Планы лекций по Методам Оптимизации,
(3 курс, 2 поток, 5 и 6 семестр, 2021/22 уч. год)
лектор профессор Потапов М.М., каф. ОУ

1. Общая информация о курсе

Курс годовой, лекции – 1 раз в неделю, в осеннем семестре упражнений нет, весной упражнения – 1 раз в неделю в академических группах.

Форма отчетности – экзамен в июне.
Итоговая экзаменационная оценка формируется с учётом персональных результатов студента:

- оценки $\boldsymbol{S}$ за **работу на семинарах** (упражнениях) в весеннем семестре,
- оценки $\boldsymbol{Y}$ за **устный экзамен** по программе курса (проводится в форме экспресс-опроса по программе курса без проверки знания доказательств; проверяется знание основных определений, постановок задач, формулировок теорем и другого программного материала),
- оценки $\boldsymbol{P}$ за **письменные работы повышенной сложности**, проводимые лектором (наличие этой оценки обязательно только для тех студентов, которые претендуют на итоговую оценку **отлично**).

С подробностями «правил игры» можно ознакомиться на сайте кафедры оптимального управления по ссылке
`http://oc.cmc.msu.ru/study/optim_methods_pot.html`

Там же размещены (и будут размещаться) и другие учебные и информационные материалы:

- программа курса в форме списка вопросов экзаменационных билетов со списком рекомендованной литературы,
- материалы лекций,
- образцы письменных заданий различного уровня сложности, критерии их оценок и др.

Сейчас на сайте размещены, в основном, прошлогодние материалы, но их базовая часть («правила игры» и программа) заметно не изменится, а за актуальностью остальных материалов я буду стараться следить.

2. Цели курса и связь с другими дисциплинами

С задачами оптимизации вы встречались и не раз: в школе, на первом курсе (МА), на втором курсе (ДУ) вы изучали основы вариационного исчисления, сейчас вам читается курс ОУ, в этом году начинается чтение курса «Методы машинного обучения», так что класс задач вам более-менее знаком. Целями данного курса является формирование у вас целостного взгляда на оптимизационные задачи и адекватного представления об основной проблематике, связанной как с корректной постановкой таких задач, так и с методами их теоретического исследования и практического решения.

3. О структуре курса

Разделы:

I. Теоремы существования (Вейерштрасса)

II. Элементы дифференциального исчисления в нормированных пространствах

III. Задачи управления линейными динамическими системами с квадратичными критериями качества

IV. Элементы выпуклого анализа

V. Итерационные методы минимизации (самый большой по объёму)

VI. Методы «снятия» ограничений (метод штрафов, правило множителей Лагранжа)

VII. Принцип максимума Понтрягина

VIII. Метод регуляризации Тихонова

Некоторые студенты (в основном, из Казахстанского филиала МГУ) в соответствии с их учебным планом должны освоить полугодовую программу курса, в которую войдут разделы I – IV и бóльшая часть раздела V. Экзамен они сдают в январе, а упражнения для них не предусмотрены.

4. Постановка задачи. Базовые обозначения

$$J(u) \to \inf, \quad u \in U \subset M. \tag{1}$$

Здесь M – некоторое пространство, U – заданное (допустимое) подмножество, $J : \ M \to R^1$ – заданная функция с числовыми значениями. Задачи *максимизации* отдельно не рассматриваются, так как они сводятся к задачам *минимизации* заменой исходной функции на функцию со значениями противоположного знака.

Что надо найти? *Нижнюю грань* функции

$$J_* = \inf_{u \in U} J(u),$$

множество *всех* оптимальных решений

$$U_* = \operatorname*{Arg\,min}_{u \in U} J(u) = \left\{ v \in U \;\middle|\; J(v) = J_* \right\}$$

или *один* из оптимальных элементов

$$u_* = \arg\min_{u \in U} J(u) \in U_*.$$

I. ТЕОРЕМЫ СУЩЕСТВОВАНИЯ

Напомним формулировку классической конечномерной теоремы Вейерштрасса.

Теорема Вейерштрасса (классическая). *Пусть в задаче* (1) *пространство* M *конечномерно:* $M = R^n$, *допустимое множество* $U \subset R^n$ *замкнуто и ограничено, а функция* $J(u)$ *непрерывна на множестве* U. *Тогда нижняя грань конечна:* $J_* > -\infty$ *и достигается, т. е.* $U_* \neq \varnothing$.

Заметим, что при тех же самых условиях в задаче (1) достигается не только *минимум*, но и *максимум*. Заметим также, что существенно ослабить условия этой теоремы нельзя. На эту тему предлагается

Упражнение 1. *Приведите примеры непрерывных функций, не достигающих своих нижних граней на ограниченном, но незамкнутом множестве; замкнутом, но неограниченном множестве.*

Перейдём к рассмотрению некоторых обобщений классической теоремы Вейерштрасса на случай пространств бесконечной размерности. Начнём с метрических пространств.

Определение 1. *Пусть* M *— некоторое множество. Функция* $\rho : M \times M \to R^1$, *определённая на декартовом произведении* $M \times M$, *называется* ***метрикой*** *или* ***расстоянием***, *если она обладает свойствами*

- $\rho(u, v) = \rho(v, u) \quad \forall\, u, v \in M$ *(симметрия),*
- $\rho(u, v) \leqslant \rho(u, w) + \rho(w, v) \quad \forall\, u, v, w \in M$ *(нер-во треугольника),*
- $\rho(u, v) \geqslant 0 \quad \forall\, u, v \in M, \quad \rho(u, v) = 0 \iff u = v.$

Множество M, *наделённое метрикой* ρ, *называется* ***метрическим пространством***.

Определение 2. *Последовательность элементов (точек)* $u_n \in M$, $n = 1, 2, \ldots$, *метрического пространства* M *называется* ***сходящейся*** *(*ρ*-сходящейся) к элементу* $u_0 \in M$, *если сходятся к нулю расстояния между* u_n *и* u_0 :

$$\lim_{n\to\infty} \rho(u_n, u_0) = 0.$$

Последовательность u_n *называется* ***фундаментальной***, *если*

$$\lim_{m,\, n\to\infty} \rho(u_n, u_m) = 0.$$

Замечание 1. *В конечномерном пространстве $M = R^n$ свойства сходимости и фундаментальности эквивалентны (по критерию Коши). Если $\dim M = \infty$, то из сходимости следует фундаментальность, а фундаментальная последовательность не обязательно сходится к некоторому элементу из M.*

Определение 3. *Метрическое пространство M, в котором любая фундаментальная последовательность сходится к некоторому элементу из M, называется* ***полным.***

Понятно, что n-мерное пространство R^n с евклидовой метрикой

$$\rho(x,y) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}, \quad x = (x_1, x_2, \ldots, x_n),\ y = (y_1, y_2, \ldots, y_n),$$

является полным.

Упражнение 2. *Докажите, что пространство $C[a,b]$ непрерывных на отрезке $[a,b]$ функций* ***является полным*** *относительно метрики (равномерной сходимости)*

$$\rho(f,g) = \max_{a \leqslant t \leqslant b} |f(t) - g(t)|$$

и ***не является полным*** *относительно интегральной метрики (сходимости в среднем)*

$$\rho(f,g) = \sqrt{\int_a^b |f(t) - g(t)|^2\, dt}\,.$$

Определение 4. *Пусть M — метрическое пространство, $J(u) : M \to R^1$ — некоторая определённая на этом пространстве функция и $u_0 \in M$ — некоторая фиксированная точка. Пусть u_n —* ***произвольная*** *последовательность, ρ-сходящаяся к точке u_0 : $\rho(u_n, u_0) \to 0$ при $n \to \infty$. Тогда в зависимости от поведения значений $J(u_n)$ функция $J(u)$ называется*

- *(секвенциально)* ***непрерывной*** *в точке u_0, если существует*

$$\lim_{n\to\infty} J(u_n) = J(u_0),$$

- *(секвенциально)* ***полунепрерывной снизу*** *(п/н снизу) в точке u_0, если*

$$\varliminf_{n\to\infty} J(u_n) \geqslant J(u_0),$$

- *(секвенциально)* ***полунепрерывной сверху*** *(п/н сверху) в точке u_0, если*

$$\varlimsup_{n\to\infty} J(u_n) \leqslant J(u_0).$$

Определение 5. *Подмножество $U \subset M$ метрического пространства M называется*

- ***замкнутым**, если из условий*

$$u_n \in U,\ n = 1, 2, \ldots, \quad и \quad \lim_{n\to\infty} \rho(u_n, u_0) = 0$$

следует, что $u_0 \in U$,

- ***ограниченным**, если существуют $u_0 \in M$ и $R > 0$, такие, что*

$$\rho(u, u_0) \leqslant R \quad \forall\, u \in U,$$

- *(секвенциально) **компактным**, если из **любой** последовательности элементов $u_n \in U$ можно выделить подпоследовательность $\{u_{n_m}\} \subset \{u_n\}$, ρ-сходящуюся к некоторому элементу $u_0 \in U$: $\rho(u_{n_m}, u_0) \to 0$ при $m \to \infty$.*

Замечание 2. *В конечномерном пространстве R^n множество U компактно тогда и только тогда когда оно замкнуто и ограничено. В бесконечномерном метрическом пространстве из компактности следует замкнутость и ограниченность, а обратного следствия, вообще говоря, нет.*

Упражнение 3. *Докажите, что единичный шар*

$$U = \left\{ f(t) \in C[a, b] \;\middle|\; \max_{t\in[a,b]} |f(t)| \leqslant 1 \right\}$$

в пространстве $C[a, b]$ является замкнутым ограниченным, но некомпактным множеством.

Чтобы разгрузить формулировку главного результата, дадим ещё одно

Определение 6. *Последовательность точек $u_n \in M$ называется **минимизирующей** для оптимизационной задачи (1), если*

$$u_n \in U \ \ \forall\, n = 1, 2, \ldots, \quad и \quad \lim_{n\to\infty} J(u_n) = J_*.$$

Теорема 1. (метрический вариант теоремы Вейерштрасса) *Пусть M — метрическое пространство, U — компактное множество из M, а функция $J(u)$ н/н снизу на множестве U. Тогда в задаче (1) $J_* > -\infty$, $U_* \neq \varnothing$ и любая минимизирующая последовательность является ρ-сходящейся ко множеству оптимальных решений U_*.*

Доказательство. Минимизирующие последовательности в задаче (1) существуют по определению $\inf$. Пусть u_n — любая из них. По условию множество U компактно, поэтому из u_n можно выделить подпоследовательность $u_{n_m} \subset u_n$, ρ-сходящуюся к некоторому элементу $u_0 \in U$. Поскольку функция $J(u)$ п/н снизу в точке u_0, то справедливы соотношения

$$\lim_{n\to\infty} J(u_n) = J_* \leqslant J(u_0) \leqslant \underline{\lim_{m\to\infty}}\, J(u_{n_m}) = \lim_{n\to\infty} J(u_n) = J_*,$$

следовательно, $J(u_0) = J_*$ и, тем самым, установлено, что $J_* > -\infty$ и $U_* \neq \varnothing$.

Под *ρ-сходимостью ко множеству U_** мы понимаем то, что *все* ρ-предельные точки *всех* минимизирующих последовательностей принадлежат множеству U_*. Для доказательства этого достаточно повторить приведённые выше рассуждения. Можно также показать, что такое понимание ρ-сходимости ко множеству равносильно её интерпретации как сходимости вида

$$\lim_{n\to\infty} \inf_{u_*\in U_*} \rho(u_n, u_*) = 0.$$

Теорема 1 доказана. ▼

Замечание 3. *В случае, когда в задаче минимизации* (1) *выполняются все утверждения теоремы 1, эту задачу принято называть* ***корректно*** *(ρ-корректно)* ***поставленной*** *в метрическом пространстве M. Наиболее стеснительным требованием в теореме 1 считают требование компактности множества U.*

Теоремы Вейерштрасса (продолжение)

Как правило, мы будем иметь дело с метрическими пространствами, над элементами которых определены линейные операции, а метрика ρ порождается нормой.

Определение 7. *Пусть L — некоторое линейное пространство. Функция $\|\cdot\| : L \to R^1$, определённая на L, называется* ***нормой****, если она обладает свойствами*

- $\|\lambda\, x\| = |\lambda|\, \|x\| \quad \forall\, \lambda \in R^1\ \forall\, x \in L$ *(положительная однородность),*
- $\|x + y\| \leqslant \|x\| + \|y\| \quad \forall\, x, y \in L$ *(нер-во треугольника),*
- $\|x\| \geqslant 0 \quad \forall\, x \in L, \quad \|x\| = 0 \iff x = 0.$

Линейное пространство L, наделённое нормой $\|\cdot\|$, называется ***нормированным пространством****. Нормированное пространство, полное относительно метрики $\rho(x, y) = \|x - y\|$, называется* ***бáнаховым****.*

Приведём пример, показывающий, что в банаховом пространстве непрерывная функция может не достигать свой нижней грани на замкнутом и ограниченном множестве.

Пример 1. *Задача минимизации:*

$$J(u) = \int_{-1}^{0} u(t)\, dt - \int_{0}^{1} u(t)\, dt \to \inf, \quad u \in U \subset C[-1, 1], \quad U = \{\|u\|_C \leqslant 1\}.$$

Здесь $C[-1, 1]$ — банахово пространство непрерывных на $[-1, 1]$ функций (см. упражнение 2), $J(u)$ — непрерывная (и к тому же линейная) на $C[-1, 1]$ функция, U — единичный шар в $C[-1, 1]$, являющийся замкнутым, ограниченным, но некомпактным в $C[-1, 1]$ множеством (см. упражнение 3), нижняя грань $J_* = -2$ и $U_* = \varnothing$. Непрерывность (на самом деле, даже липшиц-непрерывность) функции $J(u)$ следует из оценки

$$|J(u) - J(v)| = |J(u - v)| \leqslant \int_{-1}^{1} |u(t) - v(t)|\, dt \leqslant 2\, \|u - v\| \quad \forall\, u, v \in C[-1, 1],$$

а наименьшее значение функции $J_* = -2$ достигается на разрывной функции $u_*(t) = \operatorname{sign} t \notin C[-1, 1]$. Заменой функции $J(u)$ на

$$j(u) = \frac{1}{J_* - J(u)} = \frac{1}{-2 - J(u)}$$

получим пример, в котором $j_* = -\infty$ и $U_* = \varnothing$.

Следующий пример показывает, что требование компактности не является необходимым.

Пример 2. *Задача минимизации:*

$$J(u) = \int_{-1}^{1} u(t)\,dt \to \inf, \quad u \in U \subset C[-1,1], \quad U = \{\|u\|_C \leqslant 1\}.$$

Здесь также $J_* = -2$, *но теперь* $u_*(t) \equiv -1 \in U_* \neq \varnothing$.

Мы, в основном, будем иметь дело с линейными пространствами, наделёнными не только нормой, но и скалярными произведениями.

Определение 8. *Пусть* L *— некоторое линейное пространство. Функция* $\langle\cdot,\cdot\rangle : L \times L \to R^1$, *определённая на декартовом произведении* $L \times L$, *называется* ***скалярным произведением****, если она обладает свойствами*

1) $\langle x, y\rangle = \langle y, x\rangle \quad \forall\, x, y \in L$ *(симметрия),*

2) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle \quad \forall\, \alpha \in R^1\ \forall\, x, y \in L$ *(однород. по первой переменной),*

3) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle \quad \forall\, x, y, z \in L$ *(аддит. по первой переменной),*

4) $\langle x, x\rangle \geqslant 0 \quad \forall\, x \in L, \quad \langle x, x\rangle = 0 \iff x = 0.$

Линейное пространство L, *наделённое скалярным произведением* $\langle\cdot,\cdot\rangle$, *называется* ***евклидовым****. Евклидово пространство, полное относительно метрики* $\rho(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y\rangle}$, *называется* ***гильбертовым****.*

Замечание 4. *Свойства 2) и 3) означают, что скалярное произведение линейно по первой переменной. С учётом 1) оно линейно и по второй переменной, т. е. является симметричной билинейной функцией (формой), обладающей дополнительным свойством 4).*

В данном определении гильбертова пространства не упоминается его размерность. Это значит, что все конечномерные линейные пространства мы будем относить к категории гильбертовых, поскольку в них всегда можно ввести скалярное произведение, например, по хорошо известному нам правилу

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i\, y_i\,, \tag{2}$$

где x_i, y_i — координаты векторов x и y в некотором базисе. Типичным примером бесконечномерного евклидова пространства, не являющегося гильбертовым (т. е. полным), может служить пространство непрерывных функций $C[a, b]$ со скалярным произведением (см. упражнение 2)

$$\langle f, g\rangle = \int_a^b f(t)\, g(t)\,dt\,. \tag{3}$$

Одним из важных для данного курса эталонных примеров бесконечномерного гильбертова пространства является пространство Лебега $L^2(a,b)$, которое может быть получено пополнением пространства $C[a,b]$ пределами последовательностей непрерывных на $[a,b]$ функций, фундаментальных относительно интегральной нормы

$$\|f\| = \sqrt{\int_a^b f^2(t)\,dt},$$

порождённой скалярным произведением (3), в котором интеграл понимается в смысле Лебега. В курсе ФА обычно дается другое эквивалентное определение пространства $L^2(a,b)$. Так или иначе, это пространство состоит из функций $f(t)$, измеримых по Лебегу на (a,b) и интегрируемых по Лебегу на (a,b) вместе со своими квадратами $f^2(t)$.

Другим эталонным примером бесконечномерного гильбертова пространства является пространство l^2, состоящее из числовых последовательностей $x = (x_1, x_2, \dots, x_i, \dots)$, таких, что

$$\sum_{i=1}^{\infty} x_i^2 < +\infty.$$

Скалярное произведение в пространстве l^2 вводится по правилу

$$\langle x, y\rangle = \sum_{i=1}^{\infty} x_i\, y_i\,,$$

подобному конечномерному скалярному произведению (2).

Замечание 5. *Любое евклидово пространство является нормированным:*

$$\|u\| = \sqrt{\langle u, u\rangle}\,,$$

а любое нормированние пространство является метрическим:

$$\rho(u,v) = \|u - v\|\,.$$

Обсудим кратко свойство компактности множества. Разумеется, в любом (бесконечномерном) нормированном пространстве X компактными будут любые замкнутые ограниченные множества, принадлежащие конечномерным подпространствам этого пространства X. В упражнении 3 в *конкретном* бесконечномерном банаховом пространстве $X = C[a,b]$ указано замкнутое и ограниченное множество, которое не является компатктным. В следующем примере речь идёт о *произвольном* бесконечномерном гильбертовом пространстве.

Пример 3. *В любом бесконечномерном евклидовом пространстве замкнутый единичный шар $U = \{\|u\| \leqslant 1\}$ не является компактным множеством. В качестве последовательности $u_n \in U$, из которой нельзя выделить ни одной сходящейся подпоследовательности, можно взять произвольную ортонормированную систему (ОНС) элементов*

$$e_n \in U: \ \ \langle e_n, e_m\rangle = 0 \ \forall\, n \neq m, \ \ \langle e_n, e_n\rangle = \|e_n\|^2 = 1 \ \ \forall\, n = 1, 2, \dots$$

Дело в том, что никакая подпоследовательность элементов такой ОНС не может сходиться из-за отсутствия у неё свойства фундаментальности:

$$\|e_n - e_m\|^2 = \langle e_n, e_n\rangle - 2\langle e_n, e_m\rangle + \langle e_m, e_m\rangle \overset{n\neq m}{=} 2 \not\to 0 \quad \text{при} \quad n, m \to \infty.$$

В следующих упражнениях представлены примеры *бесконечномерных* компактных подмножеств.

Упражнение 4. *Докажите, что в банаховом пространстве $C[a, b]$ множество U функций, равномерно ограниченных на $[a, b]$ и удовлетворяющих на $[a, b]$ условию Липшица с одной и той же константой, является компактным:*

$$U = \left\{ f(t) \in C[a,b] \,\middle|\, \|f\|_{C[a,b]} \leqslant R, \ |f(t) - f(s)| \leqslant L|t - s| \ \forall t, s \in [a, b] \right\}.$$

Упражнение 5. *Докажите, что так называемый «гильбертов кирпич»*

$$U = \left\{ x = (x_1, x_2, \dots, x_n, \dots) \in l^2 \,\middle|\, |x_n| \leqslant 2^{-n}, \ n = 1, 2, \dots \right\}$$

является компактным множеством в гильбертовом пространстве l^2.

В формулировке следующей обобщённой теоремы Вейерштрасса мы ослабим требование ко множеству и усилим требование к функции. Для этого нам понадобится понятие *слабой* сходимости.

Определение 9. *Пусть H — гильбертово пространство. Последовательность u_n элементов из H называется* ***слабо сходящейся*** *к точке $u_0 \in H$, если для любого фиксированного $h \in H$*

$$\langle u_n, h\rangle \to \langle u_0, h\rangle \quad \text{при} \quad n \to \infty\,.$$

Обратим внимание на то, что в конечномерных пространствах разницы между сильной и слабой сходимостью нет. В бесконеномерных пространствах из сильной сходимости следует слабая сходимость:

$$|\langle u_n, h\rangle - \langle u_0, h\rangle| = |\langle u_n - u_0, h\rangle| \overset{\text{К-Б}}{\leqslant} \|u_n - u_0\|\,\|h\| \to 0 \quad \text{при} \quad n \to \infty\,.$$

Примером последовательности, сходящейся слабо, но не сильно, служит любая бесконечная ОНС, состоящая из попарно ортогональных элементов e_n единичной длины. Покажем, что она *слабо* в H сходится к нулю. Фиксируем произвольный элемент $h \in H$ и запишем для него неравенство Бесселя (МА, 2 курс):

$$\sum_{n=1}^{\infty} \langle h, e_n \rangle^2 \leqslant \|h\|^2 .$$

Из сходимости числового ряда следует, что его общий член стремится к нулю, значит,

$$\langle h, e_n \rangle \underset{n\to\infty}{\longrightarrow} 0 = \langle h, 0 \rangle .$$

При этом сильная сходимость ОНС к нулю, разумеется, отсутствует:

$$\|e_n - 0\| = \|e_n\| = 1 \not\to 0 \quad \text{при} \quad n \to \infty .$$

Определение 10. *Пусть H — гильбертово пространство, $J(u) : H \to R^1$ – некоторая определённая на этом пространстве функция и $u_0 \in H$ — некоторая фиксированная точка. Пусть u_n — **произвольная** последовательность, слабо в H сходящаяся к точке u_0. Тогда в зависимости от поведения значений $J(u_n)$ функция $J(u)$ называется*

- ***слабо непрерывной** в точке u_0, если существует*

$$\lim_{n\to\infty} J(u_n) = J(u_0),$$

- ***слабо полунепрерывной снизу** (сл п/н сн) в точке u_0, если*

$$\varliminf_{n\to\infty} J(u_n) \geqslant J(u_0),$$

- ***слабо полунепрерывной сверху** (сл п/н св) в точке u_0, если*

$$\varlimsup_{n\to\infty} J(u_n) \leqslant J(u_0).$$

Определение 11. *Подмножество $U \subset H$ гильбертова пространства H называется*

- ***слабо замкнутым**, если из условий*

$$u_n \in U,\ n = 1, 2, \ldots, \quad \text{и} \quad u_n \overset{\text{слабо}}{\longrightarrow} u_0 \quad \text{при} \quad n \to \infty,$$

  следует, что $u_0 \in U$,

- ***слабо компактным**, если из **любой** последовательности элементов $u_n \in U$ можно выделить подпоследовательность $\{u_{n_m}\} \subset \{u_n\}$, слабо в H сходящуюся при $m \to \infty$ к некоторому элементу $u_0 \in U$.*

Заметим, что в конечномерных пространствах нет разницы между слабой замкнутостью и замкнутостью, между слабой компактностью и компактностью. В бесконечномерных пространствах из компактности следует слабая компактность, но не наоборот; из слабой закнутости следует замкнутость, но не наоборот; из слабой непрерывности следует непрерывность, но не наоборот. Соответствующими *контрпримерами* в *бесконечномерном* гильбертовом пространстве могут служить:

- единичный шар $U = \{\|u\| \leqslant 1\}$, который компактен слабо, но не сильно,
- единичная сфера $U = \{\|u\| = 1\}$, которая замкнута сильно, но не слабо,
- функция $J(u) = \|u\|$, которая непрерывна в точке $u_0 = 0$ сильно, но не слабо.

