  $\{t_j\}_{j=0}^m$  для градиента

$$\nabla J(u; x_i) = \int_{t_0}^{t_1} \frac{2F_{\text{Тр}}(x_i, t)}{u(x_i)} f_1(x_i, t) dt \in U^*(S)$$

в пространственных точках  $x_i, i = 1 \dots n - 1$ .

Следует так же упомянуть, что частные производные  $\left. \frac{\partial \omega}{\partial x} \right|_Z, \left. \frac{\partial B}{\partial x} \right|_Z, \left. \frac{\partial R}{\partial Z} \right|_Z$  в свободных членах  $\mathcal{F}$  обеих задач следует реализовывать численным дифференцированием. Например, для точки  $i +$  данные производные могут иметь вид:

$$\begin{aligned} \left( \left. \frac{\partial \omega}{\partial x} \right|_Z \right)_{i+}^j &= \frac{\omega(x_{i+1}, Z_i^j) - \omega(x_i, Z_i^j)}{\Delta x}, \\ \left( \left. \frac{\partial B}{\partial x} \right|_Z \right)_{i+}^j &= \frac{B(x_{i+1}, Z_i^j) - B(x_i, Z_i^j)}{\Delta x}, \\ \left( \left. \frac{\partial R}{\partial Z} \right|_Z \right)_{i+}^j &= \frac{R(x_{i+}, 1.1Z_{i+}^j) - R(x_{i+}, Z_{i+}^j)}{0.1Z_{i+}^j}. \end{aligned}$$

#### 8.4.2. Оценка выпуклости целевого функционала

Для проведения тестовых расчётов был взят канал, соответствующий одному из участков Северо-Крымского канала со следующими характеристиками: длина канала 20 км; трапециевидальная форма поперечного сечения с шириной по дну 30 м и откосом стенок 4 (котангенс угла наклона); ровное дно с уклоном  $i = \frac{\partial Z_{\text{дно}}}{\partial x} = 0.0001$ ; боковой приток отсутствует, т.е.  $q = 0$ .

Расчёты проводились на пространственной сетке с количеством шагов  $n=40$ . При этом  $\Delta x = 500$  м, а «вычислительное» множество

$$S = (x_a, x_b) = [\Delta x, x_b - \Delta x] = [500, 19500] \text{ м.}$$

Начальные условия при  $t_0 = 0$  рассчитывались следующим образом. Задавался начальный расход воды вдоль  $Q_0(x) = 107 \text{ м}^3/\text{с}$  и уровень воды в верхнем (левом) створе  $Z_0(x_a) = Z_{\text{дно}}(x_a) + 3.6 \text{ м}$  (глубина 3.6 м). Далее рассчитывался уровень вдоль всего русла из условия установившегося течения (8.23) методом Рунге-Кутты 4-го порядка при тестовом значении  $u_* = 0.025$ . При таком коэффициенте шероховатости уровень вдоль имел физически разумное значение. На правой границе он достигал примерно 4.5 м.

После чего, на границах русла задавались граничные условия  $Z(x_a, t) = Z_0(x_a)$ ,  $Q(x_b, t) = Q_0(x_b)$  и рассчитывалось нестационарное течение. Получаем в нижнем створе русла «экспериментально» наблюдаемый уровень  $Z_e(t) = Z_0(x_b)$ , т.е. равный уровню стационарного течения справа, с точностью до вычислительных погрешностей. Шаг по времени определялся условием устойчивости и равнялся  $\Delta t \approx 1.2$  мин, а количество шагов  $m = 60$ . Теперь, задавая другие шероховатости, можно формировать различные значения целевого функционала  $J$ .

Оценим выпуклость функционала  $J$  как мы это делали в предыдущих задачах. Зададим вокруг оптимального управления  $u_* = 0.025$  два множества управлений  $\{u\}_1 = \{0.016, 0.019, 0.022\}$  и  $\{u\}_2 = \{0.028, 0.031, 0.034\}$ . Применим квадратичное интерполирование значений функции  $J_1(u)$  на множестве  $\{u\}_1$  и функции  $J_2(u)$  на множестве  $\{u\}_2$ .

На рисунке 8.5 показаны полученные интерполяционные параболы. Как и следовало ожидать, они не совпадают друг с другом, исследуемый целевой функционал *неквадратичный*, но *выпуклый*.

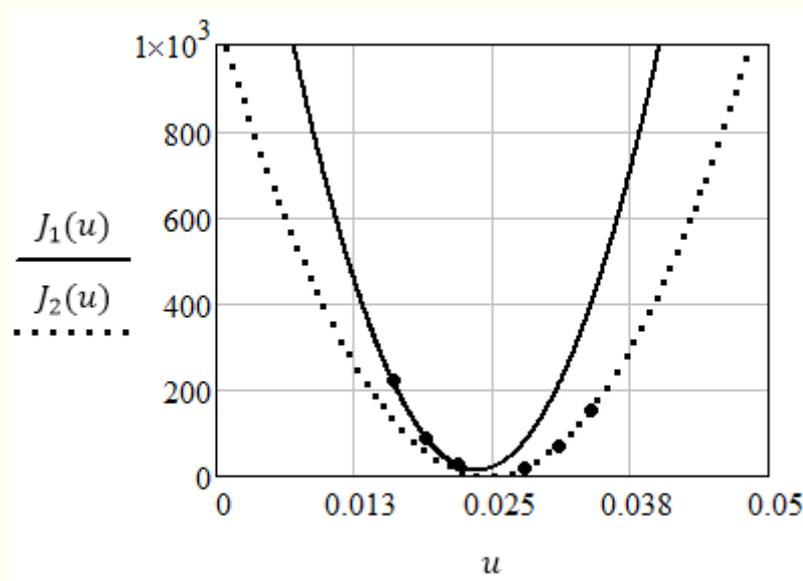


Рисунок 8.5 — Оценка квадратичности целевого функционала

При повышении шероховатости, т.е. при управлении  $u > 0.04$  уровень воды в конце русла опускается ниже линии дна, что недопустимо. Поэтому, исходя из условия (8.19) корректности решения исходной задачи, будем считать компактом корректности, достаточным для решения рассматриваемой задачи идентификации, следующее множество

$$\mathcal{U} = [0.001, 0.04].$$

Степень неквадратичности целевого функционала  $J(u)$  на  $\mathcal{U}$  представляется незначительной для возможности применения МРНС с постоянным параметром  $\alpha^k = \alpha^0$ . Такое допущение подтвердили дальнейшие расчёты. Целевой функционал  $J(u)$  на  $\mathcal{U}$  проявил себя как *локально квадратичный* на полученных траекториях спуска к  $u_*$ .

### 8.4.3. Решение тестовой задачи идентификации

Тестовое оптимальное значение коэффициента шероховатости зададим в следующем виде:

$$u_*(x) = 0.025 + 0.0075 \cos\left(\pi + \frac{2\pi x}{x_b - x_a}\right) \text{ на } S.$$

Для него из (8.23) находим начальное состояние потока. Далее по (8.2)–(8.4) пересчитываем уровень воды  $Z(x_b, t)$  на правой границе русла, который будем считать экспериментально наблюдаемым  $Z_e(t)$ . Задаём начальное приближение:

$$u^0(x) = 0.04 \in \mathcal{U}.$$

Начинается интенсивное нестационарное течение. Теперь можно решать обратную задачу идентификации шероховатости  $u(x)$  по целевому функционалу (8.5).

Время прохода характеристики  $\xi_1$  через всё русло составляло примерно 58 мин. В вычислительном эксперименте было выявлено, что для уверенного прихода волны (заметное изменение уровня воды в нижнем створе) и уверенного возвращения сопряжённой волны (заметное изменение  $f_1$  в верхнем створе) следует задавать

$$t_\varepsilon \geq 10 \text{ мин.}$$

Поэтому, с учётом условия идентифицируемости (8.21), было выбрано  $t_\varepsilon = 15$  и использовано следующее минимально достаточное для  $\Omega$  множество

$$\omega_{min} = x_b \times (t_0 = 0, t_1 = 73 \text{ мин}).$$

Именно для этого времени была задана сетка  $m = 60$ .

Уменьшение времени управляемости, например, до  $t_1 = 30$  мин не позволило сделать ни одной разумной коррекции управления  $u^k$ . Увеличение же времени до  $t_1 = 150$  мин заметно ухудшило качество

восстановления  $u_*(x)$ . В обоих примерах множество  $\omega$  существенно не соответствовало условию управляемости, что сделало область  $V^*(\Omega)$  определения оператора  $U^*$  некорректной для однозначного отображения в область значений  $U^*(S)$ . В первом случае множество  $\omega$  было *недостаточным*, во втором — *избыточным*. Везде не было выполнено первое условие теоремы 4.1 (существует область  $V^*(\Omega)$  определения оператора  $U^*$  с его значениями в области  $U^*(S)$ ).

Точность идентификации задавалась относительно высокой степенью показаний гидротехнического оборудования:

$$\Delta_{max} Z = \max_t |Z(x_b, t) - Z_e(t)| \leq 0.01 \text{ м.}$$

Были реализованы алгоритмы МНС (1.8), МСГ (3.10), МРНСГ (3.4), МРНСсг (3.11). В МСГ и МРНСсг, использующих сопряжённые направления, были заданы моменты обновления направлений спуска на каждой 15-й итерации.

Выбор шагового множителя  $b^k$  в алгоритмах осуществлялся двумя вариантами стратегий.

Первый вариант — это *четвёртая стратегия* по методу Носедала-Райта с коэффициентом Вольфе  $c_2 = 0.9$  для МНС,  $c_2 = 0.1$  для МСГ и  $c_2 = 0.5$  для МРНСГ, МРНСсг. Максимальный шаг в методе Носедала-Райта задавался равным

$$c_3 \|u^0 - u_*\|, \quad 0 < c_3 < 1.$$

Конкретное наилучшее значение  $c_3$  для каждого расчёта подбиралось индивидуально. Большое значение коэффициента  $c_3$  ускоряет сходимость, но чаще всего  $c_3$  приходилось выбирать очень малым, поскольку при движении к экстремуму, разные методы минимизации могут по-разному и значительно изменять функцию  $u^k(x)$ , выводя её из компакта корректности  $\mathcal{U}$ . Где-либо вдоль русла могут наблюдаться

$u^k < 0$  или уровень  $Z^k$  может опуститься ниже линии дна. Возможно даже появление *сверхзвуковых течений* при приближении уровня потока к линии дна, поскольку требование неразрывности потока в ситуации  $Z^k \rightarrow Z_{\text{дно}}$  (сечение потока  $\omega \rightarrow 0$ ) значительно увеличивает расход воды  $Q^k$  и в итоге оказывается скорость потока  $w = Q/\omega > c = \sqrt{g\omega/B}$ .

Второй вариант выбора шагового множителя  $b^k$  — это *вторая стратегия* адаптивной релаксации по формуле (1.14), которая применялась только для МНС и МРНСг. Параметры  $b_1 = 1.05$ ,  $b_2 = 0$ . В этой стратегии, так же, как и в предыдущей, важно не делать значительных усиления шагов минимизации посредством коэффициента  $b_1$ . При больших значениях  $b_1$  будут появляться те же негативные явления, что и при больших  $c_3$  в методе Носедала-Райта. Для традиционного МНС  $b^0 = 0.1 \frac{\|u^0\|}{\|\nabla J^0\|}$ , для МРНСг глубина первого шага определяется параметром  $\alpha(x)$  при  $b^0 = 1$ .

Для реализации МРНСг и МРНСсг зададим *шаблон*  $\tilde{u}^0$  в виде изменение управления «под 45°» (3.7) или пропорциональное изменение (3.8). Для выбранной начальной функции  $u^0(x)$  они дают одинаковый результат. Зададим  $\delta = 0.1u^0$  или  $\lambda = 0.1$  и получим параметр регулирования направления спуска

$$\alpha(x) = \left| \frac{0.1u^0}{\|\nabla J(u^0; x)\|} \right|.$$

На рисунках 8.6 и 8.7 представлены результаты идентификации коэффициента шероховатости. В таблице 8.1 сведены количественные результаты минимизации целевого функционала.

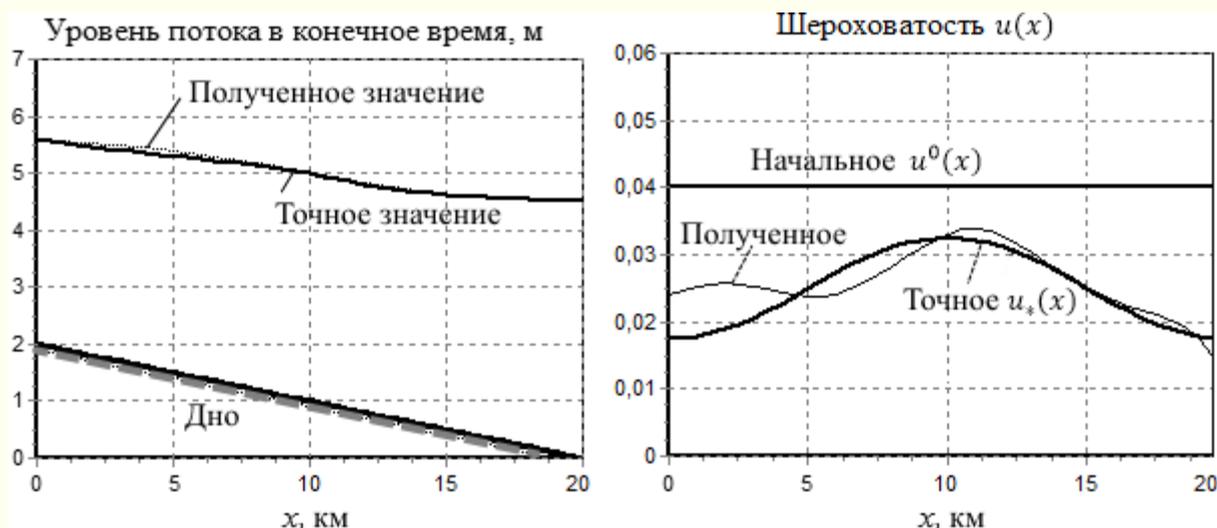


Рисунок 8.6 — Результаты идентификации шероховатости русла посредством МНС, МСГ

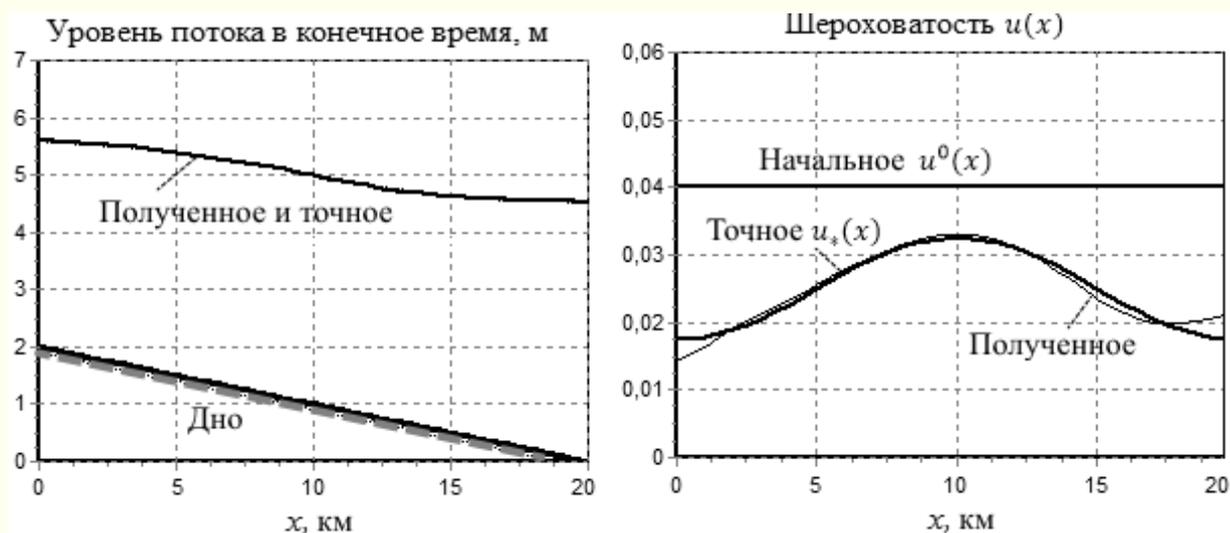


Рисунок 8.7 — Результаты идентификации шероховатости русла посредством МРНСГ, МРНССГ

Все рассмотренные методы, с шаговым множителем  $b^k$  из четвёртой стратегии, продемонстрировали равномерную сходимость к оптимуму  $u_*$ , что и следовало ожидать из анализа идентифицируемости. Диссипативные члены для выбранной длины русла и уклона дна не

проявили себя заметно негативно. При использовании адаптивной стратегии выбора  $b^k$  традиционный МНС не смог достичь заданной точности  $\Delta_{max}Z = 0.01$  м.

Таблица 8.1 — Результаты идентификации шероховатости русла

Метод	Стратегия выбора $b^k$ , значение $c_3$	Итераций $k$ / вычислений $J$ и $\nabla J$	$\ u^k - u_*\ $
МНС	2	—	—
	4, $c_3 = 0.01$	152/449	0.45
МСГ	4, $c_3 = 0.0001$	83/423	0.44
МРНСГ	2	54/111	0.16
	4, $c_3 = 0.1$	60/207	0.18
МРНСсг	4, $c_3 = 0.1$	45/634	0.18

Из рисунков и таблицы видно, что наилучшие результаты показал МРНСГ. Он потребовал меньше всех вычислений и добился наилучшего восстановления тестового значения шероховатости  $u_*$ . Методы, основанные на сопряжённых градиентах (МСГ, МРНСсг), в рассматриваемой задаче, по сути, не продемонстрировали свои потенциальные преимущества. Далее мы их не рассматривали.

Были сделаны тесты на работоспособность МНС и МРНСГ при помехах в наблюдении уровня  $Z_e$ . Были введены *абсолютные случайные помехи* (1.20) в виде:

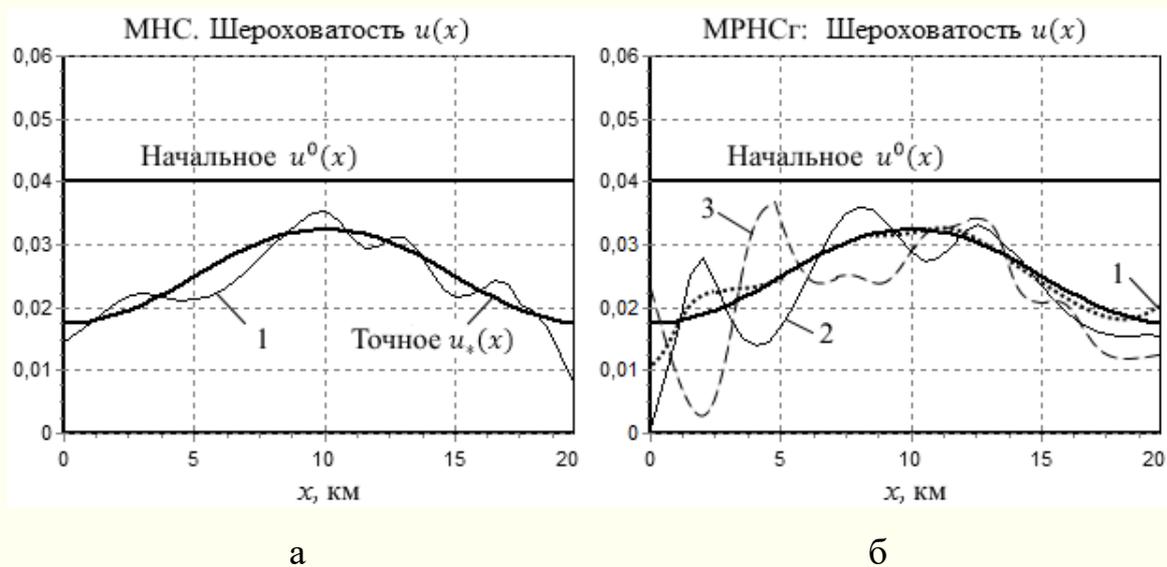
$$Z_e(t) = Z_*(t) + 2\Delta Z[\xi(t) - 0.5], \quad t \in (t_0, t_1),$$

где  $\xi(t) \in [0,1)$  — функция вещественных случайных чисел,  $\Delta Z$  — максимально возможная погрешность уровня. Для  $\Delta Z$  было сделано три варианта зашумления: 2 см; 5 см; 10 см. Это высокий уровень зашумления, поскольку измерительная аппаратура имеет погрешность

не более одного сантиметра. В этих расчётах, для завершения итераций, использовался критерий практического прекращения сходимости (1.24):

$$\frac{\|u^k - u^{k-1}\|_{L_2(S)}}{\|u^{k-1}\|_{L_2(S)}} \leq \varepsilon_4, \quad \varepsilon_4 = 10^{-6}.$$

На рисунке 8.8а показаны результаты идентификации посредством МНС. Он продемонстрировал работоспособность только при минимальном зашумлении  $\Delta Z = 2$  см. При большем зашумлении в процессе итераций управление при любом значении  $c_3$  выходило из компакта корректности, получалось  $u(x \approx x_b) < 0$ .



Помехи: 1 —  $\Delta Z = 2$  см; 2 —  $\Delta Z = 5$  см; 3 —  $\Delta Z = 10$  см

Рисунок 8.8 — Результаты идентификации шероховатости со случайными помехами

На втором рисунке 8.8б показаны результаты идентификации посредством МРНСГ с адаптивной релаксацией  $b^k$ . Он продемонстрировал прекрасную устойчивость при всех вариантах

зашумления. Количество итераций и вычислений, при разных запусках генератора  $\xi$ , в МРНСГ оказалось в несколько раз меньше чем в МНС. В МРНСГ всегда получаемое значение  $\Delta_{max}Z < \Delta Z$ , т.е. отклонения состояния модели от наблюдений с помехами во всех расчётах не превосходило сами помехи.

Теперь рассмотрим влияние длины русла и уклона дна на качество идентификации. Для оценки влияние данных параметров достаточно применения только МНС. Шаговый множитель  $b^k$  будем выбирать по методу Носедала-Райта при  $c_3 = 0.0001$ .

Начнём с длины. Возьмём её в два раза большей чем ранее, т.е. 40 км. Соответственно, зададим сетку с  $n=80$ . При этом для того, чтобы устранить влияние диффузорности потока из-за излишнего уклона дна, зададим его чуть меньшим, пусть  $i = 0.00008$ . В этих условиях, время прохода характеристики  $\xi_1$  через всё русло составляло 120 мин.

Диссипация энергии волн существенно снижала их амплитуду и размывала фронт. Это негативное явление потребовало увеличить время наблюдения прихода волны до  $t_\varepsilon = 35$  мин. То есть, время наблюдения  $t_1$  течения воды в русле было увеличено от 120 мин до 155 мин ( $m = 130$ ), что соответствует примерно середине растянувшегося фронта. То есть расчёты проводились с множеством

$$\omega_{min} = x_b \times (0, t_1 = 155 \text{ мин}).$$

Идентификация по началу фронта (120 мин) вообще не сработала.

На рисунке 8.9 показаны полученные результаты идентификации при завершении сходимости МНС.

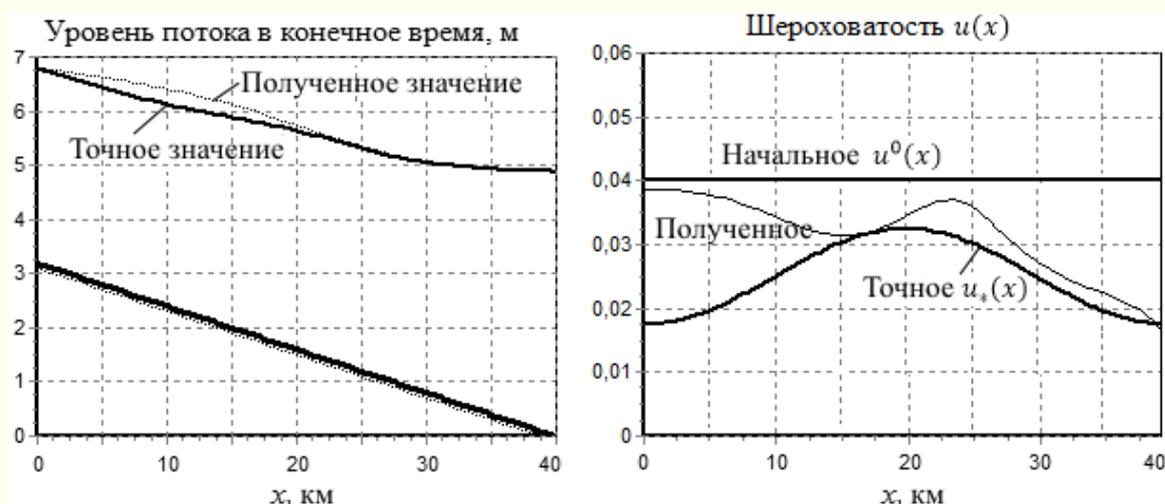


Рисунок 8.9 — Результаты идентификации шероховатости длинного русла посредством МНС

Мы видим, что качество идентификации ухудшилось, особенно в верхнем (левом) створе канала, что предсказывалось при анализе идентифицируемости. Чтобы достичь заданную точность  $\Delta_{max}Z$  для МНС потребовалось итераций  $k = 390$  и вычислений  $J$ , и  $\nabla J$  потребовалось 945. При этом была получена близость к точному управлению  $\|u^k - u_*\|_{L_2(S)} = 1.9$ , что почти на порядок хуже, чем на коротком канале.

С точки зрения идентифицируемости, мы столкнулись с ситуацией близкой к потере области  $V^*(\Omega)$  определения оператора  $\mathcal{U}^*$ , что стало следствием диссипации волновых процессов и привело к заметному ухудшению точности отображений, как минимум, при  $u \xrightarrow{\text{однозначно}} I'_Z|_{\omega}$ .

На рисунке 8.10 показаны результаты идентификации для русла предыдущей длины в 20 км ( $n=40$ ,  $m = 60$ ), но с увеличенным уклоном дна  $i = 0.00018$ . Из правой части рисунка видно, что в этом случае искомая функция шероховатости  $u_*(x)$  не идентифицируется вообще.

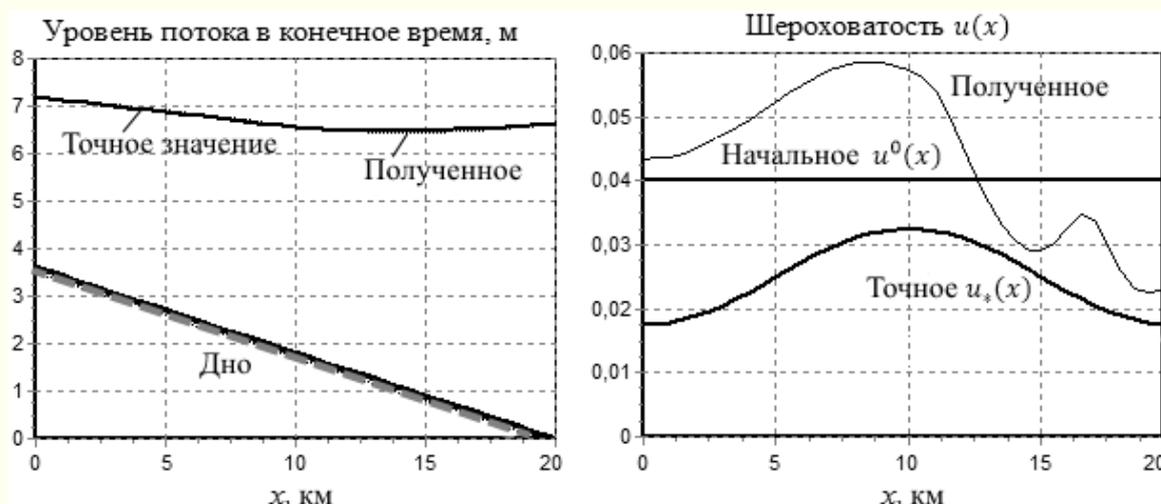


Рисунок 8.10 — Результаты идентификации шероховатости русла посредством МНС с увеличенным уклоном дна

Из левой части рисунка видно, что шероховатость в русле практически не влияет на состояние модели. Точное и полученное (при неправильном  $u$ ) значения уровня воды мало различимы. Появившаяся нечувствительность модели к параметру  $u(x)$  объясняется тем, что рассматриваемый поток не только не является одномерным, но даже и квазиодномерным. Площадь  $\omega$  поперечного сечения потока в нижнем створе русла почти в три раза больше, чем в верхнем створе. Возникающий диффузор потока (особенно в правой половине течения) существенно повышает диссипацию энергии волн. В частности, значение члена трения  $F_{\text{тр}}$  в нижнем створе канала оказывается в пять раз выше, чем в верхнем. По сути, от волн ничего не остаётся.

В данном примере, с точки зрения идентифицируемости, мы столкнулись с ситуацией полной потери области  $V^*(\Omega)$  определения оператора  $U^*$ . Наличие существенной диссипации уничтожало зависимость целевого функционала от управления, т.е. исчезло

отображение  $u \xrightarrow{\text{однозначно}} I'_Z|_{\omega}$ . Не было выполнено первое условие теоремы 4.1.

Таким образом, во-первых, во всех тестовых расчётах наилучшую эффективность и точность идентификации распределённого коэффициента шероховатости  $u(x)$  продемонстрировал МРНСт. Он так же показал прекрасную устойчивость к абсолютным случайным помехам.

Во-вторых, подтверждено, что на участках русла с высокой диссипацией волновых процессов идентификация  $u(x)$  в рассматриваемой модели невозможна.

В-третьих, изменение множества  $\omega$  в целевом функционале  $J$  больше или меньше  $\omega_{min}$  изменяет область определения оператора  $\mathbb{U}^*$  вплоть до  $J'_u \neq \nabla J$ , что принципиально влияет на возможность решения задачи идентификации.

#### **8.4.4. Идентификация шероховатости Каракумского канала**

Был выбран 128-километровый участок Каракумского канала. Это судоходный канал, он отводит воду из реки Аму-Дарьи с расходом около  $600 \text{ м}^3/\text{с}$  на расстояние 1445 км.

Исходная информация для идентификации была получена в институте «Союзгипроводхоз». Канал был представлен в трёх руслообразующих сечениях  $x = \{0 \text{ км}, 64 \text{ км}, 128 \text{ км}\}$  неправильной формы, показанных на рисунке 8.11. В этих же сечениях осуществлялись измерения уровня  $Z_e$  и расхода  $Q$  воды стационарного течения, которые приведены в таблице 8.2 для случая минимального наполнения русла. Материалом русла в начальном участке является супесь, в среднем и конечном — песок. По берегам канала имеется незначительная растительность в виде камыша. Всё это влияет на шероховатость канала.

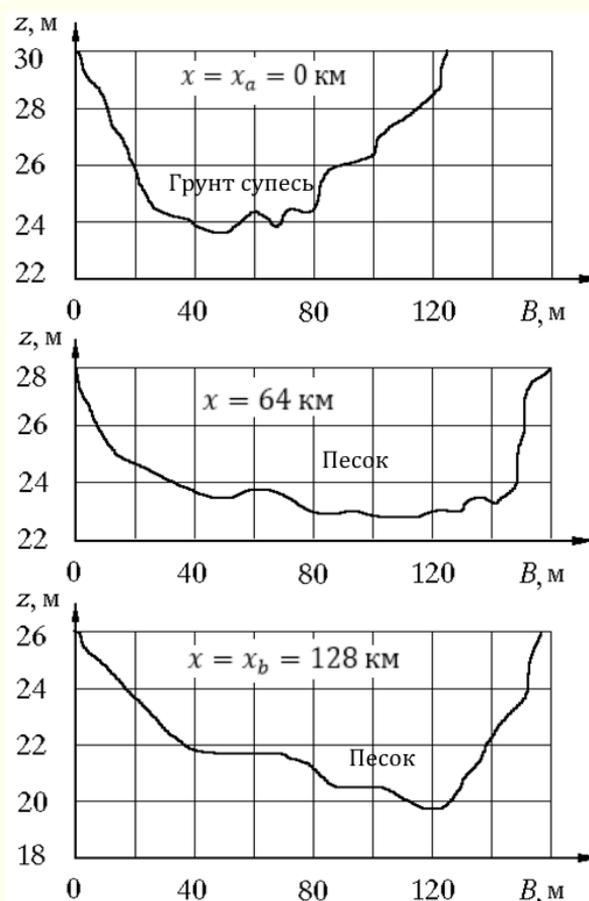


Рисунок 8.11 — Поперечные сечения исследуемого участка  
Каракумского канала

Таблица 8.2 — Данные о потоке воды на исследуемом участке  
Каракумского канала

$x, \text{ км}$	0	64	128
$Z_e, \text{ м}$	28.2	26.7	23.9
$Q, \text{ м}^3/\text{с}$	183	183	175

Канал разбивался сеткой  $\{x_i = i\Delta x, i = 0 \dots n = 8, \Delta x = 16 \text{ км}\}$ .  
Начальные значения уровня и расхода в сечениях определялись

графической интерполяцией по данным таблицы 8.2 из соображения плавности. Линия дна  $Z_{\text{дно},i}$  строилась эквидистантной относительно полученной линии уровня при глубине минимального наполнения. Форма поперечных сечений русла в виде  $B(x_i, Z_{l,i})$ ,  $i \in \{0, 4, n\}$  аппроксимировалась по рисунку 8.11 на сетке уровней  $\{Z_{l,i} = Z_{\text{дно},i} + l\Delta Z, l = 0, 1, 2, \dots, \Delta Z = 1 \text{ м}\}$ . Сечения  $B(x_i, Z_{l,i})$  в точках  $x_i$ , не представленных на рисунке 8.11, строились из условия линейного изменения ширины русла между известными сечениями.

Граничные условия, в отличие от (8.3), задавались расходом воды в обеих створах:

$$\begin{aligned} Q(x_a, t) &= Q_a(t) \text{ на } \Gamma_a = x_a \times (t_0, t_1), \\ Q(x_b, t) &= Q_b(t) \text{ на } \Gamma_b = x_b \times (t_0, t_1). \end{aligned}$$

Распределенный боковой расход определялся выражением

$$q_i = \frac{Q_i - Q_{i-1}}{\Delta x}, \quad i = 1 \dots n.$$

Целевой функционал задавался в верхнем и в нижнем створах канала, т.е. вместо (8.5), имеем

$$J(u) = \int_{t_0}^{t_1} (Z - Z_e)|_{\omega}^2 dt, \quad \omega = \Gamma_a \cup \Gamma_b. \quad (8.24)$$

Наблюдения уровня воды в среднем створе, при  $x = 64$  км, использовались только для аппроксимации начального состояния потока.

Полученная ранее первая вариация целевого функционала (8.12), теперь, для функционала (8.24), примет новый вид на левой границе  $\Gamma_a$ :

$$\langle -A^T f + I'_v, (0, \delta Z) \rangle_{V^*(\Gamma_a)} = 0.$$

Откуда следует граничное условие в верхнем створе канала в виде:

$$B(c^2 - w^2)f_1 - 2(Z - Z_e) = 0 \text{ на } \Gamma_a.$$

В нижнем створе условие останется прежним:

$$B(c^2 - w^2)f_1 + 2(Z - Z_e) = 0 \text{ на } \Gamma_b.$$

Для функционала (8.24) изменится условие идентифицируемости (8.20). Для  $t_0 = 0$ , оно примет вид [87]:

$$t_1 = \int_{x_c}^{x_a} \frac{dx}{\lambda_2} = \int_{x_c}^{x_b} \frac{dx}{\lambda_1}. \quad (8.25)$$

Смысл данного условия иллюстрируется рисунком 8.12. Координата  $x_c$  соответствует пересечению замыкающих характеристик  $\xi_2$  и  $\xi_1$ , выходящих, соответственно, из верхнего  $x_a$  и нижнего  $x_b$  граничных створов канала.

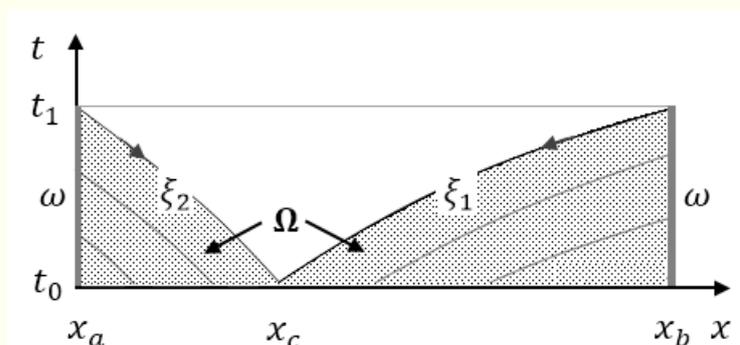


Рисунок 8.12 — Пространственно-временная диаграмма области  $\Omega$  при наблюдении уровня воды в двух створах канала

Время наблюдения течения воды в канале задавалось в соответствии с условием (8.25) и составляло пять часов. В этот промежуток включён и относительно небольшой отрезок времени  $t_\varepsilon$ , когда на обеих границах канала наблюдался заметный приход волн. Шаг по времени  $\Delta t \approx 30$  мин, а количество шагов по времени  $m = 10$ . При этом множество

$$\omega = (x_a \cup x_b) \times (0, t_1 = 5 \text{ час}) = \omega_{min},$$

что обеспечило корректное отображение оператора  $U^*: V^*(\Omega) \rightarrow U^*(S)$ .

Было задано начальное приближение коэффициента шероховатости

$$u^0(x) = 0.015.$$

При расчете нестационарного течения с таким управлением максимальное расхождение расчетных  $Z$  и экспериментальных  $Z_e$  уровней в граничных створах за пять часов расчета доходило до 1.8 м. Это очень много.

Для идентификации коэффициента шероховатости использовался МРНСг с параметром регулирования направления спуска

$$\alpha(x) = \left| \frac{0.2u^0}{\nabla J(u^0; x)} \right|$$

и адаптивным выбором шагового множителя по методу (1.14). Сходимость метода заканчивалась менее чем за 20 итераций.

На рисунке 8.13 представлено полученное оптимальное значение коэффициента шероховатости исследуемого участка Каракумского канала. Видно, что шероховатость непостоянна по длине канала.

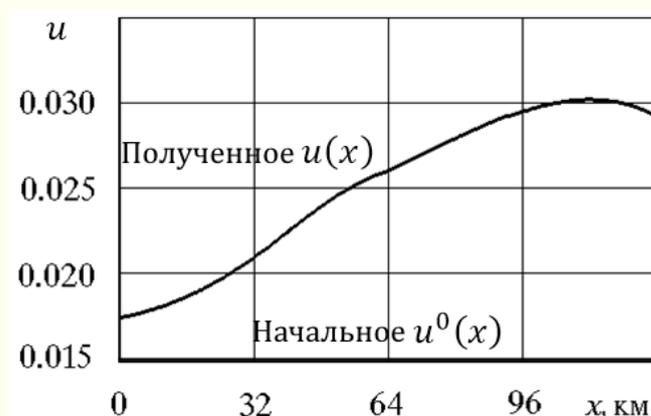


Рисунок 8.13 — Идентифицированное значение коэффициента шероховатости участка Каракумского канала

Расхождение между экспериментальным и рассчитанным уровнями в граничных створах не превосходило 5 см. Данная точность расчета уровня нестационарного потока воды, при заданной крайне малой и приблизительной исходной информации, является полностью удовлетворительной.

## 9 Оптимизация формы сопла гидропушки

### 9.1 Постановка задачи

Импульсные струи воды широко используются в промышленности, в первую очередь, для разрушения пород. Одной из установок для получения импульсных струй высокого давления (ультраструй) является гидропушка.

Простейшая схема гидропушки безударного действия приведена на рисунке 9.1. Поршень 2 под действием газа в ресивере 1 разгоняется, толкая перед собой воду 3 из цилиндрического ствола 4 в сужающееся сопло 5. Ускоряясь, передний фронт воды может достигать скорости близкой или даже превышающей скорость звука в воде. Задний фронт жидкости, замедляясь, тормозит поршень.

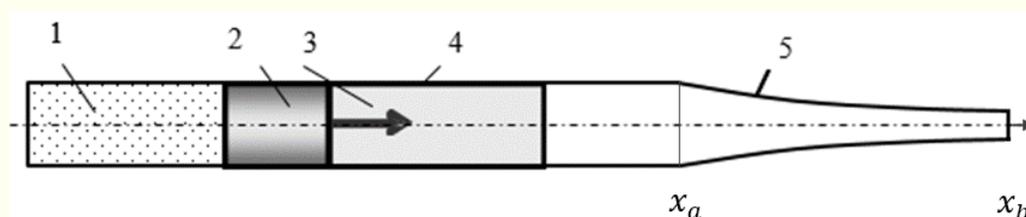


Рисунок 9.1 — Схема поршневой гидропушки

На формирование ультраструи существенное влияние оказывает форма сопла. Обычно её задают через площадь поперечного сечения  $\sigma(x)$ ,  $x \in [x_a, x_b]$ . Ствол гидропушки имеет постоянную площадь сечения  $\sigma(x) = \sigma_a$ . В струе определяющую роль играет не вся струя, а только ее начальный высокоскоростной участок от момента времени  $t_1$  начала истечения и до некоторого  $t_2$ . Для разрушения пород важна сила импульса струи, с которой она может действовать на преграду. Среднее

значение этой силы для высокоскоростного участка струи на срезе сопла  $x_b$  можно представить выражением [73]:

$$J = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \rho \sigma w^2 dt|_{x_b}, \quad (9.1)$$

где  $\rho$ ,  $w$  — плотность и скорость, которые представляют собой состояние потока воды в гидропушке.

На рисунке 9.2 изображена пространственно-временная диаграмма течения воды. Начало координат совмещено со входом в сопло  $x_a$ . Передний фронт  $\Gamma_{b_0}$  потока воды в сопле начинается с момента  $t_0$  и заканчивается истечением из сопла на  $\Gamma_{b_1} = x_b \times (t_1, t_2)$ . Правая граница потока — это  $\Gamma_b = \Gamma_{b_0} \cup \Gamma_{b_1}$ , левая граница  $\Gamma_p$  — это траектория движения поршня. Границы  $\Gamma_p(t)$  заднего и  $\Gamma_{b_0}(t)$  переднего фронта воды подвижные. Область определения состояния потока воды — это замкнутая пространственно-временная область  $\Omega$  с границей  $\partial\Omega = \Gamma_p \cup \Gamma_b \cup \Gamma_0 \cup \Gamma_2$ ,  $\Gamma_0 = [x_{p0}, x_a] \times t_0$  — линия начальных условий,  $\Gamma_2 = [x_{p2}, x_b] \times t_2$  — терминальная линия процесса. Координаты  $x_{p0}$  и  $x_{p2}$  — начальное и конечное положение поршня гидропушки,

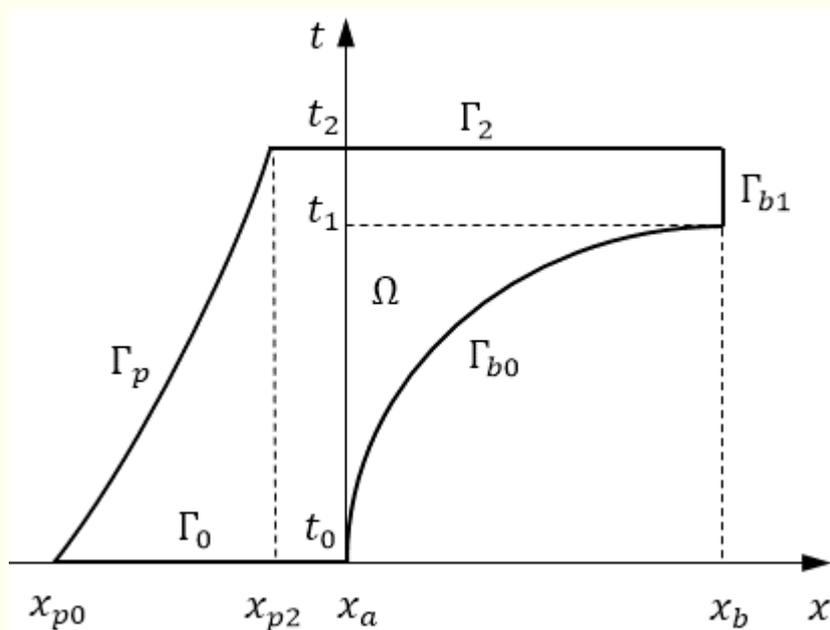


Рисунок 9.2 — Пространственно-временная диаграмма течения воды в гидропушке

Для описания процессов течения воды в сопле обычно используют уравнения неразрывности, баланса импульса и состояния, описывающие *изоэнтропические* гидродинамические процессы, в которых жидкость считается невязкой, сжимаемой. Подобное течение в сужающемся канале можно описать *квазиодномерной*, нестационарной моделью потока в виде *квазилинейной, гиперболической, неоднородной* системы уравнений [73]. Соответствующие уравнения *исходной задачи* имеют вид:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + w \frac{\partial \rho}{\partial x} + \rho \frac{\partial w}{\partial x} + \varphi &= 0, \\ \frac{\partial w}{\partial t} + \frac{c^2}{\rho} \frac{\partial \rho}{\partial x} + w \frac{\partial w}{\partial x} &= 0, \quad (x, t) \stackrel{\text{def}}{=} \tau \in \Omega. \end{aligned} \quad (9.2)$$

Здесь  $\varphi = \frac{\rho w}{\sigma} \frac{d\sigma}{dx}$ ,  $c^2 = \frac{Bn\rho^{n-1}}{\rho_0^n}$  — скорость звука в вводе,  $B$ ,  $n$  — постоянные в уравнении состояния воды в форме Тэта. Система

уравнений неоднородная, но без диссипации энергии из-за изоэнтропического течения.

Мы будем рассматривать *дозвуковые течения* (докритические течения) воды, когда

$$w < c_0 = 1475 \frac{\text{м}}{\text{с}}.$$

При сверхзвуковом истечении, когда  $w \geq c_0$ , струя разрушается. Сверхзвуковое течение является индикатором неудачной формы сопла (неудачной оптимизации).

Кроме того, принятая математическая модель не допускает появления *кавитации*<sup>1</sup>, поэтому

$$\text{если } \rho < \rho_0 \text{ тогда } \rho \leftarrow \rho_0.$$

В работе [40] в качестве управления было выбрано  $u = \sigma$  и получены необходимые условия оптимальности. Как показали дальнейшие тестовые расчёты, незначительные погрешности в форме сопла  $\sigma(x)$  приводят к значительным погрешностям значений производной  $d\sigma/dx$  и, соответственно, к значительным погрешностям в расчётах состояния потока и оптимальной формы сопла. Такие «естественные» погрешности характерны для численного дифференцирования функций с помехами.

В рассматриваемой задаче помехи оказались недопустимо высокими. Поэтому было принято использовать следующее управление в уравнениях (9.2):

$$u(x) = \frac{1}{\sigma(x)} \frac{d\sigma(x)}{dx}, \quad x \in S = (x_a, x_b). \quad (9.3)$$

При этом свободный член в уравнениях (9.2)

$$\varphi = \rho w u \theta(x - x_a), \quad x \in (x_{p0}, x_b),$$

<sup>1</sup> Простыми словами — это образование множества пузырьков пара в областях локального пониженного давления при неравномерном течении потока жидкости.

то есть  $\varphi$  определено и в стволе гидропушки, и в сопле.

Площадь сопла  $\sigma(x)$  находится через управление (9.3) по формуле:

$$\sigma(x) = \sigma_a e^{\int_{x_a}^x u(\zeta) d\zeta}, \quad x \in [x_a, x_b].$$

Уточним терминологию и обозначения поставленной задачи согласно предыдущим разделам теории прямого экстремального подхода.

Состояние потока в системе (9.2) — это *вектор-функция*  $v = (\rho, w) \in V \subset L_{2+}^m(\bar{\Omega})$ ,  $m = 2$ , замыкание  $\bar{\Omega} = \Omega \cup \partial\Omega$ . Плотность  $\rho$  и скорость потока  $w$  не могут быть отрицательными. Имеющаяся ограниченность  $V$  в виде дозвукового течения и отсутствия кавитации, не является ограничением для задачи оптимизации. Это всего лишь условия разумности решения исходной задачи. Управление  $u \in U \subset L_2(S)$ , множество  $S = (x_a, x_b)$ . Можно сказать, что  $S = P_x(\Omega|_{x>x_a})$ , т.е.,  $S$  является проекцией части (при  $x > x_a$ ) пространственно-временной области  $\Omega$  на ось  $x$ .

При итерационных коррекциях управления  $u^k$ , в принципе, может возникнуть ситуация расширения сопла, выходящее за пределы ствола гидропушки. Поэтому, необходимо ограничить область значений управления явным ограничением типа 2.1. Будем считать множеством допустимых управлений полупространство

$$U = \{u: u(x) \leq 0, \quad \forall x \in S\}. \quad (9.4)$$

Такому  $U$  соответствует  $\sigma(x) \leq \sigma_a$ , что является физически разумным.

Таким образом, исходная задача в операторном виде, аналогичном (3.20), принимает вид:

$$\mathbb{D}v + F = \frac{\partial v}{\partial t} + A \frac{\partial v}{\partial x} + F = 0 \quad \text{на } \Omega, \quad (9.5)$$

где дифференциальный оператор  $\mathbb{D} \cdot = \frac{\partial \cdot}{\partial t} + A \frac{\partial \cdot}{\partial x}$ , его матрица  $A$  и вектор-столбец свободного члена  $F$  имеют вид:

$$A(v) = \begin{pmatrix} w & \rho \\ \frac{Bn\rho^{n-2}}{\rho_0^n} & w \end{pmatrix}, \quad F(v, u) = \begin{pmatrix} \varphi \\ 0 \end{pmatrix}.$$

Оператор  $\mathbb{D}$  зависит от состояния  $v$  через матрицу  $A$ , а свободный член  $F$  — от  $v$  и управления  $u$ .

Модель состоит из двух уравнений с первыми производными и имеет соответственно два граничных условия на  $\Gamma_p$  и  $\Gamma_b$ . Слева на траектории поршня — это уравнение движения поршня, которое с учётом уравнения состояния Тэта принимает вид условия *второго рода*:

$$\frac{dw}{dt} + \frac{\sigma_a B}{m_p} \left( \left( \frac{\rho}{\rho_0} \right)^n - 1 \right) = 0 \text{ на } \Gamma_p. \quad (9.6)$$

Справа на границе  $x_b$ , взаимодействующей с атмосферой, мы имеем условие *первого рода*:

$$\rho = \rho_0 \text{ на } \Gamma_b. \quad (9.7)$$

Начальное условие:

$$v(x, t_0) = (\rho_0, w_0) \text{ на } \Gamma_0. \quad (9.8)$$

Здесь  $m_p$  — масса поршня,  $\rho_0$  — плотность воды при атмосферном давлении,  $w_0$  — скорость воды и поршня до начала втекания в сопла.