Сформулируем так называемый «слабый» вариант теоремы Вейерштрасса. Её предлагается доказать самостоятельно в качестве упражнения по схеме, аналогичной «метрической» теореме 1.

Теорема 2. (`слабый вариант теоремы Вейерштрасса`) *Пусть H – гильбертово пространство, U – слабо компактное множество из H, а функция $J(u)$ слабо п/н снизу на множестве U. Тогда в задаче* (1) *$J_* > -\infty$, $U_* \neq \varnothing$ и произвольная слабая предельная точка любой минимизирующей последовательности принадлежит множеству оптимальных решений U_*.*

Замечание 6. *В случае, когда задача минимизации* (1) *обладает всеми перечисленными в утверждения теоремы 2 свойствами, её называют **слабо корректно поставленной** в гильбертовом пространстве H.*

Приведем без доказательства достаточные условия слабой компактности множества и слабой п/н снизу функции в гильбертовом пространстве. На экзамене требуется знание этих условий.

Утверждение. *Пусть H — гильбертово пространство. Тогда*

- *любое выпуклое замкнутое ограниченное множество $U \subset H$ является слабо компактным,*
- *любая функция $J(u)$, выпуклая и п/н снизу на выпуклом множестве $U \subset H$, является слабо п/н снизу на U.*

Приведём определения выпуклых множеств и функций.

Определение 12. *Пусть L — линейное пространство.*

- *Множество $U \subset L$ называется* ***выпуклым****, если*

$$\forall\, u,\, v \in U \quad \forall\, \alpha \in [0,1] \quad \alpha u + (1-\alpha)v \in U.$$

- *Пусть $U \subset L$ — выпуклое множество. Функция $J(u) : U \to R^1$ называется* ***выпуклой*** *на U, если*

$$\forall\, u,\, v \in U \quad \forall\, \alpha \in [0,1] \quad J\big(\alpha u + (1-\alpha)v\big) \leqslant \alpha J(u) + (1-\alpha)J(v).$$

Участвующие в этих определениях линейные комбинации вида $\alpha u + (1-\alpha)v$ со значниями $\alpha \in [0,1]$ называют *выпуклыми линейными комбинациями.*

Теоремы Вейерштрасса (продолжение)

Приведём несколько типичных примеров *слабо п/н снизу* функций и *слабо компактных* множеств в гильбертовых пространствах $H,\ \dim H \leqslant \infty$.

1. *Линейный* функционал $J(u) = \langle c, u\rangle$, в котором $c \in H$ — некоторый фиксированный элемент из H, является *слабо непрерывным* на всём пространстве. Это простое следствие из определения слабой сходимости последовательности.

2. *Квадратичный* функционал $J(u) = \|Au - f\|_F^2$, в котором фиксированы *линейный ограниченный (непрерывный)* оператор $A : H \to F$, действующий в гильбертовых пространствах H и F, и элемент $f \in F$. Напомним, что линейный оператор $A : H \to F$ называется ***ограниченным (непрерывным)***, если его норма конечна:

$$\|A\| = \sup_{u\in H,\, u\neq 0} \frac{\|Au\|_F}{\|u\|_H} < +\infty\,. \tag{4}$$

В дальнейшем нам часто придётся иметь дело с линейными ограниченными операторами, поэтому зарезервируем для этого класса специальное обозначение: $L(H \to F)$. Убедимся в том, что квадратичная функция *непрерывна*. Фиксируем произвольную точку $u_0 \in H$ и рассмотрим любую сходящуюся к этой точке последовательность $u_n,\ \|u_n - u_0\|_H \to 0$. Сходимость значений $J(u_n) \to J(u_0)$ будет следовать из непрерывности оператора A и из неравенства

$$\big|\|x\| - \|y\|\big| \leqslant \|x - y\|, \tag{5}$$

являющегося следствием неравенства треугольника:

$$\Big|\|Au_n - f\|_F - \|Au_0 - f\|_F\Big| \overset{(5)}{\leqslant} \|Au_n - Au_0\|_F \overset{A\text{ лин.}}{=}$$

$$= \|A\,(u_n - u_0)\|_F \overset{A\text{ огр.}}{\leqslant} \|A\|\,\|u_n - u_0\|_H \underset{n\to\infty}{\longrightarrow} 0.$$

Теперь покажем, что квадратичная функция вида $J(u) = \|Au - f\|_F^2$ *выпукла*:

$$J(\alpha u + (1-\alpha)v) = \|A(\alpha u + (1-\alpha)v) - f\|_F^2 \overset{A\text{ лин.}}{=}$$

$$= \|\alpha(Au - f) + (1-\alpha)(Av - f)\|_F^2 \overset{\text{нер-во треуг.}}{\leqslant}$$

$$\leqslant \Big(\alpha\|Au - f\|_F + (1-\alpha)\|Av - f\|_F\Big)^2 \leqslant \big[\text{функция } y = x^2 \text{ выпукла}\big] \leqslant$$

$$\leqslant \alpha\|Au - f\|_F^2 + (1-\alpha)\|Av - f\|_F^2 = \alpha J(u) + (1-\alpha)J(v).$$

Из непрерывности и выпуклости следует *слабая п/н снизу*. Выпуклость функции одной переменной $y = x^2$ рекомендуется проверить самостоятельно по определению 12. Примером функции, слабо п/н снизу, но *не являющейся* при этом *слабо непрерывной*, может служить $J(u) = \|u\|_H^2$ в *бесконечномерном* гильбертовом пространстве H. Рекомендуется также привести пример линейного *неограниченного* оператора (в *бесконечномерном* пространстве).

3. *Невырожденный элипсоид* в гильбертовом пространстве H описывается условиями

$$U = \Big\{ u \in H \;\Big|\; \|Au - f\|_F \leqslant R \Big\},$$

где F — гильбертово пространство, данные $f \in F$ и $R > 0$ фиксированы, $A \in L(H \to F)$ — линейный ограниченный оператор, причём такой, что существует обратный к нему оператор $A^{-1} \in L(F \to H)$, который также *ограничен* (в этом проявляется *невырожденность*). Свойства *выпуклости* и *замкнутости* такого эллипсоида предлагается проверить самостоятельно. Для доказательства его ограниченности укажем шар с центром в $u_0 = 0$, содежащий множество U. Для произвольной точки $u \in U$ имеем:

$$\|u\|_H = \|A^{-1}Au\|_H = \|A^{-1}(Au - f) + A^{-1}f\|_H \overset{\text{нер-во треуг.}}{\leqslant}$$

$$\leqslant \|A^{-1}(Au - f)\|_H + \|A^{-1}f\|_H \overset{A^{-1}\text{ огр.}}{\leqslant}$$

$$\leqslant \|A^{-1}\|\,\|Au - f\|_F + \|A^{-1}f\|_H \overset{\text{опред. мн-ва } U}{\leqslant}$$

$$\leqslant \|A^{-1}\|\,R + \|A^{-1}f\|_H = R_0.$$

Значение R_0 из правой части неравенства и есть радиус искомого шара, накрывающего множество U. Таким образом, установлена *слабая компактность* рассматриваемого невырожденного эллипсоида, а наличие ограниченного обратного оператора A^{-1} позволило доказать ограниченность U. Простейшим примером невырожденного эллипсоида является шар $U = \{\|u\|_H \leqslant R\}$. Здесь мы выяснили, что он *слабо компактен*, а остутствие у шара свойства компактности было установлено ранее.

Обратим внимание на то, что в случае, когда обратный к A оператор A^{-1} не существует или существует, но неограничен, эллипсоид U может утратить свойство ограниченности, а вместе с ним и свойство слабой компактности. Примером супервырождения может служить эллипсоид с данными $A = 0,\ f = 0$, когда $U = H$. Пример, в котором обратный оператор A^{-1} существует, но неограничен, возможен только в бесконечномерном пространстве.

Рассмотрим конкретное пространство $H = L^2(0,\pi)$ и интегральный оператор $A : L^2(0,\pi) \to L^2(0,\pi)$, действующий по правилу

$$u(\cdot) \in L^2(0,\pi) \longmapsto (Au)(t) = \int_0^t u(s)\,ds, \quad t \in (0,\pi).$$

У этого оператора существует обратный, но он *определён не на всём пространстве* $L^2(0,\pi)$ и *не является ограниченным*, поскольку можно привести пример последовательности $u_n(t) = n\cos nt,\ n = 1,2,\ldots$, для которой

$$\|u_n\|_{L^2(0,\pi)} \to +\infty, \quad Au_n = \sin nt, \quad \|Au_n\|_{L^2(0,\pi)} = \sqrt{\frac{\pi}{2}}, \quad n = 1,2,\ldots$$

Эллипсоид U, заданный в пространстве $L^2(0,\pi)$ с помощью такого оператора, оказался бы *неограниченным* и, следовательно, *не мог бы быть слабо компактным.*

4. *«Параллелепипед»* в пространстве Лебега $L^2(a,b)$ задаётся условиями

$$U = \Big\{ u(\cdot) \in L^2(a,b) \;\Big|\; \alpha(t) \overset{\text{п.в.}}{\leqslant} u(t) \overset{\text{п.в.}}{\leqslant} \beta(t) \Big\},$$

в которых $\alpha(\cdot)$ и $\beta(\cdot)$ — две фиксированные функции из $L^2(a,b)$, такие, что $\alpha(t) \overset{\text{п.в.}}{\leqslant} \beta(t)$, а *«п.в.»* — сокращение для *«почти всюду»* на (a,b), означающее, что соответствующие неравенства выполняются во всех точках $t \in (a,b)$ за исключением, быть может, подмножества точек, мера Лебега которого равна нулю. Название *«параллелепипед»* объясняется сходством конструкций этого множества с классическим конечномерным параллелепипедом

$$\Pi = \Big\{ u = (u_1, u_2, \ldots, u_n) \in R^n \;\Big|\; \alpha_i \leqslant u_i \leqslant \beta_i,\ i = 1,2,\ldots,n \Big\},$$

в котором аналогом непрерывно меняющегося параметра t является номер i координаты u_i вектора u. Свойства *выпуклости* и *ограниченности* «параллелепипеда» U предлагается проверить самостоятельно. Для доказательства его *замкнутости* воспользуемся следующим фактом [3, гл.VII, §2, п.5]: *из последовательности функций* $u_n(t) \in L^2(a,b)$, *сходящейся в* $L^2(a,b)$ *по норме, т. е. в среднем, можно выделить подпоследовательность* $\{u_{n_m}(t)\} \subset \{u_n(t)\}$, *сходящуюся к тому же самому пределу* $u_0(t)$ *почти всюду на (a,b)* при $m \to \infty$. С помощью этого свойства мы сможем, переходя к поточечному пределу при $m \to \infty$, убедиться в том, что предельная функция $u_0(t)$ почти всюду на (a,b) удовлетворяет двусторонним ограничениям из определения «параллелепипеда» U и, тем самым, установить, что «параллелепипед» в $L^2(a,b)$ является *слабо компактным* множеством.

Упражнение 6. *Докажите, что «параллелепипед» с постоянными границами* $\alpha(t) \equiv \alpha < \beta \equiv \beta(t)$ *не является компактным множеством в пространстве Лебега* $L^2(a, b)$.

II. ЭЛЕМЕНТЫ ДИФФЕРЕНЦИАЛЬНОГО ИСЧИСЛЕНИЯ В НОРМИРОВАННЫХ ПРОСТРАНСТВАХ

Дифференциальное исчисление является фундаментальной составляющей математического аппарата теории экстремума. Дадим определение производной, являющееся естественным обобщением уже известных вам определений дифференцируемости на случай пространств бесконечной размерности.

Определение 13. *Пусть $F : X \to Y$ — некоторое отображение, действующее в нормированных пространствах X и Y. Это отображение называется* ***дифференцируемым по Фрешé*** *в точке $x_0 \in X$, если существует линейный ограниченный оператор $A \in L(X \to Y)$, такой, что*

$$F(x_0 + h) = F(x_0) + Ah + o(\|h\|_X) \quad \forall\, h \in X,$$
$$\text{причём} \quad \frac{\|o(\|h\|_X)\|_Y}{\|h\|_X} \to 0 \quad \textit{при} \quad \|h\|_X \to 0. \quad (6)$$

При этом оператор A называют ***производной Фрешé*** *отображения F в точке x_0 и используют обозначение $A = F'(x_0)$.*

Производные более высокого порядка определяются рекуррентно и их структура быстро усложняется с ростом порядка. Так, например, при определении второй производной функции $F : X \to Y$ мы должны будем дифференцировать отображение

$$x \in X \longmapsto F'(x) \in L(X \to Y)$$

и вторая производная по определению окажется элементом пространства

$$F''(x) \in L(X \to L(X \to Y)).$$

В рамках данного курса нам не придётся иметь дело с производными выше второго порядка. К тому же, в роли F чаще всего будут выступать функционалы $J : H \to R^1$, действующие в гильбертовом пространстве, когда $X = H$ и $Y = R^1$. Укажем на те упрощения в структуре производных, которые при этом возможны. По определению $J'(u) \in L(H \to R^1)$. Напомним, что в случае нормированного пространства X пространство $L(X \to R^1)$ линейных ограниченных функционалов над X называют *сопряжённым* к X пространством и обозначают его через X^*. В пространстве $X^* = L(X \to R^1)$ вводится обычная для пространства линейных ограниченных операторов норма

$$\|f\|_{X^*} = \sup_{x \in X,\, x \neq 0} \frac{|f(x)|}{\|x\|_X}, \qquad f \in X^*. \quad (7)$$

Таким образом, первая и вторая производные дифференцируемого функционала $J : X \to R^1$ будут элементами пространств

$$J'(u) \in X^*, \qquad J''(u) \in L(X \to X^*).$$

В случае, когда пространство $X = H$ гильбертово, есть возможность отождествления пространства H с сопряжённым к нему пространством H^* по *теореме Рисса* [3, гл.IV, §2]. В этой теореме утверждается, что существует линейное взаимно однозначное отображение (оператор Рисса) $\mathcal{R}_H : H^* \to H$, позволяющее описывать действие функционалов через скалярные произведения:

$$\forall\, f \in H^* \quad \exists!\, \mathcal{R}_H f \in H : \quad f(h) = \langle \mathcal{R}_H f, h\rangle_H \quad \forall\, h \in H, \tag{8}$$

а также сохраняющее нормы и позволяющее наделить сопряжённое пространство H^* скалярным произведением:

$$\|\mathcal{R}_H f\|_H = \|f\|_{H^*}, \quad \langle f, g\rangle_{H^*} = \langle \mathcal{R}_H f, \mathcal{R}_H g\rangle_H \quad \forall\, f, g \in H^*. \tag{9}$$

Заметим, что *возможность* отождествления $H^* \simeq H$ отнюдь *не означает обязательности* использования этой возможности, но в рамках данного курса нам будет удобно пользоваться возможностью отождествления по Риссу $H^* = H$, а тогда

$$J'(u) \in H, \qquad J''(u) \in L(H \to H). \tag{10}$$

С учётом принятого отождествления $H^* = H$ приведём два соотношения, первое из которых *эквивалентно* определению (6) дифференцируемости функционала $J : H \to R^1$:

$$J(u_0 + h) = J(u_0) + \langle J'(u_0), h\rangle_H + o(\|h\|_H),$$

а второе является *следствием* его дважды дифференцируемости:

$$J(u_0 + h) = J(u_0) + \langle J'(u_0), h\rangle_H + \frac{1}{2}\langle J''(u_0)h, h\rangle_H + o(\|h\|_H^2). \tag{11}$$

Любопытно, что из (11) существование второй производной не следует, причём даже для функций одной переменной. На эту тему предагается

Упражнение 7. *Приведите пример функции $f : R^1 \to R^1$, для которой в окрестности нуля при некоторых $a, b, c \in R^1$ выполняется соотношение*

$$f(x) = a + bx + \frac{1}{2}cx^2 + o(x^2),$$

но которая не является дважды дифференцируемой в точке $x = 0$.

Производная Фреше́ обладает обычными свойствами: если производная существует, то она единственна; производная суммы равна сумме производных; постоянный множитель выносится за знак производной и т. д. Сохраняется в силе и правило дифференцирования сложной функции, которое мы приведём в развёрнутой форме, поскольку им удобно пользоваться при решении задач.

Утверждение. [3, гл.X] (`производная сложной функции`) *Пусть* X, Y, Z — *нормированные пространства, в которых действуют отображения* $F : X \to Y$ *и* $G : Y \to Z$. *Пусть отбражение* F *дифференцируемо по Фреше в точке* $x_0 \in X$, *а отображение* G *дифференцируемо по Фреше в точке* $y_0 = F(x_0)$, *т. е. существуют производные* $F'(x_0) \in L(X \to Y)$ *и* $G'(F(x_0)) \in L(Y \to Z)$. *Тогда суперпозиция* $(GF)(x) = G(F(x)) : X \to Z$ *дифференцируема по Фреше в точке* x_0 *и*

$$(GF)'(x_0) = G'(F(x_0))\, F'(x_0) \in L(X \to Z). \tag{12}$$

В дальнейшем для классов непрерывных, липшиц-непрерывных и дифференцируемых (по Фрешé) функций (отображений) нам будет удобно пользоваться обозначениями, аналогичными тем, с которыми вы уже встречались ранее.

- $C(U)$ — класс функций, (сильно) непрерывных на множестве U,
- $C^1(U)$ — класс функций, непрерывно дифференцируемых по Фрешé на множестве U,
- $C^2(U)$ — класс функций, дважды непрерывно дифференцируемых по Фрешé на множестве U,
- $\mathrm{Lip}(U)$ — класс функций, липшиц-непрерывных на множестве U.

В этих обозначениях явно не указаны пространства, которым принадлежат *значения* рассматриваемых отображений. Как правило, это однозначно определяется из контекста. Так, например, для функционала $J(u) : H \to R^1$, действующего в гильбертовом пространстве H, запись $J(u) \in C^1(U)$ означает непрерывность на U самой функции $J(u) : U \to R^1$ и её первой производной (градиента) как отображения $J'(u) : U \to H$, а запись $J'(u) \in \mathrm{Lip}(U)$ с $L > 0$ будет означать липшиц-непрерывность градиента на множестве U с константой Липшица $L > 0$:

$$\|J'(u) - J'(v)\|_H \leqslant L\,\|u - v\|_H \quad \forall\, u, v \in U.$$

Формулы конечных приращений

Для гладких функций одной переменной $f(x) : [a, b] \to R^1$ хорошо известны формула Ньютона-Лейбница:

$$f(b) - f(a) = \int_a^b f'(t)\,dt \tag{13}$$

и формула конечных приращений Лагранжа:

$$f(b) - f(a) = f'(\xi)\,(b - a), \quad \text{где} \quad \xi \in [a, b]. \tag{14}$$

Приведём их аналоги в бесконечномерных пространствах. Так, в случае нормированных пространств X, Y и отображения $F : X \to Y$ класса $F \in C^1(X)$ аналогом формулы Ньютона-Лейбница (13) будет соотношение

$$F(u+h) - F(u) = \int_0^1 F'(u+th)\,h\,dt = \Big(\int_0^1 F'(u+th)\,dt\Big)\,h \quad \forall\, u, h \in X, \tag{15}$$

в котором отображения

$$F'(u+th) \in L(X \to Y), \quad \int_0^1 F'(u+th)\,dt \in L(X \to Y)$$

являются линейными ограниченными операторами из X в Y, а интегрирование понимается в смысле Бохнера (в конечномерном случае, когда $\dim X < \infty$ и $\dim Y < \infty$, это обычные интегралы Римана). В частном случае функционала $J(u) = F(u) : H \to R^1$, действующего в гильбертовом пространстве H, в силу принятого нами отождествления по Риссу $L(H \to R^1) = H^* = H$ присутствующие в (15) значения производных на элементе h превратятся в скалярные произведения и для $J(u) \in C^1(U)$ мы получим равенства

$$J(u+h) - J(u) = \int_0^1 \langle J'(u+th),\, h\rangle_H\,dt = \Big\langle \int_0^1 J'(u+th)\,dt,\, h\Big\rangle_H \quad \forall\, u, h \in H. \tag{16}$$

Интеграл Римана, расположенный в центре цепочки (16), можно, применяя теорему о среднем, заменить значением подынтегральной функции в промежуточной точке $\theta \in [0,1]$, что даёт возможность переписать первое равенство из (16) в виде, аналогичном (14):

$$J(u+h) - J(u) = \langle J'(u+\theta h),\, h\rangle_H \quad \forall\, u, h \in H \quad \big(\exists\, \theta \in [0,1]\big). \tag{17}$$

Заметим, что в формулах (15) перейти от интегралов к их «средним» значениям, вообще говоря, невозможно, причём уже в случае, кода $\dim Y = 2$. Соответствующим примером может служить гладкое отображение

$$F(t) = (\cos t,\, \sin t) : [0, 2\pi] \to R^2,$$

для которого

$$F(2\pi) - F(0) = (0,0) \neq F'(\theta)\,2\pi \quad \forall\, \theta \in [0, 2\pi].$$

Примеры вычисления производных

1. В *конечномерном* пространстве $H = R^n$ для гладких функций $J(u) : R^n \to R^1$ с учётом риссовского отождествления $(R^n)^* = R^n$ имеем

$$J'(u) = \nabla J(u) = \Big(\frac{\partial J(u)}{\partial u_1},\, \frac{\partial J(u)}{\partial u_2},\, \ldots,\, \frac{\partial J(u)}{\partial u_n}\Big), \quad u = (u_1, u_2, \ldots, u_n), \tag{18}$$

$$J''(u) = \begin{pmatrix} \frac{\partial^2 J(u)}{\partial u_1^2} & \frac{\partial^2 J(u)}{\partial u_2 \partial u_1} & \cdots & \frac{\partial^2 J(u)}{\partial u_n \partial u_1} \\ \frac{\partial^2 J(u)}{\partial u_1 \partial u_2} & \frac{\partial^2 J(u)}{\partial u_2^2} & \cdots & \frac{\partial^2 J(u)}{\partial u_n \partial u_2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 J(u)}{\partial u_1 \partial u_n} & \frac{\partial^2 J(u)}{\partial u_2 \partial u_n} & \cdots & \frac{\partial^2 J(u)}{\partial u_n^2} \end{pmatrix}, \quad u = (u_1, u_2, \ldots, u_n). \tag{19}$$

Вторую производную (19) в конечномерном случае принято называть *матрицей Гессе* или *гессианом*, а первую производную функционала $J(u) : H \to R^1$ независимо от размерности пространства H называют *градиентом*.

2. *Линейный* функционал $J(u) = \langle c, u \rangle$, в котором $c \in H$ — некоторый фиксированный элемент из H, является бесконечно дифференцируемым на всём пространстве H и

$$J'(u) = c \in H, \qquad J''(u) = 0 \in L(H \to H).$$

3. *Квадратичный* функционал $J(u) = \|Au - f\|_F^2$, в котором фиксированы *линейный ограниченный* оператор $A \in L(H \to F)$ и элемент $f \in F$, также является бесконечно дифференцируемым на всём пространстве H. Найдём его первую производную (градиент). Фиксируем произвольную точку $u \in H$, произвольное приращение $h \in H$ и преобразуем приращение функции:

$$J(u+h) - J(u) = \|A(u+h) - f\|_F^2 - \|Au - f\|_F^2 \overset{A \text{ лин.}}{=}$$

$$= \|(Au - f) + Ah\|_F^2 - \|Au - f\|_F^2 = \big[\text{вып. скал. умнож.}\big] =$$

$$= 2\langle Au - f,\, Ah \rangle_F + \|Ah\|_F^2 = \big[\text{определение } A^* : F \to H\big] =$$

$$= \langle 2A^*(Au - f),\, h \rangle_H + \|Ah\|_F^2.$$

Первое слагаемое в правой части данного равенства представляет собой линейный и непрерывный по переменной $h \in H$ функционал, а остаточный член $\|Ah\|_F^2$ в силу ограниченности оператора $A \in L(H \to F)$ является величиной более высокого порядка малости по сравнению с $\|h\|_H$ при $\|h\|_H \to 0$, следовательно, по определению производной Фреше́ имеем

$$J'(u) = 2A^*(Au - f) \in H, \quad J''(u) = 2A^*A \in L(H \to H), \tag{20}$$

а все старшие производные квадратичной функции будут *нулевыми* операторами.