Теперь, на основании выражения (9.1) силы импульса струи гидропушки, запишем *целевой функционал* в виде второго граничного интеграла в общем выражении (3.19):

$$J(u) = \int_{t_1}^{t_2} I(v, u)|_{\omega} dt, \quad (9.9)$$

где

$$I(v, u; t)|_{\omega} = \frac{\rho_0 \sigma_b w_b^2(t)}{t_2 - t_1}, \quad \omega = x_b \times (t_1, t_2) = \Gamma_{b1},$$

значение  $\sigma_b = \sigma_a e^{\int_{x_a}^{x_b} u(x) dx}$ , скорость истечения  $w_b(t) = w(x_b, t)$ . Мы видим, что в функции цели  $I(v, u)$  на  $\omega$  явно присутствует управление через площадь среза сопла  $\sigma_b$ .

Таким образом, задача оптимизации формы сопла гидропушки для формирования ультраструи с максимальной силой импульса формулируется следующим образом. Для распределённой системы (9.5) с краевыми условиями (9.6)–(9.8) необходимо найти управление (9.3), которое доставляет максимум функционалу (9.9) с учётом ограничения на управление (9.4).

Подобная проблема оптимизации формы сопла гидропушки ранее обсуждалась в работах [73, 45, 4, 40, 46]. Однако эти работы ограничивались либо теоретическими формулировками НУО, либо численными решениями в виде неподтверждённой, якобы оптимальной формы сопла. В конце данного раздела будет получено действительно оптимальное сопло, достоверно подтверждённое условиями управляемости и расчётами с соответствующими экстремальными алгоритмами.

### 9.1.1. Характеристическая форма исходных уравнений

Как и всякая гиперболическая система, система (9.5) имеет характеристические направления  $\lambda_{1,2}$  описывающие волновые процессы в гидропушке. Характеристические направления  $\lambda_{1,2}$  являются собственными числами матрицы  $A$ :

$$\begin{vmatrix} w - \lambda & \rho \\ \frac{c^2}{\rho} & w - \lambda \end{vmatrix} = (w - \lambda)^2 - c^2 = 0.$$

Откуда следует

$$\lambda_{1,2} = w \pm c.$$

Это означает, что *характеристики*  $\xi_1$  и  $\xi_2$  в области  $\Omega$  представляют собой семейство двух типов линий  $\xi_1 = x_1(t)$  и  $\xi_2 = x_2(t)$  вдоль которых распространяются волны в гидропушке со скоростью

$$\lambda_1 = \frac{dx_1}{dt} = w + c \text{ вдоль } \xi_1, \quad \lambda_2 = \frac{dx_2}{dt} = w - c \text{ вдоль } \xi_2.$$

При дозвуковом течении воды волны первого семейства перемещаются вдоль потока воды к соплу гидропушки, а волны второго семейства — против течения к поршню.

Вдоль характеристических направлений  $\lambda_{1,2}$  гиперболическая система (9.5) принимает характеристическую форму. Она нам далее понадобится. Для её получения запишем исходную систему на *собственных векторах*  $\Lambda_{1,2}$  матрицы  $A$ . Начнём с  $\Lambda_1$  для  $\lambda_1 = w + c$ :

$$\Lambda_1^T \begin{pmatrix} w - \lambda_1 & \rho \\ \frac{c^2}{\rho} & w - \lambda_1 \end{pmatrix} = (\Lambda_{11}, \Lambda_{12})^T \begin{pmatrix} -c & \rho \\ \frac{c^2}{\rho} & -c \end{pmatrix} = (0, 0).$$

Получаем  $-\Lambda_{11}c + \Lambda_{12}\frac{c^2}{\rho} = 0$ ,  $\Lambda_{11}\rho - \Lambda_{12}c = 0$ . Пусть первая компонента вектора будет  $\Lambda_{11} = 1$ , тогда вторая  $\Lambda_{12} = \frac{\rho}{c}$ .

Теперь найдём  $\Lambda_2$  для  $\lambda_2 = w - c$ :

$$\Lambda_2^T \begin{pmatrix} w - \lambda_2 & \rho \\ \frac{c^2}{\rho} & w - \lambda_2 \end{pmatrix} = (\Lambda_{21}, \Lambda_{22})^T \begin{pmatrix} c & \rho \\ \frac{c^2}{\rho} & c \end{pmatrix} = (0, 0).$$

Получаем  $\Lambda_{21}c + \Lambda_{22}\frac{c^2}{\rho} = 0$ ,  $\Lambda_{21}\rho + \Lambda_{22}c = 0$ , и  $\Lambda_{21} = 1$ ,  $\Lambda_{22} = -\frac{\rho}{c}$ .

Умножим слева исходную систему в форме (9.2) сначала на первый вектор  $\Lambda_1^T$ , а потом повторим умножение системы на второй вектор  $\Lambda_2^T$ . При этом вдоль двух семейств характеристик  $\xi_1$  и  $\xi_2$  получим два уравнения в виде:

$$\begin{aligned} \Lambda_1^T \left( \frac{\partial v}{\partial t} + \lambda_1 \frac{\partial v}{\partial x} \right) + \Lambda_1^T F \\ = \frac{\partial \rho}{\partial t} + \frac{\rho}{c} \frac{\partial w}{\partial t} + (w + c) \frac{\partial \rho}{\partial x} + \frac{\rho}{c} (w + c) \frac{\partial w}{\partial x} + \varphi = 0, \\ \Lambda_2^T \left( \frac{\partial v}{\partial t} + \lambda_2 \frac{\partial v}{\partial x} \right) + \Lambda_2^T F \\ = \frac{\partial \rho}{\partial t} - \frac{\rho}{c} \frac{\partial w}{\partial t} + (w - c) \frac{\partial \rho}{\partial x} - \frac{\rho}{c} (w - c) \frac{\partial w}{\partial x} + \varphi = 0. \end{aligned}$$

Мы получили *характеристическую форму* уравнений:

$$\frac{\partial \rho}{\partial t} \pm \frac{\rho}{c} \frac{\partial w}{\partial t} + (w \pm c) \frac{\partial \rho}{\partial x} \pm \frac{\rho}{c} (w \pm c) \frac{\partial w}{\partial x} + \varphi = 0,$$

где верхний знак соответствует характеристикам первого семейства  $\xi_1$ , а нижний — характеристикам второго семейства  $\xi_2$ .

Если учесть, что вдоль  $\xi_1$  и  $\xi_2$  оператор  $\frac{\partial \cdot}{\partial t} + (w \pm c) \frac{\partial \cdot}{\partial x} = \frac{d \cdot}{dt} \Big|_{\xi_{1,2}}$ , то уравнения (9.2) в частных производных примут на характеристиках вид уравнений в обыкновенных дифференциалах:

$$\frac{d\rho}{dt} \pm \frac{\rho}{c} \frac{dw}{dt} + \varphi = 0. \quad (9.10)$$

## 9.2 Градиент целевого функционала

Найдём градиент целевого функционала с учётом всей непрямоугольной области  $\Omega$ , изображённой на рисунке 9.2. Далее, при анализе управляемости покажем, что эта область для вычисления градиента будет излишней. Её следует существенно сократить и упростить до небольшого прямоугольника. При этом упростится и сопряжённая задача, и выражение градиента. Именно эта область позволит на основе производной Фреше получить градиент. Но сейчас подойдём формально (без априорных знаний об управляемости) к получению градиента и применим алгоритм преобразований раздела 3.4 для всей области  $\Omega$ .

**1. Первый шаг** алгоритма преобразований в прямом экстремальном подходе требует выделить явное ограничение (9.4) на управление, которое будет учитываться позже в экстремальных алгоритмах. Поэтому здесь будем считать, что множества состояний и управлений совпадают со своими пространствами, т.е.  $V = L_2^m(\bar{\Omega})$ ,  $m = 2$  и  $U = L_2(S)$ . Получаем задачу неограниченной максимизации целевого функционала (9.9). Перейдём ко второму шагу преобразований.

**2. Линеаризуем задачу.** Вариация управления  $\delta u$  на  $S$  приводит к вариации вектора состояния потока воды  $\delta v = (\delta \rho, \delta w)$  на  $\bar{\Omega}$ . При этом вариация целевого функционала (9.9), по аналогии с (3.22), принимает вид:

$$\begin{aligned} \delta J(u) &= \langle I'_v, \delta v \rangle_{V^*(\omega)} + \langle I'_u, \delta u \rangle_{U^*(S)} \\ &= \int_{t_1}^{t_2} I'_w \delta w dt + \int_{t_1}^{t_2} I'_u \delta u dt. \end{aligned} \quad (9.11)$$

Сопряжённое пространство  $V^*(\omega) = L_2^2(\Gamma_{b1})$ . Вектор-функция  $I'_v = (I'_\rho, I'_w)^T \Big|_{\Gamma_{b1}}$ , где  $I'_\rho \Big|_{\Gamma_{b1}} = 0$ , а  $I'_w \Big|_{\Gamma_{b1}} = \frac{2\rho_0\sigma_b}{t_2-t_1} w_b$ . Функция  $I'_u \delta u \Big|_{\Gamma_{b1}} = \frac{\rho_0}{t_2-t_1} w_b^2 \delta \sigma_b$ . С учётом того, что  $\sigma_b = \sigma_a e^{\int_{x_a}^{x_b} u dx}$  и не зависит от  $t$ , можем сделать следующие преобразования:

$$\begin{aligned} \int_{t_1}^{t_2} I'_u \delta u dt &= \int_{t_1}^{t_2} \frac{\rho_0}{t_2-t_1} w_b^2 dt \left( \sigma_a e^{\int_{x_a}^{x_b} u dx} \int_{x_a}^{x_b} \delta u dx \right) \\ &= \int_{t_1}^{t_2} \frac{\rho_0 \sigma_b}{t_2-t_1} w_b^2 dt \int_{x_a}^{x_b} \delta u dx = J \int_{x_a}^{x_b} \delta u dx = \int_S J \delta u dx. \end{aligned}$$

Таким образом, функции  $I'_v$  и  $I'_u$  на границе  $\Gamma_{b1}$  принимают вид:

$$I'_v = (0, I'_w)^T, \quad \text{где } I'_w = \frac{2\rho_0\sigma_b w_b}{t_2-t_1},$$

$$I'_u = J.$$

Линеаризация уравнений течения воды (9.5), посредством их варьирования, приводит к выражению:

$$\begin{aligned}\delta(\mathbb{D}v + F) &= \frac{\partial \delta v}{\partial t} + A \frac{\partial \delta v}{\partial x} + \frac{\partial A}{\partial v} \delta v \frac{\partial v}{\partial x} + F'_v \delta v + F'_u \delta u \\ &= \mathbb{V} \delta v + \mathbb{U} \delta u = 0 \in V(\Omega),\end{aligned}$$

или в явном виде:

$$\begin{aligned}\frac{\partial \delta \rho}{\partial t} + \frac{\partial \rho}{\partial x} \delta w + w \frac{\partial \delta \rho}{\partial x} + \rho \frac{\partial \delta w}{\partial x} + \frac{\partial w}{\partial x} \delta \rho + w u \Theta(x - x_a) \delta \rho \\ + \rho u \Theta(x - x_a) \delta w + \rho w \Theta(x - x_a) \delta u = 0\end{aligned}\tag{9.12}$$

$$\begin{aligned}\frac{\partial \delta w}{\partial t} + \frac{B n \rho^{n-2}}{\rho_0^n} \frac{\partial \delta \rho}{\partial x} + \frac{B n (n-2) \rho^{n-3}}{\rho_0^n} \frac{\partial \rho}{\partial x} \delta \rho + w \frac{\partial \delta w}{\partial x} + \frac{\partial w}{\partial x} \delta w \\ = 0\end{aligned}$$

Мы видим, что операторы  $\mathbb{V}$ ,  $\mathbb{U}$  и их отображения — это:

$$\mathbb{V} \cdot = \frac{\partial \cdot}{\partial t} + A \frac{\partial \cdot}{\partial x} + \sum_{i=1}^2 \frac{\partial A}{\partial v_i} \cdot \Big|_i \frac{\partial v}{\partial x} + F'_v \cdot \quad : \quad V(\Omega) \rightarrow V(\Omega),$$

$$\mathbb{U} \cdot = F'_u \cdot \quad : \quad U(S) \rightarrow V(\Omega).$$

Здесь  $F'_v$  и  $F'_u$  представляют собой матрицу и вектор-столбец производных от  $F$  по  $v$  и  $u$ :

$$F'_v = \begin{pmatrix} w u \Theta(x - x_a) & \rho u \Theta(x - x_a) \\ 0 & 0 \end{pmatrix}, \quad F'_u = \begin{pmatrix} \rho w \Theta(x - x_a) \\ 0 \end{pmatrix}.$$

Оператор  $\mathbb{U}$  отображает вариацию управления  $\delta u$  из области определения управления  $S$  в область  $\Omega$  функционирования распределённой системы.

**3. Отобразим** линеаризованные уравнения (9.12) в сопряжённые пространства. Возьмём произвольный линейный функционал в виде двухкомпонентного вектора  $\tilde{f} = \{\tilde{f}_1, \tilde{f}_2\}$  из пространства  $V^*(\bar{\Omega}) = L_2^2(\bar{\Omega})$  сопряжённого с пространством  $V(\bar{\Omega})$ . Запишем значение  $\tilde{f}$  на вариации системы (9.12):

$$\langle \tilde{f}, \delta(\mathbb{D}v + F) \rangle_{V^*(\Omega)} = \langle \tilde{f}, \mathbb{V} \delta v \rangle_{V^*(\Omega)} + \langle \tilde{f}, \mathbb{U} \delta u \rangle_{V^*(\Omega)} = 0.$$

Это выражение следует преобразовать к значениям линейных функционалов непосредственно на элементах  $\delta v$ ,  $\delta u$ , т.е. представить в виде тождества Лагранжа типа (3.25). В нашем случае с первой производной по времени и первой производной по пространству, оно будет иметь вид:

$$\begin{aligned} \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} + \langle \tilde{f}, \mathbb{U}\delta u \rangle_{V^*(\Omega)} & \quad (9.13) \\ \equiv \langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\Omega)} + \langle \bar{\mathbb{V}}^* \tilde{f}, \delta v \rangle_{V^*(\partial\Omega)} + \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)} \\ & = 0. \end{aligned}$$

Теперь подробно запишем данное тождество в явном виде и получим выражения для сопряжённых операторов  $\mathbb{V}^*$  и  $\mathbb{U}^*$ . Начнём с первого слагаемого, содержащего оператор  $\mathbb{V}$ :

$$\begin{aligned} \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} & = \iint_{\Omega} \left[ \tilde{f}_1 \frac{\partial \delta \rho}{\partial t} + \tilde{f}_1 \frac{\partial \rho}{\partial x} \delta w + \tilde{f}_1 w \frac{\partial \delta \rho}{\partial x} + \tilde{f}_1 \rho \frac{\partial \delta w}{\partial x} + \tilde{f}_1 \frac{\partial w}{\partial x} \delta \rho \right. \\ & + \tilde{f}_1 (w u \delta \rho + \rho u \delta w) \Theta(x - x_a) + \tilde{f}_2 \frac{\partial \delta w}{\partial t} + \tilde{f}_2 \frac{B n \rho^{n-2}}{\rho_0^n} \frac{\partial \delta \rho}{\partial x} \\ & \left. + \tilde{f}_2 \frac{B n (n-2) \rho^{n-3}}{\rho_0^n} \frac{\partial \rho}{\partial x} \delta \rho + \tilde{f}_2 w \frac{\partial \delta w}{\partial x} + \tilde{f}_2 \frac{\partial w}{\partial x} \delta w \right] dx dt. \end{aligned}$$

Сделаем необходимые преобразования, используя правило дифференцирования произведений:

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} \\
&= \iint_{\Omega} \left[ \frac{\partial(\tilde{f}_1 \delta \rho)}{\partial t} - \frac{\partial \tilde{f}_1}{\partial t} \delta \rho + \tilde{f}_1 \frac{\partial \rho}{\partial x} \delta w + \frac{\partial(\tilde{f}_1 w \delta \rho)}{\partial x} - w \frac{\partial \tilde{f}_1}{\partial x} \delta \rho \right. \\
&\quad - \tilde{f}_1 \frac{\partial w}{\partial x} \delta \rho + \frac{\partial(\tilde{f}_1 \rho \delta w)}{\partial x} - \rho \frac{\partial \tilde{f}_1}{\partial x} \delta w - \tilde{f}_1 \frac{\partial \rho}{\partial x} \delta w + \tilde{f}_1 \frac{\partial w}{\partial x} \delta \rho \\
&\quad + \tilde{f}_1 (wu \delta \rho + \rho u \delta w) \Theta(x - x_a) + \frac{\partial(\tilde{f}_2 \delta w)}{\partial t} - \frac{\partial \tilde{f}_2}{\partial t} \delta w \\
&\quad + \frac{\partial}{\partial x} \left( \frac{Bn \rho^{n-2} \delta \rho}{\rho_0^n} \tilde{f}_2 \right) - \frac{Bn \rho^{n-2}}{\rho_0^n} \frac{\partial \tilde{f}_2}{\partial x} \delta \rho \\
&\quad - \tilde{f}_2 \frac{Bn(n-2) \rho^{n-3}}{\rho_0^n} \frac{\partial \rho}{\partial x} \delta \rho + \tilde{f}_2 \frac{Bn(n-2) \rho^{n-3}}{\rho_0^n} \frac{\partial \rho}{\partial x} \delta \rho \\
&\quad \left. + \frac{\partial(\tilde{f}_2 w \delta w)}{\partial x} - w \frac{\partial \tilde{f}_2}{\partial x} \delta w - \tilde{f}_2 \frac{\partial w}{\partial x} \delta w + \tilde{f}_2 \frac{\partial w}{\partial x} \delta w \right] dx dt
\end{aligned}$$

Четыре пары слагаемых  $\tilde{f}_1 \frac{\partial w}{\partial x} \delta \rho$ ,  $\tilde{f}_1 \frac{\partial \rho}{\partial x} \delta w$ ,  $\tilde{f}_2 \frac{Bn(n-2) \rho^{n-3}}{\rho_0^n} \frac{\partial \rho}{\partial x} \delta \rho$ ,

$\tilde{f}_2 \frac{\partial w}{\partial x} \delta w$  сокращаются. Слагаемые в виде интегралов от производных

сразу же интегрируются в соответствующих пределах. Получаем:

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} \\
&= \iint_{\Omega} \left[ \left( -\frac{\partial \tilde{f}_1}{\partial t} - w \frac{\partial \tilde{f}_1}{\partial x} - \frac{Bn \rho^{n-2}}{\rho_0^n} \frac{\partial \tilde{f}_2}{\partial x} + wu \Theta(x - x_a) \tilde{f}_1 \right) \delta \rho \right. \\
&\quad \left. + \left( -\frac{\partial \tilde{f}_2}{\partial t} - \rho \frac{\partial \tilde{f}_1}{\partial x} - w \frac{\partial \tilde{f}_2}{\partial x} + \rho u \Theta(x - x_a) \tilde{f}_1 \right) \delta w \right] dx dt \\
&\quad + \int_{\Gamma_p \cup \Gamma_2} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx - \int_{\Gamma_0 \cup \Gamma_{b_0}} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx \\
&\quad + \int_{\Gamma_b} \left[ \left( \tilde{f}_1 w + \frac{Bn \rho^{n-2}}{\rho_0^n} \tilde{f}_2 \right) \delta \rho + (\tilde{f}_1 \rho + \tilde{f}_2 w) \delta w \right] dt \\
&\quad - \int_{\Gamma_p} \left[ \left( \tilde{f}_1 w + \frac{Bn \rho^{n-2}}{\rho_0^n} \tilde{f}_2 \right) \delta \rho + (\tilde{f}_1 \rho + \tilde{f}_2 w) \delta w \right] dt.
\end{aligned}$$

Одиночные интегралы только с нижним пределом означают интегрирование по указанной части контура  $\partial\Omega$ .

При интегрировании по  $x$  на участках  $\Gamma_a$  и  $\Gamma_{b_0}$  следует учесть, что здесь  $dx = wdt$ , поэтому:

$$\begin{aligned} \int_{\Gamma_p \cup \Gamma_2} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx \\ &= \int_{\Gamma_p} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) w dt + \int_{\Gamma_2} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx, \\ \int_{\Gamma_0 \cup \Gamma_{b_0}} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx \\ &= \int_{\Gamma_0} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx + \int_{\Gamma_{b_0}} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) w dt. \end{aligned}$$

Объединим эти интегралы по времени на  $\Gamma_p$  и  $\Gamma_{b_0}$  с другими, ранее полученными интегралами на этих же участках, и сократим соответствующие члены. Тогда

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{V}\delta v \rangle_{V^*(\Omega)} \\
&= \iint_{\Omega} \left[ \left( -\frac{\partial \tilde{f}_1}{\partial t} - w \frac{\partial \tilde{f}_1}{\partial x} - \frac{Bn\rho^{n-2}}{\rho_0^n} \frac{\partial \tilde{f}_2}{\partial x} \right. \right. \\
&\quad \left. \left. + wu\Theta(x - x_a)\tilde{f}_1 \right) \delta \rho \right. \\
&\quad \left. + \left( -\frac{\partial \tilde{f}_2}{\partial t} - \rho \frac{\partial \tilde{f}_1}{\partial x} - w \frac{\partial \tilde{f}_2}{\partial x} \right. \right. \\
&\quad \left. \left. + \rho u\Theta(x - x_a)\tilde{f}_1 \right) \delta w \right] dx dt \\
&\quad + \int_{\Gamma_2} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx - \int_{\Gamma_0} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx \\
&\quad + \int_{\Gamma_{b0}} \left[ \frac{Bn\rho^{n-2}}{\rho_0^n} \tilde{f}_2 \delta \rho + \tilde{f}_1 \rho \delta w \right] dt \\
&\quad + \int_{\Gamma_{b1}} \left[ \left( \tilde{f}_1 w + \frac{Bn\rho^{n-2}}{\rho_0^n} \tilde{f}_2 \right) \delta \rho + (\tilde{f}_1 \rho + \tilde{f}_2 w) \delta w \right] dt \\
&\quad - \int_{\Gamma_p} \left[ \frac{Bn\rho^{n-2}}{\rho_0^n} \tilde{f}_2 \delta \rho + \tilde{f}_1 \rho \delta w \right] dt.
\end{aligned} \tag{9.14}$$

Сравнивая полученное выражение в области  $\Omega$ , с тождеством (9.13) при  $\delta v$ , мы видим, что сопряжённый оператор  $\mathbb{V}^*$ , действующий на вектор  $\tilde{f}$ , имеет вид

$$\mathbb{V}^* \cdot = -\frac{\partial \cdot}{\partial t} - A^T \frac{\partial \cdot}{\partial x} + F_v'^T \cdot : V^*(\Omega) \rightarrow V^*(\Omega).$$

Для второго слагаемого в тождестве Лагранжа получаем:

$$\begin{aligned}
& \langle \tilde{f}, \mathbb{U}\delta u \rangle_{V^*(\Omega)} \\
&= \iint_{\Omega} \tilde{f}_1 \rho w \Theta(x - x_a) \delta u dx dt \\
&= \int_{x_a}^{x_b} \left( \int_{\Gamma_{b0}}^{t_2} \tilde{f}_1 \rho w dt \right) \delta u dx = \langle \mathbb{U}^* \tilde{f}, \delta u \rangle_{U^*(S)},
\end{aligned} \tag{9.15}$$

где сопряжённый оператор

$$\mathbb{U}^* \cdot = \int_{\Gamma_{b_0}}^{t_2} (\rho w, 0)^T \cdot dt = \int_{\Gamma_{b_0}}^{t_2} F_u'^T \cdot dt : V^*(\Omega) \rightarrow U^*(S).$$

Мы видим, что матричный дифференциальный оператор  $\mathbb{V}^*$  характеризуется транспонированной матрицей  $A^T$  и имеет свободный член в виде транспонированной матрицы  $F_v'^T$ . Второй оператор  $\mathbb{U}^*$  из вектор-столбца  $F_u'$  превратился в вектор-строку  $F_u'^T$ , с интегрированием по времени на  $\Omega$  (на части  $\Omega$  где  $\delta u \neq 0$ ), что обеспечивает переход отображения данного оператора из пространства  $V^*(\Omega)$  в пространство  $U^*(S)$ . То есть в своём отображении оператор  $\mathbb{U}^*$  осуществляет *проецирование*  $P_x(\Omega|_{x>x_a})$ , которое реализуется интегрированием по времени от линии  $\Gamma_{b_0}$  до момента  $t_2$ .

**4. Объединим** линейные функционалы из (9.11), (9.14), (9.15) в соответствующих пространствах на элементах, к которым они относятся. Получим выражение первой вариации целевого функционала (без учёта краевых условий задачи), аналогичное (3.30):

$$\begin{aligned} \delta J(u) &= \langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\Omega)} + \langle \mathbb{U}^* \tilde{f} + I_u', \delta u \rangle_{U^*(S)} \\ &+ \int_{\Gamma_2} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx - \int_{\Gamma_0} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx \\ &+ \int_{\Gamma_{b_0}} \left( \frac{Bn\rho^{n-2}}{\rho_0^n} \tilde{f}_2 \delta \rho + \tilde{f}_1 \rho \delta w \right) dt \\ &+ \int_{\Gamma_{b_1}} \left[ \left( \tilde{f}_1 w + \frac{Bn\rho^{n-2}}{\rho_0^n} \tilde{f}_2 \right) \delta \rho \right. \\ &\quad \left. + (\tilde{f}_1 \rho + \tilde{f}_2 w + I_w') \delta w \right] dt \\ &- \int_{\Gamma_p} \left( \frac{Bn\rho^{n-2}}{\rho_0^n} \tilde{f}_2 \delta \rho + \tilde{f}_1 \rho \delta w \right) dt. \end{aligned} \tag{9.16}$$

Здесь, при  $U^*(S)$  добавлена производная  $I_u'$ , а на границе  $\Gamma_{b_1}$  — производная  $I_w'$ .