Упражнение 8. *Найдите первую и вторую производные $J'(u)$ и $J''(u)$ квадратичного (необязательно выпуклого) функционала вида*

$$J(u) = \frac{1}{2}\langle Au, u \rangle_H - \langle f, u \rangle_H, \quad u \in H,$$

где оператор $A \in L(H \to H)$ и элемент $f \in H$ заданы, а пространство H — гильбертово.

Упражнение 9. *Пусть H — гильбертово пространство и $g(t) : R^1 \to R^1$ — достаточно гладкая скалярная функция. Найдите первую и вторую производные Фреше́ суперпозиции $J(u) = g\big(\|u\|_H\big),\ u \in H$. Существуют ли производные в нуле $J'(0)$ в случаях, когда $g(t) = t$ и $g(t) = t^3$? Если производная существует, найдите её.*

III. ЗАДАЧИ УПРАВЛЕНИЯ ЛИНЕЙНЫМИ ДИНАМИЧЕСКИМИ СИСТЕМАМИ С КВАДРАТИЧНЫМИ КРИТЕРИЯМИ КАЧЕСТВА

В этом разделе будут рассмотрены некоторые приложения прочитанного теоретического материала, относящегося к вопросам существования решений оптимизационных задач и проблемам дифференцируемости. Мы изучим некоторые постановки задач оптимального управления процессами, динамика которых описывается обыкновенными дифференциальными уравнениями, уравнениями с частными производными параболического типа и системами дискретных соотношений.

Управляемые процессы с обыкновенными дифференциальными уравнениями

Пусть движение управляемого объекта описывается задачей Коши для линейной системы обыкновенных дифференциальных уравнений (ОДУ):

$$x'(t) = D(t)x(t) + B(t)u(t) + F(t), \quad t \in (0,T); \quad x(0) = x_0. \tag{21}$$

Здесь $x = (x_1, x_2, \ldots, x_n)$ — фазовые координаты, $u = (u_1, u_2, \ldots, u_r)$ — управляющие параметры, $u(t) = (u_1(t), u_2(t), \ldots, u_r(t))$ — управление, $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))$ — фазовая траектория системы, $D(t)$, $B(t)$ и $F(t)$ — заданные матрицы размеров $n \times n$, $n \times r$ и $n \times 1$ соответственно, $T > 0$ — заданный финальный момент времени, x_0 — заданное начальное положение траектории. Будем предполагать, что элементы матриц $D(t)$, $B(t)$ и $F(t)$ являются *измеримыми по Лебегу* и *ограниченными* на промежутке $t \in (0,T)$ функциями и записывать это в виде

$$D(t),\ B(t),\ F(t) \in L^\infty(0,T).$$

В данной записи для краткости мы явно не указываем размерность описываемых объектов. Одномерное пространство $L^\infty(0,T)$ состоит из измеримых скалярных функций $f(t) : (0,T) \to R^1$, ограниченных в смысле:

$$\exists\, M = \text{const} > 0 : \quad |f(t)| \leqslant M \quad \text{для п.в.} \quad t \in (0,T).$$

Пространство $L^\infty(0,T)$ содержит в себе все классы Лебега $L^p(0,T)$, $p \geqslant 1$, состоящие из измеримых на $(0,T)$ функций с конечными нормами

$$\|f\|_{L^p(0,T)} = \Big(\int_0^T |f(t)|^p \, dt \Big)^{1/p}$$

и являющиеся *банаховыми* пространствами. Само пространство $L^\infty(0,T)$ также является *банаховым* относительно нормы

$$\|f\|_{L^\infty(0,T)} = \lim_{p\to+\infty}\Big(\int_0^T |f(t)|^p\,dt\Big)^{1/p} = \inf_{M>0:\,|f(t)|\overset{\text{п.в.}}{\leqslant} M} M.$$

Вернёмся к постановке оптимизационной задачи. Допустимыми будем считать управления из множества

$$u(\cdot)\in U\subset L^2(0,T). \tag{22}$$

Целью управления будет минимизация одного из двух критериев: *интегрального квадратичного* функционала

$$J_I(u)=\int_0^T |x(t;u)-f(t)|^2_{R^n}\,dt\to\inf \tag{23}$$

или *терминального квадратичного* функционала

$$J_T(u)=|x(T;u)-f|^2_{R^n}\to\inf\,. \tag{24}$$

В (23) и (24) использовано развёрнутое обозначение $x(t;u)$ для фазовой траектории, указывающее на её зависимость от управления u. Целевая функция $f(t)\in L^2(0,T)$ в (23) и целевая позиция $f\in R^n$ в (24) предполагаются заданными.

По-видимому, требует пояснения вопрос существования и единственности решения $x(t)=x(t;u)$ задачи Коши, соответствующего *измеримому* управлению $u(t)\in L^2(0,T)$. Под решением будем понимать непрерывную на $[0,T]$ функцию $x(t)$, удовлетворяющую интегральному уравнению

$$x(t)=x_0+\int_0^t \big(D(s)x(s)+B(s)u(s)+F(s)\big)\,ds,\quad t\in[0,T]. \tag{25}$$

При сформулированных требованиях к матрицам $D(t),B(t),F(t)$ и управлениям $u(t)$ с помощью *метода сжимающих отображений* или его модифицированной версии [3, гл. II, §4] можно доказать существование и единственность решения $x(t;u)$ интегрального уравнения (25) для любого управления $u(t)\in L^2(0,T)$. Далее при желании можно показать, что это решение является не только непрерывной, но и абсолютно непрерывной на $[0,T]$ функцией, которая удовлетворяет дифференциальному уравнению (21) для п.в. $t\in(0,T)$.

Проведём редукцию задач минимизации функционалов (23) и (24) к задачам минимизации квадратичного функционала вида $J(u)=\|Au-f\|^2_F$, для

которого нам уже известны свойства выпуклости, слабой п/н снизу, дифференцируемости и выражения для его производных. Возможность такой трансформации определяется *линейностью* системы (21) и *квадратичной* зависимостью функционалов от траектории. Представим решение $x(t)$ задачи (21) в виде суммы $x(t) = x_1(t) + x_2(t)$, где

$$x_1'(t) = D(t)x_1(t) + B(t)u(t), \quad t \in (0,T); \quad x_1(0) = 0,$$
$$x_2'(t) = D(t)x_2(t) + F(t), \quad t \in (0,T); \quad x_2(0) = x_0.$$

Составляющую $x_2(t)$ такого представления можно считать *известной*, поскольку она *не зависит от управления* $u(t)$ и может быть найдена аналитически или численно как решение соответствующей задачи Коши. Чтобы сформировать искомую квадратичную конструкцию для интегрального квадратичного функционала (23), выберем пространства, оператор и целевой элемент следующим образом:

$$H = L^2(0,T), \;\; F = L^2(0,T), \;\; A_I u = x_1(t;u) : H \to F, \;\; f := f(t) - x_2(t) \in F.$$

Заметим, что здесь пространства H и F отличаются друг от друга только размерностью: $\dim u = r$, $\dim x = n$. Для квадратичного терминального функционала (24) эти данные заменяются следующими:

$$H = L^2(0,T), \quad F = R^n, \quad A_T u = x_1(T;u) : H \to F, \quad f := f - x_2(T) \in F.$$

Линейный характер зависимости операторов A_I и A_T от управления u не вызывает сомнений. Чтобы иметь возможность в полной мере воспользоваться изученными выше свойствами квадратичных функционалов вида $J(u) = \|Au - f\|_F^2$, следует убедиться в том, что эти операторы *ограничены* (непрерывны) и для них имеются оценки

$$\|Au\|_F \leqslant C\,\|u\|_H \quad \forall\, u \in H \quad (C = \text{const}{>}0). \tag{26}$$

В рассматриваемом случае наличие таких оценок связано с возможностью оценивания траекторий $x_1(t)$ через управления $u(t)$. Для систем ОДУ чаще всего в таких целях используется лемма Гронуолла-Беллмана, с которой вы уже должны были встречаться ранее в курсе ДУ. Будем считать, что описанная выше редукция исходной задачи уже произведена и динамическая система имеет вид

$$x'(t) = D(t)x(t) + B(t)u(t), \quad t \in (0,T); \quad x(0) = 0, \tag{27}$$

а неоднородности $F(t)$ и x_0 из исходной системы (21) через вспомогательную траекторию $x_2(t)$ уже внесли соответствующие изменения в целевые элементы $f(t)$ и f функционалов (23) и (24). Перейдём от дифференциальной постановки (27) к интегральному уравнению

$$x(t) = \int_0^t \big(D(s)x(s) + B(s)u(s)\big)\,ds \quad t \in [0,T],$$

а от него — к оценке

$$|x(t)|_{R^n} = \Big|\int_0^t \big(D(s)x(s) + B(s)u(s)\big)\,ds\Big|_{R^n} \leqslant$$

$$\leqslant \big[\text{вносим норму под знак интеграла и применяем нер-во треугольника}\big] \leqslant$$

$$\leqslant \int_0^t \big(|D(s)|_{R^{n\times n}}\,|x(s)|_{R^n} + |B(s)|_{R^{n\times r}}|u(s)|_{R^r}\big)\,ds \leqslant \big[D(t),\, B(t) \in L^\infty(0,T)\big] \leqslant$$

$$\leqslant C_D \int_0^t |x(s)|_{R^n}\,ds + C_B \int_0^t |u(s)|_{R^r}\,ds \leqslant \big[t \leqslant T\big] \leqslant$$

$$\leqslant C_D \int_0^t |x(s)|_{R^n}\,ds + C_B \int_0^T |u(s)|_{R^r}\,ds, \quad t \in [0,T].$$

К интегральным неравенствам именно такого типа применима упомянутая выше лемма Гронуолла [1, гл. 6, §3], с помощью которой можно перейти к оценке вида

$$|x(t)|_{R^n} \leqslant C_B\Big(\int_0^T |u(s)|_{R^r}\,ds\Big)\,e^{C_D\,t}, \quad t \in [0,T],$$

а от неё — к оценке

$$\|x(\cdot)\|_{C[0,T]} = \max_{t\in[0,T]} |x(t)|_{R^n} \leqslant C_B\,e^{C_D\,T}\,\|u(\cdot)\|_{L^1(0,T)}\,, \tag{28}$$

которая отражает факт непрерывной зависимости траектории $x(t)$ системы (27) в норме пространства $C[0,T]$ от возмущений управлений $u(t)$ в норме пространства Лебега $L^1(0,T)$. Из (28) следует также и непрерывная зависимость от $u(t)$ по норме гильбертова пространства $H = L^2(0,T)$, поскольку

$$\|u\|_{L^1(0,T)} = \int_0^T |u(s)|_{R^r}\,ds = \int_0^T 1\cdot|u(s)|_{R^r}\,ds \overset{\text{К-Б}}{\leqslant}$$

$$\leqslant \Big(\int_0^T 1^2\,ds\Big)^{1/2}\Big(\int_0^T |u(s)|^2_{R^r}\,ds\Big)^{1/2} = \sqrt{T}\,\|u\|_{L^2(0,T)}\,. \tag{29}$$

Из (28) и (29) следуют искомые оценки ограниченности вида (26) для операторов A_T и A_I со следующими значениями оценочных констант:

$$C \overset{A_T}{=} C_B\,\sqrt{T}\,e^{C_D\,T}, \qquad C \overset{A_I}{=} C_B\,T\,e^{C_D\,T}\,.$$

Установленное свойство ограниченности означает, что оба функционала (23) и (24) являются слабо п/н снизу на всём пространстве $H = L^2(0,T)$ и мы можем воспользоваться утверждением «слабого» варианта теоремы Вейерштрасса (теоремы 2) при формулировке следующего результата.

Теорема 3. *Пусть в системе ОДУ* (27) $D(t)$, $B(t) \in L^\infty(0,T)$, *допустимое множество управлений* U *слабо компактно в пространстве* $L^2(0,T)$. *Тогда для любых целевых элементов* $f(t) \in L^2(0,T)$ *в* (23) *и* $f \in R^n$ *в* (24) *каждая из оптимизационных задач* (22),(23),(27) *и* (22),(24),(27) *имеет решение, т. е.* $J_* > -\infty$ *и* $U_* \neq \varnothing$.

Управляемые процессы с обыкновенными дифференциальными уравнениями (продолжение)

Перейдём к исследованию функционалов (23) и (24) на предмет дифференцируемости. Как и с существованием решений, здесь никаких принципиальных проблем тоже не возникнет, поскольку для квадратичных функционалов вида $J(u) = \|Au - f\|_F^2$ производные нам уже известны (см. (20)):

$$J'(u) = 2A^*(Au - f) \in H, \qquad J''(u) = 2A^*A \in L(H \to H) \quad \forall\, u \in H.$$

Остаётся только объяснить, как именно действует сопряженный к A оператор A^*. Приведём соответствующие подробности для случая *терминального* квадратичного функционала (24), в котором оператор $A = A_T$ действует из $H = L^2(0,T)$ в $F = R^n$ по правилу $A_T u = x(T;u)$. По определению сопряжённого оператора и с учётом риссовских отождествлений $H^* = H$, $F^* = F$ оператор A^* будет действовать из $F = R^n$ в $H = L^2(0,T)$, подчиняясь тождеству

$$\langle Au,\, v\rangle_F = \langle u,\, A^*v\rangle_H \quad \forall\, u \in H\ \forall v \in F. \tag{30}$$

Левая часть (30) для оператора $A = A_T$ нам известна:

$$\langle A_T u,\, v\rangle_F = \langle x(T;u),\, v\rangle_{R^n}$$

и мы должны представить эти значения в виде, который диктуется правой частью (30), а именно:

$$\langle u,\, A_T^* v\rangle_H = \int_0^T \big\langle u(t),\, (A_T^* v)(t)\big\rangle_{R^r}\, dt,$$

подобрав подходящую функцию на место $(A_T^* v)(t)$. Для проведения такой трансформации вычтем из правой части дифференциального уравнения (27) для траектории $x(t)$ его левую часть, умножим полученную нулевую вектор-функцию скалярно в R^n на (пока) произвольную вектор-функцию $\psi(t) \in L^2(0,T)$, проинтегрируем полученный нулевой результат по $(0,T)$ и добавим ноль, записанный в указанном «экзотическом» виде, к левой части (30):

$$\langle A_T u,\, v\rangle_F = \langle x(T;u),\, v\rangle_{R^n} + \underbrace{\int_0^T \langle D(t)x(t) + B(t)u(t) - x'(t),\, \psi(t)\rangle_{R^n}\, dt}_{=0}. \tag{31}$$

Далее, учтём, что

$$\langle D(t)x(t) + B(t)u(t),\, \psi(t)\rangle_{R^n} = \langle x(t),\, D^\top(t)\psi(t)\rangle_{R^n} + \langle u(t),\, B^\top(t)\psi(t)\rangle_{R^r}$$

и, предполагая функцию $\psi(t)$ гладкой, проинтегрируем по частям последнее слагаемое из (31):

$$\int_0^T \langle -x'(t),\, \psi(t)\rangle_{R^n}\, dt = -\langle x(T),\, \psi(T)\rangle_{R^n} + \langle x(0),\, \psi(0)\rangle_{R^n} + \int_0^T \langle x(t),\, \psi'(t)\rangle_{R^n}\, dt.$$

Поскольку $x(0) = 0$, то от (31) можно будет перейти к равенству

$$\langle A_T u,\, v\rangle_F = \langle x(T),\, v - \psi(T)\rangle_{R^n} + \int_0^T \big\langle x(t),\; D^\top(t)\psi(t) + \psi'(t)\big\rangle_{R^n}\, dt + \\ + \int_0^T \langle u(t),\, B^\top(t)\psi(t)\rangle_{R^r}\, dt. \quad (32)$$

В нём последний интеграл имеет нужный нам вид, а от всех остальных мы можем избавиться, предъявив к функции $\psi(t)$ дополнительные требования:

$$\psi'(t) = -D^\top(t)\psi(t), \quad 0 < t < T; \quad \psi(T) = v. \quad (33)$$

Условия (33) формируют для функции $\psi(t)$ задачу Коши для линейной системы из n дифференциальных уравнений с обратным течением времени, поскольку «стартовое» условие $\psi(T) = v$ задаётся не в начальный, а в финальный момент времени. Саму задачу (33) принято называть *сопряжённой* по отношению к исходной дифференциальной задаче (27). При выборе $\psi(t)$ из условий (33) предыдущее равенство (32) приводит нас к выводу:

$$A_T^* v = B^\top(t)\psi(t), \quad t \in (0, T), \quad \forall\, v \in R^n. \quad (34)$$

Отсюда, возвращаясь к порождаемому оператором A_T *терминальному* квадратичному функционалу $J_T(u)$, получаем формулу для его градиента:

$$J_T'(u) = 2A_T^*(A_T u - f) = 2\, B^\top(t)\psi(t; v)\,\big|_{\,v = x(T;u) - f}\,, \quad \forall\, u \in L^2(0, T). \quad (35)$$

Для *интегрального* квадратичного функционала $J_I(u)$ при выводе формулы градиента выполняются аналогичные действия, поэтому ограничимся формулировкой результата:

$$J_I'(u) = 2A_I^*(A_I u - f) = 2\, B^\top(t)\psi(t; v)\,\big|_{\,v = x(t;u) - f(t)}\,, \quad \forall\, u \in L^2(0, T), \quad (36)$$

где

$$A_I^* v = B^\top(t)\psi(t), \quad t \in (0, T), \quad \forall\, v \in L^2(0, T), \quad (37)$$

а $\psi(t) = \psi(t; v)$ — решение следующей *сопряжённой* задачи Коши:

$$\psi'(t) = -D^\top(t)\psi(t) - v(t), \quad 0 < t < T; \quad \psi(T) = 0. \quad (38)$$

Соответствующие выкладки студентам рекомендуется воспроизвести самостоятельно. Установленные свойства дифференцируемости содержит следующая

Теорема 4. *Пусть в системе ОДУ* (27) $D(t),\ B(t) \in L^\infty(0,T)$. *Тогда для любых целевых элементов* $f(t) \in L^2(0,T)$ *в* (23) *и* $f \in R^n$ *в* (24) *каждый из двух квадратичных функционалов* (23) *и* (24) *бесконечно дифференцируем по Фрешé на всём пространстве* $L^2(0,T)$ *и их первые производные имеют вид* (36), (38) *и* (35), (33) *соответственно.*

Упражнение 10. *В пространстве* $L^2(0,l)$ *найдите градиент функционала*

$$J(u) = \int_0^l \rho(x)\,|y(x;u) - z(x)|^2\,dx, \quad u \in L^2(0,l),$$

где $\rho(x) \in C[0,l],\ \rho(x) \geqslant 0$, — *заданная весовыя функция, а* $y(x) = y(x;u)$ — *решение краевой задачи для линейного обыкновенного дифференциального уравнения второго порядка*

$$\begin{cases} (k(x)y'(x))' - q(x)y(x) = -u(x), \quad 0 < x < l, \\ y(0) = 0, \qquad y'(l) = 0, \end{cases}$$

с заданными непрерывными на $[0,l]$ *коэффициентами* $k(x) \geqslant k_0 > 0,\ q(x) \geqslant 0$.

Задачи управления для параболического уравнения

Рассмотрим управляемый процесс, который описывается первой краевой задачей для уравнения теплопроводности (параболического типа):

$$\begin{gathered} y_t = y_{xx}, \quad 0 < t < T, \quad 0 < x < l, \\ y\,\big|_{x=0} = 0, \quad y\,\big|_{x=l} = 0, \quad 0 < t < T, \\ y\,\big|_{t=0} = u(x), \quad 0 < x < l. \end{gathered} \tag{39}$$

Фазовой траекторией здесь является функция $y = y(t,x) = y(t,x;u)$ двух независимых переменных (t,x), определённая на прямоугольнике $Q = (0,T)\times(0,l)$ и зависящая от управления $u = u(x)$, являющегося в данном случае начальным состоянием системы. Функцию $y(t,x)$ можно интерпретировать как температуру в момент t в точке x однородного стержня длины l. При этом граничные условия означают, что на концах этого стержня поддерживается нулевая температура, а управляющие воздействия на процесс передаются через начальное распределение температуры $u(x)$. Управления будем выбирать из гильбертова пространства Лебега:

$$u(\cdot) \in U \subset H = L^2(0,T). \tag{40}$$

В качестве критериев качества управления возьмем, как и в случае ОДУ, *интегральный* квадратичный функционал

$$J_I(u) = \iint\limits_Q \left|y(t,x;u) - f(t,x)\right|^2 dt\,dx \tag{41}$$

или *терминальный* квадратичный функционал

$$J_T(u) = \int_0^l \left|y(T,x;u) - f(x)\right|^2 dx\,. \tag{42}$$

Задача минимизации функционала $J_I(u)$ отражает желание приблизить распределение температуры к заданному целевому её распределению $f = f(t,x)$, $(t,x) \in Q$, а в задаче минимизации критерия $J_T(u)$ важно приблизиться к заданному распределению температуры $f = f(x),\ x \in (0,l)$, лишь в финальный момент времени $t = T$. Формально оба функционала (41) и (42) записываются в стандартной квадратичной форме $J(u) = \|Au - f\|_F^2$, если взять

$$Au = A_I u = y(t,x;u),\ (t,x) \in Q,\ A_I : H \to F = F_I = L^2(Q),\ f = f(t,x) \in F,$$
$$Au = A_T u = y(T,x;u),\ x \in (0,l),\ A_T : H \to F = F_T = L^2(0,l),\ f = f(x) \in F.$$

Чтобы иметь возможность в полной мере использовать известные нам свойства квадратичного функционала вида $J(u) = \|Au - f\|_F^2$, как и в задачах управления системами ОДУ, следует установить свойства *линейности* и *ограниченности* операторов A_I и A_T. Их *линейность* обусловлена линейностью рассматриваемой дифференциальной задачи (39).

Для доказательства *ограниченности* воспользуемся техникой *энергетических оценок*, весьма распространённой в теории и методах уравнений математической физики. Сначала предположим, что управления $u = u(x)$ являются достаточно гладкими функциями, например, бесконечно гладкими: $u(x) \in C^\infty[0,l]$ и принимающими *нулевые* значения на концах орезка $[0,l]$, т. е. удовлетворяющими заданным однородным граничным условиям. Для таких данных в курсе «Уравнения математической физики» (УМФ) доказывается, что начально-краевая задача (39) имеет единственное классическое решение $y(t,x)$, допускающее явное аналитическое представление в виде ряда Фурье. Умножим обе части дифференциального уравнения (39) на само решение $y(t,x)$ и проинтегрируем полученное равенство по усечённому прямоугольнику $Q_{t_0} = (0,t_0) \times (0,l)$:

$$\iint\limits_{Q_{t_0}} y_t y\,dt\,dx = \iint\limits_{Q_{t_0}} y_{xx} y\,dt\,dx\,, \quad t_0 \in [0,T]\,. \tag{43}$$

Преобразуем левую и правую части (43):

$$\iint\limits_{Q_{t_0}} y_t y\,dt\,dx = \big[\, y_t y = \frac{1}{2}\left(y^2\right)_t \,\big] \overset{\text{Н-Л}}{=} \frac{1}{2}\int_0^l y^2(t_0,x)\,dx - \frac{1}{2}\int_0^l y^2(0,x)\,dx =$$

$$= \big[y(0,x) = u(x)\big] = \frac{1}{2}\int_0^l y^2(t_0,x)\,dx - \frac{1}{2}\int_0^l u^2(x)\,dx,$$

$$\iint\limits_{Q_{t_0}} y_{xx} y\,dt\,dx = \big[\text{интегр. по } x \text{ по частям}\big] = \int_0^{t_0} y_x\, y\,\Big|_{x=0}^{x=l} dt - \iint\limits_{Q_{t_0}} y_x^2\,dt\,dx =$$

$$= \big[\, y(t,0) = y(t,l) = 0 \,\big] = -\iint\limits_{Q_{t_0}} y_x^2\,dt\,dx\,.$$

В результате получаем энергетическое равенство

$$\frac{1}{2}\int_0^l y^2(t_0,x)\,dx + \iint\limits_{Q_{t_0}} y_x^2\,dt\,dx = \frac{1}{2}\int_0^l u^2(x)\,dx \quad \forall\, t_0 \in [0,T],$$

из которого, после удаления из левой части двойного интеграла, следует неравенство

$$\int_0^l y^2(t_0,x)\,dx \leqslant \int_0^l u^2(x)\,dx \quad \forall\, t_0 \in [0,T]. \tag{44}$$

Взяв в (44) $t_0 = T$, получим ограниченность оператора A_T :

$$\|A_T u\|^2_{L^2(0,l)} \leqslant \|u\|^2_H \quad \forall\, u \in H \quad \Longrightarrow \quad \|A_T\| \leqslant 1.$$

Проинтегрировав (44) по переменной t_0 от 0 до T, получим ограниченность оператора A_I :

$$\|A_I u\|^2_{L^2(Q)} \leqslant T\,\|u\|^2_H \quad \forall\, u \in H \quad \Longrightarrow \quad \|A_I\| \leqslant \sqrt{T}.$$

Полученные оценки ограниченности операторов A_T и A_I позволяют корректно ввести понятие обобщённого решения $y(t,x)$ начально-краевой задачи (39) для негладких начальных состояний (управлений) $u(x) \in L^2(0,l)$, используя принцип «доопределения по непрерывности». Поскольку множество функций класса $C^\infty[0,l]$, принимающих нулевые граничные значения, *всюду плотно* в $L^2(0,l)$, то любую функцию $u(x) \in L^2(0,l)$ можно аппроксимировать последовательностью гладких функций:

$$u_k(x) \in C^\infty[0,l], \quad u_k(0) = u_k(l) = 0, \quad \|u_k - u\|_{L^2(0,l)} \to 0 \quad \text{при} \quad k \to \infty.$$

Гладким данным $u_k(x) \in C^\infty[0,l]$ будут соответствовать классические решения $y_k(t,x) = y(t,x;u_k)$ задачи (39), у которых, в силу полученных оценок ограниченности, будут существовать (единственные) пределы в соответствующих операторам A_T и A_I пространствах $F_T = L^2(0,l)$ и $F_I = L^2(Q)$, которые интерпретируются как обобщённые решения $y(t,x) \in L^2(Q)$ или их следы $y(T,x) \in L^2(0,l)$ на финальном слое $t = T$, соответствующие негладким управлениям $u(x)$. Эти, пусть не совсем строгие, рассуждения дают возможность считать, что оба квадратичных функционала $J_I(u)$ и $J_T(u)$ определены на всём пространстве $H = L^2(0,l)$ и являются на всём этом пространстве выпуклыми, бесконечно дифференцируемыми и слабо п/н снизу, а в известной нам формуле градиента $J'(u) = 2A^*(Au - f)$ нам остаётся расшифровать правила действия сопряжённых операторов A_I^* и A_T^*. Как и в задаче управления с ОДУ, остановимся подробнее на описании действия оператора A_T^*, а для A_I^* приведём лишь окончательный результат.

Исходим из определения сопряжённого оператора тождеством:

$$\langle A_T u,\, v\rangle_{L^2(0,l)} = \langle u,\, A_T^* v\rangle_{L^2(0,l)} \quad \forall\, u \in H = L^2(0,l)\ \forall v \in F = L^2(0,l).$$

Левая часть для оператора $A = A_T$ нам известна:

$$\langle A_T u,\, v\rangle_{L^2(0,l)} = \int_0^l y(T,x)\, v(x)\, dx$$

и мы должны представить её в виде

$$\int_0^l u(x)\, (A_T^* v)(x)\, dx\,,$$

подобрав подходящую функцию на место $(A_T^* v)(x)$. Для этого вычтем из правой части дифференциального уравнения (39) его левую часть, умножим полученную нулевую разность на (пока) произвольную функцию $\psi(t,x) \in L^2(Q)$, проинтегрируем полученный нулевой результат по прямоугольнику Q и добавим полученный ноль к левой части тождества:

$$\langle A_T u,\, v\rangle_{L^2(0,l)} = \int_0^l y(T,x)\, v(x)\, dx + \underbrace{\iint\limits_Q \big(y_{xx} - y_t\big)\, \psi\, dt\, dx}_{=0}\,. \qquad (45)$$

В двойном интеграле производим интегрирование по частям, предполагая

функцию $\psi(t,x)$ гладкой:

$$\iint\limits_Q \bigl(y_{xx}-y_t\bigr)\,\psi\,dt\,dx = \int_0^T y_x\,\psi\,\Big|_{x=0}^{x=l} dt - \iint\limits_Q y_x\,\psi_x\,dt\,dx -$$

$$- \int_0^l y\,\psi\,\Big|_{t=0}^{t=T} dx + \iint\limits_Q y\,\psi_t\,dt\,dx =$$

$$= \bigl[\,y(0,x)=u(x);\ \text{требуем, чтобы } \psi(t,0)=\psi(t,l)=0\bigr] =$$

$$= -\int_0^T y\,\psi_x\,\Big|_{x=0}^{x=l} dt + \iint\limits_Q y\,\bigl(\psi_{xx}+\psi_t\bigr)\,dt\,dx -$$

$$- \int_0^l y(T,x)\,\psi(T,x)\,dx + \int_0^l u(x)\,\psi(0,x)\,dx =$$

$$= \bigl[y(t,0)=y(t,l)=0;\ \text{требуем, чтобы } \psi_t+\psi_{xx}=0 \text{ на } Q\bigr] =$$

$$- \int_0^l y(T,x)\,\psi(T,x)\,dx + \int_0^l u(x)\,\psi(0,x)\,dx\,. \quad (46)$$

В (46) требуемый для описания действия A_T^* вид имеет только последний интеграл. Чтобы избавиться от оставшейся пары «мешающих слагаемых», потребуем дополнительно, чтобы $\psi(T,x)=v(x)$. В результате получаем, что сопряжённый оператор A_T^* каждой функции $v=v(x)\in F=L^2(0,l)$ ставит в соответствие след

$$A_T^* v = \psi(0,x), \quad x\in(0,l), \quad (47)$$

при $t=0$ решения $\psi=\psi(t,x)=\psi(t,x;v)$ следующей начально-краевой задачи с обратным течением времени, которую называют *сопряжённой* по отношению к (39):

$$\begin{gathered}\psi_t+\psi_{xx}=0, \quad 0<t<T, \quad 0<x<l,\\ \psi\,\big|_{x=0}=0, \quad \psi\,\big|_{x=l}=0, \quad 0<t<T,\\ \psi\,\big|_{t=T}=v(x), \quad 0<x<l.\end{gathered} \quad (48)$$

В результате формулу градиента *терминального* квадратичного функционала $J_T(u)$ можно будет записать в виде

$$J_T'(u)=2A_T^*(A_Tu-f)=2\,\psi(0,x;v)\,\big|_{v=y(T,x;u)-f(x)}\,, \quad \forall\,u\in L^2(0,l). \quad (49)$$

Внешне от неё почти не отличается и формула градиента *интегрального* квадратичного функционала $J_I(u)$:

$$J_I'(u)=2A_I^*(A_Iu-f)=2\,\psi(0,x;v)\,\big|_{v=y(t,x;u)-f(t,x)}\,, \quad \forall\,u\in L^2(0,l), \quad (50)$$

в которой оператор A_I^* действует по тому же правилу (47), а функция $\psi = \psi(t,x) = \psi(t,x;v)$ будет решением сопряжённой задачи, похожей на (48):

$$\begin{gathered} \psi_t + \psi_{xx} = -v(t,x), \quad 0 < t < T, \quad 0 < x < l, \\ \psi\,\big|_{x=0} = 0, \quad \psi\,\big|_{x=l} = 0, \quad 0 < t < T, \\ \psi\,\big|_{t=T} = 0, \quad 0 < x < l. \end{gathered} \tag{51}$$

Теорема 5. *Пусть допустимое множество управлений U слабо компактно в пространстве $L^2(0,l)$. Тогда для любых целевых элементов $f(t,x) \in L^2(Q)$ в (41) и $f(x) \in L^2(0,l)$ в (42) каждая из оптимизационных задач* (39),(40),(41) *и* (39),(40),(42) *имеет решение, т. е. $J_* > -\infty$ и $U_* \neq \varnothing$. Каждый из двух квадратичных функционалов* (41) *и* (42) *бесконечно дифференцируем по Фрешé на всём пространстве $L^2(0,l)$ и их первые производные имеют вид* (50), (52) *и* (49), (48) *соответственно.*

Упражнение 11. *В пространстве $L^2(0,l)$ найдите первые производные двух квадратичных функционалов*

$$J_I(u) = \iint\limits_Q \big|y(t,x;u) - f(t,x)\big|^2\,dt\,dx \quad \text{и} \quad J_T(u) = \int_0^l \big|y(T,x;u) - f(x)\big|^2\,dx,$$

заданных на решениях $y = y(t,x;u)$ третьей краевой задачи для уравнения теплопроводности с заданными граничными коэффициентами $\sigma_0 > 0,\ \sigma_1 > 0$:

$$\begin{gathered} y_t = y_{xx}, \quad (t,x) \in Q = (0,T) \times (0,l), \\ -y_x + \sigma_0 y\,|_{x=0} = 0, \quad y_x + \sigma_1 y\,|_{x=0} = 0, \quad 0 < t < T, \\ y\,|_{t=0} = u(x), \quad 0 < x < l. \end{gathered}$$

Задачи управления линейной дискретной системой

Рассмотрим дискретный аналог непрерывного управяемого процесса (27):

$$\frac{x_{i+1}-x_i}{h} = D_i x_i + B_i u_i, \quad i = 0,1,\ldots,N-1; \quad x_0 = 0, \tag{52}$$

где $h = T/N$ — шаг сетки на промежутке $[0,T]$ с узлами $t_i = i \cdot h$, $i = 0,1,\ldots,N$, $u = (u_0, u_1, \ldots, u_{N-1})$, $u_i \in R^r$, — дискретные управления, а $x = x(u) = (x_0, x_1, \ldots, x_N)$, $x_i \in R^n$, — соответствующие им дискретные траектории. Заметим, что привязка дискретного процесса (52) к его непрерывному аналогу (27) совершенно необязательна (при этом вполне можно быо бы обойтись без h), но, на мой взгляд, это целесообразно с точки зрения удобства сопоставления непрерывных и дискретных результатов и выводов. По той же причине в конечномерных пространствах сеточных управлений $R^{N\times r}$ и сеточных траекторий $R^{N\times n}$ введём скалярные произведения, устроенные по образцу и подобию принятых для пространства Лебега $L^2(0,T)$ интегральных конструкций, а сами пространства дискретных управлений и траекторий будем обозначать через L^2_h :

$$\langle u, v\rangle_{L^2_h} = \sum_{i=0}^{N-1} \langle u_i, v_i\rangle_{R^r} h, \qquad \langle x, y\rangle_{L^2_h} = \sum_{i=1}^{N} \langle x_i, y_i\rangle_{R^n} h\,.$$

Управления выбираются из допустимого множества

$$u \in U \subset L^2_h, \tag{53}$$

а минимизации подлежит, как и в предыдущих примерах, либо *«интегральный»* квадратичный функционал

$$J_I(u) = \sum_{i=1}^{N} \big|x_i(u) - y_i\big|^2_{R^n} h = \|x(u) - y\|^2_{L^2_h}\,, \tag{54}$$

либо *терминальный* квадратичный функционал

$$J_T(u) = \big|x_N(u) - y\big|^2_{R^n}\,. \tag{55}$$

В (54) задана целевая дискретная траектория $y = (y_1, y_2, \ldots, y_N) \in L^2_h$, а в (55) — целевая финальная позиция $y \in R^n$. Оба функционала (54) и (55)

записываются в стандартной квадратичной форме $J(u) = \|Au - f\|_F^2$, если ввести операторы

$$Au = A_I u = x(u), \quad A_I : H = L_h^2 \to F = F_I = L_h^2, \quad f = y \in F_I,$$
$$Au = A_T u = x_N(u), \quad A_T : H \to F = F_T = R^n, \quad f = y \in F_T.$$

Эти операторы *линейны*, а пространства, в которых они действуют, *конечномерны*, поэтому они *ограничены* (непрерывны) и никакого отдельного доказательства этих свойств не требуется.

Формула для градиента $J'(u) = 2A^*(Au - f)$ нам хорошо известна и в ней следует лишь расшифровать правило действия сопряжённого оператора A^*. Подробности вывода изложим лишь для *терминального* оператора $A_T : L_h^2 \to R^n$, а для *«интегрального»* оператора $A_I : L_h^2 \to L_h^2$ приведём лишь окончательный результат. Как обычно, отправляемся от определения:

$$\langle A_T u,\, v\rangle_{R^n} = \langle u,\, A_T^* v\rangle_{L_h^2} \quad \forall\, u \in L_h^2\ \forall v \in R^n.$$

Левая часть этого тождества нам известна:

$$\langle A_T u,\, v\rangle_{R^n} = \langle x_N(u),\, v\rangle_{R^n}$$

и мы должны представить её в виде, который предписывается правой частью:

$$\langle u,\, A_T^* v\rangle_H = \sum_{i=0}^{N-1} \langle u,\, A_T^* v\rangle_{R^r}\, h,$$

подобрав подходящую сеточную функцию на место A^*v. Схема действий типичная:

$$\langle A_T u,\, v\rangle_{R^n} = \langle x_N(u),\, v\rangle_{R^n} + \underbrace{\sum_{i=0}^{N-1} \langle D_i x_i + B_i u_i - \frac{x_{i+1} - x_i}{h},\, \psi_i\rangle_{R^n}\, h}_{=0}. \qquad (56)$$

Здесь $\psi = \big(\psi_0, \psi_1, \ldots, \psi_N\big)$ — произвольная (пока) сеточная функция со значениями $\psi_i \in R^n$. Далее, преобразуем по отдельности суммы, из которых формируется нулевая добавка:

$$\sum_{i=0}^{N-1} \langle D_i x_i,\, \psi_i\rangle_{R^n}\, h = \sum_{i=0}^{N-1} \langle x_i,\, D_i^\top \psi_i\rangle_{R^n}\, h \overset{(x_0=0)}{=} \sum_{i=1}^{N-1} \langle x_i,\, D_i^\top \psi_i\rangle_{R^n}\, h =$$

$$= \big[\, \pm \text{слаг. с номером } i = N\big] = \sum_{i=1}^{N} \langle x_i,\, D_i^\top \psi_i\rangle_{R^n}\, h - \langle x_N,\, D_N^\top \psi_N\rangle_{R^n}\, h\,; \quad (57)$$

$$\sum_{i=0}^{N-1}\langle B_i u_i,\, \psi_i\rangle_{R^n} h = \sum_{i=0}^{N-1}\langle u_i,\, B_i^\top \psi_i\rangle_{R^r} h\,; \tag{58}$$

$$\begin{aligned}\sum_{i=0}^{N-1}\Big\langle -\frac{x_{i+1}-x_i}{h},\, \psi_i\Big\rangle_{R^n} h &= -\sum_{i=0}^{N-1}\langle x_{i+1},\, \psi_i\rangle_{R^n} + \sum_{i=0}^{N-1}\langle x_i,\, \psi_i\rangle_{R^n} = \\ &= \big[\text{сдвиг нумерации, } x_0 = 0,\ \pm\langle x_N,\, \psi_N\rangle_{R^n}\big] = \\ &= -\sum_{i=1}^{N}\langle x_i,\, \psi_{i-1}\rangle_{R^n} + \sum_{i=1}^{N}\langle x_i,\, \psi_i\rangle_{R^n} - \langle x_N,\, \psi_N\rangle_{R^n} = \\ &= \sum_{i=1}^{N}\Big\langle x_i,\, \frac{\psi_i-\psi_{i-1}}{h}\Big\rangle_{R^n} h - \langle x_N,\, \psi_N\rangle_{R^n}\,. \end{aligned} \tag{59}$$

Из всех слагаемых, появившихся в правой части (56) после выполненных преобразований (57) – (59), нас устраивает лишь одно, находящееся в правой части (58), а от всех остальных мы сможем избавиться, если выберем специальную сеточную функцию ψ из условий

$$\frac{\psi_i-\psi_{i-1}}{h} = -D_i^\top \psi_i, \quad i = 1, 2, \ldots, N; \quad \big(I + h\, D_N^\top\big)\psi_N = v\,. \tag{60}$$

При этом равенство (56) примет вид

$$\langle A_T u,\, v\rangle_{R^n} = \sum_{i=0}^{N-1}\langle u_i,\, B_i^\top \psi_i\rangle_{R^r} h\,,$$

означающий, что

$$\begin{aligned} A_T^* v = \big(B_0^\top\psi_0, B_1^\top\psi_1, \ldots, B_{N-1}^\top\psi_{N-1}\big) \in L_h^2 \quad \forall\, v \in R^n\,, \\ J_T'(u) = 2\,\big(B_0^\top\psi_0, B_1^\top\psi_1, \ldots, B_{N-1}^\top\psi_{N-1}\big)\,\Big|_{\,v=x_N(u)-f}\,. \end{aligned} \tag{61}$$

Действие оператора A_I^*, сопряжённого к *«интегральному»* оператору A_I, внешне почти не отличается от (61), только на вход ему подаётся не вектор $v \in R^n$, а сеточная функция $v = (v_1, v_2, \ldots, v_N) \in L_h^2$ и функция ψ будет решением другой сеточной задачи, похожей на (60):

$$\begin{aligned} A_I^* v = \big(B_0^\top\psi_0, B_1^\top\psi_1, \ldots, B_{N-1}^\top\psi_{N-1}\big) \in L_h^2 \quad \forall\, v \in L_h^2\,, \\ J_I'(u) = 2\,\big(B_0^\top\psi_0, B_1^\top\psi_1, \ldots, B_{N-1}^\top\psi_{N-1}\big)\,\Big|_{\,v=x(u)-f}\,, \end{aligned} \tag{62}$$

$$\frac{\psi_i-\psi_{i-1}}{h} = -D_i^\top \psi_i - v_i, \quad i = 1, 2, \ldots, N; \quad \psi_N = 0\,. \tag{63}$$

Подытожим результаты нашей работы с дискретной управляемой системой.

Теорема 6. *Для любой целевой траектории* $y \in L_h^2$ *из* (54) *и любой целевой позиции* $y \in R^n$ *из* (55) *каждый из двух квадратичных функционалов* (54) *и* (55) *бесконечно дифференцируем по Фрешé на всём пространстве* L_h^2 *и их первые производные имеют вид* (62), (63) *и* (61), (60) *соответственно. Если допустимое множество* U *из* (53) *замкнуто и ограничено в пространстве* L_h^2, *то обе задачи минимизации* (52), (53), (54) *и* (52), (53), (55) *имеют оптимальные решения, т. е.* $J_* > -\infty$ *и* $U_* \neq \varnothing$.

При сопоставлении результатов, полученных для систем с дискретной и непрерывной динамикой, обнаруживается их несомненное сходство (подобие).

Упражнение 12. *В пространстве* L_h^2 *сеточных функций* $u = (u_1, u_2, \ldots, u_{N-1})$, $u_i \in R^1$, *со скалярным произведением* $\langle u, v \rangle_{L_h^2} = \sum\limits_{i=1}^{N-1} u_i v_i \, h$, $h > 0$, *найдите первую производную* $J'(u)$ *квадратичного функционала*

$$J(u) = \sum_{i=1}^{N-1} |y_i(u) - z_i|^2 h,$$

в котором $z = (z_1, z_2, \ldots, z_{N-1}) \in L_h^2$ *— заданная сеточная функция, а* $y = y(u) = (y_0, y_1, \ldots, y_N)$ *— решение разностной краевой задачи*

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} - y_i = -u_i, \quad i = 1, 2, \ldots, N-1,$$
$$y_0 = 0, \qquad y_N = 0.$$

IV. ЭЛЕМЕНТЫ ВЫПУКЛОГО АНАЛИЗА

Дадим определение, позволяющее ранжировать выпуклые функции, выделяя из них *«более выпуклые»* и *«менее выпуклые»*.

Определение 14. *Пусть H — гильбертово пространство и множество $U \subset H$ выпукло. Функция $J(u) : U \to R^1$ называется*

- ***выпуклой*** *на U, если*

$$\forall\, u,\, v \in U \quad \forall\, \alpha \in [0,1] \quad J\big(\alpha u + (1-\alpha)v\big) \leqslant \alpha J(u) + (1-\alpha)J(v); \quad (64)$$

- ***строго выпуклой*** *на U, если она выпукла и при $u \neq v$ неравенство* (64) *является* ***строгим*** $\forall\, \alpha \in (0,1)$;

- ***сильно выпуклой*** *на U с коэффициентом $\varkappa > 0$, если*

$$\forall\, u, v \in U\ \forall\, \alpha \in [0,1] \quad J\big(\alpha u + (1-\alpha)v\big) \leqslant$$
$$\leqslant \alpha J(u) + (1-\alpha)J(v) - \frac{\varkappa}{2}\,\alpha(1-\alpha)\,\|u-v\|_H^2. \quad (65)$$

Простейшим примером функции, являющейся *выпуклой*, но *не строго выпуклой* на любом выпуклом множестве U в гильбертовом пространстве H, служит любая линейная функция $J(u) = \langle c, u\rangle$, а также норма $J(u) = \|u\|$. Квадратичная функция вида $J(u) = \|u\|^2$ будет *сильно выпуклой* на всём пространстве H (проверьте, что её коэффициент сильной выпуклости равен $\varkappa = 2$), а функция одной переменной $f(x) = e^x$ *строго выпукла* на всей числовой прямой, но при этом *не является* на R^1 *сильно выпуклой* ни при каких $\varkappa > 0$, хотя на любом отделённом от $-\infty$ множестве $[a, +\infty)$ она будет сильно выуклой с коэффициентом $\varkappa = \varkappa(a) > 0$, зависящим от a, $\varkappa(a) \to 0$ при $a \to -\infty$.

Простые и вместе с тем важные свойства задач минимизации *выпуклых* функций содержит следующая

Теорема 7. (о локальном минимуме выпуклой функции) *Пусть в задаче минимизации*

$$J(u) \to \inf, \quad u \in U \subset X,$$

X — линейное нормированное пространство, U — выпуклое множество и $J(u)$ — выпуклая на множестве U функция. Тогда

1) ***любая*** *точка* ***локального*** *минимума функции $J(u)$ на множестве U является точкой её* ***глобального*** *минимума;*

2) *если множество U_* оптимальных решений непусто, то оно выпукло;*

3) *если множество U_* оптимальных решений непусто, а функция $J(u)$* ***строго выпукла*** *на U, то оптимальное решение u_** ***единственно****.*

Доказательство. *1)* Напомним, что точку $u_* \in U$ называют точкой *локального минимума* функции $J(u)$ на множестве U, если

$$\exists\, \varepsilon > 0 : \ \forall u \in U \cap \big\{\|u - u_*\| < \varepsilon\big\} \quad J(u) \geqslant J(u_*).$$

Покажем, что для *выпуклой* функции $J(u)$ неравенство $J(u) \geqslant J(u_*)$ будет выполняться *глобально*, т. е. на всём множестве U, а не только вблизи точки u_*. Фиксируем произвольную точку $u \in U$ и рассмотрим точки прямолинейного отрезка, соединяющего u с u_* :

$$u_\alpha = u_* + \alpha\,(u - u_*) = \alpha u + (1 - \alpha)u_*, \quad \alpha \in [0, 1].$$

В силу выпуклости множества U имеем $u_\alpha \in U \ \forall\, \alpha \in [0, 1]$, а при всех достаточно малых α, именно, для $0 \leqslant \alpha < \frac{\varepsilon}{\|u - u_*\|}$, точки u_α окажутся внутри ε-окрестности точки u_*, следовательно, для этих достаточно малых α

$$J(u_*) \leqslant J(u_\alpha).$$

Пользуясь определением выпуклости функции $J(u)$, оценим сверху правую часть данного неравенства:

$$J(u_*) \leqslant J(u_\alpha) \leqslant \alpha J(u) + (1 - \alpha)J(u_*) \quad \Longrightarrow \quad \alpha J(u_*) \leqslant \alpha J(u).$$

После деления обеих частей на $\alpha > 0$, получим искомое *глобальное* неравенство $J(u_*) \leqslant J(u) \ \forall\, u \in U$. Утверждения *2)* и *3)* предлагается доказать самостоятельно.

Сформулируем заключительный вариант из серии обобщённых теорем Вейерштрасса, содержащий *усиленные* требования к функции и *ослабленные* требования к допустимому множеству.