Далее всё, что находится при вариации  $\delta u$ , обозначим оператором

$$\mathbb{U}_{\emptyset}^* \cdot = \mathbb{U}^* \cdot + I'_u = \int_{\Gamma_{b_0}}^{t_2} F'_u{}^T \cdot dt + J,$$

где  $\mathbb{U}_{\emptyset}^*$  линейный сопряжённый, но *неоднородный оператор*.

**5. Учёт краевых условий.** Поварьируем граничные и начальные условия (9.6)–(9.8):

$$\begin{aligned} \frac{d\delta w}{dt} + \frac{\sigma_a B n \rho^{n-1}}{m_p \rho_0^n} \delta \rho &= 0 && \text{на } \Gamma_p, \\ \delta \rho &= 0 && \text{на } \Gamma_b, \\ \delta w = 0, \quad \delta \rho &= 0 && \text{на } \Gamma_0. \end{aligned}$$

Из первого граничного условия мы получаем явное выражение вариации

$$\delta \rho = -\frac{m_p \rho_0^n}{\sigma_a B n \rho^{n-1}} \frac{d\delta w}{dt} \quad \text{на } \Gamma_p.$$

Теперь все значения граничных и начальных вариаций следует подставить в вариацию функционала (9.16). При этом интеграл на границе  $\Gamma_p$  может быть преобразован к виду:

$$\begin{aligned} - \int_{\Gamma_p} \left( \frac{B n \rho^{n-2}}{\rho_0^n} \tilde{f}_2 \delta \rho + \tilde{f}_1 \rho \delta w \right) dt &= \int_{\Gamma_p} \left( \tilde{f}_2 \frac{m_p}{\sigma_a \rho} \frac{d\delta w}{dt} - \tilde{f}_1 \rho \delta w \right) dt \\ &= \int_{\Gamma_p} \left[ \frac{d}{dt} \left( \tilde{f}_2 \frac{m_p}{\sigma_a \rho} \delta w \right) - \frac{d}{dt} \left( \tilde{f}_2 \frac{m_p}{\sigma_a \rho} \right) \delta w - \tilde{f}_1 \rho \delta w \right] dt \\ &= \tilde{f}_2 \frac{m_p}{\sigma_a \rho} \delta w \Big|_{(x_{p2}, t_2)} - \int_{\Gamma_p} \left[ \frac{d}{dt} \left( \tilde{f}_2 \frac{m_p}{\sigma_a \rho} \right) + \tilde{f}_1 \rho \right] \delta w dt. \end{aligned}$$

Здесь мы учли, что на  $\Gamma_p$  в начальной точке  $(x_{p0}, t_0) \in \Gamma_0$  вариация  $\delta w = 0$ .

В итоге получим *первую вариацию* целевого функционала с учётом граничных и начальных условий:

$$\begin{aligned}
\delta J(u) = & \langle \mathbb{V}^* \tilde{f}, \delta v \rangle_{V^*(\Omega)} + \langle \mathbb{U}_{\emptyset}^*, \delta u \rangle_{U^*(S)} \\
& + \int_{\Gamma_2} (\tilde{f}_1 \delta \rho + \tilde{f}_2 \delta w) dx + \int_{\Gamma_{b0}} \tilde{f}_1 \rho_0 \delta w dt \\
& + \int_{\Gamma_{b1}} (\tilde{f}_1 \rho + \tilde{f}_2 w + I'_w) \delta w dt \\
& - \int_{\Gamma_p} \left[ \frac{d}{dt} \left( \tilde{f}_2 \frac{m_p}{\sigma_a \rho} \right) + \tilde{f}_1 \rho \right] \delta w dt \\
& + \tilde{f}_2 \frac{m_p}{\sigma_a \rho} \delta w \Big|_{(x_{p2}, t_2)}.
\end{aligned} \tag{9.17}$$

**6. Выделение градиента.** Зададим в выражении (9.17) значение вектор-функции  $\tilde{f} = f$ , такое, чтобы все линейные функционалы при  $\delta v$  обратились в ноль. Получаем в области  $\Omega$ :

$$\mathbb{V}^* f = - \frac{\partial f}{\partial t} - A^T \frac{\partial f}{\partial x} + F_v'^T f = 0 \quad \text{на } \Omega,$$

или в развёрнутом виде:

$$\begin{aligned}
- \frac{\partial f_1}{\partial t} - w \frac{\partial f_1}{\partial x} - \frac{c^2}{\rho} \frac{\partial f_2}{\partial x} + w u \Theta(x - x_a) f_1 &= 0, \\
- \frac{\partial f_2}{\partial t} - \rho \frac{\partial f_1}{\partial x} - w \frac{\partial f_2}{\partial x} + \rho u \Theta(x - x_a) f_1 &= 0
\end{aligned} \quad \text{на } \Omega. \tag{9.18}$$

На границах получаем:

$$\frac{d}{dt} \left( \frac{m_p f_2}{\sigma_a \rho} \right) + f_1 \rho = 0 \quad \text{при } \delta w \text{ на } \Gamma_p, \tag{9.19}$$

$$f_1 = 0 \quad \text{при } \delta w \text{ на } \Gamma_{b0}, \tag{9.20}$$

$$\rho_0 f_1 + w f_2 + I'_w = 0 \quad \text{при } \delta w \text{ на } \Gamma_{b1} \tag{9.21}$$

$$f_1 = 0, \quad f_2 = 0 \quad \text{при } \delta w, \delta \rho \text{ на } \Gamma_2. \tag{9.22}$$

Мы получили сопряжённую задачу (9.18)–(9.22), аналогичную (3.33), которая решается в обратном направлении по времени, начиная с условия (9.22). Обыкновенное дифференциальное уравнение (9.19) на

траектории поршня  $\Gamma_p$  в угловой точке  $(x_{p2}, t_2) \in \Gamma_2$  решается с нулевым «начальным» условием, согласно (9.22).

При таком значении сопряжённой переменной  $f$ , первая вариация целевого функционала принимает вид:

$$\delta J(u) = \langle \mathbb{U}^* f, \delta u \rangle_{U^*(S)} = \langle \nabla J, \delta u \rangle_{U^*(S)}.$$

Откуда следует, что *градиент целевого функционала*  $\nabla J = \mathbb{U}^* f$  имеет вид:

$$\nabla J(u; x) = \mathbb{U}^* f + J = \int_{\Gamma_{b0}}^{t_2} \rho w f_1 dt + J \in U^*(S). \quad (9.23)$$

Полученное выражение соответствует первой строке общей формы градиента (3.35), когда линейный сопряжённый оператор градиента неоднородный (функция цели  $I(v, u)$  явно содержит управление), и искомый параметр (управление) присутствует в дифференциальном уравнении модели.

### 9.3 Управляемость

Найдём условия управляемости, при которых выражение (9.23) действительно будет градиентом  $\nabla J$ , а не просто некоторой производной Фреше  $J'_u$ . Это сделает обоснованным применение методов направленного поиска оптимальной формы сопла. Согласно теореме 3.4 об управляемости, сначала необходимо убедиться в корректности исходной и сопряжённой задач. Затем следует уточнить области определения и значений оператора  $\mathbb{U}^*$  и сделать оценку возможной вырожденности  $\mathbb{U}^*$ .

Исходная и сопряжённая задачи являются гиперболическими, поэтому их корректность определяется поведением характеристик  $\xi_{1,2}$ . Напомним, что характеристики  $\xi_{1,2}$  в обеих задачах одинаковые.

При течении без ударных волн (именно такие течения мы рассматриваем) характеристики одного и того же семейства никогда не пересекаются. Все характеристики в области  $\Omega$  выходят из участков границы  $\partial\Omega$  с известными решениями (заданными краевыми условиями). Следовательно, на пересечении двух характеристик разного семейства  $\xi_1$  и  $\xi_2$  в любой точке  $\Omega$  может быть найдено решение гиперболической системы двух уравнений [70, 44] в виде двумерных вектор-функций  $v$  и  $f$  для обеих задач.

Найдём область определения  $V^*(\Omega)$  для однородного оператора  $\mathbb{U}^*$  в градиенте  $\nabla J = \mathbb{U}^* f + J$ . Для этого воспользуемся определением 3.5.

Из корректности исходной прямой задачи следует, что любые функции  $u(x) \in U(S)$  будут однозначно влиять на значение производной целевой функции  $I'_w$  на  $\omega = \Gamma_{b1}$  через характеристики  $\xi_1$ , если хотя бы одна из них прошла через всё сопло. То есть имеет место левая ветка отображений в определении 3.5:

$$U(S) \ni u \xrightarrow{\text{однозначно}} I'_w|_{\omega}.$$

Перейдём к выявлению множества  $\Omega$ , используемого в следующих ветках поредения, для формирования  $V^*(\Omega)$ .

Источником возмущений сопряжённого решения  $f$  где-либо на  $\bar{\Omega}$ , является член  $I'_w|_{\omega}$ , присутствующий в граничном условии (9.21) на срезе сопла гидропушки, на  $\Gamma_{b1}$ . От среза сопла возмущения  $f$  при  $t \rightarrow t_0$  распространяются в виде волн по характеристикам первого семейства  $\xi_1$  в сторону поршня, как это показано на рисунке 9.3. Эти возмущения распространяются вдоль всего сопла и переносят информацию о целевом функционале в виде производной  $I'_w$  из каждой точки  $\omega$  в разные точки  $S = (x_a, x_b)$ . На рисунке 9.3 показан пример возможного множества  $\Omega$  (вся заштрихованная с разной плотностью область под верхней характеристикой  $\xi_1$  от  $x_a$  до  $x_b$ ).

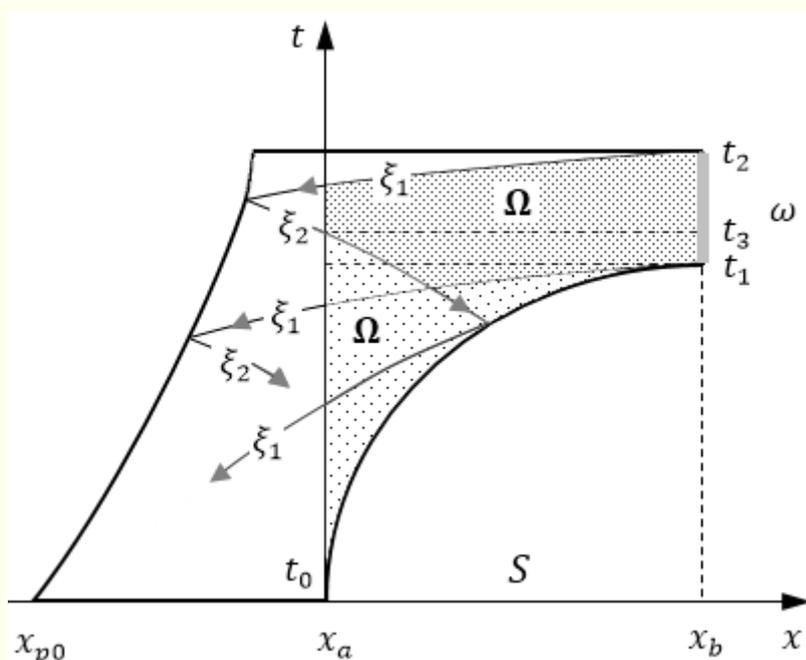


Рисунок 9.3 — Пространственно-временная диаграмма области  $\Omega$  с избыточной областью  $\omega$

На поршне характеристики  $\xi_1$  отражаются характеристиками второго семейства  $\xi_2$  и продолжают переносить информацию, полученную от  $\xi_1$  о возмущениях  $I'_w|_\omega$ , добавляя при этом новую информацию о движении поршня. Такой процесс отражений обеих характеристик и от поршня, и от внутренней части сопла продолжается до момента  $t_0$ .

Начиная с момента времени  $t_3$  и ниже, в одну и ту же точку некоторых участков множества  $S$  будет приходиться две сопряжённые волны  $\xi_1$  и  $\xi_2$ , порождённые разными значениями  $I'_w$  и с ненужной информацией (помехами), как минимум, от поршня. А ниже характеристики  $\xi_1$ , вышедшей из сопла ниже  $t_1$ , будет добавлена ещё и ненужная информация от сопла. Эта информация ненужная потому, что не содержит ничего о цели оптимизации из  $I'_w|_\omega$ .

В области  $\Omega$ , под характеристикой  $\xi_2$  (отражение  $\xi_1$ , вышедшей при  $t_2$ ) и под  $\xi_1$  (вышедшей при  $t_1$ ), начинает образовываться светлая область неоднозначного влияния значений функции  $I'_w|_\omega$  на сопряжённое состояние  $f$ . Очевидно, что бессмысленно решать сопряжённую задачу и рассчитывать градиент в такой области  $\Omega$ .

Разумно на рисунке 9.3 ограничиться рассмотрением сопряжённого состояния  $f$  на части  $\Omega$ , заключённой в прямоугольнике

$$\Omega_1 = (x_a, x_b) \times (t_1, t_2).$$

Если поршень будет находится относительно близко к началу сопла и влиять на сопряжённое состояние, то прямоугольник  $\Omega_1$  следует считать слишком большим. В таком прямоугольнике множество  $\Omega$  (заштрихованная с разной плотностью область от  $t_1$  до  $t_2$ ) будет соответствовать *избыточному* множеству  $\omega$ . То есть в целевом функционале  $J$  интервал  $(t_1, t_2)$  будет избыточным. При этом в рассматриваемой области  $\Omega$  могут присутствовать неприемлемые помехи слева (малая плотность штриховки в  $\Omega_1$ ) для вычисления градиента  $\nabla J$ .

Избыточность  $\omega$  устраняется дальнейшим сокращением  $\Omega_1$ , когда  $t_3 = t_1$ , т.е. когда  $t_3$  соответствует началу истечения, как это показано на рисунке 9.4а. В этом случае вся область достаточного решения сопряжённой задачи сужается до ещё меньшего прямоугольника

$$\Omega_2 = (x_a, x_b) \times (t_3, t_2).$$

Здесь поршень не будет «мешать» отображению  $I'_w|_\omega \xrightarrow{\text{однозначно}} f|_\Omega$ .

Если, по техническим условиям конструирования гидропушки, допускается сокращение времени истечения  $t_2 - t_1$ , то можно было бы уменьшить прямоугольник решения сопряжённой задачи ещё сильнее, до прямоугольника  $\Omega_3$ , показанного на рисунке 9.4б,

$$\Omega_3 = (x_a, x_b) \times (t_1, t_2),$$



$$\rho_0 f_1 + w f_2 + I'_w = 0 \quad \text{на } \Gamma_{b1} = \omega_{min}, \quad (9.26)$$

где  $\Gamma_a = x_a \times (t_1, t_2)$ . Выражение градиента (9.23) также упростится, получим

$$\nabla J = \int_{t_1}^{t_2} \rho w f_1 dt + J \in U^*(S). \quad (9.27)$$

Из полученного выражения сопряжённой задачи (9.24)–(9.26) следует, что изначально вариацию  $\delta J$  целевого функционала достаточно находить только в  $\Omega_{2,3}$ , а не во всей области  $\Omega$ .

Если учесть, что исходная и сопряжённая системы не содержат диссипативные члены (изоэнтропические течения), то потери чувствительности сопряжённого состояния  $f \in V^*(\Omega)$  к управлению  $u$  не будет. Полученное множество  $\Omega \subset \Omega_{2,3}$  будет корректно задавать область  $V^*(\Omega)$  определения оператора  $\mathbb{U}^*$  со значениями в  $U^*(S)$ . При этом получим выражение времени для формирования такого  $\Omega$ :

$$t_2 \geq t_1 + \int_{x_a}^{x_b} \frac{dx}{\lambda_1}, \quad t_1 \geq t_3, \quad (9.28)$$

которое будет *условием управляемости* в рассматриваемой задаче. Данное условие означает, что, во-первых, верхняя характеристика  $\xi_1$  должна пройти через всё сопло от  $x_b$  до  $x_a$ , во-вторых, начало истечения  $t_1$  не должно быть меньше момента  $t_3$  начала попадания в сопло волн, отражённых от поршня.

В этих условиях выполняются оставшиеся ветки отображения в определении 3.5:

$$I'_w|_\omega \xrightarrow{\text{однозначно}} f|_\Omega \xrightarrow{\mathbb{U}^*} U^*(S).$$

Мы нашли область определения оператора  $\mathbb{U}^*$ . Это означает, что первое требование теоремы 3.4 «существует область  $V^*(\Omega)$  определения оператора  $\mathbb{U}^*$  с его значениями в области  $U^*(S)$ » выполняется.

Несмотря на то, что диссипации в уравнениях задач нет, но, как показали тестовые расчёты, в обеих задачах могут присутствовать значительные вычислительные помехи. Значительность помех обусловлена высокой чувствительностью функционала к форме сопла, которая входит в управление как производная. Данное обстоятельство, и другие вычислительные помехи, приводят к значительной разнице в точности численного решения обеих задач, поскольку одна задача — нелинейная, с существенно подвижной правой и левой границами расчётной сеткой (смотрите далее в разделе 9.4.1 рисунок 9.5, от  $t_0$  до  $t_2$ ), другая — линейная с практически неподвижной сеткой после начала истечения (рисунок 9.5, от  $t_1$  до  $t_2$ ).

Заметная разница в точности решений обеих задач, в свою очередь, приводит к необходимости согласованного учёта (коррекции) *помех* в формуле (9.27) градиента. То есть, погрешности решения сопряжённой задачи в первом члене  $\int_{t_1}^{t_2} \rho w f_1 dt$  должны быть «согласованы» с погрешностями решения исходной задачи при расчёте второго члена  $J$ . Иначе говоря, необходимо юстировать (точно подгонять) члены  $\int_{t_1}^{t_2} \rho w f_1 dt$  и  $J$  в значении градиента.

Таким образом, нам придётся вместо (9.27) использовать следующую формулу градиента целевого функционала:

$$\nabla J = \int_{t_1}^{t_2} \rho w f_1 dt + \kappa J \in U^*(S), \quad (9.29)$$

где  $\kappa$  — весовой коэффициент выравнивания вычислительных помех решения исходной (9.5)–(9.8) и сопряжённой (9.24)–(9.26) задач.

Юстировку  $\kappa$  можно производить на какой-нибудь «эталонной» форме сопла, например, на конусе, неразрывно исходящем из ствола гидропушки. Форма такого сопла будет зависеть только от радиуса  $R_b$  среза сопла. При этом  $J(u)$  будет функцией  $J(R_b)$ . Необходимо в

вычислительном эксперименте для этой функции найти  $\max J(R_b)$ , а затем, для полученного аргумента  $R_b$ , подобрать  $\kappa$  при котором градиент  $\nabla J$  будет минимальным. Нулевое значение  $\nabla J$ , опять-таки из-за вычислительных помех, вряд ли удастся получить.

Теперь обсудим возможную вырожденность однородного оператора  $\mathbb{U}^*$ , входящего в  $\mathbb{U}_\emptyset^*$  в градиенте (9.29)

$$\nabla J = \mathbb{U}_\emptyset^* f = \mathbb{U}^* f + \kappa J.$$

Согласно условиям теоремы 3.4, достаточно исследовать невырожденность однородного оператора  $\mathbb{U}^*$ . Напомним, что неоднородность оператора  $\mathbb{U}_\emptyset^*$  приводит к смещению нулевого ядра однородного оператора  $\mathbb{U}^*$ . В  $\mathbb{U}_\emptyset^*$  ненулевое ядро  $f_{ker} \neq 0$  на  $S$ , при котором возможно НУО  $\|\mathbb{U}_\emptyset^* f_{ker}\| = 0$ , не делает оператор  $\mathbb{U}_\emptyset^*$  вырожденным.

Очевидно, что при произвольных значениях  $\rho$  и  $w$  результат интегрирования в однородном операторе  $\mathbb{U}^* f = \int_{t_1}^{t_2} \rho w f_1 dt$  может стать нулевым на  $S$  только при  $f_1 = 0$  на  $\Omega$ . В этом случае выполняется второе требование теоремы 3.4 «оператор  $\mathbb{U}^*$  — невырожденный».

Таким образом, все требования теоремы 3.4 для обеспечения управляемости выполняются. Распределённая система (9.2) с краевыми условиями (9.6)–(9.8) управляема посредством  $u(x)$  на  $S$  по функционалу  $J$  (9.1) при условии (9.28). Остаётся корректно применить экстремальные алгоритмы.

Сделаем заключение. Условие управляемости (9.28) позволило сформулировать корректную задачу оптимизации в малой прямоугольной области  $\Omega_{2,3}$  (рисунок 9.4), а не в большой и сложной области  $\Omega$ . Теперь мы знаем, как из производной Фреше  $J'_u$  получить градиент  $\nabla J$ :

$$J'_u = \mathbb{U}_\emptyset^* f \rightarrow \mathbb{U}_\emptyset^* f|_\Omega = \nabla J.$$

Производная Фреше вне  $\Omega$  из  $\Omega_{2,3}$  приводит к выражению градиента, неоднозначно соответствующему целевому функционалу  $J$ . Это один из главных факторов (второй — необходимые экстремальные алгоритмы), который предыдущим исследователям в данной области [73, 45, 4, 46] не позволил найти оптимальную форму сопла гидропушки. Ими использовался подход классического вариационного исчисления для  $\delta J$  во всей области  $\Omega$ . Такой подход привёл к НУО в виде нелинейного бесконечномерного уравнения  $J'_u(u_*; x) = \int_{\Gamma_{b_0}}^{t_2} \rho w f_1 dt + J = 0$  на  $S$ . Попытки поиска корня  $u_*$  этого уравнения численными методами релаксации не привели к положительным результатам. Сейчас понятно, что этого и невозможно было сделать в рамках применяемого подхода, без анализа управляемости.

Если бы мы сделали анализ управляемости перед выкладками первой вариации  $\delta J$  на  $\Omega$  в разделе 9.2, то существенно облегчили бы себе жизнь при формировании сопряжённой задачи в небольшом прямоугольнике  $\Omega_{2,3} \subset \Omega$ . Поэтому, следует рекомендовать сначала делать анализ управляемости до выявления области  $\Omega$ , а потом, для полученной  $\Omega$ , выделять область решения сопряжённой задачи (в нашем случае — это  $\Omega_2$  или  $\Omega_3$ ) и находить вариацию  $\delta J$ . После чего можно продолжить окончательный анализ управляемости и найти градиент  $\nabla J$ .

### 9.3.1. Характеристическая форма сопряжённых уравнений

Анализ управляемости привёл нас к новой сопряжённой задаче (9.24)–(9.26). Для её решения нам потребуется характеристическая форма уравнений (9.24). Давайте её получим.

Матрицы  $A$  и  $A^T$ , в исходных (9.5) и сопряжённых (9.24) уравнениях, имеют одинаковые собственные числа, поэтому сопряжённая задача имеет те же характеристики  $\xi_{1,2} = w \pm c$ , что и

исходная. То есть, сопряжённые волны движутся по тем же характеристикам  $\xi_{1,2}$ , но в обратном направлении.

*Собственные векторы* для транспонированной матрицы в сопряжённой задаче будут другими. Найдём первый собственный вектор  $\Lambda_1$  для  $\lambda_1 = w + c$ :

$$\Lambda_1^T \begin{pmatrix} w - \lambda_1 & \frac{c^2}{\rho} \\ \rho & w - \lambda_1 \end{pmatrix} = (\Lambda_{11}, \Lambda_{12})^T \begin{pmatrix} -c & \frac{c^2}{\rho} \\ \rho & -c \end{pmatrix} = (0, 0).$$

Получаем  $-\Lambda_{11}c + \Lambda_{12}\rho = 0$ ,  $\Lambda_{11}\frac{c^2}{\rho} - \Lambda_{12}c = 0$ . Пусть первая компонента вектора будет  $\Lambda_{11} = 1$ , тогда  $\Lambda_{12} = \frac{c}{\rho}$ .

Второй собственный вектор  $\Lambda_2$  для  $\lambda_2 = w - c$ :

$$\Lambda_2^T \begin{pmatrix} w - \lambda_2 & \frac{c^2}{\rho} \\ \rho & w - \lambda_2 \end{pmatrix} = (\Lambda_{21}, \Lambda_{22})^T \begin{pmatrix} c & \frac{c^2}{\rho} \\ \rho & c \end{pmatrix} = (0, 0).$$

Получаем  $\Lambda_{11}c + \Lambda_{12}\rho = 0$ ,  $\Lambda_{11}\frac{c^2}{\rho} + \Lambda_{12}c = 0$  и  $\Lambda_{21} = 1$ ,  $\Lambda_{22} = -\frac{c}{\rho}$ .

Умножим сопряжённую систему (9.24) сначала на  $\Lambda_1^T$ , потом на  $\Lambda_2^T$ . Получим вдоль характеристик  $\xi_1$  и  $\xi_2$  два характеристических уравнения в виде:

$$\begin{aligned} \Lambda_1^T \left( -\frac{\partial f}{\partial t} - \lambda_1 \frac{\partial f}{\partial x} \right) + \Lambda_1^T (wuf_1, \rho uf_1) \\ = -\frac{\partial f_1}{\partial t} - \frac{c}{\rho} \frac{\partial f_2}{\partial t} - (w + c) \frac{\partial f_1}{\partial x} - \frac{c}{\rho} (w + c) \frac{\partial f_2}{\partial x} + (w + c)uf_1 \\ = 0, \end{aligned}$$

$$\begin{aligned} \Lambda_2^T \left( -\frac{\partial f}{\partial t} - \lambda_2 \frac{\partial f}{\partial x} \right) + \Lambda_2^T (wuf_1, \rho uf_1) \\ = -\frac{\partial f_1}{\partial t} + \frac{c}{\rho} \frac{\partial f_2}{\partial t} - (w - c) \frac{\partial f_1}{\partial x} + \frac{c}{\rho} (w - c) \frac{\partial f_2}{\partial x} + (w - c)uf_1 \\ = 0 \end{aligned}$$

Можно записать короче:

$$-\frac{\partial f_1}{\partial t} - (w \pm c) \frac{\partial f_1}{\partial x} \mp \frac{c}{\rho} \frac{\partial f_2}{\partial t} \mp \frac{c}{\rho} (w \pm c) \frac{\partial f_2}{\partial x} + (w \pm c) u f_1 = 0.$$

где верхний знак соответствует характеристикам первого семейства  $\xi_1$ , а нижний — характеристикам второго семейства  $\xi_2$ .