Теорема 8. (`«сильно выпуклый» вариант теоремы Вейерштрасса`)
Пусть в задаче минимизации

$$J(u) \to \inf, \quad u \in U \subset H,$$

H — гильбертово пространство, U — выпуклое замкнутое множество и $J(u)$ — п/н снизу и ***сильно выпуклая*** *на множестве U функция с коэффициентом $\varkappa > 0$. Тогда*

1) $J_* > -\infty$, а множество оптимальных решений U_* непусто и состоит из ***единственной*** *точки: $U_* = \{u_*\}$;*

2) справедлива оценка

$$\frac{\varkappa}{2}\,\|u - u_*\|^2 \leqslant J(u) - J(u_*) \quad \forall\, u \in U, \tag{66}$$

обеспечивающая ***сильную*** *сходимость всех минимизирующих последовательностей.*

Доказательство. *1)* Фиксируем произвольную точку $u_0 \in U$ и рассмотрим соответствующее ей множество (Лебега)

$$M_0 = \big\{\, u \in U \;\big|\; J(u) \leqslant J(u_0) \big\}. \tag{67}$$

По определению именно во множестве M_0 содержатся оптимальные решения поставленной задачи минимизации, т. е.

$$J_* = \inf_{u \in M_0} J(u), \quad U_* = \big\{\, u \in M_0 \;\big|\; J(u) = J_* \big\},$$

и поэтому доказательство существования оптимальных решений мы можем проводить не на всём допустимом множестве U, а на выделенной его части M_0. В соответствии с теоремой 2 («слабый» вариант теоремы Вейерштрасса) для этого достаточно, чтобы множество M_0 было слабо компактным в H, а функция $J(u)$ была слабо п/н снизу на M_0. Свойство слабой п/н снизу следует из выпуклости и п/н снизу функции $J(u)$. Для слабой компактности множества M_0 достаточно, чтобы оно было выпуклым, замкнутым и ограниченным в H. *Выпуклость* M_0 следует из выпуклости множества U и выпуклости функции $J(u)$. *Замкнутость* M_0 следует из замкнутости множества U и п/н снизу функции $J(u)$.

На проверке свойства *ограниченности* множества M_0 остановимся подробнее. Именно здесь нам понадобится затребованное свойство *сильной выпуклости* функции $J(u)$. Представим M_0 в виде объединения двух непересекающихся множеств:

$$M_0 = M_1 \cup M_2, \quad M_1 = M_0 \cap \big\{\, \|u - u_0\| \leqslant 2\big\}, \quad M_2 = M_0 \cap \big\{\, \|u - u_0\| > 2\big\}.$$

Понятно, что множество M_1 ограничено. Ограниченность множества M_2 установим по определению, указав конкретный шар с центром в точке u_0, содержащий множество M_2. Фиксируем произвольную точку $u \in M_2$ и возьмём соответствующее ей конкретное значение $\alpha = \frac{1}{\|u-u_0\|}$, для которого будут выполняться два условия:

$$0 < \alpha < \frac{1}{2}, \quad \frac{1}{2} < 1 - \alpha < 1. \tag{68}$$

Составим с участием данного α выпуклую линейную комбинацию точек u и u_0 :

$$u_\alpha = u_0 + \alpha\,(u - u_0) \in M_1$$

и запишем для неё определение сильной выпуклости (65):

$$J\big(u_\alpha\big) \leqslant \alpha J(u) + (1-\alpha)J(u_0) - \frac{\varkappa}{2}\,\alpha(1-\alpha)\,\|u - u_0\|^2 \Longrightarrow$$
$$\Longrightarrow \quad \frac{\varkappa}{2}\,\alpha(1-\alpha)\,\|u - u_0\|^2 \leqslant \alpha J(u) + (1-\alpha)J(u_0) - J\big(u_\alpha\big)\,.$$

В левую часть подставим $\alpha = \frac{1}{\|u-u_0\|}$ и учтём, что $1 - \alpha > \frac{1}{2}$. Правую часть можно будет оценить сверху разностью $J(u_0) - J\big(u_\alpha\big)$, т. к. $u \in M_2 \subset M_0$ и $J(u) \leqslant J(u_0)$. В результате получаем оценку

$$\|u - u_0\| \leqslant \frac{4}{\varkappa}\,\big(J(u_0) - J(u_\alpha)\big).$$

Присутствующий в её правой части элемент u_α находится в пределах *ограниченного*, выпуклого и замкнутого, т. е. *слабо компактного* множества M_1, на котором по теореме 2 нижняя грань функции $J(u)$ конечна:

$$J(u_\alpha) \geqslant J_{1*} = \inf_{u \in M_1} J(u) > -\infty,$$

а тогда мы можем предъявить искомый шар, содержащий множество M_2 :

$$\|u - u_0\| \leqslant R = \frac{4}{\varkappa}\,\big(J(u_0) - J_{1*}\big).$$

Тем самым, существование оптимального решения доказано, а его единственность следует их сильной выпуклости функции $J(u)$.

2) Для доказательства оценки (66) фиксируем произвольный допустимый элемент $u \in U$ и рассмотрим отрезок, соединяющий u с единственным оптимальным решением u_* : $u_\alpha = \alpha u + (1-\alpha)u_*,\ \alpha \in [0,1]$. По определению сильной выпуклости имеем

$$
\begin{aligned}
J\big(u_\alpha\big) &\leqslant \alpha J(u) + (1-\alpha)J(u_*) - \frac{\varkappa}{2}\,\alpha(1-\alpha)\,\|u-u_*\|^2 \Longrightarrow \\
&\Longrightarrow \quad \frac{\varkappa}{2}\,\alpha(1-\alpha)\,\|u-u_*\|^2 \leqslant \alpha J(u) + (1-\alpha)J(u_*) - J\big(u_\alpha\big) = \\
&\qquad = \alpha\big(J(u) - J\big(u_*\big)\big) + \underbrace{\big(J(u_*) - J\big(\overbrace{u_\alpha}^{\in U}\big)\big)}_{\leqslant 0} \leqslant \alpha\big(J(u) - J\big(u_*\big)\big)\,.
\end{aligned}
$$

После деления обеих частей полученного неравенства на $\alpha > 0$ устремляем $\alpha \to 0$ и приходим к (66). Терема 8 доказана.

Элементы выпуклого анализа (продолжение)

В задачах *минимизации* свойства *выпуклости* имеют немалую ценность, поэтому полезно иметь для их проверки какие-то другие инструменты помимо определения. В случае *дифференцируемых* (по Фрешé) функций такие инструменты в виде критериев представлены в следующих двух теоремах.

Теорема 9. (критерии выпуклости)
Пусть H — гильбертово пространство, U — выпуклое множество из H и $J(u) \in C^1(U)$. Тогда эквивалентны следующие три утверждения (a), (b), (c) :

(a) *функция $J(u)$ выпукла на множестве U*;

(b) *справедлива оценка*

$$J(u) \geqslant J(v) + \langle J'(v),\, u - v\rangle \quad \forall\, u, v \in U;$$

(c) *справедлива оценка*

$$\langle J'(u) - J'(v),\, u - v\rangle \geqslant 0 \quad \forall\, u, v \in U.$$

Если, кроме того, функция $J(u) \in C^2(U)$, а $\operatorname{int} U \neq \varnothing$, то каждому из условий (a), (b), (c) *будет эквивалентно условие*

(d) $$\langle J''(u)h,\, h\rangle \geqslant 0 \quad \forall\, u \in U\ \forall h \in H.$$

Замечание 7. *Геометрический смысл условия (b) заключается в том, что график гладкой выпуклой функции лежит выше всех без исключения проведённых к нему касательных. Свойство (c) принято называть свойством* ***монотонности градиента*** *выпуклой функции (в одномерном случае это свойство неубывания первой производной $f'(x)$). Условие (d) означает, что вторая производная выпуклой функции является (симметричным) неотрицательно определённым оператором $J''(u) \in L(H \to H)$ (в одномерном случае это свойство неотрицательности второй производной $f''(x) \geqslant 0$).*

Доказательство теоремы 9.
(a) $\Longrightarrow$ (b): Запишем неравенство из определения выпуклой функции:

$$J(\alpha u + (1-\alpha)v) \leqslant \alpha J(u) + (1-\alpha)J(v) \quad \Longrightarrow$$
$$\Longrightarrow \quad \alpha J(u) \geqslant \alpha J(v) + \big(J(\alpha u + (1-\alpha)v) - J(v)\big) \overset{\text{конеч. приращ.}}{=}$$
$$\overset{\exists\,\theta\in[0,1]}{=} \alpha J(v) + \alpha \left\langle J'(v + \theta\alpha(u-v)),\, u - v\right\rangle.$$

Отсюда после деления обеих частей на $\alpha > 0$, а затем предельного перехода при $\alpha \to 0$, с учётом непрерывности градиента $J'(u) \in C(U)$, получаем (b).

(b) $\Longrightarrow$ (c): Здесь для доказательства достаточно сложить два неравенства, отражающие свойство (b), в которых точки u и v меняются местами.

(c) $\Longrightarrow$ (a): Фиксируем произвольные $u, v \in U$, $\alpha \in [0,1]$ и покажем, что для них выполняется условие выпуклости (a), записанное в виде неравенства

$$\alpha J(u) + (1-\alpha)J(v) - J(\alpha u + (1-\alpha)v) \geqslant 0\,. \tag{69}$$

Введём обозначение

$$w = \alpha u + (1-\alpha)v$$

и заметим, что

$$u - w = (1-\alpha)(u-v), \quad v - w = \alpha(v-u)\,. \tag{70}$$

Преобразуем левую часть неравенства (69):

$$\begin{aligned}
&\alpha J(u) + (1-\alpha)J(v) - J(\alpha u + (1-\alpha)v) = \\
&\qquad = \alpha\big(J(u) - J(w)\big) + (1-\alpha)\big(J(v) - J(w)\big) \overset{\text{кон. прир.}}{=} \\
&= \alpha \int_0^1 \langle J'(w + t(u-w)),\, u-w\rangle\, dt + (1-\alpha)\int_0^1 \langle J'(w+t(v-w)),\, v-w\rangle\, dt = \\
&\qquad \overset{(70)}{=} \alpha(1-\alpha)\int_0^1 \langle J'(w+t(u-w)) - J'(w+t(v-w)),\, u-v\rangle\, dt \geqslant \\
&\geqslant \Big[(w+t(u-w)) - (w+t(v-w)) = t(u-v), \quad u-v = \frac{t(u-v)}{t}\Big] \geqslant \\
&\qquad \overset{\text{(c)}}{\geqslant} \alpha(1-\alpha)\int_0^1 \frac{1}{t}\cdot 0\, dt \geqslant 0\,.
\end{aligned}$$

Тем самым, эквивалентность всех трёх условий (a), (b) и (c) для функций класса $C^1(U)$ доказана.

Теперь для более гладких функций из $C^2(U)$ установим равносильность условий (c) и (d).

(c) $\Longrightarrow$ (d): Как раз здесь будет использовано предположение о непустоте внутренности множества U. Фиксируем произвольную точку $u \in \operatorname{int} U$ и произвольный элемент $h \in H$. Так как точка u — внутренняя для множества U, то

$$\exists\, \varepsilon_0 > 0 : \quad \forall\, \varepsilon \in [0, \varepsilon_0] \quad u + \varepsilon h \in U,$$

и для пары u, $u+\varepsilon h$ по условию (c) будет выполняться неравенство

$$\langle J'(u+\varepsilon h) - J'(u),\, \varepsilon h\rangle \geqslant 0 \quad \forall\, \varepsilon \in [0, \varepsilon_0].$$

Его левая часть по формуле конечных приращений (15) записывается в виде

$$\int_0^1 \langle J''(u+t\varepsilon h)\varepsilon h,\, \varepsilon h\rangle\, dt \overset{\exists\,\theta\in[0,1]}{=} \varepsilon^2 \cdot \langle J''(u+\theta\varepsilon h)h,\, h\rangle \geqslant 0 \quad \forall\, \varepsilon \in [0, \varepsilon_0].$$

После деления обеих частей неравенства на $\varepsilon^2 > 0$ устремляем $\varepsilon \to 0$ и, с учётом ограниченности значений θ и непрерывности второй производной $J''(u)$, получаем искомое условие (d), в котором $u \in \operatorname{int} U$, $h \in H$. Для перехода к *произвольным* $u \in U$ воспользуемся следующим специфическим свойством *выпуклых* множеств *с непустой внутренностью* [4, гл. 2, §2.6]:

$$\overline{\operatorname{int} U} = \overline{U} \quad \big(\overline{M} - \text{замыкание множества } M\big).$$

Это означает, что для любой точки $u \in \overline{U}$ и, в частности, $\forall\, u \in U$ существует последовательность $u_n \in \operatorname{int} U,\ n = 1, 2, \ldots$, для которой выполнены и условие (d) и условие аппроксимации $\|u_n - u\| \to 0$ при $n \to \infty$, что позволяет по непрерывности распространить действие условия (d) на любые $u \in U$.

(d) $\Longrightarrow$ (c): По формуле конечных приращений представим левую часть (c) в виде

$$\langle J'(u) - J'(v),\, u - v\rangle \overset{\exists\,\theta\in[0,1]}{=} \langle J''\big(\underbrace{v + \theta(u-v)}_{\in U}\big)(\underbrace{u-v}_{=h}),\, \underbrace{u-v}_{=h}\rangle \overset{\text{(d)}}{\geqslant} 0.$$

Теорема 9 доказана.

В следующей теореме приведены *критерии сильной выпуклости* для гладких функций. Её доказательство проводится по той же схеме, что и в теореме 9, поэтому ограничимся формулировкой.

Теорема 10. (критерии сильной выпуклости)
Пусть H — гильбертово пространство, U — выпуклое множество из H и $J(u) \in C^1(U)$. Тогда эквивалентны следующие три утверждения $(\mathrm{a}'), (\mathrm{b}'), (\mathrm{c}')$:

(a') *функция $J(u)$ сильно выпукла на множестве U с коэффициентом $\varkappa > 0$;*

(b') *справедлива оценка*

$$J(u) \geqslant J(v) + \langle J'(v),\, u - v\rangle + \frac{\varkappa}{2}\,\|u - v\|^2 \quad \forall\, u, v \in U;$$

(c') *справедлива оценка*

$$\langle J'(u) - J'(v),\, u - v\rangle \geqslant \varkappa\,\|u - v\|^2 \quad \forall\, u, v \in U.$$

Если, кроме того, функция $J(u) \in C^2(U)$, а $\operatorname{int} U \neq \varnothing$, то каждому из условий $(\mathrm{a}'), (\mathrm{b}'), (\mathrm{c}')$ *будет эквивалентно условие*

(d') $$\langle J''(u)h,\, h\rangle \geqslant \varkappa\,\|h\|^2 \quad \forall\, u \in U\ \forall h \in H.$$

Замечание 8. *Условие непустоты внутренности* $\operatorname{int} U \neq \varnothing$, *относящееся к свойствам* (d) *и* (d′), *является важным, но в то же время и не абсолютно необходимым. Отказ от этого требования возможен, но должен сопровождаться ограничением области вариации переменной h квадратичной формы $\langle J''(u)h, h\rangle$: условие $h \in H$ заменяется условием $h \in H_U$, где H_U — подпространство в H минимальной размерности, содержащее множество U (подпространство, «натянутое» на U). Для иллюстрации сказанного предлагается рассмотреть двумерный пример*

$$H = R^2, \quad u = (x, y), \quad J(u) = x^2 - y^2, \quad U = \{y = 0\}, \quad \operatorname{int} U = \varnothing.$$

Вернёмся к задачам минимизации

$$J(u) \to \inf, \quad u \in U \subset H,$$

и в случае *гладких* функций $J(u)$ и *выпуклых* множеств U приведём для них условия *первого порядка* (содержащие первые производные $J'(u)$), которые необходимы и достаточны для оптимальности элемента u_*.

Теорема 11. (условия оптимальности)
Пусть H — гильбертово пространство, U — выпуклое множество из H и $J(u) \in C^1(U)$. Тогда

1) если $u_* \in U_*$ — оптимальное решение, то выполняется вариационное неравенство

$$\langle J'(u_*),\, u - u_* \rangle \geqslant 0 \quad \forall\, u \in U; \tag{71}$$

2) если $\operatorname{int} U \neq \varnothing$ и $u_* \in U_* \cap \operatorname{int} U$, то

$$J'(u_*) = 0; \tag{72}$$

3) если выполняется вариационное неравенство (71) и функция $J(u)$ выпукла на множестве U, то $u_* \in U_*$ — оптимальное решение задачи минимизации.

Доказательство. *1)* Для оптимального решения u_* и любых $u \in U$, $\alpha \in [0,1]$ записываем соотношения

$$0 \leqslant J(\underbrace{\alpha u + (1-\alpha)u_*}_{\in U}) - J(\underbrace{u_*}_{\in U_*}) = \big[\text{конеч. приращ.}, \exists\, \theta \in [0,1]\big] =$$
$$= \alpha \left\langle J'\big(u_* + \theta\alpha(u - u_*)\big),\, u - u_* \right\rangle \quad \forall \alpha \in [0,1]\,.$$

После деления на $\alpha > 0$ устремляем $\alpha \to 0$ и, учитывая непрерывность первой производной $J'(u)$, получаем (71).

2) Если $u_* \in U_* \cap \operatorname{int} U$, то все достаточно малые смещения из точки u_* по любым направлениям не будут выводить за пределы допустимого множества U, а тогда при достаточно малых $\varepsilon > 0$ можно выбрать для подстановки в вариационное неравенство (71) точки вида

$$u := u_* - \varepsilon\, J'(u_*)\,,$$

что приведёт к неравенству

$$-\varepsilon \langle J'(u_*),\, J'(u_*) \rangle \geqslant 0 \quad \overset{\varepsilon > 0}{\Longrightarrow} \quad \|J'(u_*)\|^2 \leqslant 0 \quad \Longrightarrow \quad J'(u_*) = 0\,.$$

3) Пусть выполнено вариационное неравенство (71) и функция $J(u)$ выпукла. Тогда оптимальность элемента u_* устанавливается следующим образом:

$$\forall\, u \in U \quad J(u) - J(u_*) \overset{\text{теор.9, (b)}}{\geqslant} \langle J'(u_*),\, u - u_* \rangle \overset{(71)}{\geqslant} 0.$$

Теорема 11 доказана.

Замечание 9. *Условие оптимальности в форме вариационного неравенства* (71) *является естественным обобщением условия Ферма́* (72) *на случай задач минимизации с ограничениями — задач* ***условной*** *минимизации. Каждое из условий* (71) *и* (72) *для выпуклых функций является критерием оптимальности для задач условной и безусловной минимизации соответственно.*

Примеры применения теорем 9 и 10

1. *Линейный* функционал $J(u) = \langle c, u \rangle$ является *выпуклым*, но *не строго* и, тем более, *не сильно*, поскольку $J''(u) = 0 \in L(H \to H)$, неравенство (d) выполняется тривиально на всём пространстве $h \in H$ в виде тождества, а неравенство (d′) не будет при этом выполняться ни при каких $\varkappa > 0$.

2. *Квадратичный* функционал $J(u) = \|Au - f\|_F^2,\ A \in L(H \to F),\ f \in F$, является *выпуклым*, поскольку $J''(u) = 2A^*A \in L(H \to H)$ и условие (d) выполняется:

$$\langle J''(u)h,\, h\rangle_H = \langle 2A^*Ah,\, h\rangle_H = 2\langle Ah,\, Ah\rangle_F = 2\|Ah\|_F^2 \geqslant 0 \quad \forall\, h \in H.$$

Что касается свойства его *сильной выпуклости* с коэффициетом $\varkappa > 0$, то в соответствии с (d′) оно равносильно условию

$$\langle J''(u)h,\, h\rangle_H = 2\|Ah\|_F^2 \geqslant \varkappa\, \|h\|_H^2 \quad \forall\, h \in H$$

и обязывает оператор иметь *ограниченный обратный*:

$$\exists\, A^{-1} \in L(\operatorname{Im} A \to H), \quad \|A^{-1}\| \leqslant \sqrt{\frac{2}{\varkappa}}\,.$$

В конечномерном случае, когда $H = R^n,\ F = R^m$ и A является прямоугольной матрицей размера $m \times n$, свойство *сильной выпуклости* квадратичной функции $J(u) = \|Au - f\|_{R^m}^2$ n переменных $u = (u_1, u_2, \ldots, u_n)$ превращается в условие *положительной определённости* квадратной симметричной $n \times n$ матрицы $A^\top A$:

$$\langle 2A^\top Ah,\, h\rangle_{R^n} \geqslant \varkappa\, \|h\|_{R^n}^2 \quad \forall\, h \in R^n\,,$$

означающее, что *все собственные числа* λ_i матрицы $A^\top A$ *положительны*, а коэффициент сильной выпуклости квадратичной функции определяется наименьшим из них:

$$\varkappa = 2 \min_{1 \leqslant i \leqslant n} \lambda_i > 0 \,.$$

В том случае, когда $H = F = R^n$, свойство *сильной выпуклости* квадратичной функции будет равносильно *условию невырожденности* $\det A \neq 0$ самой матрицы A.

3. Конкретный квадратичный функционал в гильбертовом пространстве Лебега $H = L^2(0, \pi)$:

$$J(u) = \int_0^\pi \Big(\int_0^t u(s)\, ds \Big)^2 dt = \|Au\|_H^2, \quad (Au)(t) = \int_0^t u(s)\, ds, \quad t \in (0, \pi) \,.$$

Здесь $F = H = L^2(0, \pi)$, $A \in L(H \to H)$, $f = 0$ и, как следует, из *2)*, функционал $J(u)$ является *выпуклым* на всём пространстве $L^2(0, \pi)$. При этом он *не является сильно выпуклым* ни при каких $\varkappa > 0$. Действительно, свойство сильной выпуклости эквивалентно тому, что при некотором $\varkappa > 0$ выполняется оценка

$$2\|Ah\|_H^2 \geqslant \varkappa \, \|h\|_H^2 \quad \forall\, h \in H \,,$$

которая в случае $H = L^2(0, \pi)$ принимает вид

$$2 \int_0^\pi \Big(\int_0^t h(s)\, ds \Big)^2 dt \geqslant \varkappa \int_0^\pi h^2(t)\, dt \quad \forall\, h = h(t) \in L^2(0, \pi) \,. \qquad (73)$$

После подстановки в (73) на место $h(t)$ элементов $h_k(t) = \cos kt,\ k = 1, 2, \ldots$, получаем неравенства

$$2\, \frac{\pi}{2\, k^2} \geqslant \varkappa\, \frac{\pi}{2}, \quad k = 1, 2, \ldots$$

которые не оставляют для $\varkappa$ возможности быть положительным. Заметим, что в этом примере у оператора A существует обратный A^{-1}, который определён не на всём пространстве $L^2(0, \pi)$ и, что важнее, *не является ограниченным* (непрерывным). Заметим также, что функция $J(u)$ является *строго выпуклой* на всём пространстве $L^2(0, \pi)$, а на любой прямой, т. е. на любом одномерном линейном аффинном многообразии в $L^2(0, \pi)$, она даже *сильно выпукла* (коэффициент $\varkappa > 0$ зависит от прямой).

Метрическая проекция

Решение задачи минимизации специального вида, а именно, задачи минимизации расстояния от точки до множества играет заметную роль как в теории, так и в вычислениях.

Определение 15. *Пусть M — метрическое пространство, $U \subset M$ — некоторое множество, $x \in M$ — некоторая точка.* ***Метрической проекцией*** *точки x на множество U называется точка $p \in U$:*

$$p = \arg\min_{u \in U} \rho(u, x)\,. \tag{74}$$

Для обозначения проекции точки x на множество U будем использовать символ $\mathrm{pr}_U x$.

Замечание 10. *Если $x \in U$, то $\mathrm{pr}_U x = x$. Если множество U невыпукло, проекций может быть много, а незамкнутость множества U может привести к отсутствию проекции.*

Теорема 12. `(о свойствах проекции в гильбертовых пространствах)` *Пусть H — гильбертово пространство, $U \subset H$ — выпуклое замкнутое множество. Тогда*

1) *для любого $x \in H$ существует единственная проекция* $\mathrm{pr}_U x$;

2) *точка p является проекцией элемента $x \in H$ на множество U тогда и только тогда когда*

$$p \in U \quad \text{и} \quad \langle p - x,\, u - p\rangle \geqslant 0 \quad \forall\, u \in U; \tag{75}$$

3) *оператор* $\mathrm{pr}_U : H \to U$ *проектирования на U обладает свойством нестрогой сжимаемости:*

$$\|\mathrm{pr}_U x - \mathrm{pr}_U y\| \leqslant \|x - y\| \quad \forall\, x,\, y \in H\,. \tag{76}$$

Доказательство. 1) Напомним, что в гильбертовом пространстве $\rho(u, v) = \|u - v\|$ и заметим, что

$$\mathrm{pr}_U x \overset{(74)}{=} \arg\min_{u \in U} \|u - x\| = \arg\min_{u \in U} \|u - x\|^2\,.$$

Появившаяся здесь квадратичная функция $J(u) = \|u - x\|^2$ является *сильно выпуклой* на всём пространстве H и, в частности, на выпуклом замкнутом

множестве U, поэтому по теореме 8 («сильно выпуклый» вариант теоремы Вейерштрасса) она достигает своей нижней грани в единственной точке множества U, которая по определению 15 является проекцией.

2) Проекция $p = \mathrm{pr}_U x$ является решением задачи минимизации

$$\|u - x\|^2 \to \inf, \quad u \in U\,,$$

в которой функционал обладает свойствами выпуклости и гладкости, а допустимое множество U выпукло. Для таких задач представленное в теореме 11 вариационное неравенство (71)

$$\langle J'(u_*),\, u - u_* \rangle \geqslant 0 \quad \forall\, u \in U,$$

является *критерием оптимальности*. В нашем случае $J'(u_*) = 2(u_* - x)$, а $u_* = p$, откуда следует равносильность соотношений (75) определению проекции.

3) Запишем вариационные неравенства (75) для точек $x,\, y \in H$ и подставим в первое из них $u := \mathrm{pr}_U y$, а во второе $u := \mathrm{pr}_U x$:

$$\langle \mathrm{pr}_U x - x,\, \mathrm{pr}_U y - \mathrm{pr}_U x \rangle \geqslant 0, \quad \langle \mathrm{pr}_U y - y,\, \mathrm{pr}_U x - \mathrm{pr}_U y \rangle \geqslant 0\,.$$

После сложения этих неравенств получим соотношение

$$\langle \mathrm{pr}_U x - x - \mathrm{pr}_U y + y,\, \mathrm{pr}_U x - \mathrm{pr}_U y \rangle \geqslant 0\,,$$

которое переписывается в виде

$$\|\mathrm{pr}_U y - \mathrm{pr}_U x\|^2 \leqslant \langle \mathrm{pr}_U y - \mathrm{pr}_U x,\, y - x \rangle \overset{\text{К-Б}}{\leqslant} \|\mathrm{pr}_U y - \mathrm{pr}_U x\|\, \|y - x\|\,,$$

приводящем к (76). Теорема 12 доказана.

Заметим, что в зависимости от конфигурации границы множества U, на которое производится проектирование, в неравенстве (76) может наблюдаться как суперсжатие $\|\mathrm{pr}_U x - \mathrm{pr}_U y\| = 0 < \|x - y\|$, так и точное сохранение расстояния: $\|\mathrm{pr}_U x - \mathrm{pr}_U y\| = \|x - y\| \neq 0$.

Примеры вычисления проекций

1. *Проектирование на шар* $U = \{u \in H \;\big|\; \|u - u_0\| \leqslant R\}$ с центром в u_0 радиуса $R > 0$ в гильбертовом пространстве H. Из геометрических соображений ответ можно попробовать угадать:

$$\mathrm{pr}_U x = \begin{cases} u_0 + R\,\frac{x - u_0}{\|x - u_0\|}, & \text{если} \quad x \notin U, \\ x, & \text{если} \quad x \in U. \end{cases} \tag{77}$$

Для проверки правильности данной версии воспользуемся вариационным неравенством (75). Рассматриваем содержательный случай $x \notin U$, когда $\|x - u_0\| > R$. Для любого элемента $u \in U$ левая часть (75) примет вид

$$\left\langle u_0 + R\,\frac{x-u_0}{\|x-u_0\|} - x,\; u - u_0 - R\,\frac{x-u_0}{\|x-u_0\|} \right\rangle .$$

Мы должны убедиться в том, что для всех без исключения $u \in U$ значения таких скалярных произведений *неотрицательны*:

$$\begin{aligned}
&\left\langle u_0 + R\,\frac{x-u_0}{\|x-u_0\|} - x,\; u - u_0 - R\,\frac{x-u_0}{\|x-u_0\|} \right\rangle = \\
&\qquad = \Big(\frac{R}{\|x-u_0\|} - 1\Big) \left\langle x - u_0,\; u - u_0 - R\,\frac{x-u_0}{\|x-u_0\|} \right\rangle = \\
&\qquad = \underbrace{\Big(1 - \frac{R}{\|x-u_0\|}\Big)}_{>0,\ \text{т.к. } x\notin U} \Big(R\,\|x-u_0\| - \underbrace{\langle x-u_0,\; u-u_0\rangle}_{\leqslant R\,\|x-u_0\|,\ \text{т.к. } u\in U}\Big) \geqslant 0,\ \text{ч.т.д.}
\end{aligned}$$

2. *Проектирование* в гильбертовом пространстве Лебега $L^2(0,T)$ *на «параллелепипед»*

$$U = \Big\{ u(t) \in L^2(0,T) \;\Big|\; a(t) \overset{\text{п.в.}}{\leqslant} u(t) \overset{\text{п.в.}}{\leqslant} b(t) \Big\},$$

где $a(t),\, b(t) \in L^2(0,T)$ — заданные функции, удовлетворяющие условию $a(t) \leqslant b(t)$ для п. в. $t \in (0,T)$. В данном случае явное выражение для проекции $\mathrm{pr}_U f$ функции $f(t) \in L^2(0,T)$ на параллелепипед U удобнее получить непосредственно из определения проекции, минимизируя расстояние

$$\|u - f\|_{L^2(0,T)} = \Big(\int_0^T \big(u(t) - f(t)\big)^2\, dt \Big)^{1/2}$$

от функций $u(\cdot) \in U$ до функции $f(\cdot)$. Понятно, что наименьшее значение данного интеграла будет достигаться при выборе

$$u(t) = \mathrm{pr}_U f(t) = \begin{cases} a(t), & \text{если} \quad f(t) < a(t), \\ f(t), & \text{если} \quad a(t) \leqslant f(t) \leqslant b(t), \\ b(t), & \text{если} \quad f(t) > b(t)\,. \end{cases}$$

Поскольку все три функции $f(t), a(t), b(t)$, участвующие в данной конструкции, *измеримы* по Лебегу, то будут *измеримы* и множества, на которых происходит коррекция значений функции $f(t)$, а, тем самым, будет *измерима* по Лебегу и результирующая функция $\mathrm{pr}_U f(t)$. Её принадлежность пространству $L^2(0,T)$ следует из свойств интеграла Лебега.

Для задач минимизации гладких выпуклых функций на выпуклых замкнутых множествах в гильбертовых пространствах критерий оптимальности, записанный в теореме 11 в форме *вариационного неравенства*, можно переписать в *проекционных терминах*. Такую возможность описывает

Теорема 13. (проекционная форма критерия оптимальности)
Пусть H — гильбертово пространство, U — выпуклое замкнутое множество из H, а функция $J(u) \in C^1(U)$ и выпукла на множестве U. Тогда точка $u_* \in U$ является оптимальным решением задачи

$$J(u) \to \inf, \quad u \in U$$

тогда и только тогда когда для некоторого $\alpha > 0$

$$u_* = \mathrm{pr}_U\big(u_* - \alpha\, J'(u_*)\big)\,. \tag{78}$$

Доказательство. При выполнении условий теоремы 13 критерием оптимальности элемента $u_* \in U$ является вариационное неравенство (71):

$$\langle J'(u_*),\, u - u_* \rangle \geqslant 0 \quad \forall\, u \in U.$$

После умножения обеих частей на $\alpha > 0$ добавим и вычтем в левой части скалярного произведения элемент u_* :

$$\langle u_* - \big(u_* - \alpha\, J'(u_*)\big),\, u - u_* \rangle \geqslant 0 \quad \forall\, u \in U,$$

и в соответствии с приведённой в условии (75) теоремы 12 характеризацией проекции придём к заключению об эквивалентности данного неравенства соотношению (78). Теорема 13 доказана.

Упражнение 13. *Пусть H – гильбертово пространство, L – его замкнутое подпространство. Докажите, что оператор pr_L метрического проектирования из H на L совпадает с оператором P ортогонального проектирования из H на L, т. е. является линейным ограниченным и самосопряженным оператором.*

Упражнение 14. *Пусть x_0 – фиксированный элемент из H, $x_0 + L$ – соответствующее замкнутое линейное аффинное многообразие. Докажите, что для любого $x \in H$ $p = pr_{x_0+L}x$ в том и только в том случае, когда*

$$p \in x_0 + L \quad \text{и} \quad \langle p - x,\, l \rangle_H = 0 \quad \forall\, l \in L.$$

Упражнение 15. *Найдите проекции точек на гиперплоскость $\{u \in H \mid \langle c, u \rangle_H = \beta\}$ в гильбертовом пространстве H (заданы $c \in H$, $c \neq 0$ и число β) и на параллелепипед $\{u = (u_1, ..., u_n) \in R^n \,|\, \alpha_i \leqslant u_i \leqslant \beta_i,\ i = 1, ..., n\}$ в R^n (заданы числа α_i, β_i, $i = 1, 2, \ldots, n$).*

V. ИТЕРАЦИОННЫЕ МЕТОДЫ МИНИМИЗАЦИИ

Рассмотрим и обсудим ряд конкретных итерационных методов, которые применяются для практического решения оптмимизационных задач. В рамках данного курса основное внимание будет уделено условиям сходимости этих процессов и оценкам скорости сходимости. Рассматриваемые методы будут ориентированы на решение задач минимизации вида

$$J(u) \to \inf, \quad u \in U \subset H,$$

в гильбертовом пространстве H, $\dim H \leqslant \infty$. Ограничения на переменные u могут отсутствовать, т. е. случай $U = H$ не исключается. В основном, мы будем рассматривать итерационные процессы вида

$$u_{k+1} = u_k + \alpha_k\, p_k, \quad \alpha_k \in R^1, \quad p_k \in H, \quad k = 0, 1, \ldots, \tag{79}$$

для запуска которых требуется задать некоторое *начальное приближение* $u_0 \in U$. Коэффициент α_k в (79) принято называть *шагом спуска*, вектор p_k — *направлением спуска*, а k — номером итерации. Существующие подходы к выбору значений α_k и p_k обычно разделяют на два класса. Названия этих классов имеют зарубежное происхождение: **«Line Search»** и **«Trust-Region»**. Для первого из них имеется более-менее устоявшееся русское наименование **«одномерный поиск»**. Для второго класса общепринятые варианты перевода на русский мне неизвестны, а использовать термины типа «доверительный» мне не хочется, поскольку он занят в другой математической дисциплне. Основное различие между этими двумя подходами заключается в порядке выбора параметров α_k и p_k. При *одномерном поиске* сначала выбирается направление спуска p_k, после чего решается вспомогательная задача одномерной минимизации по α. При применении подхода *Trust-Region* сначала выбирают некоторую *модельную* функцию $m_k(u)$, являющуюся хорошим приближением к функции $J(u)$ в окрестности текущей итерации u_k, и задают вызывающий доверие размер окрестности $d_k > 0$, а затем ищут направление спуска p_k, минимизируя в этой окрестности модель $m_k(u)$. Если полученное при этом новое приближение

$$u_{k+1} = u_k + p_k$$

не даёт удовлетворительного убывания самой функции $J(u)$, процедура поиска p_k повторяется с той же моделью $m_k(u)$ в окрестности *ме́ньшего* размера.

В качестве модели для гладких функций $J(u)$ чаще всего выбирают функции вида

$$m_k(u) = J(u_k) + \langle J'(u_k),\, u - u_k\rangle + \frac{1}{2}\,\langle B_k\,(u - u_k),\, u - u_k\rangle\,,$$

где линейный оператор $B_k : H \to H$ либо является нулевым (в *градиентных* методах), либо $B_k = J''(u_k)$ (в методе *Ньютона*), либо B_k является некоторым приближением к гессиану $J''(u_k)$ (в *квазиньютоновских* методах). Действия, реализующие *Trust-region* метод, можно было бы назвать, используя недословный перевод, «поиском окрестности». Детальное описание и анализ подходов *Line Search* и *Trust-Region* содержатся в книге [5].

В данном курсе не будут обсуждаться практически важные проблемы выбора *начального приближения* u_0 и *правила останова*. Удачный выбор стартовой точки во многом зависит от уровня понимания той предметной области, в которой возникла данная оптимизационная постановка и представления об уровне точности рассматриваемой модели. Чтобы остановить итерации (79), часто задают некоторый (достаточно малый) уровень погрешности $\varepsilon > 0$ и используют условия вида

$$\|u_{k+1} - u_k\| \leqslant \varepsilon, \quad \|J'(u_k)\| \leqslant \varepsilon, \quad \big|J(u_{k+1}) - J(u_k)\big| \leqslant \varepsilon,$$

$$\frac{\|u_{k+1} - u_k\|}{\|u_k\|} \leqslant \varepsilon, \quad \frac{\|J'(u_k)\|}{\|J'(u_0)\|} \leqslant \varepsilon, \quad \frac{\big|J(u_{k+1}) - J(u_k)\big|}{\big|J(u_k)\big|} \leqslant \varepsilon,$$

или им подобные в различных логических сочетаниях.

Медод скорейшего спуска (МСС)

Начнём с классического *метода скорейшего спуска (МСС)*, относящегося к категории *Line Search* и ориентированного на решение задач *безусловной минимизации (без ограничений)* дифференцируемой функции:

$$J(u) \to \inf, \quad u \in U = H. \tag{80}$$

В качестве направления спуска на каждом шаге этого метода выбирается $p_k = -J'(u_k)$ — направление *наискорейшего убывания* функции $J(u)$ в точке u_k, а при выборе шага спуска выполняется *точный одномерный поиск*:

$$u_{k+1} = u_k + \alpha_k \big(-J'(u_k)\big), \tag{81}$$

$$\alpha_k = \arg\min_{\alpha \geqslant 0} J\big(u_k + \alpha\,\big(-J'(u_k)\big)\big), \quad k = 0, 1, \ldots \tag{82}$$

Замечание 11. *Для выпуклой гладкой функции $J(u)$ до тех пор, пока минимум ещё не достинут, т. е. пока $J'(u_k) \neq 0$, точное решение α_k задачи одномерного поиска* (82) *является корнем скалярного уравнения*

$$\frac{df_k(\alpha)}{d\alpha} = 0, \quad \text{где} \quad f_k(\alpha) = J\big(u_k + \alpha\left(-J'(u_k)\right)\big), \quad \frac{df_k(\alpha)}{d\alpha} = \left\langle J'\big(u_k + \alpha\left(-J'(u_k)\right)\big),\, -J'(u_k)\right\rangle.$$

В результате новое направление поиска $p_{k+1} = -J'(u_{k+1})$ получается ортогональным предыдущему: $\langle p_{k+1},\, p_k\rangle = 0$ и итерационный процесс (81) *движется в пространстве H по зигзагообразной траектории. При практическом применении МСС следует критически оценивать целесообразность выполнения процедуры одномерного поиска с высокой точностью, соизмеряя вычислительные затраты на поиск шага α_k с затратами на поиск нового направления спуска p_{k+1}.*

Медод скорейшего спуска (продолжение)

Теорема 14. (о сходимости МСС)
Пусть H — гильбертово пространство, $J(u) \in C^1(H)$, $J'(u) \in \mathrm{Lip}(H)$ с $L > 0$ и, кроме того, функция $J(u)$ сильно выпукла на H с коэффициентом $\varkappa > 0$. Тогда МСС (81), (82) сходится к единственному оптимальному решению u_* задачи (80) из любого начального приближения $u_0 \in H$, причём

$$\frac{\varkappa}{2}\,\|u_k - u_*\|^2 \leqslant J(u_k) - J(u_*) \leqslant q^k \left(J(u_0) - J(u_*)\right), \tag{83}$$

где $q = 1 - \frac{\varkappa}{L} \in [0, 1)$.

Доказательство. Существование и единственность оптимального решения u_* следует из теоремы 8 («сильно выпуклый» вариант теоремы Вейерштрасса). Убедимся в том, что $q = 1 - \frac{\varkappa}{L} \in [0, 1)$. Принимая во внимание утверждение (c′) теоремы 10 и липшиц-непрерывность градиента $J'(u)$, запишем двустороннюю оценку

$$\varkappa\,\|u - v\|^2 \leqslant \langle J'(u) - J'(v),\, u - v\rangle \leqslant L\,\|u - v\|^2 \quad \forall\, u, v \in H.$$

Отсюда имеем

$$0 < \varkappa \leqslant L \quad \Longrightarrow \quad 0 \leqslant q = 1 - \frac{\varkappa}{L} < 1.$$

Из сильной выпуклости функции $J(u)$ в случае, когда $J'(u_k) \neq 0$, следует сильная выпуклость функции одной переменной $f_k(\alpha) = J\big(u_k + \alpha\left(-J'(u_k)\right)\big)$ и, тем самым, существование и единственность шага спуска α_k, являющегося решением одномерной задачи (82).

Введём обозначения для текущих уклонений до оптимума по функции:

$$a_k = J(u_k) - J(u_*)$$

и оценим a_{k+1} через a_k :

$$a_{k+1} - a_k = J(u_{k+1}) - J(u_k) = J\big(u_k - \alpha_k J'(u_k)\big) - J(u_k) =$$
$$\overset{(82)}{=} \min_{\alpha\geqslant 0} J\big(u_k - \alpha J'(u_k)\big) - J(u_k) \overset{(\forall\alpha\geqslant 0)}{\leqslant} J\big(u_k - \alpha J'(u_k)\big) - J(u_k) =$$
$$\overset{\text{конеч. прир.}}{=} \int_0^1 \langle J'\big(u_k - t\alpha J'(u_k)\big),\, -\alpha J'(u_k)\rangle\, dt = \big[\mp J'(u_k)\big] =$$
$$= -\alpha\, \|J'(u_k)\|^2 + \int_0^1 \langle J'\big(u_k - t\alpha J'(u_k)\big) - J'(u_k),\, -\alpha J'(u_k)\rangle\, dt \leqslant$$
$$\overset{J'(u)\in\mathrm{Lip}}{\leqslant} -\alpha\, \|J'(u_k)\|^2 + L\int_0^1 \|t\alpha J'(u_k)\|\, \|-\alpha J'(u_k)\|\, dt =$$
$$= \|J'(u_k)\|^2 \Big(-\alpha + \frac{L}{2}\,\alpha^2\Big) \quad \forall\,\alpha\geqslant 0\,. \quad (84)$$

Пользуясь произволом в выборе $\alpha \geqslant 0$, возьмём наименьшее из возможных значений в правой части (84), которое достигается при $\alpha = \frac{1}{L}$:

$$a_{k+1} - a_k \leqslant \|J'(u_k)\|^2 \min_{\alpha\geqslant 0}\Big(-\alpha + \frac{L}{2}\,\alpha^2\Big) = \|J'(u_k)\|^2\Big(-\frac{1}{2L}\Big)\,. \quad (85)$$

Чтобы связать значения $\|J'(u_k)\|^2$ с a_k, , воспользуемся утверждением (b′) теоремы 10:

$$a_k = J(u_k) - J(u_*) \leqslant \langle J'(u_k),\, u_k - u_*\rangle - \frac{\varkappa}{2}\,\|u_k - u_*\|^2,$$

с помощью неравенства Коши-Буняковского оценим сверху скалярное произведение:

$$\langle J'(u_k),\, u_k - u_*\rangle \leqslant \|J'(u_k)\|\,\|u_k - u_*\|$$

и, обозначив $y = \|u_k - u_*\| \geqslant 0$, придём к неравенству

$$a_k \leqslant \|J'(u_k)\|\, y - \frac{\varkappa}{2}\, y^2\,,$$

в правой части которого можно взять максимум по $y \geqslant 0$:

$$a_k \leqslant \max_{y\geqslant 0}\Big(\|J'(u_k)\|\, y - \frac{\varkappa}{2}\, y^2\Big) \overset{(y_{\max}=\|J'(u_k)\|/\varkappa)}{=} \frac{1}{2\varkappa}\,\|J'(u_k)\|^2\,. \quad (86)$$

Из (85) и (86) извлекается рекуррентная оценка для a_k :

$$a_{k+1} \leqslant \Big(1 - \frac{\varkappa}{L}\Big)\, a_k = q\, a_k, \quad k = 0, 1, \ldots, \quad (87)$$

из которой следует правое неравенство (83). Что касается левого неравенства (83), то оно просто дублирует ранее доказанное утверждение (66) теоремы 8. Теорема 14 доказана.

Замечание 12. *Сходимость по функции или по аргументу вида* (83), *т. е. сходимость со скоростью геометрической прогресии или, другими словами, с экспоненциальной скоростью, принято называть* ***линейной****. Если наряду с самим фактом сходимости по функции* $J(u_k) \to J(u_*)$ *или по аргументу* $\|u_k - u_*\| \to 0$ *при некотором* $q \in (0,1)$ *выполняются условия типа* (87)*:*

$$|J(u_{k+1} - J(u_*)| \leqslant q\,|J(u_k - J(u_*)| \quad \textit{или} \quad \|u_{k+1} - u_*\| \leqslant q\,\|u_k - u_*\|, \quad k = 0, 1, \ldots,$$

то такую сходимость обычно называют q***-линейной.***

Замечание 13. *Теоретически скорость сходимости МСС повышается при уменьшении* q, *а в идеальной и крайне редкой ситуации, когда* $\varkappa = L$ *и* $q = 0$, *МСС сойдётся к точному решению за один шаг. Примером такой функции является квадратичная сильно выпуклая функция специального вида*

$$J(u) = \frac{\varkappa}{2}\,\|u\|^2 + \langle b, u\rangle + c, \quad \varkappa > 0,\ b \in H,\ c \in R^1,$$

единственная точка минимума которой находится в позиции $u_* = -\frac{b}{\varkappa}$. *Для сильно выпуклых квадратичных функций более общего вида в конечномерном пространстве* $H = R^n$:

$$J(u) = \frac{1}{2}\,\langle Au, u\rangle + \langle b, u\rangle + c, \quad A^\top = A > 0,$$

значения $\varkappa$ *и* L *коэффициента сильной выпуклости и константы Липшица градиента связаны со значениями минимального и максимального собственного числа матрицы* A, *так что* $\varkappa/L \simeq \lambda_{\min}/\lambda_{\max}$ *и в достаточно распространённом случае, когда матрица* A ***плохо обусловлена*** *и отношение* $\lambda_{\min}/\lambda_{\max} \simeq 0$, *значение* $q \simeq 1$ *и ухудшается качество не только теоретической оценки* (83), *но, как правило, существенно замедляется сходимость и реального вычислительного процесса.*

Иногда задачу одномерного поиска (82) удаётся решить точно и даже аналитически. Важным примером функций, для которых это возможно, являются квадратичные функции вида

$$J(u) = \|Au - f\|_F^2, \quad A \in L(H \to F), \quad f \in F.$$

Для них

$$f_k(\alpha) = \|A\big(u_k - \alpha\,J'(u_k)\big) - f\|_F^2 =$$
$$= \|Au_k - f\|_F^2 - 2\alpha\,\langle Au_k - f,\ A\,J'(u_k)\rangle_F + \alpha^2\,\|A\,J'(u_k)\|_F^2$$

и при $\|A\,J'(u_k)\|_F^2 > 0$ минимум квадратичной функции $f_k(\alpha)$ достигается в стационарной точке

$$\alpha_k = \frac{2\,\langle Au_k - f,\ A\,J'(u_k)\rangle_F}{2\,\|A\,J'(u_k)\|_F^2} = \big[J'(u_k) = 2A^*(Au_k - f)\big] =$$
$$= \frac{\langle Au_k - f,\ A\,2A^*(Au_k - f)\rangle_F}{\|A\,2A^*(Au_k - f)\|_F^2} = \frac{\|A^*(Au_k - f)\|_H^2}{2\,\|AA^*(Au_k - f)\|_F^2} \geqslant 0.$$

Если же $\|A\,J'(u_k)\|_F^2 = 4\,\|AA^*(Au_k - f)\|_F^2 = 0$, то $Au_k - f \in \operatorname{Ker} AA^*$, а, как известно,

$$\operatorname{Ker} AA^* = \operatorname{Ker} A^*, \tag{88}$$

поэтому

$$Au_k - f \in \operatorname{Ker} A^* \quad \Longleftrightarrow \quad A^*(Au_k - f) = 0 \quad \Longleftrightarrow \quad J'(u_k) = 0,$$

а это означает, что $u_k \in U_*$ — оптимальное решение задачи (80) и итерационный процесс (81), (82) следует остановить. Равенство ядер (88) следует из их взаимных вложений:

$$\operatorname{Ker} A^* \subset \operatorname{Ker} AA^* \quad \text{и} \quad \operatorname{Ker} AA^* \subset \operatorname{Ker} A^*,$$

первое из которых очевидно, а второе подтверждается следующими рассуждениями:

$$v \in \operatorname{Ker} AA^* \quad \Longrightarrow \quad AA^*v = 0 \quad \Longrightarrow \quad \langle AA^*v, v\rangle_F = 0 \quad \Longrightarrow$$
$$\Longrightarrow \quad \|A^*v\|_H^2 = 0 \quad \Longrightarrow \quad A^*v = 0 \quad \Longrightarrow \quad v \in \operatorname{Ker} A^*.$$

Упражнение 16. *Найдите явное выражения для шага спуска α_k из условия точного одномерного поиска* (82) *в случае квадратичной функции вида*

$$J(u) = \frac{1}{2}\,\langle Au, u\rangle_H - \langle f, u\rangle_H, \quad A \in L(H \to H), \quad A^* = A \geqslant 0, \quad f \in H.$$

Непрерывный аналог градиентного метода

На мой взгляд, имеет смысл наряду с дискретными рассмотреть и непрерывные процессы минимизации, динамика которых описываются дифференциальными уравнениями. В случае задачи безусловной минимизации (80):

$$J(u) \to \inf, \quad u \in H,$$

для предложенного в (81) итерационного процесса

$$u_{k+1} = u_k + \alpha_k \big(-J'(u_k)\big),$$

легко угадывается ОДУ, являющееся его аналогом, если воспринимать точки u_k как положения $u_k = u(t_k)$ некоторой непрерывной траектории $u(t)$ в моменты t_k :

$$u'(t) = -\beta(t)\,J'\big(u(t)\big), \quad t > 0; \quad u(0) = u_0. \tag{89}$$

Задача (89) относится к классу задач Коши для ОДУ в гильбертовом пространстве H. Для «запуска» непрерывного процесса (89) должна быть задана стартовая точка $u_0 \in H$ и коэффициент $\beta(t)$, который играет роль шага спуска и в задачах минимизации должен быть положительным. Аналог процедуры одномерного поиска (82) мы во избежание осложнений в процесс (89) встраивать не будем, так что его можно считать непрерывным аналогом итерационного процесса (81) с априорно заданными шагами α_k.