Если учесть, что  $\frac{\partial \cdot}{\partial t} + (w \pm c) \frac{\partial \cdot}{\partial x} = \frac{d \cdot}{dt} \Big|_{\xi_{1,2}}$ , то полученные уравнения

можно представить в обыкновенных дифференциалах:

$$-\frac{df_1}{dt} \mp \frac{c}{\rho} \frac{df_2}{dt} + (w \pm c) u f_1 = 0, \quad (9.30)$$

## 9.4 Аппроксимация задач для численного решения

### 9.4.1. Сетка аппроксимаций

Аппроксимация задачи осуществлялась на двух сетках.

Для управления  $u$  задавалась неподвижная пространственная сетка в сопле гидропушки.

$$\{x_{i_s} = i_s \Delta_s x, \quad i_s = 0 \dots n_s, \quad \Delta_s x = (x_b - x_a)/n_s\},$$

где  $i_s$  — номера узлов сетки сопла,  $\Delta_s x$  шаг сетки сопла,  $n_s$  — количество узлов неподвижной сетки сопла.

Исходная и сопряжённая задачи решались на подвижной пространственно-временной сетке, которая изображена на рисунке 9.5. Причём, для сопряжённой задачи (9.24)–(9.26), согласно условиям управляемости, использовалась только верхняя, практически неподвижная, часть сетки от  $t_1$  до  $t_2$ .

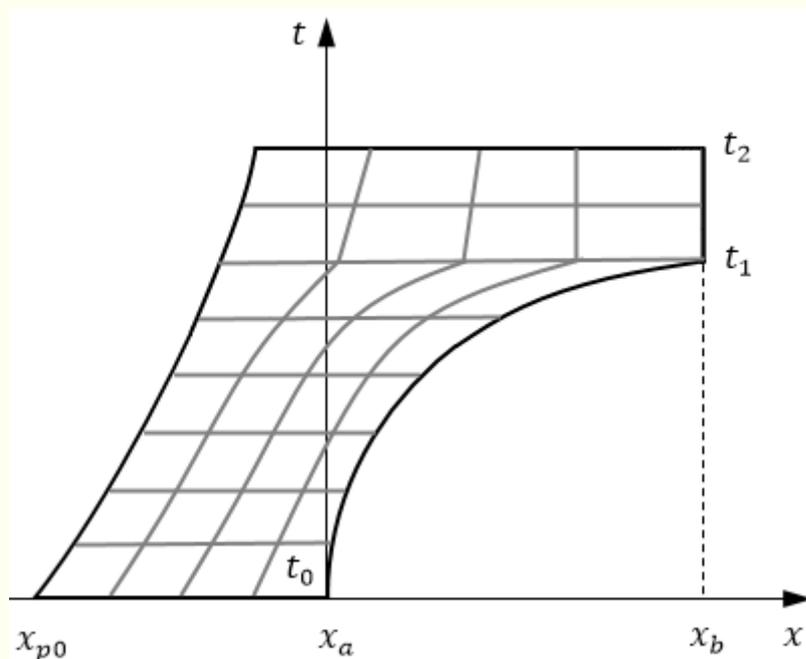


Рисунок 9.5 — Принципиальный вид подвижной расчётной сетки

Поскольку квазилинейные уравнения обеих гиперболических задач не являются однородными, полулинейными, то их *инварианты*  $r_{1,2}$  будут определяться сложным образом и не будут постоянными. Поэтому, задачи решались методом с использованием характеристик, описанным в предыдущем разделе. Текущая реализация метода оказалась заметно сложнее из-за наличия подвижной сетки. Уравнения в характеристической форме с обыкновенными дифференциалами аппроксимировались по *явной конечно-разностной схеме*, показанной на рисунке 9.6.

Границы подвижной пространственной сетки были привязаны к границам потока  $\Gamma_p(t)$  и  $\Gamma_b(t)$ , поэтому сетка меняла свой шаг на разных временных слоях:

$$\begin{cases} x_i^j = \Gamma_p^j + i\Delta x^j, & i = 0 \dots n, & \Delta x^j = (\Gamma_b^j - \Gamma_p^j)/n, \\ t^j = t^{j-1} + \Delta t^j, & j = 1 \dots m, \end{cases}$$

где  $n$  — количество узлов подвижной сетки,  $\Gamma_p^j \equiv x_0^j$ ,  $\Gamma_b^j \equiv x_n^j$ . Для исходной задачи новый временной слой — это  $j + 1$ , а для сопряжённой — слой  $j - 1$ . Точки  $i \pm$  (белые точки на рисунке 9.6) соответствуют месту выхода характеристик  $\xi_1, \xi_2$  слева или справа от точки  $i$ .

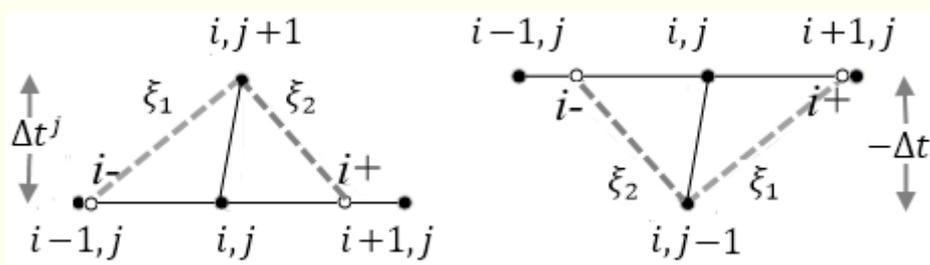


Рисунок 9.6 — Шаблоны конечно-разностных аппроксимаций для решения исходной (слева) и сопряжённой (справа) задач

Конечно-разностные аппроксимации предполагают докритические течения со скоростью потока  $w < c$ . Условие устойчивости на каждом временном слое для всех характеристик на подвижной сетке имеет вид:  $\Delta t^j \leq \min_{\forall x} \frac{\Delta x^j}{|(w \pm c)^j \mp \Delta \dot{x}^j|}$ , где  $\Delta \dot{x}^j$  — изменение скорости узлов сетки на соседних слоях времени. Все характеристики, пересекающиеся на новом временном слое в центральном узле  $i$ , не должны выходить со старого временного слоя из-за пределов отрезка  $[x_i^j - \Delta x^j, x_i^j + \Delta x^j]$ .

Оценку шага  $\Delta t^j$  можно делать по правой границе. Именно правая граница обладает наибольшей скоростью характеристик, в частности —  $\xi_1$ , поэтому здесь будет получен наименьший шаг по времени, удовлетворяющий всем точкам пространственной сетки.

До начала истечения:

$$\Delta t^j = k_z \frac{\Delta x^j}{(w + c)_n^j - \Delta \dot{x}_n^j}, \quad \Delta \dot{x}_n^j = \begin{cases} w_n^j - w_n^{j-1}, & x_n^j < x_b, \\ 0, & x_n^j \geq x_b \end{cases}$$

где  $k_z = 0.9$  — коэффициент запаса.

Границы потока задавались выражениями:

$$\begin{aligned} x_0^{j+1} &= x_0^j + w_0^j \Delta t^j, \\ x_n^{j+1} &= x_n^j + w_n^j \Delta t^j. \end{aligned}$$

При первом наступлении события  $x_n^{j+1} \geq x_b$  фиксируем  $j_m \leftarrow j + 1$  и

$$x_n^{j+1} \leftarrow x_b, \quad \forall j \geq j_m,$$

где индекс  $m$  обозначает слой времени  $j_m$  начала истечения воды из сопла, что соответствует времени  $t_1$ . Такое фиксирование правой границы потока в момент начала истечения приводит к требованию уменьшения текущего шага по времени.

В момент начала истечения

$$\Delta t^j = k_z \frac{x_b - x_n^{j_m-1}}{w_n^{j_m-1}}, \quad j = j_m - 1.$$

Далее, при продолжении истечения,

$$\Delta t^j = k_z \frac{\Delta x^j}{(w + c)_n^j}, \quad \forall j \geq j_m.$$

Для приблизительного определения местоположения точек  $i \pm$  необходимо решать уравнения характеристик  $\frac{dx}{dt} = w \pm c$ , например,

$$x_{i+}^j = x_i^{j+1} - (w - c)_i^j \Delta t^j, \quad x_{i-}^j = x_i^{j+1} - (w + c)_i^j \Delta t^j.$$

Далее в точках  $x_{i\pm}$  можно определять параметры потока  $w_{i\pm}^j$ ,  $\rho_{i\pm}^j$  или сопряжённое состояние  $f_{1i\pm}^j$ ,  $f_{2i\pm}^j$  посредством интерполяции, например, квадратичной, полиномом Лагранжа (8.22). Кроме этого, на каждом временном слое придётся делать и интерполяцию значений управления  $u(x_{i_s})$  из узлов  $i_s = 0 \dots n_s$  неподвижной сетки в точки  $i \pm$  узлов  $i = 0 \dots n$  подвижной сетки потока.

### 9.4.2. Аппроксимация исходной задачи

Аппроксимируем обыкновенные дифференциальные уравнения (9.10) исходной задачи по методу Эйлера. Этот метод имеет первый порядок точности относительно шага аппроксимаций  $\Delta t$ . Коэффициенты, свободные члены берутся со старого временного слоя. Тогда на характеристиках первого семейства мы получим:

$$\frac{\rho_i^{j+1} - \rho_{i-}^j}{\Delta t^j} + e_- \frac{w_i^{j+1} - w_{i-}^j}{\Delta t^j} + \mathcal{F}_- = 0,$$

где  $e_- = \left(\frac{\rho}{c}\right)_{i-}^j$ ,  $\mathcal{F}_- = \varphi_{i-}^j$ .

На характеристиках второго семейства получим:

$$\frac{\rho_i^{j+1} - \rho_{i+}^j}{\Delta t^j} - e_+ \frac{w_i^{j+1} - w_{i+}^j}{\Delta t^j} + \mathcal{F}_+ = 0,$$

где  $e_+ = \left(\frac{\rho}{c}\right)_{i+}^j$ ,  $\mathcal{F}_+ = \varphi_{i+}^j$ .

Из приведенной системы двух уравнений для двух неизвестных  $\rho_i^{j+1}$  и  $w_i^{j+1}$  находим:

$$\begin{aligned} w_i^{j+1} &= \frac{1}{e_- + e_+} (e_- w_{i-}^j + e_+ w_{i+}^j + \rho_{i-}^j - \rho_{i+}^j + (\mathcal{F}_+ - \mathcal{F}_-) \Delta t^j), \\ \rho_i^{j+1} &= \rho_{i-}^j - e_+ (w_i^{j+1} - w_{i-}^j) - \mathcal{F}_- \Delta t^j, \\ & i = 1 \dots n - 1, \quad j = 0 \dots m - 1. \end{aligned} \tag{9.31}$$

На правой границе  $\Gamma_b$ , при  $i = n$ , будем использовать характеристическое уравнение, соответствующее приходящей туда характеристике  $\xi_1$ . Если учесть, что плотность потока взаимодействующего с атмосферой  $\rho_n^{j+1} = \rho_0$ , то во втором уравнении (9.31) мы имеем  $\rho_0 = \rho_{n-}^j - e_- (w_n^{j+1} - w_{n-}^j) - \mathcal{F}_- \Delta t^j$ . Откуда, с учётом первого уравнения (9.31), следует значение граничной скорости потока:

$$w_i^{j+1} = w_{i-}^j + \frac{1}{e_-} (\rho_{i-}^j - \rho_0 - \mathcal{F}_- \Delta t^j), \quad i = n.$$

На левой границе  $\Gamma_p$ , при  $i = 0$ , будем использовать характеристическое уравнение, соответствующее характеристике  $\xi_2$ . Скорость поршня на новом временном слое находим после аппроксимации уравнения его движения (9.6), а плотность — из аппроксимации условий потока на второй характеристике. Получаем:

$$w_i^{j+1} = w_i^j + \left( \frac{uB}{m_p} \left( 1 - \left( \frac{\rho}{\rho_0} \right)^n \right) \right)_i^j \Delta t^j,$$

$$\rho_i^{j+1} = \rho_{i+}^j + e_+ (w_i^{j+1} - w_{i+}^j) - \mathcal{F}_+^j \Delta t^j, \quad i = 0.$$

Мы получили конечно-разностные аппроксимации первого порядка точности для исходной задачи на пространственно-временной сетке, характеризуемой тремя множествами  $\{t^j\}_{j=0}^m$ ,  $\{x_0^j\}_{j=0}^m$ ,  $\{x_n^j\}_{j=0}^m$ .

Поскольку исходная задача нелинейная, с быстро меняющимся состоянием потока, то целесообразно улучшать аппроксимации до второго порядка, например, по методу предиктор-корректор. Метод основан на дополнительном втором шаге после метода Эйлера, когда коэффициенты, свободные члены берутся с нового временного слоя. То есть, решение  $\rho_i^{j+1} \stackrel{\text{def}}{=} \hat{\rho}$  и  $w_i^{j+1} \stackrel{\text{def}}{=} \hat{w}$ , полученное ранее методом Эйлера, называется предсказанием (предиктор) решения. На втором шаге метода повторяется (корректор) решение на основе имеющихся  $\hat{\rho}$ ,  $\hat{w}$ , после чего получаем новое решения  $\check{\rho}$  и  $\check{w}$ . Результирующим решением считается среднеарифметическое двух предыдущих решений.

Для второго шага корректора задаём  $e_- = e_+ = \hat{e} = \frac{\hat{\rho}}{c(\hat{w})}$  и  $\mathcal{F}_+ = \mathcal{F}_- = \hat{\mathcal{F}} = \hat{\phi}_i$  и получаем:

$$\check{w} = \frac{w_{i-}^j + w_{i+}^j}{2} + \frac{\rho_{i-}^j - \rho_{i+}^j}{2\hat{e}},$$

$$\check{\rho} = \rho_{i-}^j - \hat{e}(\hat{w} - w_{i-}^j) - \hat{\mathcal{F}} \Delta t^j, \quad i = 1 \dots n - 1.$$

На правой границе для скорости потока:

$$\begin{aligned}\tilde{w} &= w_{i-}^j + \frac{1}{\hat{e}}(\rho_{i-}^j - \rho_0 - \hat{\mathcal{F}}\Delta t^j), \\ \check{\rho} &= \rho_0, \quad i = n.\end{aligned}$$

Напомним,  $\rho_0$  — это плотность воды при атмосферном давлении.

На левой границе, где имеется поршень:

$$\begin{aligned}\tilde{w} &= w_i^j + \left( \frac{uB}{m_p} \left( 1 - \left( \frac{\hat{\rho}}{\rho_0} \right)^n \right) \right)_i^j \Delta t^j, \\ \check{\rho} &= \rho_{i+}^j + \hat{e}(\hat{w} - w_{i+}^j) - \hat{\mathcal{F}}\Delta t^j, \quad i = 0.\end{aligned}$$

В итоге новое решение методом предиктор-корректор — это:

$$\begin{aligned}w_i^{j+1} &= \frac{1}{2}(\hat{w} + \tilde{w}), \\ \rho_i^{j+1} &= \frac{1}{2}(\hat{\rho} + \check{\rho}), \quad i = 0 \dots n, \quad j = 1 \dots m.\end{aligned}$$

Мы получили необходимые конечно-разностные аппроксимации метода предиктор-корректор второго порядка точности. Для этого метода, в дополнение к методу Эйлера, необходимо непосредственно после вычислений  $\rho$ ,  $w$  в каждом узле сетки сделать их перерасчёт с новыми значениями коэффициентов  $e$  и свободного члена  $\mathcal{F}$ . При этом новые интерполяции состояния потока не требуются.

При выполнении расчётов течения необходимо отметить две особенности. Первая, — условие «антикавитации»:

$$\text{если } \rho_i^{j+1} > \rho_0, \text{ тогда } \rho_i^{j+1} \leftarrow \rho_0, \quad \forall i, j.$$

Вторая особенность — это возможное появление незначительных по амплитуде осцилляций в решении, что связано с накоплением интерполяционных погрешностей в быстро меняющемся состоянии потока при втекании воды в сопло. Данные погрешности на каждом новом временном слое легко устраняются среднеарифметическим сглаживанием по трём точкам, типа:

$$y_i^j = \frac{(y_{i-1}^j + y_i^j + y_{i+1}^j)}{3}, \quad i = 1 \dots n - 1, \forall j.$$

Безусловно, результат такого сглаживания не будет точно отражать очень резкие изменения в потоке, но зато устранил осцилляции состояния потока.

### 9.4.3. Аппроксимация сопряжённой задачи

Будем использовать пространственно-временную сетку, полученную на  $\Omega_2$  при решении исходной задачи. При решении сопряжённой задачи (9.24)–(9.26) надо двигаться по этой сетке в обратном по времени направлении от  $t_2$  до  $t_1$ . То есть, решение сопряжённой задачи следует заканчивать при первом пересечении замыкающей характеристикой  $\xi_1$  координаты  $x_a$ . Обозначим узел  $i, j$  подвижной сетки, соответствующий такому пересечению, символами  $i_a, j_a$ .

Для приблизительного определения местоположения точек  $i \pm$  на старом временном слое  $j$ , согласно правому рисунку 9.6, необходимо решать уравнения  $\frac{dx}{dt} = w \pm c$  между слоями  $j$  и  $j - 1$ :

$$x_{i+}^j = x_i^{j-1} + (w + c)_i^j \Delta t^j, \quad x_{i-}^j = x_i^{j-1} + (w - c)_i^j \Delta t^j.$$

Аппроксимируем уравнения сопряженной задачи в характеристической форме (9.30), аналогично исходной задаче. Коэффициенты и свободные члены будем брать со старого слоя  $j$ .

На характеристиках первого семейства, с учётом движения в обратном направлении по времени, имеем:

$$-\frac{f_{1i}^{j-1} - f_{1i+}^j}{-\Delta t^j} - e_+ \frac{f_{2i}^{j-1} - f_{2i+}^j}{-\Delta t^j} + \mathcal{F}_+ = 0,$$

где  $e_+ = \left(\frac{c}{\rho}\right)_{i+}^j$ ,  $\mathcal{F}_+ = ((w + c)uf_1)_{i+}^j$ .

На характеристиках второго семейства:

$$-\frac{f_{1i}^{j-1} - f_{1i-}^j}{-\Delta t^j} + e_- \frac{f_{2i}^{j-1} - f_{2i-}^j}{-\Delta t^j} + \mathcal{F}_- = 0,$$

где  $e_- = \left(\frac{c}{\rho}\right)_{i-}^j$ ,  $\mathcal{F}_- = ((w - c)uf_1)_{i-}^j$ .

Из приведенной системы двух уравнений для двух неизвестных  $f_{1i}^{j-1}$  и  $f_{2i}^{j-1}$  находим:

$$f_{2i}^{j-1} = \frac{1}{e_+ + e_-} (e_+ f_{2i+}^j + e_- f_{2i-}^j + f_{1i+}^j - f_{1i-}^j - (\mathcal{F}_+ - \mathcal{F}_-) \Delta t^j),$$

$$f_{1i}^{j-1} = f_{1i-}^j + e_- (f_{2i}^{j-1} - f_{2i-}^j) - \mathcal{F}_- \Delta t^j,$$

$$i = i_a + 1 \dots n - 1, \quad j = m - 1 \dots j_a + 1.$$

где  $e_+ = \left(\frac{c}{\rho}\right)_{i+}^j$ ,  $\mathcal{F}_+ = ((w + c)uf_1)_{i+}^j$ .

На правой границе при  $i = n$  будем использовать характеристическое уравнение, соответствующее приходящей туда характеристике  $\xi_2$ :

$$-\frac{f_{1i}^{j-1} - f_{1i-}^j}{-\Delta t^j} + e_- \frac{f_{2i}^{j-1} - f_{2i-}^j}{-\Delta t^j} + \mathcal{F}_- = 0$$

совместно с граничным условием (9.27) на  $\Gamma_{b1}$ :

$$f_{1i}^{j-1} + e_i f_{2i}^{j-1} + \frac{I'_w}{\rho_0} = 0, \quad i = n,$$

где  $e_i = \frac{w_i^{j-1}}{\rho_0}$ . Решая полученную систему уравнений, находим на правой границе:

$$f_{2i}^{j-1} = \frac{e_- f_{2i-}^j - f_{1i-}^j - I'_w / \rho_0 + \mathcal{F}_- \Delta t^j}{e_- + e_i},$$

$$f_{1i}^{j-1} = -e_i f_{2i}^{j-1} - I'_w / \rho_0, \quad i = n.$$

На левой границе решение известно:

$$f_{1i}^{j-1} = 0, \quad f_{2i}^{j-1} = 0, \quad i = i_a.$$

Терминальное условие так же известно:

$$f_{1i}^m = 0, \quad f_{2i}^m = 0, \quad i = i_a \dots n.$$

Мы получили все необходимые конечно-разностные аппроксимации сопряженной задачи для её решения в прямоугольнике  $\bar{\Omega}_3 = [x_a, x_b] \times [t_1, t_2]$ . Поскольку сопряжённая задача линейная, то полученные аппроксимации можно далее не уточнять пересчётами второго порядка.

#### 9.4.4. Аппроксимация градиента

При решении исходной и сопряжённой задач нам приходилось интерполировать управление  $u(x_{i_s})$  на неподвижной сетке в координаты  $x_{i_{\pm}}$  подвижной сетки. При расчёте градиента (9.29) придётся делать «обратную» интерполяцию.

Значения градиента (9.29) в узлах неподвижной сетки сопла имеют вид:

$$\nabla J(u; x_{i_s}) = \int_{t_1}^{t_2} \rho(x_{i_s}, t) w(x_{i_s}, t) f_1(x_{i_s}, t) dt + \kappa J, \quad i_s = 1 \dots n_s - 1.$$

В каждой точке  $x_{i_s}$  необходимо проводить численное интегрирование по времени. При этом, значения  $\rho(x_{i_s}, t_j)$ ,  $w(x_{i_s}, t_j)$ ,  $f_1(x_{i_s}, t_j)$  на каждом слое  $j = j_a \dots m - 1$  надо интерполировать в координаты точек сопла  $x_{i_s}$ .

Площадь сопла для  $x \in [x_a, x_b]$  на основе управления  $u(x)$  определяется как

$$\sigma(x_{i_s}) = \sigma_a e^{\int_{x_a}^{x_{i_s}} u(\zeta) d\zeta}, \quad i_s = 1 \dots n_s.$$

Поскольку начало сопла закреплено, то здесь  $\sigma(x_{i_s=0}) = \sigma_a$ .

## 9.5 Численное решение

Расчёты течения воды в гидропушке проводились по описанным выше конечно-разностным алгоритмам со следующими исходными данными [73]. Начало координат  $x_a = 0$ , длина сопла  $x_b = 0.253$  м,

начальное положение поршня с левой границей воды  $x_{p0} = -0.28$  м, масса поршня  $m_p = 2.25$  кг, начальная скорость поршня и воды  $w_0 = 76.2$  м/с, плотность воды при атмосферном давлении  $\rho_0 = 10^3$  кг/м<sup>3</sup>, радиус ствола гидропушки  $R_a = 33 \cdot 10^{-3}$  м. Константы состояния воды  $B = 304.5 \cdot 10^6$ ,  $n = 7.15$ . Начало втекания воды в сопло  $t_0 = 0$ .

Начало истечения струи из сопла наблюдалось при  $t_1 \approx 2.7 \cdot 10^{-3}$  с. Среднее значение силы импульса полученной струи оценивалось по целевому функционалу  $J(u)$  в интервале  $t_2 - t_1 = 3 \cdot 10^{-4}$  с. Данный интервал составляет примерно  $2 \int_{x_a}^{x_b} \frac{dx}{\lambda_1}$ . При этом время начала истечения  $t_1 > t_3$ , т.е. поршень не влияет на истечение. Таким образом заданное время  $t_2$  соответствует минимально достаточной области

$$\omega_{min} = x_b \times (t_1, t_2),$$

показанной на рисунке 9.4а, что удовлетворяет условию управляемости (9.28).

Исходная задача решалась в сложной замкнутой области  $\bar{\Omega}$  рисунка 9.2, а сопряжённая — на небольшом прямоугольнике  $\bar{\Omega}_2 = [x_a, x_b] \times [t_1, t_2]$ . При этом, пространственные сетки задавались количеством шагов для расчёта течения воды  $n = 50$ , для формы сопла  $n_s = 50$ . Шаги по времени  $m$  — до  $10^3$ . Количество шагов по времени зависит от формы сопла и определяется конечным временем  $t_2$ . Шаг сетки сопла  $\Delta_s x = 5.06$  мм, а «вычислительное» множество

$$S = (x_a, x_b) = [\Delta_s x, x_b - \Delta_s x].$$

### 9.5.1. Оценка выпуклости целевого функционала

Для получения представления о выпуклости функционала (9.9), оценки размеров компакта  $\mathcal{U}$  возможного корректного решения задачи оптимизации, давайте рассмотрим управление в классе функций-

конусов. При этом  $J(u) = J(R_b)$ , т.е.  $J$  будет функцией радиуса среза сопла  $R_b$ . Площадь  $\sigma(x) = \pi \left( \frac{R_b - R_a}{x_b - x_a} x + R_a \right)^2$ ,  $R_a = \sqrt{\sigma_a / \pi}$ .