Теорема 15. (о сходимости непр. вар-та градиентного метода)
Пусть H — гильбертово пространство, $J(u) \in C^1(H)$, $J'(u) \in \mathrm{Lip}(H)$ и, кроме того, функция $J(u)$ сильно выпукла на H с коэффициентом $\varkappa > 0$, а $\beta(t) \in C[0,+\infty)$, $\beta(t) \geqslant \beta_0 > 0$. Тогда решение $u(t)$ задачи Коши (89) сходится к единственному оптимальному решению u_* задачи (80) из любого начального приближения $u_0 \in H$, причём

$$\|u(t) - u_*\| \leqslant \|u_0 - u_*\|\, e^{-\varkappa\beta_0 t} \quad \forall\, t \geqslant 0. \tag{90}$$

Доказательство. Обратим внимание на то, что в условии теоремы присутствует требование липшиц-непрерывности градиента, но в оценке скорости сходимости (90) константа Липшица не просматривается. Дело в том, что в данной теореме условие липшиц-непрерывности градиента затребовано для гарантии существования и единственности решения $u(t)$ задачи Коши для любого начального условия $u_0 \in H$ (условие такого типа было обычным в курсе ДУ). Единственное оптимальное решение u_* задачи минимизации существует по тем же причинам, что и в теореме 14. Чтобы вывести оценку (90), рассмотрим функцию (Ляпунова)

$$V(t) = \|u(t) - u_*\|^2, \quad t \geqslant 0,$$

и оценим её производную:

$$V'(t) = 2\,\langle u(t) - u_*,\, u'(t)\rangle \overset{(89)}{=} 2\,\langle u(t) - u_*,\, -\beta(t)\,J'\big(u(t)\big)\rangle \overset{(J'(u_*)=0)}{=}$$
$$= -2\,\underbrace{\beta(t)}_{\geqslant\beta_0}\,\underbrace{\langle u(t) - u_*,\, J'\big(u(t)\big) - J'(u_*)\rangle}_{\geqslant\varkappa\,\|u(t)-u_*\|^2} \leqslant -2\beta_0\varkappa\,\|u(t)-u_*\|^2 = -2\beta_0\varkappa\,V(t)\,.$$

После интегрирования этого неравенства (используем интегрирующий множитель $e^{2\beta_0\varkappa t}$) получаем оценку

$$V(t) \leqslant V(0)\,e^{-2\beta_0\varkappa t}, \quad t \geqslant 0,$$

фактически совпадающую с искомой оценкой (90). теорема 15 доказана.

Замечание 14. *Непрерывные способы описания итерационных процедур типа* (89) *открывают широкие возможности их дискретизации с помощью разнообразных методов: Адамса, Рунге-Кутта и др. С этой точки зрения на дискретный процесс* (81) *из МСС можно смотреть как на аппроксимацию непрерывной версии* (89) *с помощью простейшей явной схемы Эйлера.*

Замечание 15. *Если использовать терминологию, принятую в теории ДУ, то установленный в теореме 15 результат можно назвать* ***асимптотической устойчивостью по Ляпунову*** *стационарного решения (неподвижной точки)* $u(t) \equiv u_*$ *системы* (89). *В терминах той же теории искомое оптимальное решение* u_* *задачи минимизации является* ***глобальным аттрактором*** *системы* (89), *притягивающим траектории, исходящие из любых стартовых позиций* $u_0 \in H$.

Метод проекции градиента (МПГ)

Этот метод ориентирован на решение задач минимизации ***с ограничениями*** (*задач условной минимизации*):

$$J(u) \to \inf, \quad u \in U \subset H\,. \tag{91}$$

В качестве направления спуска на k-ой итерации выбирается, как и в МСС, $p_k = -J'(u_k)$, а затем, поскольку точка $u_k + \alpha_k\, p_k$ может выйти за пределы допустимого множества U, производится её коррекция посредством проектирования:

$$u_{k+1} = \mathrm{pr}_U\big(u_k - \alpha_k\, J'(u_k)\big), \quad k = 0, 1, \ldots \tag{92}$$

Заметим, что если на каждой итерации МПГ выполнять процедуру одномерного поиска с целью оптимизации значения шага спуска α_k, то придётся иметь дело с функцией одной переменной

$$f_k(\alpha) = J\big(\mathrm{pr}_U\big(u_k - \alpha\, J'(u_k)\big)\big), \quad \alpha \geqslant 0.$$

При этом внутренние итерации по α при фиксированном k из-за наличия операции проектирования на U могут оказаться слишком дорогостоящими (разумеется, их цена существенно зависит от конфигурации множества U и от стоимости вычисления значений функции J). По этим причинам для простоты при теоретическом анализе МПГ мы ограничимся рассмотрением случая, когда шаги спуска α_k в (92) выбираются *априорно* и, более того, даже постоянными: $\alpha_k = \alpha > 0,\ k = 0, 1, \ldots$

Теорема 16. (о сходимости МПГ с постоянным шагом)
Пусть H — гильбертово пространство, множество U выпукло и замкнуто, функция $J(u) \in C^1(U)$, $J'(u) \in \mathrm{Lip}(U)$ с $L > 0$ и, кроме того, $J(u)$ сильно выпукла на U с коэффициентом $\varkappa > 0$. Пусть шаг спуска в (92) выбирается постоянным: $\alpha_k = \alpha$, причём $\alpha \in \big(0,\, 2\varkappa/L^2\big)$. Тогда МПГ (92) сходится к единственному оптимальному решению u_* задачи (91) из любого начального приближения $u_0 \in U$ и справедлива оценка

$$\|u_k - u_*\| \leqslant \|u_0 - u_*\| \big(q(\alpha)\big)^{k/2}, \quad k = 0, 1, \ldots, \tag{93}$$

где $0 < q(\alpha) = 1 - 2\varkappa\alpha + \alpha^2 L^2 < 1$.

Доказательство. Сначала убедимся в том, что при $\alpha \in \big(0,\, 2\varkappa/L^2\big)$ значения $q(\alpha)$ находятся в интервале $(0, 1)$. Это действительно так, поскольку квадратичная и сильно выпуклая по α функция $q(\alpha) = 1 - 2\varkappa\alpha + \alpha^2 L^2$ достигает своего минимума в центральной точке $\alpha_* = \varkappa/L^2$ промежутка $\big(0,\, 2\varkappa/L^2\big)$, а в его концевых точках принимает одинаковые значения $q(0) = q\big(2\varkappa/L^2\big) = 1$.

По теореме 8 («сильно выпуклый» вариант теоремы Вейерштрасса) оптимальное решение u_* задачи (91) существует и единственно. По теореме 13 для точки u_* справедлив критерий оптимальности в проекционной форме:

$$u_* = \mathrm{pr}_U\big(u_* - \alpha\, J'(u_*)\big)\,,$$

в котором на месте α может стоять любое положительное число, в частности, любое $\alpha \in \big(0,\, 2\varkappa/L^2\big)$. На данное равенство можно посмотреть как на операторное уравнение вида

$$u = Au, \tag{94}$$

где оператор $Au = \mathrm{pr}_U\big(u - \alpha\, J'(u)\big)$ действует из U в U, а искомая точка минимума u_* будет неподвижной точкой этого оператора на множестве U. Заметим, что относительно той самой метрики, которой наделено гильбертово пространство H, *замкнутое* множество U является *полным метрическим пространством*. В таких пространствах в соответствии с *принципом сжимающих отображений* [3, гл. 2, §4] у сжимающего оператора существует единственная неподвижная точка u_*, к которой метод простой итерации $u_{k+1} = Au_k$ сходится из любого начального приближения $u_0 \in U$ со скоростью геометрической прогресии (с линейной скоростью). Убедимся в том, что оператор $Au = \mathrm{pr}_U\big(u - \alpha\, J'(u)\big) : U \to U$ действительно является сжимаю-

щим:

$$\begin{aligned}
\|Au - Av\|^2 = \|\mathrm{pr}_U\big(u - \alpha\, J'(u)\big) - \mathrm{pr}_U\big(v - \alpha\, J'(v)\big)\|^2 \leqslant \\
\leqslant \big[\text{нестрогая сжим. оператора проектирования, теор. 12}\big] \leqslant \\
\leqslant \|\big(u - \alpha\, J'(u)\big) - \big(v - \alpha\, J'(v)\big)\|^2 = \\
= \|u - v\|^2 - 2\,\alpha\, \underbrace{\langle u - v,\, J'(u) - J'(v)\rangle}_{\geqslant \varkappa\, \|u-v\|^2} + \alpha^2\, \underbrace{\|J'(u) - J'(v)\|^2}_{\leqslant L^2\, \|u-v\|^2} \leqslant \\
\leqslant \big(1 - 2\varkappa\alpha + \alpha^2 L^2\big)\, \|u - v\|^2 = q(\alpha)\, \|u - v\|^2 \quad \forall\, u,\, v \in U.
\end{aligned}$$

Таким образом, показано, что оператор сжимающий и его коэффициент сжатия равен $\sqrt{q(\alpha)} \in (0, 1)$. Остаётся заметить, что МПГ (92) с постоянным шагом α *буквально* совпадает с методом простой итерации $u_{k+1} = Au_k$ для уравнения (94) с оператором $Au = \mathrm{pr}_U\big(u - \alpha\, J'(u)\big)$, поэтому и оценка (93) — это известное свойство метода простой итерации (см. [3, гл. 2, §4]). Теорема 16 доказана.

Замечание 16. *Представляется интересным сравнить скорости сходимости МСС и МПГ на классе задач минимизации без ограничений, в которых $U = H$ и нет необходимости в проектировании. Напомним, что в теореме 14 была доказана оценка*

$$\|u_k^{MCC} - u_*\| = O\big(q^{k/2}\big), \quad q = 1 - \frac{\varkappa}{L},$$

а в теореме 16 для МПГ для специального $\alpha_* = \frac{\varkappa}{L^2}$, для которого $q_* = q(\alpha_*) = 1 - \frac{\varkappa^2}{L^2} = \min q(\alpha)$, получена оценка

$$\|u_k^{МПГ} - u_*\| = O\big(q_*^{k/2}\big), \quad q_* = 1 - \frac{\varkappa^2}{L^2}.$$

Асимптотически при больших k скорость сходимости определяется значением q. Понятно, что $q < q_*$, но для достаточно типичной на практике ситуации, когда $\varkappa/L \simeq 0$, преимущество МСС перед МПГ становится неочевидным, особенно с учётом трудоёмкости выполнения в МСС процедуры одномерного поиска.

Непрерывные аналоги метода простой итерации (МПИ) и метода проекции градиента (МПГ)

МПГ (92) с постоянным шагом является частным случаем МПИ $u_{k+1} = Au_k$, сходящегося к неподвижной точке сжимающего оператора A. Подходящий вид его непрерывного аналога нетрудно угадать, отправляясь от записи МПИ в форме $u_{k+1} - u_k = -u_k + Au_k$ и используя те же соображения, что и при реконструкции непрерывного аналога градиентного метода:

$$u'(t) = \beta(t)\big(-u(t) + Au(t)\big), \quad t > 0; \quad u(0) = u_0 \in U. \tag{95}$$

Теорема 17. (непрерывный аналог МПИ)
Пусть H — гильбертово пространство, $U \subset H$ — выпуклое замкнутое множество, на котором определён оператор $A : U \to U$, являющийся сжимающим с коэффициентом $q \in [0, 1)$:

$$\|Au - Av\| \leqslant q\,\|u - v\| \quad \forall\, u,\, v \in U. \tag{96}$$

Пусть коэффициент $\beta(t)$ в (95) непрерывен и отделён от нуля: $0 < \beta_0 \leqslant \beta(t)$. Тогда траектория $u(t)$ системы (95) сходится к неподвижной точке $u_* \in U$ оператора A из любого начального положения $u_0 \in U$ с экспоненциальной скоростью:

$$\|u(t) - u_*\| \leqslant \|u_0 - u_*\|\, e^{-\beta_0\,(1-q)\,t}, \quad t \geqslant 0. \tag{97}$$

Доказательство. Существование и единственность неподвижной точки $u_* \in U$ следуют из сжимаемости оператора A. Свойство непрерывности коэффициента $\beta(t)$ обеспечивает липшиц-непрерывность по u правой части ОДУ (95), а этого достаточно для существования и единственности решения $u(t)$ системы (95) для любой стартовой точки $u_0 \in U$. Свойства выпуклости и замкнутости множества U, принадлежности значений $Au \in U$ при $u \in U$ и положительности $\beta(t) > 0$ гарантируют, что траектория $u(t)$, исходящая из $u_0 \in U$, не выйдет за пределы множества $U : u(t) \in U \;\forall\, t \geqslant 0$. Нестрогое объяснение такого свойства можно дать, предположив, что в момент t состояние $u(t) \in U$ и оценить её положение $u(t + \Delta t)$ в момент $t + \Delta t$, считая $\Delta t > 0$ достаточно малым:

$$u(t + \Delta t) \simeq \big(1 - \beta(t)\Delta t\big)\, u(t) + \beta(t)\Delta t\, Au(t) \in U.$$

Здесь мы использовали малость Δt, включения $u(t) \in U$, $Au(t) \in U$, $\beta(t)\,\Delta \in [0, 1]$ и выпуклость множества U.

Для вывода оценки (97) сначала заметим, что функция-константа $u_*(t) \equiv u_*$ является решением ОДУ (95):

$$u_*'(t) = \beta(t)\,\big(-u_* + Au_*\big) = 0, \quad t > 0.$$

Введём обозначение $x(t) = u(t) - u_*$ и составим для функции $x(t)$ задачу Коши:

$$x'(t) = \beta(t)\,\big(-x(t) + Au(t) - Au_*\big), \quad t > 0; \quad x(0) = u_0 - u_*. \tag{98}$$

Рассмотрим функцию (Ляпунова) $V(t) = \|x(t)\|^2$ и оценим её производную:

$$\begin{aligned} V'(t) = 2\,\langle x(t),\, x'(t)\rangle \overset{(98)}{=} 2\,\big\langle x(t),\, \beta(t)\,\big(-x(t) + Au(t) - Au_*\big)\big\rangle = \\ = 2\,\beta(t)\,\big(-\|x(t)\|^2 + \langle x(t),\, Au(t) - Au_*\rangle\big) \overset{\text{К-Б, сжим.}}{\leqslant} \\ \leqslant 2\,\underbrace{\beta(t)}_{\geqslant\beta_0>0}\,\underbrace{(q-1)}_{<0}\,\|x(t)\|^2 \leqslant -2\,\beta_0\,(1-q)\,V(t), \quad t > 0. \end{aligned}$$

Отсюда, как и в теореме 15, следует оценка

$$V(t) \leqslant V(0)\, e^{-2\,\beta_0\,(1-q)\,t}, \quad t \geqslant 0,$$

совпадающая с утверждением (97). Теорема 17 доказана.

Замечание 17. *В МПГ оператор имеет вид* $Au = \mathrm{pr}_U\big(u - \alpha\, J'(u)\big)$ *и в соответствии с* (95) *непрерывным аналогом дискретного процесса* (92) *будет задача Коши*

$$u'(t) = \beta(t)\,\Big(- u(t) + \mathrm{pr}_U\big(u - \alpha\, J'(u)\big)\Big), \quad t > 0; \quad u(0) = u_0 \in U,$$

в которой коэффициент $\beta(t)$ *следует подчинять требованиям теоремы 17, а параметр* α *— требованию теоремы 16:* $\alpha \in \big(0,\, 2\varkappa/L^2\big)$, *гарантирующему сжимаемость отображения* A *с коэффициентом* $q(\alpha) = 1 - 2\varkappa\alpha + \alpha^2 L^2 \in (0, 1)$.

Метод условного градиента (МУГ) (метод линейной аппроксимации функции)

Этот метод ориентирован на решение задач минимизации с ограничениями (91) :

$$J(u) \to \inf, \quad u \in U \subset H\,.$$

В рамках данного курса будет рассматриваться *классическая версия* МУГ, которая относится к категории *Line Search* и по скорости сходимости будет заметно уступать МПГ. По-видимому, за счёт *Trust-Region*-реконструкции итерационного процесса от данных недостатков можно было бы избавиться, но добиться при этом какого-либо качественного превосходства над МПГ вряд ли удалось бы, поэтому мы воздержимся от подобных упражнений.

Итак, на каждой итерации в МУГ сначала выбирается направление спуска p_k по правилу

$$p_k = \overline{u}_k - u_k, \quad \overline{u}_k = \arg\min_{u\in U} m_k(u), \quad m_k(u) = J(u_k) + \langle J'(u_k),\, u - u_k\rangle\,, \quad (99)$$

т. е. из условия минимума линейной модели (аппроксиманта) $m_k(u)$ функции $J(u)$ в окрестности точки u_k. Затем ищется шаг спуска α_k как точное решение задачи одномерной минимизации:

$$\alpha_k = \arg\min_{\alpha\in[0,1]} J\big(u_k + \alpha\, p_k\big) = \arg\min_{\alpha\in[0,1]} J\big(u_k + \alpha\,(\overline{u}_k - u_k)\big)\,, \qquad (100)$$

и определяется следующее приближение

$$u_{k+1} = u_k + \alpha_k\, p_k = u_k + \alpha_k\,(\overline{u}_k - u_k)\,. \qquad (101)$$

Так как нижняя грань m_{k*} *линейной* модели $m_k(u)$ на *неограниченном* множестве U может оказаться *бесконечной*, в теореме о сходимости МУГ будут рассматриваться только *ограниченные* множества U.

Теорема 18. (о сходимости МУГ)
Пусть H — гильбертово пространство, множество U выпукло, замкнуто, ограничено и его диаметр равен

$$D = \operatorname{diam} U = \sup_{u,v\in U} \|u - v\| < \infty.$$

Пусть функция $J(u)$ выпукла на U, а также $J(u) \in C^1(U)$ и $J'(u) \in \operatorname{Lip}(U)$ с $L > 0$. Тогда МУГ (99) – (101) *сходится по функции из любого начального приближения $u_0 \in U$:*

$$J(u_k) - J_* \leqslant \frac{J(u_0) - J_*}{1 + \left(\frac{J(u_0)-J_*}{2LD^2}\right) k} = O\Big(\frac{1}{k}\Big), \quad k = 0, 1, \ldots, \tag{102}$$

Если, кроме того, $J(u)$ сильно выпукла на U с коэффициентом $\varkappa > 0$, то в дополнение к (102) *МУГ будет сходиться и по аргументу к единственному оптимальному решению u_* задачи* (91) *:*

$$\|u_k - u_*\| \leqslant \sqrt{\frac{2}{\varkappa}\big(J(u_k) - J_*\big)} = O\Big(\sqrt{\frac{1}{k}}\Big), \quad k = 0, 1, \ldots \tag{103}$$

Доказательство теоремы. Заметим, что при дополнительном предположении о сильной выпуклости функции $J(u)$ оценка (102) не улучшается (к сожалению), а (103) является простым следствием оценки (102) и теоремы 8 («сильно выпуклый» вариант теоремы Вейерштрасса). Из теоремы 8 следует также существование и единственность оптимального решения u_*.

Вывод оценки (102) проводится примерно по той же схеме, что и доказательство теоремы 14 о сходимости МСС, однако здесь, в отличие от МСС, для уклонений по функции $a_k = J(u_k) - J_*$ вместо «сжимающей» оценки вида $a_{k+1} \leqslant \big(1 - \frac{\varkappa}{L}\big)\, a_k$ получается более слабая оценка

$$a_{k+1} \leqslant a_k - \frac{1}{2LD^2}\, a_k^2, \quad k = 0, 1, \ldots \tag{104}$$

Более слабой она является из-за присутствия в её правой части a_k^2 вместо a_k, что и приводит к заметному замедлению сходимости. На выводе самой оценки (104) мы останавливаться не будем, а поясним, как из неё получается сходимость (102) со скоростью $O\Big(\frac{1}{k}\Big)$ с помощью следующей леммы.

Лемма. *Пусть имеется монотонно невозрастающая числовая последовательность* a_k :

$$a_k > 0, \quad a_{k+1} \leqslant a_k, \quad k = 0, 1, \ldots, \tag{105}$$

обладающая при некоторм $C > 0$ *свойством*

$$a_k - a_{k+1} \geqslant C\, a_k^2, \quad k = 0, 1, \ldots \tag{106}$$

Тогда справедлива оценка

$$a_k \leqslant \frac{a_0}{1 + C\, a_0\, k} = O\Big(\frac{1}{k}\Big), \quad k = 0, 1, \ldots \tag{107}$$

Доказательство леммы. Для обратных по отношению к a_k величин выполняется оценка

$$\frac{1}{a_{k+1}} - \frac{1}{a_k} = \frac{a_k - a_{k+1}}{a_{k+1}\, a_k} \overset{(106)}{\geqslant} \frac{C\, a_k^2}{a_{k+1}\, a_k} = \frac{C\, a_k}{a_{k+1}} \overset{(105)}{\geqslant} C, \quad k = 0, 1, \ldots$$

После суммирования этих неравенств от 0 до $k - 1$ получим оценку

$$\frac{1}{a_k} - \frac{1}{a_0} \geqslant C\, k,$$

совпадающую с (107). Лемма доказана.

Применяя лемму к последовательности уклонений $a_k = J(u_k) - J_*$ из теоремы 18, обладающей свойством (104), т. е. удовлетворяющей условию (106) с постоянной $C = \frac{1}{2LD^2}$, получим оценку (102). Теорема доказана.