Полученная функция  $J(R_b)$  и норма её градиента  $\|\nabla J\|$  представлены на рисунке 9.7. Правые точки на графике соответствуют соплу в виде трубы, когда  $R_b = R_a$ . Левые точки соответствуют минимально возможному радиусу среза сопла, при котором ещё не наступает сверхзвуковое истечение. Из поведения  $J$  (чёрные точки), мы видим, что целевой функционал не является ни выпуклым, ни вогнутым.

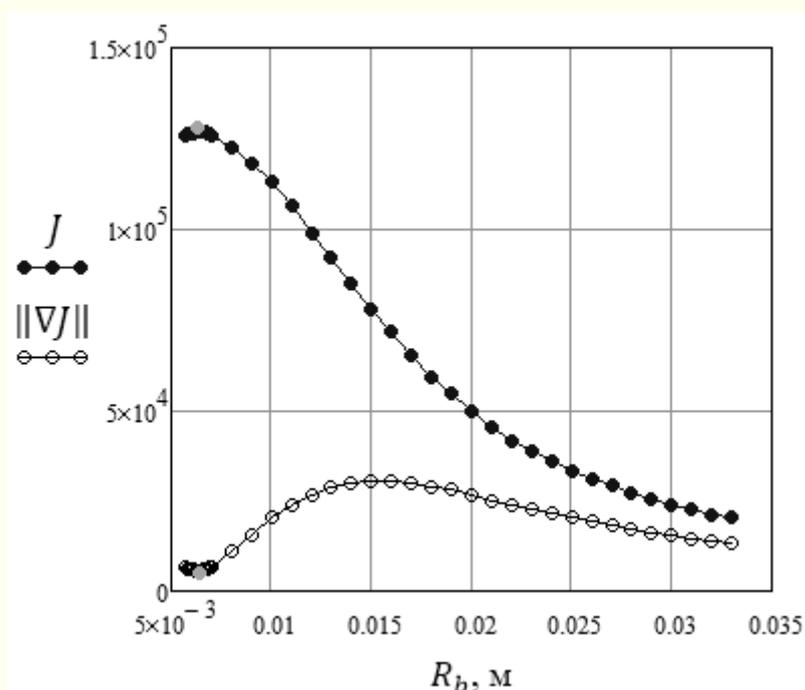


Рисунок 9.7 — Значения целевого функционала  $J(R_b)$  и нормы его градиента

Рядом с левой границей имеется  $\max J$  (серая точка), с большой вогнутостью, минимальной величиной градиента и очень малой окрестностью дозвукового течения. Функционал, в процессе изменения радиуса сопла от трубы до минимально допустимого сужения, меняет

совую начальную выпуклость на вогнутость. Причём, вогнутость на много сильнее начальной выпуклости.

Если максимизировать функцию  $J(R_b)$ , начиная с сопла-трубы при  $R_b = R_a$ , классическим конечномерным градиентным методом с первой стратегией выбора  $b = b^0$ , то небольшие начальные шаги  $b\nabla J^k$  при движении влево сменяться на большие шаги (из-за возрастания нормы градиента). Поскольку вогнутость имеет место в относительно малой окрестности максимума, то большие шаги могут приводить к перепрыгиванию через экстремум вплоть до появления сверхзвукового истечения. Это недопустимо. Попасты в  $\max J$ , возможно, удастся только с очень малым  $b$  и, естественно, при чрезмерно большом количестве итераций.

Применение популярных МНС или МСГ и других методов, использующих *линейный поиск* наилучшего шагового множителя  $b^k$ , здесь невозможно, поскольку такой поиск, в общем случае, будет начинаться со сверхзвукового истечения. Здесь также нельзя применить МРНС с постоянным параметром регулирования направления спуска  $\alpha^k = \alpha$ , выбранным на первой итерации. Слишком радикально меняется выпуклость целевого функционала в области значений управления.

Для всех рассмотренных вариантов конической формы сопла градиент  $J(u) = J(R_b)$ , изображённый на рисунке 9.7 нормой  $\|\nabla J(u)\|$ , корректно отражает *чувствительность* целевого функционала к управлению при дозвуковом истечении:

$$w(x_b, t) < c_0, \quad \forall t.$$

В этих условиях о компакте  $\mathcal{U}$  можно сказать следующее. Фактически,  $\mathcal{U} = U$  в выражении (9.4) с дополнительным требованием дозвукового истечения:

$$\mathcal{U} = \{u: u(x) \leq 0, \quad \forall x; \quad w(x_b, t) < c_0, \quad \forall t\}. \quad (9.32)$$

В таком множестве  $\mathcal{U}$  может существовать корректное значение градиента и отыскиваться оптимальное управление. Напомним, что функции  $u(x)$  из  $U = \mathcal{U}$  не должны слишком резко изменяться. Будем далее это иметь в виду.

Перейдём к реальной оптимизации произвольной формы сопла.

### 9.5.2. Первый локальный максимум

Для максимизации целевого функционала  $J(u)$ , с переменной выпуклостью и вогнутостью, придётся использовать разные версии обычного градиентного метода, различающиеся стратегиями выбора шагового множителя. Будем далее применять бесконечномерный градиентный метод (1.5), когда  $p = \nabla J$ :

$$u^{k+1}(x) = u^k(x) + b^k \nabla J^k(u; x) \quad \text{на } S, \quad k = 0, 1, \dots \quad (9.33)$$

Уточним процедуру вычисления градиента  $\nabla J$ . Согласно (9.29)  $\nabla J = \mathbb{U}^* f + \kappa J$ . Для его вычисления необходимо знание весового коэффициента  $\kappa$ , который выравнивает вычислительные помехи решения исходной и сопряжённой задач. Следовательно, перед применением алгоритма (9.33) необходимо провести юстировку  $\kappa$ . Сделаем это, как и предлагалось ранее, на сопле-конусе.

Из расчётов следует, что  $\max J(R_b) = 1.27 \cdot 10^5$ , которое достигается при оптимуме  $R_b = 6.4$  мм. Теперь необходимо провести вычислительный эксперимент по подбору коэффициента  $\kappa$ , такого, чтобы абсолютное значение градиента, при установленном оптимуме  $R_b$ , было минимальным. Численный эксперимент дал значение  $\kappa = 0.17$ , при  $\min \|\nabla J\| = 5.74 \cdot 10^3$ . Именно такое поведение градиента, при отъюстированном  $\kappa$ , изображено на рисунке 9.7.

Теперь перейдём к решению бесконечномерной задачи оптимизации. Пусть начальное приближение управления

$$u^0(x) = 0,$$

что соответствует соплу-трубе с площадью сечения  $\sigma^0(x) = \pi R_a^2$ .

Зададим начальное значение шагового множителя

$$b^0 = \frac{0.5}{\|\nabla J^0\|},$$

что соответствует первому шагу  $\|u^1 - u^0\| = 0.5$ . Остановку итераций будем осуществлять при практическом прекращении сходимости по критерию (1.24):

$$\frac{\|u^k - u^{k-1}\|}{\|u^{k-1}\|} \leq \varepsilon_4, \quad \varepsilon_4 = 10^{-6}.$$

Сначала рассмотрим задачу оптимизации без ограничения (9.4) контроля расширения сопла.

Шаговый множитель  $b^k$  будем выбирать согласно *второй стратегии* с адаптивной релаксацией по алгоритму (1.15), при  $r = 1$ . При этом градиентный метод (9.33) для максимизации принимает вид:

$$\left\{ \begin{array}{l} \text{Если } J^k > J^{k-1} \text{ тогда:} \\ \quad \cdot \text{ если } \|\nabla J^k\| < \|\nabla J^{k-1}\| \text{ тогда } b^k = b_1 b^{k-1}, \\ \quad \text{иначе } b^k = \frac{b^{k-1} \|\nabla J^{k-1}\|}{b_3 \|\nabla J^k\|}; \\ \quad \cdot u^{k+1} = u^k + b^k \nabla J^k. \\ \text{Иначе повторять до } J^k > J^{k-1}: \\ \quad b^{k-1} \leftarrow b_2 b^{k-1}, \quad u^{k+1} = u^{k-1} + b^{k-1} \nabla J^{k-1}, \\ \quad \text{если } b^{k-1} \approx 0, \text{ тогда стоп итерации } k. \end{array} \right. \quad (9.34)$$

Были подобраны значения параметров:  $b_1 = 1.05$ ,  $b_2 = 0.3$ ,  $b_3 = 1.05$ . Подбор осуществлялся из соображений:  $b_1$  должен быть небольшим для «осторожного» подхода к ближайшему сильно вогнутому экстремуму;  $b_2$  во всех расчётах удовлетворительно выполнял свою задачу при 0.3;  $b_3$  — чуть больше единице, иначе, при больших значениях сходимость сильно замедлялась, а при  $b_3 = 1$  происходил слишком большой переход (из-за резкой смены выпуклости) через экстремум.

Результаты оптимизации представлены на рисунке 9.8 в виде радиуса сопла для полученной оптимальной площади сечения  $\sigma^{32}$ . Сходимость градиентного метода в виде (9.34) закончилась на итерации  $k = 32$ . Оптимальным оказалось сопло с формой, практически совпадающей с конусом. Найденный *локальный максимум* целевого функционала имеет значение  $\max J = 1.26 \cdot 10^5$ .

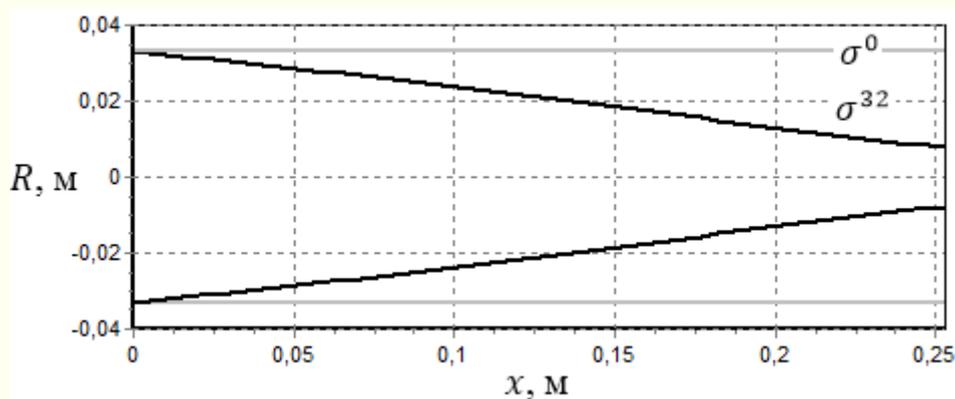


Рисунок 9.8 — Оптимальное сопла в первом экстремуме

Потребовалось общее количество 76 вычислений  $J$  и  $\nabla J$ . Если учесть, что на каждой итерации необходимо дважды вычислять  $J$  и  $\nabla J$ , то блок «Иначе повторять» в (9.34) потребовал дополнительных вычислений  $J$  всего 12 раз. Причём, эти вычисления, для уменьшения шагового множителя, происходили вблизи локального  $\max J$  на некоторых итерациях, начиная с  $k = 29$ .

Градиентным методом (9.33), с постоянным шаговым множителем *первой стратегии*, не удалось попасть в рассматриваемый локальный  $\max J$ . При  $b \gtrsim 10b^0$  перешагивание экстремума сопровождалось сверхзвуковым истечением из-за чрезмерного сужения сопла, т.е. мы выходили за пределы установленного компакта  $\mathcal{U}$  (9.32). При уменьшении  $b$  до  $10^{-2}b^0$  сходимость не заканчивалась даже поле

нескольких тысяч итераций. В конце итераций всегда наблюдалось неконтролируемое расширение сопла за пределы ствола гидропушки с аварийным завершением сверхзвукового истечения. Опять-таки, наблюдался выход из компакта  $\mathcal{U}$ .

### 9.5.3. Второй локальный максимум

Выясним, является ли первый максимум целевого функционала глобальным. Для этого необходимо посредством метода (9.34) грубо перейти через найденный локальный максимум и далее — через близлежащий за ним минимум, чтобы опять выполнилось условие роста функционала  $J^k > J^{k-1}$ .

Увеличим шаги в окрестности максимума посредством параметра  $b_1$ , изменив его начальное значение 1.05 на новое  $b_1 = 1.16$ . При этом реализуется переход через локальные максимум и минимум в область роста  $J$ . Но, если не учитывать ограничение (9.4), то далее происходит расширение сопла за пределы ствола гидропушки мы выходим из компакта  $\mathcal{U}$ . На рисунке 9.9 для наглядности, на итерации  $k = 41$ , показано точечными линиями такое расширение сопла.

Ограничение (9.4) легко реализуется проецированием  $P_U(u^{k+1})$  нового приближения управления на допустимое множество  $U$ . После шага любым алгоритмом полученное управление  $u^{k+1}$  дополнительно корректируется:

$$\text{Если } u^{k+1}(x) > 0, \text{ тогда } u^{k+1}(x) = 0.$$

Дальнейшие расчёты оптимизации ограниченного сопла методом (9.34), а также традиционным методом (9.33) при любых  $b^0$ , приводили к сверхзвуковому истечению. Опять происходил выход из компакта  $\mathcal{U}$ . Для (9.34) этот «неудачный результат» означает, что параметр увеличения шагов  $b_1$  был недопустимо большим, и возможный второй локальный  $\max J$  был грубо пройден. По-видимому, ожидаемый

максимум имеет очень высокую вогнутость функционала  $J(u)$ . Здесь вместо метода (9.34) необходимо применять адаптивную релаксацию шагов типа (1.16), позволяющую уменьшать  $b_1$  в окрестности ожидаемого  $\max J$ .

Окрестность предполагаемого максимума можно определить по максимальной скорости истечения  $w_b = \max_{t \in (t_1, t_2)} w(x_b, t)$ . Она должна быть близкой к скорости звука  $c_0$ . В частности, такой скоростью была принята  $w_b = 800$  м/с. При её превышении усиление шагов было отменено, т.е. было принято  $b_1 = 1$ .

Метод (1.16), для максимизации функционала  $J$ , имеет вид:

$$\left\{ \begin{array}{l} \text{Если } J^k > J^{k-1} \text{ тогда:} \\ \quad \cdot \text{ если } \|\nabla J^k\| < \|\nabla J^{k-1}\| \text{ тогда} \\ \quad \quad \{ \text{если } w_b > 800 \text{ тогда } b_1 = 1.0 \} \\ \quad \quad b^k = b_1 b^{k-1}, \\ \quad \text{иначе } b^k = \frac{b^{k-1} \|\nabla J^{k-1}\|}{b_3 \|\nabla J^k\|}; \\ \quad \cdot u^{k+1} = u^k + b^k \nabla J^k. \\ \text{Иначе повторять до } J^k > J^{k-1}: \\ \quad b^{k-1} \leftarrow b_2 b^{k-1}, \quad u^{k+1} = u^{k-1} + b^{k-1} \nabla J^{k-1}, \\ \quad \text{если } b^{k-1} \approx 0 \text{ тогда стоп итерации } k. \end{array} \right. \quad (9.35)$$

Параметры были теми же:  $b_2 = 0.3$ ,  $b_3 = 1.05$ , начальное  $b_1 = 1.16$ .

Метод (9.35) закончил сходимость на итерации  $k = 68$ . Был найден второй *локальный максимум* целевого функционала  $J(u)$ , который является *краевым максимумом*. На рисунке 9.9 сплошными линиями показан радиус сопла полученной оптимальной площади  $\sigma^{68}$ . Блок «Иначе повторять» в (9.35) срабатывал только в конце итераций при  $k \geq 66$ , что потребовало дополнительно 12 вычислений  $J$ . Значение целевого функционала составило  $\max J = 3.75 \cdot 10^5$ , т.е. в 3 раза больше (лучше) чем в первом локальном максимуме.

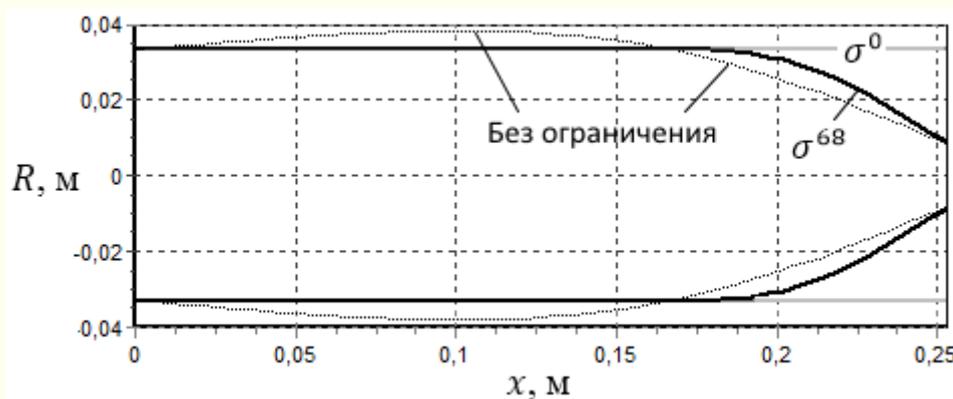


Рис. 9.9 — Оптимальное соплѣ (сплошная линия) во втором экстремуме

Таким образом, второй *максимум* следует считать *глобальным*, а соответствующую форму сопла — оптимальной.

Предложенные градиентные методы (9.34), (9.35) максимизации целевых функционалов (для минимизации — это методы (1.15), (1.16)), с адаптивной релаксацией шагов, позволили решить сложную многоэкстремальную задачу оптимизации формы сопла гидропушки. Были выявлены два локальных максима целевого функционала, характеризующего среднюю силу действия импульса струи на преграду. Применение традиционного градиентного метода (9.33), не позволило найти ни одного локального экстремума. То есть, решить поставленную задачу градиентным методом с постоянным шаговым множителем не удалось, и по всей вероятности — невозможно.

Добавим, что численные методы подхода классического вариационного исчисления [4, 46] в виде методов релаксации (внешне напоминающие градиентные методы с постоянным шаговым множителем) для поиска корня  $u_*$  из уравнения НУО  $\|J'_u(u; \tau)\|_{L_2(S)} = 0$ , не смогли привести к удовлетворительному решению. Более того, этот подход потребовал от исследователей ввести дополнительное ограничение на площадь среза сопла для предотвращения его

схлопывания. Нам этого не потребовалось. Прямой экстремальный подход [92] с анализом управляемости [91] и оригинальными градиентными алгоритмами [85] позволил найти оптимальное сопло, включая его срез.

# Литература

1. Alexandrov N.M. Optimization of Engineering Systems Governed by Differential Equations // SIAG/OPT Views-and-News, Vol. 11, No. 2, 2000. pp. 1–4.
2. Andrei N. An Unconstrained Optimization Test Functions // Advanced Modeling and Optimization, Vol. 10, No. 1, 2008. pp. 147-161.
3. Atanov G.A. Tolstykh V.K. Optimization problems for nonstationary wave processes // Journal of Mathematical Sciences, No. 77, 1995. pp. 3540–3542.
4. Atanov G.A. The optimal control problem of profiling the hydro-cannon nozzle to obtain the maximum outlet speed // Proc. Inst. Mech. Engrs. 1997. Vol. 221. pp. 541-547.
5. Bongartz I., Conn A.R., Gould N.I.M., Toint P. CUTE: constrained and unconstrained testing environments // ACM Transactions on Mathematical Software (TOMS), Vol. 21, No. 1, 1995. pp. 123-160.
6. Dennis J.E., Schnabel R.B. Numerical methods for Unconstrained Optimization and Nonlinear Equations. NJ: SIAM, 1996. 378 pp.
7. Fletcher R. Practical Methods of Optimization. 2nd ed. New York: John Wiley & Sons, 1987. 456 pp.
8. Hope R.H., Petrova S.I. Applications of the Newton interior-point method for Maxwell's equations // Proc. 16-th IMACS World Congress on Sc. Computat., Appl. Math. and Simulat. Lausanne-Switzerland. 2000.
9. Kelley C.T. Iterative Methods for Optimization. Philadelphia: SIAM, 1999. 180 pp.
10. Kubrusly G.S. Distributed parameter system identification. A survey. // Int. J. Control, Vol. 26, No. 4, 1977. pp. 509-535.
11. Lewis R.M. The Adjoint Approach in a Nutshell // SIAG/OPT Views-and-News. 2000. Vol. 11. No. 2. pp. 9-11.
12. Nocedal J., Wright S. Numerical Optimization. New York: Springer-Verlag, 1999. 640 pp.
13. Polak E. Optimization: algorithms and consistent approximations. New York: Springer, 1997. 782 pp.
14. Ronald Glowinski J.L.L.A.J.H. Exact and Approximate Controllability for Distributed Parameter Systems. Cambridge Univ. Press, 2008.
15. Schwefel H.P. Evolution and Optimum Seeking. New York: Wiley Interscience, 1995. 456 pp.
16. Tolstykh V.K., Volodin N.A. Optimal Control By Heat Flow In Continuous Casting Steel // Operations Research Proceedings. Braunschweig. 1996. pp. 480-483.
17. Tolstykh V.K. Collinear gradients method for minimizing smooth functions // SN Operations Research Forum, Vol. 4,20, No. 1, 2023.
18. Tolstykh V.K. New first-order algorithms for optimal control under large and infinite-dimensional objective functions // 16th IMACS World Congress on Scientific Computation, Applied Mathematics and Simulation. Lausanne-Switzerland. 2000.
19. Tolstykh V.K. Optimality Conditions and Algorithms for Direct Optimizing the Partial Differential Equations // Engineering, Vol. 4, No. 7, 2012. pp. 390-393.
20. Tolstykh V.K. Optimization for Systems Governed by Partial Differential Equations // Advanced Modeling and Optimization, Vol. 14, No. 3, 2012. pp. 703-716.
21. Авдонин С.А., Иванов С.А. Управляемость систем с распределёнными параметрами. Киев: УМК ВО, 1989.
22. Алексеев В.М., Тихомиров В.М., Фомин С.В. Оптимальное управление. 3-е-е изд. М.: Физматлит, 2007. 408 с.
23. Андерсен Д., Таннехилл Д., Плетчер Р. Вычислительная гидромеханика и теплообмен. Т. 1. М.: Мир, 1990. 384 с.
24. Беллман Р. Энджел Э. Динамическое программирование и уравнения в частных производных. М.: Мир, 1974.
25. Бертсекас Д. Условная оптимизация и методы множителей Лагранжа. М.: Радио и связь, 1987. 398 с.

26. Бородин В.С., Володин Н.А., Толстых В.К. Идентификация параметров в моделях формирования отливок // Процессы литья, № 1, 1995. С. 96-101.
27. Бородин В.С., Мелихов В.М., Петренко Л.П., Гридин С.В. Математическое моделирование теплопереноса при затвердевании под давлением чугунных отливок. // Процессы литья. 1992. № 3. С. 29-33.
28. Бутковский А.Г. Методы управления системами с распределенными параметрами. М: Наука, 1975. 568 с.
29. Васильев Ф.П. Методы оптимизации. Т. 2. М.: МЦНМО, 2011. 433 с.
30. Васильев Ф.П. Численные методы решения экстремальных задач. М: Наука, 1988. 552 с.
31. Владимиров В.С., Жаринов В.В. Уравнения математической физики. М.: Физматлит, 2004. 400 pp.
32. Владимиров С.В. Обобщенные функции в математической физике. М.: Наука, 1979. 320 с.
33. Володин Н.А., Толстых В.К. Развитие теоретических основ оптимизации и идентификации параметров в слитках и отливках. Донецк: ИПШ "Наука і освіта", 2008. 128 с.
34. Воронин С.Т., Толстых В.К. Вариационный метод определения коэффициента шероховатости открытого русла // Труды Гидрометцентра СССР. 1986. № 283. С. 54-59.
35. Гасников А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска : учебное пособие. М.: МФТИ, 2018. 272 с.
36. Голичев И.И. Модифицированный градиентный метод наискорейшего спуска решения линеаризованной задачи для нестационарных уравнений навье-стокса // Уфимский математический журнал, Т. 5, № 3, 2013. С. 60-67.
37. Дейч А.М. Методы идентификации динамических объектов. М: Энергия, 1979. 240 с.
38. Дж. С.Д. Волны на воде. Математическая теория и приложения. М.: Государственное издательство иностранной литературы, 1959. 620 с.
39. Дилигенская А.Н. Идентификация объектов управления. Самара: Самар. гос. техн. ун-т., 2009. 136 с.
40. Дмитрук Ю.А., Толстых В.К. Условия оптимальности формы сопла гидропушки // Вестник ДонНУ. Серия Г: Технические науки, 2022. С. 54-63.
41. Евтушенко Ю.Г. Методы решения экстремальных задач и их применение в системах оптимизации. М.: Наука, 1982. 482 с.
42. Егоров А.И., Знаменская Л.Н. Введение в теорию управления системами с распределенными параметрами. СПб: Лань, 2017. 288 pp.
43. Егоров А.И. Оптимальное управление тепловыми и диффузионными процессами. М: Наука, 1978. 464 с.
44. Емельянов В.Н. Численные методы: введение в теорию разностных схем. М.: Издательство Юрайт, 2018. 188 с.
45. Зубов В.И. 3.3.Г. Об одном классе решений задачи оптимизации сопла гидропушки // Ж. вычисл. матем. и матем. физ. 1994. Т. 34. № 10. С. 1541-1550.
46. Зуйкова З.Г. Вариационная задача о втекании сжимаемой жидкости в сужающийся канал, Донецк, Дис. на соискание уч. степени канд. физ.-матем. наук. 1984.
47. Ильин А.М. Уравнения математической физики. М.: Физматлит, 2009. 192 с.
48. Канторович Л.В., Акилов Г.П. Функциональный анализ. М.: Наука, 1984. 752 с.
49. Карманов В.Г. Математическое программирование. М.: ФИЗМАТЛИТ, 2004. 264 с.
50. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. М.: Физматлит, 2006. 572 с.
51. Крайко А.Н. Вариационные задачи газовой динамики. М: Наука, 1979. 448 с.
52. Лионс Ж.Л. Оптимальное управление системами, описываемыми уравнениями в частных производных. М: Мир, 1972. 416 с.
53. Лионс Ж.Л. Управление сингулярными распределенными системами. М.: Наука, 1987. 368 с.
54. Лятхер В.М., Яковлев Ю.С. (ред.) Динамика сплошных сред в расчётах гидротехнических сооружений. М.: Энергия, 1976. 391 с.
55. Марчук Г.И. Методы вычислительной математики. М.: Наука, 1989. 536 с.

56. Миеле А. Теория оптимальных аэродинамических форм. М: Мир, 1969. 508 с.
57. Минюк С.А., Искра К.К. Об управляемости и наблюдаемости линейных систем в частных производных первого порядка // Докл. АН БССР, Т. 32, № 4, 1988. С. 300-302.
58. Михайлов В.П., Гушин А.К. Дополнительные главы курса "Уравнения математической физики". М.: МИАН, 2007. 146 с.
59. Моисеев Н.Н. Численные методы в теории оптимальных систем. М.: Наука, 1971. 424 pp.
60. Недопёкин Ф.В. Математическое моделирование гидродинамики и тепломассопереноса в слитках. Ижевск: Издательство Удмуртского университета, 1995. 236 с.
61. Немировский А.С., Юдин Д.Б. Сложность задач и эффективность методов оптимизации. М.: Наука, 1979. 384 с.
62. Нестеров Ю.Е. Методы выпуклой оптимизации. М.: МЦНМО, 2010. 282 с.
63. Огурцов А.П., Гридин С.В., Недопёкин Ф.В. Математическое моделирование теплофизических процессов при затвердевании отливок из высокопрочного чугуна. Донецк: Юго-Восток, 1998. 228 с.
64. Огурцов А.П., Недопёкин Ф.В., Толстых В.К., Володин Н.А. Прямая оптимизация теплофизических процессов. Донецк: Юго-Восток, 1997. 150 с.
65. Петровский И.Г. Лекции об уравнениях с частными производными. М.: Физматлит, 1961. 401 с.
66. Поляк Б.П. Введение в оптимизацию. М.: Наука, 1983. 384 с.
67. Пшеничный Б.Н., Данилин Ю.М. Численные методы в экстремальных задачах. М: Наука, 1975. 320 с.
68. Пясковский Р.В., Померанец К.С. Наводнения. Математическая теория и предсказания. Л.: Гидрометеоздат, 1982.
69. Ракитский Ю.В., Устинов С.М., Черноруцкий И.Г. Численные методы решения жестких систем. М.: Наука, 1979. 208. с.
70. Рождественский Б.Л., Яненко Н.Н. Системы квазилинейных уравнений и их приложения к газовой динамике. М.: Наука, 1979. 688 с.
71. Самарский А.А. Теория разностных схем. М.: Наука, 1989. 616 с.
72. Сеа Ж. Оптимизация. Теория и алгоритмы. М.: Мир, 1973. 244 с.
73. Семко А.Н. Импульсные струи жидкости высокой скорости и их применение. Донецк: ДонНУ, 2014. 370 с.
74. Сиразетдинов Т.К. Оптимизация систем с распределенными параметрами. М: Наука, 1977. 480 с.
75. Системы с распределёнными параметрами // Большая российская энциклопедия. URL: <https://old.bigenc.ru/physics/text/3666483> (дата обращения: 20.03.2024).
76. Смирнов И.В. Курс высшей математики, том 4, часть 2. М.: Наука, 1981. 550 pp.
77. Соболев В.В., Трефилов П.М. Оптимизация тепловых режимов затвердевания расплавов. Красноярск: Красноярский ун-т, 1986. 154 с.
78. Соболев В.В., Трефилов П.М. Теплофизика затвердевания металла при непрерывном литье. М.: Металлургия, 1988. 160 с.
79. Соболев Л.С. Некоторые применения функционального анализа в математической физике. М.: Наука, 1988. 333 с.
80. Темам Р. Уравнения Навье-Стокса. Теория и численный анализ. 2-е изд. М.: Мир, 1981. 408 с.
81. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. М: Наука, 1979. 285 pp.
82. Тихонов А.Н., Гончарский А.В., Степанов В.В., Ягола А.Г. Численные методы решения некорректных задач. М.: Наука, 1990. 232 с.
83. Тихонов А.Н., Самарский А.А. Уравнения математической физики. М.: МГУ Наука, 2004. 798 с.
84. Тихонов А.Н. О решении некорректно поставленных задач и методе регуляризации // ДАН СССР, Т. 151, № 3, 1963. С. 501-504.
85. Толстых В.К. Алгоритмы оптимизации систем с многоэкстремальными функционалами // Журн. вычисл. матем. и матем. физики, Т. 64, № 3, 2024. С. 392-400.
86. Толстых В.К. Градиентный метод оптимального управления распределенными системами // Диф. уравнения, Vol. 27, No. 2, 1991. pp. 303-312.

87. Толстых В.К. Идентифицируемость систем с распределёнными параметрами // Автоматика и телемеханика, № 10, 1989. С. 49-56.
88. Толстых В.К. Метод коллинеарных градиентов. Wikipedia, 2015. Электронные данные.
89. Толстых В.К. О выборе критерия качества идентификации распределённых систем // Автоматика и телемеханика, № 11, 1990. С. 187-189.
90. Толстых В.К. О применении градиентного метода к задачам оптимизации систем с распределёнными параметрами // Журн. вычисл. матем. и матем. физики, Т. 26, № 1, 1986. С. 137-140.
91. Толстых В.К. Об управляемости распределённых систем // Журн. вычисл. матем. и матем. физики, Т. 65, № 6, 2024. С. 959–972.
92. Толстых В.К. Прямой экстремальный подход для оптимизации систем с распределёнными параметрами. Донецк: Юго-Восток, 1997. 178 с.
93. Толстых В.К. Управляемость пространственно-распределённых систем // Математичне та комп'ютерне моделювання. Серія: Технічні науки: зб. наук. пр. 2012. № 6. С. 207-214.
94. Толстых В.К. Эффективный метод оптимизации физических процессов // Инженерно-физический журнал, Т. 76, № 2, 2003. С. 424-427.
95. Трифонов А.Г. Постановка задачи оптимизации и численные методы ее решения // Сообщество Экспонента. 2019. URL: [http://matlab.exponenta.ru/optimiz/book\\_2/index.php](http://matlab.exponenta.ru/optimiz/book_2/index.php) (дата обращения: 10.04.2024).
96. Уизем Д. Линейные и нелинейные волны. М.: Мир, 1977. 624 с.
97. Химмельблау Д. Прикладное нелинейное программирование. М: Мир, 1975. 536 с.
98. Цыпкин Я.З. Информационная теория идентификации. М.: Наука. Физматлит, 1995. 336 с.
99. Шукаев Н.Д. Прикладные методы оптимизации. М.: Издательский дом Академии Естествознания, 2017. 212 с.

# Предметный указатель

<b>А</b>			
Аддитивность		41, 44	
Алгоритм экстремальный	См. Экстремальный		
алгоритм			
Антиградиент		21, 75	
<b>Б</b>			
Базовое направление минимизации		36, 108	
Безусловная оптимизация		22	
Билинейная форма функционала		68, 69	
<b>В</b>			
Вариации			
зависимые		147, 152	
независимые		134, 146, 151	
Вариация			
исходной системы		44, 132	
краевых (граничных и начальных) условий		146	
управления		41, 218	
целевого функционала		44, 131	
вторая		48	
первая	41, 45, 151, 214, 265, 303, 348, 398		
Векторизация		38, 115, 126	
Вектор-столбец		24	
управления		18, 22	
Вектор-строка		24	
градиента		39	
Выпуклая оптимизация		72	
Выпуклость		62	
<b>Г</b>			
Гессиан		34, 48, 62, 78, 80	
конечно-разностный		83	
Градиент		27, 34	
бесконечномерный (функция)		43, 46	
знакопостоянный		116, 119	
квадратичного функционала		69	
квадратичной функции		62	
конечномерный (вектор)		39, 74	
конечно-разностный		78	
целевого функционала	153, 215, 265, 304, 349, 400		
Граничные условия			
1 первого рода	147, 290, 338, 387		
с управлением		149	
со связью компонент состояния		150	
2 второго рода	148, 206, 257, 290, 387		
с управлением		150, 206	
3 третьего рода		257, 290	
4 четвёртого рода		258	
<b>Д</b>			
Диссипация энергии		163, 172, 336	
Дифференциал Гато		41	
Дифференциальное неравенство		188	
Дифференциальные уравнения		14, 20, 128	
гиперболического типа		156, 165, 334, 384	
второго порядка		158	
первого порядка		150, 160	
квазилинейные		156, 334, 384	
линейные		153, 157, 206, 256	
нелинейные		289	
неоднородные		172, 334, 384	
нестационарные		158, 160, 164	
однородные		163, 206, 256, 289	
параболического типа	156, 159, 164, 206, 256, 305		
полулинейные		157	
смешанного типа		160	
стационарные		156, 164	
эллиптического типа		156, 158	
Дифференциальный оператор	См. Оператор		
Дифференцирование произведения		136, 139	
Допустимое множество	См. Множество		
Достаточное условие оптимальности	См. Условие оптимальности достаточное		
<b>З</b>			
Задача			
идентификации		190, 194	
исходная (прямая)	20, 128, 206, 256, 289, 334, 384		
линейного поиска		53, 72, 422	
нелинейной оптимизации		21, 32	
обратная		122, 167, 200	
оптимизации		20	
бесконечномерная		22, 127	
бузусловная		22	
конечномерная		22, 74	
ретроспективная		130	
синтеза оптимального управления	200, 205, 309		
сопряжённая	45, 153, 215, 265, 303, 349, 399		
Замыкание		19	
<b>И</b>			
Идентификация		190	
off-line		192	
on-line		191	
оценка значений параметров		195	
параметрическая		192	
структурная		192	
Идентифицируемость		194, 201	
Измерения	194, 199, 203, 284, 338, 376		
косвенные		193	

Изоэнтропический	384		
Инвариант Римана	162, 360, 411		
<b>К</b>			
Кавитация	385		
Квадратичная форма			
функции	62		
функционала	69		
Квазидномерная модель	335, 384		
Количество вычислений	См. Эффективность		
Коллинеарность			
бесконечномерных градиентов	110		
векторов (понятие)	86		
конечномерных градиентов	88, 93		
Компакт корректности	226, 353, См. Множество компактное		
Координата			
пространственно-временная	18		
Корректность			
классическая по Адамару	122, 168		
условная по Тихонову	123, 168, 201		
Краевые условия	18, 129, 146		
Критерий			
завершения итераций	59		
близость градиента к нулю	59		
близость к оптимуму	59		
прекращение сходимости	60		
качества идентификации	195		
качества оптимизации	19		
<b>М</b>			
Максимизация	19, 21, 388		
Максимум			
глобальный	428		
краевой	427		
локальный	425, 427		
Матрица			
Гессе	См. гессиан		
квадратичной формы	62		
положительно определённая	63, 82		
симметричная	62		
Метод			
Trust-region	55		
Гаусса	82		
градиентный	39		
бесконечномерный	39		
конечномерный	75		
МНС — наискорейшего спуска			
бесконечномерный	39		
конечномерный	76		
казиньютоновский	83		
BFGS	83		
L-BFGS	84		
линейного поиска	53		
дихотомии	53		
золотого сечения	53		
Носедала-Райта	55		
МКГ — коллинеарных градиентов	90		
многошаговый	92		
одношаговый	92		
МНСг — inexact Newton method	85, 94		
МРНС — с регулируемым направлением спуска			
бесконечномерный	38, 113		
конечномерный	96		
МРНСг (относительно градиента)			
бесконечномерный	114		
конечномерный	96		
МРНСг (относительно сопряжённых градиентов) бесконечномерный	120		
МСГ — сопряжённых градиентов			
бесконечномерный	120		
конечномерный	78		
Ньютона	80		
truncated Newton methods	85		
inexact — МНСг	85		
МКГ	90, 96		
демпфированный (Рафсона)	81		
конечно-разностный	83		
Минимизация	19, 21		
Минимум			
глобальный	17, 21, 62		
краевой	71, 178, 236, 237, 320, 324, 330		
локальный	62, 63		
строгий	62, 63		
Множество			
возможной равномерной сходимости $\Delta$	31, 109		
выпуклое	71, 182		
допустимых состояний	177		
допустимых управлений	22, 71, 123, 178, 208		
замкнутое	18, 129		
избыточное $\omega$	174, 355, 403		
компактное	123, 168, 200		
меры нуль	31, 112		
наблюдений состояния объекта	196, 258, 338		
невыпуклое	71, 178		
недостаточное $\omega$	174, 368		
ограниченное	18		
открытое	18		
пространственно-временное	19		
пространственное	129		
Модель			
объекта идентификации	190, 194		
объекта, процесса	13		
распределённой системы	18, 20		
функции (функционала)	33, 34, 55		
Моменты обновления	79, 120		
<b>Н</b>			
Наблюдаемость	194, 196		
Направление			
базовое	См. Базовое направление		
кратчайшего пути к оптимуму	89, 110		
Направление минимизации	33		
бесконечномерное	36, 37		
конечномерное	37, 74		
Направление на оптимум			
бесконечномерное	110, 320		
конечномерное	87, 88		
Невязка ортов градиентов	93		
Невязка состояний	195, 258, 338		
Невязка управлений	228		
Нелинейное программирование	21		

Необходимое условие оптимальности	См. Условие оптимальности необходимое		
Норма		24	
<b>О</b>			
Область	См. Множество		
Область определения			
состояния системы $\Omega$		19, 129	
управления $S$		22, 130	
функции цели $I - \omega$		19, 127	
целевого функционала $J$		22, 127	
целевой функции $J$		22	
Объект, процесс		13, 190, 194	
Ограничение		71	
активное		71, 178, 210	
на состояние системы		176	
штрафы		179	
на управление		22, 43, 177	
в виде диапазона		182, 208	
изопериметрическое		183, 208	
кусочно-постоянное		186, 209	
по скорости изменения		187, 209	
Ограничения			
типа 1		179, 181, 208, 292	
типа 2		182, 183, 186, 187, 208, 209, 386	
Однородность		41, 44	
Оператор		20	
вырожденный		172, 271	
дифференциальный		20, 129, 141, 156	
в квадратичной форме		157	
линейный		44, 48, 69, 132, 135	
неоднородный		145, 173, 398	
наблюдений		196, 259, 338	
невыврожденный		172	
положительный, симметричный		69, 109	
проецирования		130, 182, 188, 264, 348, 397	
сопряжённый		45, 135, 139, 141, 143, 163	
сужения		127, 168, 179, 224, 291, 294	
Операторы			
$U^*$ 215, 303			
$V^*, V^*$		135	
$U^*$ 45, 144, 264, 347			
$U\emptyset^*$		145, 173, 398	
$V^*$ 45, 139, 141, 143			
$D$ 20, 129			
$U$ 44, 133			
$V$ 44, 132, 141			
Определение 1.1. Равномерная сходимость		29	
Определение 1.2. Локальная квадратичность		70	
Определение 2.1. (конечномерное)			
Коллинеарность, пропорциональность		88	
Определение 3.1. (бесконечномерное)			
Коллинеарность, пропорциональность		110	
Определение 3.2. (бесконечномерное)			
Направление на оптимум		110	
Определение 3.3. Управляемость		168	
Определение 3.4. Удовлетворительные параметры регуляризации		169	
Определение 3.5. Область определения и значений оператора $U^*$		171	
Определение 4.1. Идентифицируемость		201	
Оптимальное управление, оптималь, оптимум		21	
Оптимизация	См. Задача оптимизации		
Оракул	См. Эффективность		
Отображение		44	
в исходных пространствах		44, 132, 133	
в сопряжённых пространствах		45, 143, 145	
на границе		149, 150	
обратное		122, 200	
прямое		122	
Ошибка		195, 254, 282	
<b>П</b>			
Параметр			
в условии Вольфе		55	
регулирования направления спуска			
бесконечномерный		37, 114, 116, 126	
конечномерный		98, 99	
регуляризации		124, 126	
удовлетворительный		169	
шага (шаговый множитель)		33	
Первая вариация	См. Вариация		
Переворот краевых вариаций состояния		149, 153	
Погрешность наблюдений		198	
Полупространство			
бесконечномерное		114	
конечномерное		98	
Помехи	27, 50, 57, 59, 79, 81, 172, 174, 198, 406		
абсолютные детерминированные		58, 250, 254	
абсолютные случайные		58, 248, 254, 371	
относительные детерминированные		58, 252, 254, 282	
относительные случайные		58	
Порядок			
малости, $o$ -малое		41, 48	
метода минимизации			
второй		81	
нулевой		78	
первый		76, 79, 83	
Почти всюду на $S$ (п.в. $S$ )	См. Сходимость		
Произведение			
прямое		129	
скалярное		24, 26	
в цилиндрических координатах		295	
Производная		138	
вторая		34, 48, 80, 93, 111	
квадратичного функционала		69	
квадратичной функции		62	
Гато		41	
обобщённая		25	
первая		34, 38	
по направлению		55	
пространственно-временная		135	
смешанная		141	
Фреше		39, 46, 176	
функции цели		44	
частная		39, 131	
штрафной функции		181, 210, 296	
Пропорциональность	См. Определение		
Пространство			
банахово		24	
гильбертово		25	
евклидово		22, 24	
линейное		24	

самосопряжённое	46, 75, 134
Соболева	25
сопряжённое	44
Прямой экстремальный подход	20, 74, 108, 195

## Р

Распределённая система	См. Система
Релаксация	49
Робастный	51, 254

## С

Система	
бесконечномерная	14
распределённая	13
сосредоточенная	14
Система (оси) координат	23
Системы с распределёнными параметрами	15
Скорость сходимости	
квадратичная	81, 119
линейная	76
сверхлинейная	79
Собственные	
вектора	161, 340, 349, 389, 409
числа	161, 339, 388
Состояние	
объекта наблюдений	196, 339
системы, модели	
вектор-функция	26, 337, 386
исходное	18, 43, 128
сопряжённое	45, 153
Стратегии выбора длины шага	
вторая	49, 236, 239, 242, 246, 326, 369, 424
первая	49, 425
пятая (Trust-region)	55
третья	53, 120
четвёртая	54, 75, 120, 227, 244, 280, 318, 368
Схема конечно-разностная	
Кранка-Николсона	221
неявная	221
явная	274, 311, 359, 411
Сходимость	
очень сильная	113
покомпонентная	35, 61
поточечная	29, 59
почти всюду (п.в.S)	31
пропорциональная	
в бесконечномерном пространстве	111
в конечномерном пространстве	86
равномерная	29, 31, 37, 38, 108, 111, 112, 235, 283, 319
сильная	113, 119
слабая	113
среднеквадратичная	29, 59, 108

## Т

Теорема 2.1 (конечномерное покомпонентное условие оптимальности)	88
Следствие 2.1	89
Следствие 2.2	93

Теорема 3.1 (Очень сильное бесконечномерное условие оптимальности)	110
Следствие 3.1	111
Теорема 3.2 (Сильное необходимое условие)	112
Теорема 3.3 (Слабое необходимое условие)	112
Теорема 3.4 (Управляемость)	175
Теорема 4.1 (Идентифицируемость)	201
Терминальное условие	153
Течение	
в диффузоре	163, 336
в конфузоре	163, 336
дозвуковое	163, 336, 385
сверхзвуковое	336, 369
Тип ограничений	См. Ограничения типа
Транспонирование	24, 136

## У

Управление	15
вектор (конечномерное)	22, 74, 186
вектор-функция (бесконечномерное)	26, 114, 130
оптимальное	21, 62
Управляемость	47, 155, 166, 168, 175
Уравнения в частных производных	См.
Дифференциальные уравнения	
Условие	
Вольфе	54
граничное	См. Граничные условия
идентифицируемости	273, 355, 357, 379
сопряжённости	50, 78
управляемости	175, 220, 309, 405
Условие оптимальности	
достаточное	42
бесконечномерное очень сильное	117
для квадратичной формы	
бесконечномерное	110
конечномерное	88, 93, 99
необходимое	42, 86, 90
бесконечномерное	
очень сильное	110, 117
сильное	112, 117
слабое	112
конечномерное	88, 93, 98
Условия 1-4 рода	См. Граничные условия

## Ф

Формула	
Грина	136
парметра регулирования в МРНС	98, 116
на первом шаге	100, 118
практического завершения сходимости	60
шага коллинеарных градиентов	92
Функционал	18
билинейный	68
вогнутый	51
выпуклый	69, 278, 317, 366
квадратичный	68, 109, 226, 315
линейный в $L_2$	41, 44, 131, 134
локально квадратичный	70, 113, 119, 279, 317, 366
невогнутый	421
невыпуклый	69, 421

- неквадратичный 278, 317, 366  
 овражный 69  
 сглаживающий 125  
 стабилизирующий 124  
 строго выпуклый 69  
 целевой 17, 19, 127, 206, 210, 260, 291, 293, 294, 339, 387  
   зависит от управления  
     неявно 20, 43, 127  
     явно (пример) 42  
     явно, но частично 388  
   линейный 41, 68  
 штрафной 124, 179, 209, 293  
 Функциональная производная См. Производная  
 Фреше  
 Функция  
   о-малое См. Порядок малости  
    $\delta$  (Дирака) 197, 259, 276, 297, 314  
    $\Theta$  (Хевисайда) 257, 276, 291, 313  
    $\Pi$  (прямоугольная) 224, 294, 321  
   барьерная 180  
   вогнутая 51, 63  
   выпуклая 63, 64  
   жёсткая См. – овражная  
   квадратичная 62  
   локально квадратичная 70, 92, 98, 101  
   неопределённой выпуклости 63  
   нестрого выпуклая 63  
   овражная (жёсткая) 65, 77, 80  
   передаточная 191  
   седловая 63  
   сильно выпуклая 65  
   строго вогнутая 63  
   строго выпуклая 63  
   тестовая  
     HilbertA 102  
     QF1 102  
     Staircase2, Staircase2sin 103  
     Tolstykh 103  
     Tridia 103  
     квадратичная двумерная 65  
     Розенброка 66  
   униmodalная 53  
   целевая J конечномерная 18, 22, 74, 187, 244  
   цели I на  $\omega$  19, 127  
   штрафная 179, 209, 294
- Х**
- Характеристики 157, 339, 389
- Характеристическая форма гиперболической системы уравнений 162, 341, 390  
 Характеристические направления 158  
 Характеристическое уравнение второго порядка 158
- Ц**
- Целевая функция См. Функция  
 Целевой функционал См. Функционал
- Ч**
- Чувствительность J к состоянию 44, 47, 154  
 Чувствительность J к управлению 37, 39, 65, 171, 194, 199, 422
- Ш**
- Шаблон, шаблонное приближение  
   бесконечномерное 116  
     под  $45^\circ$  118, 233, 281, 318, 369  
     пропорциональное 118, 233, 318, 369  
     эвристическое 118  
   конечномерное 98, 100  
     под  $45^\circ$  94, 100  
 Шаг См. Стратегии выбора длины шага  
 Шаговый множитель 33  
   в методе коллинеарных градиентов 92  
   в методе Носедала-Райта 54  
   постоянный 49  
   с адаптивной релаксацией 49  
   с полной релаксацией (оптимальный) 53
- Э**
- Эквивалентность 35, 85  
 Экстремальная задача 21  
 Экстремальные методы 17  
 Экстремальный алгоритм  
   бесконечномерный 33, 36  
   МРНС 38  
   конечномерный 74  
   формально обобщённый на бесконечномерное пространство 36  
 Экстремум 17, 34, 52, 422  
 Эффективность метода оптимизации 27, 107, 283
- Я**
- Ядро оператора 146, 172

## Обозначения разделов 1–4

### Основные

$\stackrel{\text{def}}{=}$  — равно по определению, равносильно;

$\equiv$  — тождественно равно, равно во всех случаях, при всех исходных данных;

$\gtrsim$  — больше и приблизительно равно;

$\forall$  — квантор общности, т.е. «для всех», «для любых»;

$\exists$  — квантор существования;

$\Rightarrow$  — знак логического следования;

$\Leftrightarrow$  — символ эквивалентности, т.е. выражения с обеих сторон от этого символа «по существу одинаковы»;

$\cdot$  — умножение там, где без него может оказаться двусмысленное понимание выражения. Например,  $b \cdot (u - d)$ , где  $b$  не есть функция от разности  $u - d$ . Смотрите также обозначения для операторов;

$(a, b)$  — интервал на числовой оси в границах  $a$  и  $b$ , не включая эти границы;

$[a, b]$  — отрезок на числовой оси в границах  $a$  и  $b$ , включая эти границы;

$(a_1, \dots, a_n)$  — вектор (вектор-столбец)  $a$  с перечисленными  $n$  компонентами;

$(a_1, \dots, a_n)^T$  — вектор-строка  $a$  с перечисленными  $n$  компонентами;

$b^k$  — шаговый множитель, параметр шага на итерации  $k$ ;

$d^k \stackrel{\text{def}}{=} d(u^k)$  — направление спуска к минимуму  $J(u)$  из точки  $u^k$  в пространстве управлений;

$\tilde{d}^0 = \tilde{u}^0 - u^0$  — направление шаблонного шага из точки  $u^0$  на точку  $\tilde{u}^0$ ;

$F$  — свободный член в дифференциальных уравнениях модели распределённой системы, например,  $\mathbb{D}v + F = 0$ ;

$I(v)$  или  $I(v, u)$  — сложная функция цели  $I(\tau)$ , определённая на  $\omega$  в целевом функционале  $J = \int_{\omega} I d\omega$ , явно зависящая от внутренней функции  $v(\tau)$  или от двух внутренних функций  $v(\tau), u(\tau)$ , где пространственно-временная переменная  $\tau \in \omega$ ;

$I_F(v, v_{\min}, v_{\max})$  — штрафная функция, явно зависящая от  $v(\tau)$  и от параметров-ограничений  $v_{\min}, v_{\max}$ . Функция  $I_F(\tau)$  — это сложная функция, определённая на  $\omega_F$ ;

$I'_u$  — частная производная от функции цели по явно заданной внутренней функции  $u$ , если управление  $u$  присутствует в  $I(v, u)$ ;

$I'_v, I'_{Fv}$  — частные производные от функций  $I, I_F$  по явно заданной внутренней функции  $v$ ;

$J(u)$  — целевой функционал или функция (если аргумент  $u$  — конечномерный вектор) с областью определения  $U$ ;

$J(u) = \frac{1}{2} \langle Au, u \rangle_{E^n}$  — квадратичная функция, где  $A$  — квадратная матрица размерности  $n \times n$ ;

$J(u) = \frac{1}{2} \langle \mathcal{A}u, u \rangle_{L_2(S)}$  — квадратичный функционал, где  $\mathcal{A}$  — линейный оператор, действующий из  $L_2(S)$  в  $L_2(S)$ ;

$J'_u$  — производная целевого функционала или функции  $J(u)$  по управлению. Если управление является функцией, то  $J'_u$  — производная Фреше;

$\nabla J(u) \stackrel{\text{def}}{=} \{\nabla_i J(u)\}_{i=1}^n = \left\{ \frac{\partial J(u)}{\partial u_1}, \dots, \frac{\partial J(u)}{\partial u_n} \right\}$  — вектор градиента целевой функции  $J(u)$ ,  $u \in E^n$ . Если функция квадратичная и матрица  $A = A^T$ , то  $\nabla J = Au$ ;

$\nabla J(u) \stackrel{\text{def}}{=} \nabla J(u; \tau)$  — функция градиента целевого функционала  $J(u)$ ,  $u(\tau) \in L_2(S)$ . Если функционал квадратичный и линейный оператор  $\mathcal{A}$  симметричный, то  $\nabla J = \mathcal{A}u$ ;

$\delta J = \langle \nabla J(u), \delta u \rangle$  — первая вариация (дифференциал Гато, главная линейная часть приращения) целевого функционала  $J(u)$ ;

$\delta^2 J = \langle \nabla J(u) \delta u, \delta u \rangle$  — вторая вариация целевого функционала  $J(u)$ ;

$\nabla^2 J$  — вторая производная целевого функционала или функции. Если  $J(u)$  функция, то  $\nabla^2 J \stackrel{\text{def}}{=} H$  — матрица Гессе, гессиан, для

квадратичной функции  $\nabla^2 J = A = H$ . Если  $J(u)$  квадратичный функционал, то  $\nabla^2 J = \mathcal{A}$ ;

$x, t$  — пространственная и временная независимые переменные;

$u = (u_1, \dots, u_n) \in U \subseteq E^n$  — вектор управления;

$u = (u_1(\tau), \dots, u_n(\tau)) \in U(S) \subseteq L_2^n(S)$  — вектор-функция управления с  $n$  компонентами в виде функций из  $L_2(S)$ . В частности, если  $n = 1$ , то управление  $u(\tau) \in U(S) \subseteq L_2(S)$ ;

$u_*$  — оптимальное управление;

$\tilde{u}^0$  — шаблон, шаблонное приближение, обеспечивающее необходимое изменение управления  $u$  и градиента  $\nabla J(u)$  после первой итерации в направлении  $\tilde{d}^0$ ;

$\delta u$  — вариация (небольшое изменение, отклонение) управления  $u$ ;

$v(\tau) \in V(\bar{\Omega}) \subseteq L_2^m(\bar{\Omega})$  — вектор-функция состояния распределённой системы из  $m$  уравнений;

$v_e(\tau) \in V_e(\omega) = L_2^{m_e}(\omega)$  — экспериментально наблюдаемое состояние распределённого объекта,  $m_e \leq m$ ;

$m$  — размерность вектора  $v$  состояния распределённой системы;

$m_e \leq m$  — количество наблюдаемых компонент вектора состояния  $v$ ;

$n$  — размерность вектора управления  $u$ ;

$\hat{n}$  — количество множеств  $\hat{S}_i$ , определения кусочно-постоянных сегментов функции управления;

$N$  — размерность пространственной области  $X \ni x$  функционирования системы,  $1 \leq N \leq 3$ ;

$p^k$  — базовое направление минимизации на итерации  $k$ , которое используются традиционными методами. Это может быть, например, направление антиградиента  $p^k = -\nabla J^k$ , направление сопряжённых градиентов  $p^k = -\nabla J^k + \beta^k p^{k-1}$ ;

$r$  — погрешности.

## Индексы

$e$  — нижний индекс  $e$  означает принадлежность к экспериментальным данным;

$i$  — нижний индекс  $i = 1 \dots n$  указывает компоненты вектора управления и пробегает значения от 1 до  $n$ . Этим индексом также обозначаются узлы и переменные на пространственной конечно-разностной сетке;

$j$  — нижний индекс  $j = 1 \dots m$  указывает компоненты вектора состояния и пробегает значения от 1 до  $m$ ;

$j$  — верхний индекс  $j = 1 \dots m$ , как правило, обозначает узлы и переменные на временной конечно-разностной сетке;

$k$  — верхний индекс  $k$  означает принадлежность к итерации  $k$  и может пробегать значения от 0 до бесконечности;

$l \equiv k_l$  — индекс  $l$  означает принадлежность к подитерации  $l$  на итерации  $k$  и может пробегать значения от 0 до бесконечности;

$r$  — верхний индекс  $r$  означает порядок дифференцирования;

$T$  — верхний индекс  $T$  означает транспонирование или вектор-строку. Например, если  $y$  — вектор и  $Z$  — матрица, то  $y^T Z = (Z^T y)^T$ ;

$*$  — верхний индекс  $*$  означает принадлежность сопряжённому пространству;

$*$  — нижний индекс  $*$  означает оптимальное значение;

$\emptyset$  — нижний индекс  $\emptyset$  у оператора означает отсутствие у него ядра (отсутствие ненулевого ядра);

— — индекс черта над множеством означает замыкание множества, например,  $\bar{\Omega} = \Omega \cup \partial\Omega$ ;

## Пространства и множества

$\{ , \}$  — множество из элементов, которые перечисляются через запятую;

$\{a_i\}_{i=1}^n$  — множество элементов  $a_i$ , пронумерованных от 1 до  $n$ ;

$\{a|B\}$  — множество элементов  $a$ , удовлетворяющих условию  $B$ ;

$C^1(S)$  — пространство гладких функций (функции с непрерывной первой производной), определённых на множестве  $S$ ;

$C_+^1(S)$  — ограниченное положительное полупространство функций в  $C^1(S)$ . Например, для функций  $\alpha(\tau)$  полупространство  $C_+^1(S) = \{\alpha \in C^1(S) \mid 0 < \alpha(\tau) < \infty \ \forall \tau \in S\}$ ;

$C^{1,n}(S)$  — пространство вектор-функций с  $n$  компонентами в виде функций из  $C^1(S)$ .

$E^n$  — евклидово пространство размерности  $n$  со скалярным произведением векторов  $\langle a, b \rangle_{E^n} = a^T b = \sum_{i=1}^n a_i b_i$  и нормой  $\|a\|_{E^n} = \sqrt{\langle a, a \rangle_{E^n}}$ ;

$E^{n*} = E^n$ , т.е. они самосопряжённые;

$E_+^n$  — ограниченное положительное полупространство в  $E^n$ . Например, для вектора  $\alpha$  полупространство  $E_+^n = \{\alpha \in E^n \mid 0 < \alpha_i < \infty \ \forall i = 1 \dots n\}$ ;

$L_2(S)$  — евклидово пространство функций, определённых на множестве  $S$  со скалярным произведением функций  $\langle a, b \rangle_{L_2(S)} = \int_S a(\tau) b(\tau) dS$  и нормой  $\|a\|_{L_2} = \sqrt{\langle a, a \rangle_{L_2(S)}}$ ;

$L_2^* = L_2$ , т.е. они самосопряжённые;

$L_2^n(S)$  — пространство вектор-функций с  $n$  компонентами в виде функций из  $L_2(S)$ . Скалярное произведение вектор-функций  $\langle a, b \rangle_{L_2^n(S)} = \int_S \sum_{i=1}^n a_i(\tau) b_i(\tau) dS$ ;

$L_{2+}(S)$  — ограниченное положительное полупространство функций в  $L_2(S)$ . Например, для функций  $u(\tau)$  полупространство  $L_{2+}(S) = \{u(\tau) \in L_2(S) \mid 0 < u(\tau) < \infty \ \forall \tau \in S\}$ ;

$S$  — область определения управления;

п.в. $S$  — почти всюду на  $S$  за исключением множеств в  $S$  нулевой меры;

$S_0$  — множество нулевой меры, на котором отсутствует поточечная сходимость функций  $u^k(\tau) \xrightarrow{k \rightarrow \infty} u_*(\tau)$  или имеется неопределённость (разрывность) функций  $u^k(\tau)$ ;

$\Delta S \subset S$  — множество ненулевой меры, на котором отсутствует равномерная сходимость функций управления  $u^k(\tau) \xrightarrow{k \rightarrow \infty} u_*(\tau)$ .  
Множество  $\Delta S$  включает в себя множество  $S_0$ ;

$S_\Delta = S \setminus \Delta S$  — множество возможной равномерной сходимости функций управления  $u^k(\tau) \xrightarrow{k \rightarrow \infty} u_*(\tau)$ , т.е.  $S_\Delta$  — это множество  $S$  с исключённым из него множеством ненулевой меры  $\Delta S$ ;

$\hat{S}_i, \hat{\Delta}t_i$  — множества, интервалы определения кусочно-постоянных значений функции управления;

$U$  — пространство управлений, или допустимое множество управлений в случае, если  $U$  ограниченное;

$\partial U$  — граница допустимого множества  $U$ ;

$U^*$  — сопряжённое пространстве управлений, где существует градиент  $\nabla J$ ;

$\text{int } U$  — внутренность (без границы  $\partial U$ ) допустимого множества  $U$ ;

$\mathcal{U} \subseteq U$  — компактное множество (компакт корректности) управлений, где существует единственное и устойчивое решение обратной задачи оптимизации. Ограниченное, замкнутое множество с не слишком сильным колебанием функций управления в нём, функции равностепенно непрерывны;

$V$  — пространство состояний распределённой системы;

$V_e$  — пространство наблюдений состояния системы;

$X$  — ограниченная пространственная область в виде открытого (без границы  $\partial X$ ) множества;

$\partial X, \partial \Omega$  — границы множества  $X$  и  $\Omega$  соответственно;

$\bar{\Omega}$  — замкнутое (замыкание) множества  $\Omega$ , например,  $\bar{\Omega} = \Omega \cup \partial \Omega$ ;

$\emptyset$  — пустое множество, т.е. не содержащее ни одного элемента.

### Операции с множествами

$a \in Y$  — элемент  $a$  принадлежит множеству или пространству  $Y$ ,

$Y \ni a$  — то же самое;

- $Y \subset Z$  — множество  $Y$  включено (содержится) в множестве  $Z$ , т.е. каждый элемент из  $Y$  также является элементом и из  $Z$ , но  $Y \neq Z$ ;
- $Y \supset Z$  — множество  $Z$  включено (содержится) в множестве  $Y$ , т.е. каждый элемент из  $Z$  также является элементом и из  $Y$ , но  $Y \neq Z$ ;
- $Y \subseteq Z$  — множество  $Y$  включено (содержится) в множестве  $Z$  и возможно совпадает с  $Z$ , т.е. каждый элемент из  $Y$  также является элементом и из  $Z$ , возможно даже  $Y = Z$ ;
- $Y \cup Z$  — объединение двух множеств  $Y$  и  $Z$ . Результат объединения — это множество, содержащее все элементы из  $Y$  и  $Z$ ;
- $Y \cap Z$  — пересечение двух множеств  $Y$  и  $Z$ . Результат пересечения — это множество одинаковых элементов, принадлежащих и  $Y$ , и  $Z$ ;
- $Y \setminus Z$  — разность двух множеств  $Y$  и  $Z$ . Результат разности — это множество элементов, принадлежащих  $Y$ , но не принадлежащих  $Z$ ;
- $Y \times Z$  — прямое (декартово) произведение двух множеств  $Y$  и  $Z$ . Результатом произведения являются множество из упорядоченной пары  $(y, z)$  для всех  $y \in Y$  и  $z \in Z$ ;
- $Y \rightarrow Z$  — отображение множества (пространства)  $Y$  в множество (пространство)  $Z$ ;
- $\text{mes } Y = \int_Y dZ$  — мера множества  $Y$  в множестве  $Z$ .

## Операторы

- — место расположения элемента, выражения на которое действует впереди стоящий оператор. Например,  $A \cdot$  может иметь вид  $Af$  или  $Av$ , т.е.  $A$  действует на  $f$  или на  $v$ ;
- Символ  $\cdot$  может означать и умножение там, где без него может оказаться двусмысленное понимание выражения. Например,  $b \cdot (u - d)$ , где  $b$  не есть функция от разности  $u - d$ ;
- $\leftarrow$  — стрелка-оператор присваивания. Например,  $a \leftarrow b$ , где переменной  $a$  присваивается значение переменной  $b$ ;
- $\rightarrow$  — стрелка-оператор отображения. Например,  $Y \rightarrow Z$ , где множество  $Y$  отображается в множество  $Z$ ;
- $f : X \rightarrow Y$  — означает, что оператор (функция)  $f$  отображает элементы пространства (множества)  $X$  в элементы пространства (множества)  $Y$ ;

$\overrightarrow{ab} = \{a_i b_i\}_{i=1}^n$  — векторизация для произведения векторов  $a$  и  $b$ . Если  $a$  и  $b$  векторы из  $E^n$ , то и результат векторизации — вектор из  $E^n$ . Если  $a$  и  $b$  вектор-функции из  $L_2^n(S)$ , то и результат векторизации — вектор-функция из  $L_2^n(S)$ ;

$\langle a, b \rangle$  — скалярное произведение элементов  $a$  и  $b$ ;

$arg \min \cdot$  — оператор поиска аргумента, который минимизирует заданную функцию  $\cdot$  или функционал  $\cdot$ , например,  $u_* = arg \min J(u)$ ;

$A$  — линейный симметричный оператор, действующий из  $E^n$  в  $E^n$ , который может быть представлен квадратной матрицей  $A = A^T$  размерности  $n \times n$ . Например, для квадратичной функции  $J(u) = \frac{1}{2} \langle Au, u \rangle_{E^n}$ , действие оператора  $Au = \nabla J \in E^n$ ;

$\mathcal{A} \cdot$  — линейный положительный, симметричный оператор, действующий из  $L_2(S)$  в  $L_2(S)$ . Например, для квадратичного функционала  $J(u) = \frac{1}{2} \langle \mathcal{A}u, u \rangle_{L_2(S)}$ , действие оператора  $\mathcal{A}u = \nabla J \in L_2(S)$ ;

$\mathbb{C}_r \cdot = \sum_{i=1}^m \frac{\partial A}{\partial v_i} \cdot \Big|_i \frac{\partial^r v}{\partial x^r}$  — линейный недифференциальный матричный оператор для вариации  $\delta v$ , действующий из  $V$  в  $V$ . Нижний индекс  $r$  означает, что оператор содержит производные от состояния системы порядка  $r$ , а именно,  $\frac{\partial^r v}{\partial x^r}$ ;

$\mathbb{C}$  — аналог  $\mathbb{C}_r$  для смешанной производной от состояния системы,  $\mathbb{C} \cdot = \sum_{i=1}^m \frac{\partial A}{\partial v_i} \cdot \Big|_i \frac{\partial^2 v}{\partial x_1 \partial x_2}$ .

$\mathbb{D} \cdot$  — линейный дифференциальный оператор для состояния  $v$  распределённой системы, действующий из  $V$  в  $V$ . Например,  $\mathbb{D} \cdot = A \frac{\partial^r \cdot}{\partial x^r}$ , где  $r$  — порядок дифференцирования;

$\mathbb{V} \cdot$  — линейный оператор для вариации  $\delta v$ , действующий из  $V$  в  $V$ , представляет собой частную производную от уравнений исходной распределённой системы по состоянию  $v$ . Например,  $\mathbb{V} \delta v$ ;

$\mathbb{U} \cdot$  — линейный оператор для вариации  $\delta u$ , действующий из  $U$  в  $V$ , представляет собой частную производную от уравнений исходной распределённой системы по управлению  $u$ . Например,  $\mathbb{U} \delta u$ ;

$\mathbb{V}^* \cdot$  — сопряжённый линейный оператор при вариации  $\delta v$ , действующий из  $V^*$  в  $V^*$ . Например,  $\langle \mathbb{V}^* f, \delta v \rangle_{V^*(\Omega)}$ ;

$\mathbb{U}^* \cdot$ ,  $\bar{\mathbb{U}}^* \cdot$ ,  $\bar{\bar{\mathbb{U}}}^* \cdot$  — сопряжённые линейные операторы при вариации управления  $\delta u$ , заданного в дифференциальных уравнениях системы ( $\mathbb{U}^*$ ), в граничном условии первого ( $\bar{\mathbb{U}}^*$ ) или второго рода ( $\bar{\bar{\mathbb{U}}}^*$ ). Операторы действуют из  $V^*$  в  $U^*$ . Например,  $\langle \bar{\bar{\mathbb{U}}}^* f, \delta u \rangle_{U^*(S)}$ ;

$\mathbb{U}_{\emptyset}^* \cdot = \mathbb{U}^* \cdot + I'_u|_{\omega}$  — сопряжённый линейный неоднородный оператор, полученный добавлением к линейному  $\mathbb{U}^*$  слагаемого  $I'_u|_{\omega}$ . Нижний индекс  $\emptyset$  означает отсутствие нулевого ядра у такого оператора. Если  $\mathbb{U}^*$  невырожденный, то и  $\mathbb{U}_{\emptyset}^*$  невырожденный;

$\text{Ker } \mathbb{U}^* \equiv \{f\}_{\text{Ker}}$  — ядро сопряжённого оператора  $\mathbb{U}^*$ . Это множество функций  $\{f(\tau)\}_{\text{Ker}}$ ,  $\tau \in S$  на которых значение оператора  $\mathbb{U}^* f$  обращается в ноль на  $S$ , или хотя бы на некотором его подмножестве ненулевой меры. Если ядро имеется при  $u \neq u_*$ , то такой оператор вырожденный;

$M(v)$  — оператор наблюдений (измерений) состояния  $v$  системы. Действует из  $V(\Omega)$  в  $V_e(\omega)$ ;

$P_U(\cdot)$  — оператор проецирования величины  $\cdot$  на множество  $U$ ;

$\cdot|_{\omega} = \begin{cases} \cdot & \text{если } \tau \in \omega, \\ 0 & \text{если } \tau \notin \omega \end{cases}$  — оператор сужения действия функции  $\cdot$  на множество  $\omega$  из некоторого большого множества  $\Omega \supset \omega$ ;

$\frac{\partial \cdot}{\partial \tau} = \sum_{k=1}^{N+1} \frac{\partial \cdot}{\partial \tau_k}$  — оператор дифференцирования по вектору  $\tau$  пространственно-временных переменных;

$\frac{\partial^r \cdot}{\partial x^r}$  — оператор дифференцирования порядка  $r$ ;

$\delta \cdot$  — оператор варьирования.

### Греческие символы

$\alpha$  — параметр регулирования направления спуска для МРНС;

$\Gamma = \Gamma_{\partial} \cup \Gamma_0$  — область определения краевых (граничных и начальных) условий;

$\Gamma_{\partial} = \partial X \times (t_0, t_1)$  — область определения граничных условий. В одномерном случае  $\Gamma_{\partial} = \Gamma_a \cup \Gamma_b$ ;

$\Gamma_{\partial 1}$  — область определения граничных условий первого рода;

$\Gamma_{\partial 2}$  — область определения граничных условий второго рода;

$\Gamma_0 = \bar{X} \times t_0$  — область определения начальных условий;

$\Gamma_1 = \bar{X} \times t_1$  — область определения терминальных условий;

$\Gamma^* \subseteq \partial\Omega$  — множество определения «перевёрнутых» краевых вариаций  $\delta v, \frac{\partial \delta v}{\partial \tau}$ . Например, для одномерной нестационарной задачи множество  $\Gamma^*$  может состоять из областей  $\Gamma_a^* \equiv \Gamma_b, \Gamma_b^* \equiv \Gamma_a, \Gamma_0^* \equiv \Gamma_1$ ;

$\delta$  — символ варьирования,

символ дельта-функции Дирака, смотрите далее специальные функции,

длина шаблонного шага ( $\delta$ -окрестность) в пространстве управлений относительно точки  $u^k$ . Например, в шаблоне «под 45°» шаг  $\tilde{u}^0(\tau) = u^0(\tau) \pm \delta$ ;

$\varepsilon_j$  — множество малых положительных чисел;

$\eta$  — коэффициент усиления сходимости МНсг;

$\vartheta$  — скорость изменения функции управления;

$\kappa$  — весовой коэффициент;

$\lambda$  — коэффициент пропорциональности задания шаблонного шага  $\tilde{u}^0 = \lambda u^0(\tau)$ ,

$\lambda = \frac{dx}{dt}$  — уравнение характеристических направлений. Например, для одномерной нестационарной гиперболической системы с двумя характеристиками  $\lambda_{1,2} = \frac{d\xi_{1,2}}{dt}$ ;

$\xi$  — характеристики системы. Например, для одномерной нестационарной гиперболической системы характеристики  $\xi_1(x, t), \xi_2(x, t)$  — это траектории распространения возмущений состояния системы в пространственно-временной области  $\bar{\Omega}$ .

$\tau$  — пространственно-временная независимая переменная,  $\tau \stackrel{\text{def}}{=} (x_1, \dots, x_N, t)$ ;

$\varphi$  — шаблонный шаг для  $\tilde{u}^0(\tau) = \varphi(\tau)$ , задаваемый исследователем;

$\omega \subseteq \Omega$  — пространственно-временная область, на которой задаётся целевой функционал  $J = \int_{\omega} I d\omega$ ;

$\omega_{min} \subset \omega$  — минимально достаточная (не избыточная) область задания целевого функционала для управляемости (идентифицируемости);

$\omega_F \subset \Omega$  — пространственно-временная область определения штрафной функции  $I_F$ ;

$\Omega = X \times (t_0, t_1)$  — ограниченная пространственно-временная область функционирования системы, её замыкание — это  $\bar{\Omega}$ ;

$\mathbf{\Omega} \subset \bar{\Omega}$  — множество определения сопряжённых функций  $f(\tau) \in V^*(\mathbf{\Omega})$ , на которых определён оператор  $\mathbb{U}^*$  (или  $\bar{\mathbb{U}}^*$ ,  $\bar{\bar{\mathbb{U}}}^*$ ) такой, что имеет место отображение  $V^*(\mathbf{\Omega}) \rightarrow U^*(S)$  для вычисления градиента на основе производной Фреше, когда  $J'_u = \nabla J = \mathbb{U}^* f \in U^*(S)$  (или  $\nabla J = \bar{\mathbb{U}}^* f$ ,  $\nabla J = \bar{\bar{\mathbb{U}}}^* f$ ).

### Специальные функции

$\sigma(x)$  —  $\sigma$ -малое от  $x$ , пренебрежимо малая величина по сравнению с  $x$ , такая, что  $\lim_{x \rightarrow 0} \frac{\sigma(x)}{x} = 0$ ,  $\sigma(0) = 0$ ;

$\text{sgn } y = \begin{cases} 1, & y > 0 \\ 0, & y = 0 \\ -1, & y < 0 \end{cases}$  — функция знак числа  $y$ ;

$\delta(x) = \begin{cases} +\infty, & x = 0, \\ 0, & x \neq 0 \end{cases}$  — дельта-функция Дирака, обладающая

свойствам нормировки  $\int_{-\infty}^{+\infty} \delta(x) dx = 1$  и фильтрующим свойством  $\int_{-\infty}^{+\infty} f(x) \delta(x - x_0) dx = f(x_0)$ ;

$\Theta(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 0 \end{cases}$  — тета-функция Хевисайда, является первообразной функцией для  $\delta$ -функции Дирака, т.е.  $\Theta' = \delta$ . Иногда неопределённость  $\Theta$ -функции в точке  $x = 0$  устраняют как  $\Theta(0) = 0$ , или  $\Theta(0) = \frac{1}{2}$ , или  $\Theta(0) = 1$ ;

$\Pi(\tau; \omega_F) = \begin{cases} 1, & \tau \in \omega_F, \\ 0, & \tau \notin \omega_F \end{cases}$  — многомерная прямоугольная функция.

## Сокращения

МКГ — метод коллинеарных градиентов;

МНС — метод наискорейшего спуска;

МН — метод Ньютона;

МНкр — конечно-разностный метод Ньютона;

МНсг — приближенный метод Ньютона (inexact Newton method) с сопряжёнными градиентами;

МРНС — методы с регулируемым направлением спуска;

МРНСг — метод с регулируемым направлением спуска относительно градиента;

МРНСсг — метод с регулируемым направлением спуска относительно сопряжённых градиентов;

МСГ — метод сопряжённых градиентов;

НУО — необходимое условие оптимальности;

п.в. $S$  — почти всюду на  $S$ , т.е. всюду на  $S$  кроме конечного числа множеств в  $S$  нулевой меры;

BFGS — метод Бroyдена-Флетчера-Гольдфарба-Шанно;

L-BFGS — метод BFGS с ограниченной памятью (Limited-memory BFGS);