Замечание 18. *На устном экзамене* не требуется *детально и точно воспроизводить правые части итоговой оценки* (102) *скорости сходимости МУГ. Вполне достаточно правильно формулировать условия сходимости и указывать порядки* $O\Big(\frac{1}{k}\Big)$ *и* $O\Big(\sqrt{\frac{1}{k}}\Big)$ *скорости сходимости по функции и по аргументу.*

Метод Ньютона

Метод Ньютона (МН) ориентирован на решение задач минимизации

$$J(u) \to \inf, \quad u \in U \subset H,$$

в которых случай отсутствия ограничений $U = H$ не исключается. На каждом шаге этого метода, как и в методе условного градиента (МУГ), ищется вспомогателное приближение $\overline{u}_k$, доставляющее минимум квадратичной модели (квадратичного аппроксиманта) $m_k(u)$ (в МУГ модель была *линейной*):

$$\overline{u}_k = \arg\min_{u \in U} m_k(u),$$

$$m_k(u) = J(u_k) + \langle J'(u_k),\, u - u_k \rangle + \frac{1}{2} \langle J''(u_k)(u - u_k),\, u - u_k \rangle . \quad (108)$$

В качестве направления спуска выбирается

$$p_k = \overline{u}_k - u_k, \quad (109)$$

а затем либо включается процедура одномерного поиска (*Line Search*) либо можно попробовать обойтись без неё и довериться вспомогательному приближению, взяв $u_{k+1} = \overline{u}_k$. В случае МН, который весьма востребован в современной вычислительной практике, мы достаточно подробно рассмотрим оба подхода. Начнём с более простого, в котором *процедура одномерного поиска отсутствует* и который обычно называют *классическим* вариантом МН:

$$u_{k+1} = \overline{u}_k = \arg\min_{u \in U} m_k(u). \quad (110)$$

Прежде чем формулировать теорему сходимости, посмотрим на *классический* МН в случае, когда ограничения отсутствуют и обе задачи минимизации: исходной гладко-выпуклой функции $J(u)$ и её квадратичной модели $m_k(u)$ будут эквивалентны соответствующим уравнениям Ферма́:

$$J'(u) = 0 \quad \text{и} \quad m_k'(u) = 0 .$$

Найдём производную модельной функции:

$$m_k'(u) = J'(u_k) + J''(u_k)\,(u - u_k)$$

и, учитывая, что в *классическом* МН (110) $m_k'(u_{k+1}) = 0$, находим

$$u_{k+1} = u_k - \big(J''(u_k)\big)^{-1} J'(u_k) .$$

Этот результат буквально совпадает с расчётными формулами для МН (метода касательных), который рассматривался у вас в курсе МА как итерационный метод решения скалярных уравнений $f(x) = 0$:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}\,.$$

Теорема 19. (о локальной сходимости классического МН)
Пусть H — гильбертово пространство, множество U выпукло и замкнуто, функция $J(u) \in C^2(U)$, $J''(u) \in \mathrm{Lip}(U)$ с $L > 0$ и её вторая производная $J''(u) \in L(H \to H)$ является положительно определённым оператором с коэффициентом $\varkappa > 0$ (см. условие (d′) *теоремы 10):*

$$\langle J''(u)h, h\rangle \geqslant \varkappa \|h\|^2 \quad \forall\, u \in U\ \forall\, h \in H\,. \tag{111}$$

Пусть начальное приближение $u_0 \in U$ выбрано из достаточно малой окрестности оптимального решения u_*, такой, что выполнено условие

$$q = \frac{L}{2\varkappa}\,\|u_0 - u_*\| < 1\,. \tag{112}$$

Тогда классический МН (108), (110) *сходится к u_* с* ***квадратичной*** *скоростью:*

$$\|u_k - u_*\| \leqslant \frac{2\varkappa}{L}\,q^{2^k}\,, \quad k = 0, 1, \ldots \tag{113}$$

Замечание 19. *Сходимость по функции $J(u_k) \to J(u_*)$ или по аргументу $\|u_k - u_*\| \to 0$ принято называть* ***квадратичной*** *или q-**квадратичной**, если наряду с самим фактом сходимости при некотором $q \in (0, 1)$ для всех достаточно больших номеров k выполняются условия*

$$|J(u_{k+1}) - J(u_*)| \leqslant q\,|J(u_k) - J(u_*)|^2 \quad \textit{или} \quad \|u_{k+1} - u_*\| \leqslant q\,\|u_k - u_*\|^2\,.$$

Доказательство теоремы. Принятое нами условие (111) обеспечивает свойство сильной выпуклости функции $J(u)$ на множестве U с коэффициентом $\varkappa$ и вместе с тем не обязывает множество U иметь непустую внутренность. При этом квадратичная модель $m_k(u)$ вида (108) будет обладать аналогичным свойством:

$$\langle m_k''(u)h, h\rangle = \langle J''(u_k)h, h\rangle \geqslant \varkappa \|h\|^2 \quad \forall\, h \in H\ \forall\, k = 0, 1, \ldots,$$

т. е. будет сильно выпуклой на множестве U с тем же коэффициентом $\varkappa$. Отсюда, в частности, следует, что оптимальные решения u_* и $u_{k+1} = \overline{u}_k$ исходной

и модельной задачи существуют и единственны, т. е. итерационный процесс (110) полностью определён. Запишем для исходной и модельной задачи критерии оптимальности в форме вариационных неравенств (теорема 11):

$$\langle J'(u_*),\, u - u_* \rangle \geqslant 0 \quad \forall\, u \in U, \tag{114}$$

$$\begin{gathered} \langle m_k'(u_{k+1}),\, u - u_{k+1} \rangle = \\ = \langle J'(u_k) + J''(u_k)(u_{k+1} - u_k),\, u - u_{k+1} \rangle \geqslant 0 \quad \forall\, u \in U. \end{gathered} \tag{115}$$

Подставим в (114) $u := u_{k+1}$, в (115) $u := u_*$ и сложим полученные неравенства:

$$\langle J'(u_*) - J'(u_k) - J''(u_k)(u_{k+1} - u_k),\, u_{k+1} - u_* \rangle \geqslant 0\,. \tag{116}$$

По формуле конечных приращений имеем

$$\begin{gathered} \langle J'(u_*) - J'(u_k),\, u_{k+1} - u_* \rangle = \\ = \int_0^1 \langle J''\big(u_k + t\,(u_* - u_k)\big)\,(u_* - u_k),\, u_{k+1} - u_* \rangle\, dt\,. \end{gathered} \tag{117}$$

Представим слагаемое $J''(u_k)(u_{k+1} - u_k)$ в виде суммы

$$J''(u_k)(u_{k+1} - u_k) = J''(u_k)(u_{k+1} - u_*) + J''(u_k)(u_* - u_k)$$

и, учитывая (117), перепишем (116):

$$\begin{gathered} \int_0^1 \Big\langle \Big(J''\big(u_k + t\,(u_* - u_k)\big) - J''(u_k)\Big)\,(u_* - u_k),\, u_{k+1} - u_* \Big\rangle\, dt \geqslant \\ \geqslant \langle J''(u_k)(u_{k+1} - u_*),\, u_{k+1} - u_* \rangle\,. \end{gathered} \tag{118}$$

Левая часть (118) оценивается *сверху* величиной ($J''(u) \in \mathrm{Lip}(U)$ с $L > 0$):

$$L \int_0^1 \|t\,(u_* - u_k)\|\, \|u_* - u_k\|\, \|u_{k+1} - u_*\|\, dt = \frac{L}{2}\, \|u_* - u_k\|^2\, \|u_{k+1} - u_*\|\,,$$

а правая часть (118) оценивается *снизу* (см. (111)):

$$\langle J''(u_k)(u_{k+1} - u_*),\, u_{k+1} - u_* \rangle \geqslant \varkappa\, \|u_{k+1} - u_*\|^2\,.$$

В результате получается рекуррентное соотношение

$$\|u_{k+1} - u_*\| \leqslant \frac{L}{2\varkappa}\, \|u_* - u_k\|^2\,,$$

из него по индукции выводится искомая оценка (113), которая вместе с условием (112) подтверждает *квадратичную* скорость *локальной* сходимости *классического* МН (110). Теорема доказана.

Замечание 20. *Заметим, что итерационная процедура даже классического МН, в которой отсутствует одномерный поиск, и даже для задач без ограничений, в которых $U = H$, в вычислительном плане весьма затратна, поскольку требует на каждом шаге при нахождении очередного направления спуска $p_k = -\big(J''(u_k)\big)^{-1} J'(u_k)$ обращения гессиана $J''(u_k)$. Понятно, что при наличии ограничений ($U \neq H$) трудоёмкость реализации МН лишь возрастёт. Одним из главных недостатков классического МН является отсутствие у него свойства* ***глобальной*** *сходимости.*

Метод Ньютона (продолжение)

Чтобы гарантировать *глобальную* сходимость МН, не следует полностью доверять минимизаторам $\overline{u}_k$ квадратичной модели. Рекомендуется при выборе очередного приближения u_{k+1} включать процедуру одномерного поиска (не обязательно точного). Эту процедуру можно организовывать по-разному. Рассмотрим один из подходящих вариантов, известный в зарубежной литературе под названием **«backtracking approach».** Будем предполагать, что функция $J(u)$ и множество U в рассматриваемой задаче минимизации

$$J(u) \to \inf, \quad u \in U \subset H,$$

обладают всеми перечисленными в теореме 19 свойствами, и дополним эти условия требованием *ограниченности гессиана*:

$$\exists\, M = \mathrm{const} > 0 : \quad \langle J''(u)h, h\rangle \leqslant M\,\|h\|^2 \quad \forall\, u \in U\ \forall\, h \in H\,. \tag{119}$$

От требования (112) к выбору стартовой точки u_0 мы отказываемся. На k-ой итерации очередное приближение u_{k+1} ищется в виде

$$u_{k+1} = u_k + \alpha_k\, p_k, \quad p_k = \overline{u}_k - u_k\,, \tag{120}$$

где $\overline{u}_k$ — минимизатор квадратичной модели $m_k(u)$, а шаг $\alpha_k \in (0, 1]$ ищется с помощью вспомогательной итерационной *процедуры дробления (backtracking).* На каждом шаге этих внутренних итераций по α проверяется условие, оценивающее прогресс в убывании функции $J(u)$ по отношению к уменьшению значения модели $m_k(u)$:

$$J(u_k + \alpha\, p_k) - J(u_k) \leqslant \frac{\alpha}{2}\left(m_k(\overline{u}_k) - m_k(u_k)\right). \tag{121}$$

Если условие (121) выполняется, значение α считается подходящим, в противном случае производится его дробление с помощью заранее выбранного коэффициента дробления $\lambda \in (0, 1)$, например, $\lambda = 1/2$. В МН последовательно перебираются значения

$$\alpha = \lambda^0 = 1,\ \lambda^1,\ \lambda^2, \ldots$$

находят *наименьший* номер

$$m = \min_{j=0,1,\ldots} j\ : \quad \text{выполнено (121) при } \alpha = \lambda^j\,, \tag{122}$$

после чего полагают

$$\alpha_k = \lambda^m \tag{123}$$

и переходят к следующему приближению (120).

Убедимся в том, что указанный в (122), (123), (121) номер m обязательно найдётся. Заметим, что если вдруг $\overline{u}_k = u_k$, то критерий оптимальности $\langle m_k'(\overline{u}_k), u - \overline{u}_k\rangle \geqslant 0 \;\forall\, u \in U$ для гладко-выпуклой задачи (110) примет вид $\langle J'(u_k), u - u_k\rangle \geqslant 0 \;\forall\, u \in U$, означающий, что оптимальное решение $u_k = u_*$ уже найдено и процесс останавливается. Если же $\overline{u}_k \neq u_k$, то правая часть неравенства (121) будет отрицательной:

$$m_k(\overline{u}_k) - m_k(u_k) < 0\,. \tag{124}$$

Оценим левую часть (121) для значений $\alpha \in [0,1]$:

$$\begin{aligned} J(u_k + \alpha\, p_k) - J(u_k) &= \int_0^1 \big\langle J'(u_k + t\alpha\, p_k) \mp J'(u_k),\, \alpha\, p_k\big\rangle\, dt = \\ &= \big\langle J'(u_k),\, \alpha\, p_k\big\rangle + \int_0^1 \big\langle J'(u_k + t\alpha\, p_k) - J'(u_k),\, \alpha\, p_k\big\rangle\, dt = \\ &= \big\langle J'(u_k),\, \alpha\, p_k\big\rangle + \int_0^1 dt \int_0^1 \big\langle J''(u_k + st\alpha\, p_k)\,(t\alpha\, p_k),\, \alpha\, p_k\big\rangle\, ds \overset{(119)}{\leqslant} \\ &\leqslant \big\langle J'(u_k),\, \alpha\, p_k\big\rangle + \frac{M\,\alpha^2}{2}\,\|p_k\|^2\,. \end{aligned} \tag{125}$$

Заметим, что

$$\begin{aligned} m_k(u_k + \alpha\, p_k) - m_k(u_k) &\overset{(108)}{=} \big\langle J'(u_k),\, \alpha\, p_k\big\rangle + \frac{\alpha^2}{2}\,\langle J''(u_k)\, p_k,\; p_k\rangle \overset{(111)}{\geqslant} \\ &\geqslant \big\langle J'(u_k),\, \alpha\, p_k\big\rangle + \frac{\varkappa\alpha^2}{2}\,\|p_k\|^2\,. \end{aligned} \tag{126}$$

С учётом (126) переходим от (125) к оценке

$$\begin{aligned} J(u_k + \alpha\, p_k) - J(u_k) &\leqslant m_k(u_k + \alpha\, p_k) - m_k(u_k) + \frac{(M-\varkappa)\alpha^2}{2}\,\|p_k\|^2 \leqslant \\ &\leqslant \big[m_k(u) \text{ выпукла } \Rightarrow m_k(u_k + \alpha\, p_k) \leqslant \alpha m_k(\overline{u}_k) + (1-\alpha) m_k(u_k)\big] \leqslant \\ &\leqslant \alpha\,\big(m_k(\overline{u}_k) - m_k(u_k)\big) + \frac{(M-\varkappa)\alpha^2}{2}\,\|p_k\|^2 \leqslant \\ &\leqslant \Big[m_k(u) \text{ сильно выпукла } \Rightarrow \frac{\varkappa}{2}\,\|p_k\|^2 \leqslant \big(m_k(u_k) - m_k(\overline{u}_k)\big)\Big] \leqslant \\ &\leqslant \Big(\alpha - \frac{(M-\varkappa)\alpha^2}{\varkappa}\Big)\,\big(m_k(\overline{u}_k) - m_k(u_k)\big) \quad \forall\,\alpha \in [0,1]\,. \end{aligned} \tag{127}$$

Сопоставляя (127) с (121), заключаем, что для выполнения условия (121) достаточно потребовать выполнения неравенства

$$\alpha - \frac{(M-\varkappa)\,\alpha^2}{\varkappa} \geqslant \frac{\alpha}{2} \quad\Longleftrightarrow\quad 0 \leqslant \alpha \leqslant \frac{\varkappa}{2(M-\varkappa)}\,. \tag{128}$$

При фиксированном $\lambda \in (0,1)$ мы ищем шаг α_k в степенной форме (123), поэтому ясно, что найдётся такой достаточно большой номер j, для которого

$$\lambda^j \leqslant \frac{\varkappa}{2(M-\varkappa)} < \lambda^{j-1}\,. \tag{129}$$

На практике, скорее всего, процесс поиска подходящего шага α_k, удовлетворяющего условию (121), остановится раньше на значении

$$\alpha_k = \lambda^m \geqslant \lambda^j \overset{(129)}{>} \frac{\lambda\varkappa}{2(M-\varkappa)} = \mathrm{const} > 0\,. \tag{130}$$

Условие (130) гарантирует отделённость шагов α_k от нуля и позволяет извлечь из (121) для $\alpha = \alpha_k$ следствие

$$J(u_{k+1}) - J(u_k) \leqslant \frac{\lambda\varkappa}{4(M-\varkappa)}\left(m_k(\overline{u}_k) - m_k(u_k)\right), \quad k = 0, 1, \ldots \tag{131}$$

Отсюда и из (124) следует *монотонность* числовой последовательности $J(u_k)$, которая к тому же *ограничена снизу* оптимальным значением $J(u_*)$, поэтому она *сходится*, а, значит, *фундаментальна* и тогда

$$J(u_{k+1}) - J(u_k) \to 0 \quad \text{при} \quad k \to \infty.$$

Учитывая, что обе части (131) *отрицательны*, получаем из (131) сходимость

$$m_k(\overline{u}_k) - m_k(u_k) \to 0 \quad \text{при} \quad k \to \infty.$$

Воспользовавшись, как и в (127), свойством $\frac{\varkappa}{2}\|p_k\|^2 \leqslant \left(m_k(u_k) - m_k(\overline{u}_k)\right)$ сильной выпуклости модельной функции $m_k(u)$, получим сходимость

$$\|p_k\| \to 0 \quad \text{при} \quad k \to \infty. \tag{132}$$

Теперь покажем, что обе последовательности u_k и $\overline{u}_k$ сходятся к оптимальному решению u_*. Обратимся к неравенству (116) из доказательства теоремы 19, в котором точки u_{k+1} совпадали с $\overline{u}_k$:

$$\langle J'(u_*) - J'(u_k) \pm J'(\overline{u}_k) - J''(u_k)(\overline{u}_k - u_k),\, \overline{u}_k - u_*\rangle \geqslant 0$$

и перепишем его в виде

$$\langle J'(\overline{u}_k) - J'(u_k),\, \overline{u}_k - u_*\rangle + \langle J''(u_k)(u_k - \overline{u}_k),\, \overline{u}_k - u_*\rangle \geqslant$$
$$\geqslant \langle J'(\overline{u}_k) - J'(u_*),\, \overline{u}_k - u_*\rangle \geqslant \varkappa\,\|\overline{u}_k - u_*\|^2\,. \tag{133}$$

Левую часть (133) оценим сверху следующим образом:

$$\langle J'(\overline{u}_k) - J'(u_k),\, \overline{u}_k - u_*\rangle \leqslant \|J'(\overline{u}_k) - J'(u_k)\|\, \|\overline{u}_k - u_*\| \leqslant$$
$$\leqslant \frac{1}{\varkappa}\, \|J'(\overline{u}_k) - J'(u_k)\|^2 + \frac{\varkappa}{4}\, \|\overline{u}_k - u_*\|^2 \leqslant \frac{M^2}{\varkappa}\, \|\overline{u}_k - u_k\|^2 + \frac{\varkappa}{4}\, \|\overline{u}_k - u_*\|^2 ;$$
$$\langle J''(u_k)(u_k - \overline{u}_k),\, \overline{u}_k - u_*\rangle \leqslant M\, \|u_k - \overline{u}_k\|\, \|\overline{u}_k - u_*\| \leqslant \frac{M^2}{\varkappa}\, \|u_k - \overline{u}_k\|^2 + \frac{\varkappa}{4}\, \|\overline{u}_k - u_*\|^2 .$$

Отсюда и из (133) следует, что

$$\frac{\varkappa}{2}\, \|\overline{u}_k - u_*\|^2 \leqslant \frac{2\, M^2}{\varkappa}\, \|u_k - \overline{u}_k\|^2 = \frac{2\, M^2}{\varkappa}\, \|p_k\|^2 \overset{(132)}{\to} 0 ; \quad \|u_k - u_*\| \to 0 . \tag{134}$$

Имея сходимость (134), покажем, что при использовании правила (121) для выбора шагов α_k их значения через конечное число итераций станут постоянными:

$$\exists\, k_0 : \quad \alpha_k = 1 \quad \forall\, k \geqslant k_0 . \tag{135}$$

Вернёмся к (126), преобразуем использованое в нём представление

$$m_k(u_k + \alpha\, p_k) - m_k(u_k) = \big\langle J'(u_k),\, \alpha\, p_k\big\rangle + \frac{\alpha^2}{2}\, \langle J''(u_k)\, p_k,\ p_k\rangle =$$
$$= \big\langle J'(u_k),\, \alpha\, p_k\big\rangle + \int_0^1 dt \int_0^1 \big\langle J''(u_k)\, (t\alpha\, p_k),\, \alpha\, p_k\big\rangle\, ds$$

и подставим его в (125):

$$J(u_k + \alpha\, p_k) - J(u_k) = m_k(u_k + \alpha\, p_k) - m_k(u_k) +$$
$$+ \int_0^1 dt \int_0^1 \big\langle \big(J''(u_k + st\alpha\, p_k) - J''(u_k)\big)\, (t\alpha\, p_k),\, \alpha\, p_k\big\rangle\, ds \leqslant$$
$$\leqslant \big[\ \text{см. (127);}\ J'' \in \mathrm{Lip}\ \big] \leqslant \alpha\, \big(m_k(\overline{u}_k) - m_k(u_k)\big) + \frac{L\, \alpha^3}{6}\, \|p_k\|^3 \leqslant$$
$$\leqslant \big[\ m_k(u) \text{ сильно вып. }\big] \leqslant \alpha\, \Big(-1 + \frac{L\, \alpha^2}{3\varkappa}\, \|p_k\|\Big)\, \big(m_k(u_k) - m_k(\overline{u}_k)\big) =$$
$$= \alpha\, \Big(1 - \frac{L\, \alpha^2}{3\varkappa}\, \|p_k\|\Big)\, \big(m_k(\overline{u}_k) - m_k(u_k)\big) . \tag{136}$$

Обратим внимание на то, что неравенство (136) выполняется для всех $k = 0, 1, \ldots$ и для любых $\alpha \in [0, 1]$, в частности, для $\alpha = 1$, а тогда в силу сходимости (132) это будет означать, что для всех достаточно больших номеров $k \geqslant k_0$ в (136) при $\alpha = 1$ мы будем иметь

$$1 - \frac{L\, \alpha^2}{3\varkappa}\, \|p_k\| \geqslant \frac{1}{2}$$

и, тем самым, правило выбора шага (121) будет для $k \geqslant k_0$ вырабатывать значения $\alpha_k = 1$. В таком случае при $k \geqslant k_0$ приближения $u_{k+1} = \overline{u}_k$ будут определяться по тому же правилу, что и в классическом МН из теоремы 19 и обладать установленным в этой теореме свойством сходимости с *квадратичной* скоростью. Параметр q в соответствующей оценке скорости сходимости, который в теореме 19 был равен $q = \frac{L}{2\varkappa}\|u_0 - u_*\|$ и обязан был быть меньше 1, при этом замещается значением $q_m = \frac{L}{2\varkappa}\|u_m - u_*\|$, которое в силу сходимости (134) будет заведомо меньше 1 при достаточно больших m.

Теорема 20. `(о глобальной сходимости МН)`
Пусть исходные данные H, U и $J(u)$ удовлетворяют всем условиям теоремы 19 и дополнительному условию (119). *Тогда итерационный процесс МН u_k, организованный по правилам* (120) – (123), *сходится к оптимальному решению u_* из* ***любого*** *начального приближения $u_0 \in U$ с* ***квадратичной*** *скоростью.*

Замечание 21. *Обратим внимание на то, что* ***для практической реализации*** *описанных вариантов МН* ***не требуется*** *знать значения постоянных L, $\varkappa$, M, которые использовались нами в доказательствах сходимости. Что касается устного экзамена, то требуется знать конструкцию приближений обоих итерационных процессов: классического и глобально сходящегося, а также уметь характеризовать скорости их сходимости.*

Замечание 22. *В настоящее время по мере возрастания возможностей вычислительной техники МН становится всё более востребованным, тем не менее, пока более популярными остаются так называемые* ***квазиньютоновкие методы****, в которых используются модельные квадратичные функции $m_k(u)$, отличающиеся от ньютоновских квадратичным фрагментом, в котором точный гессиан $J''(u_k)$ замещается некоторыми его приближениями B_k, позволяющими заметно снизить трудоёмкость вычислений. Разумеется, при этом по сравнению с ньютоновской квадратичной снижается скорость сходимости, но, как правило, её удаётся сохранить на* ***сверхлинейном*** *уровне, превосходящем методы градиентного типа. Освоение конструктивных особенностей квазиньютоновских методов и исследование их сходимости являются достаточно сложной материей, выходящей за рамки данного курса. Для более близкого знакомства с ними рекомендуется книга [5].*

Метод сопряжённых градиентов

Метод сопряжённых градиентов (МСГ) остаётся *наиболее популярным* в современной вычислительной практике *среди методов градиентного типа* и, как правило, превосходит их и по скорости и по точности и по устойчивости, однако теоретически обладает такой же, как и они, *линейной скоростью сходимости*, которая к тому же технически сложнее доказывается и поэтому в эти вопросы мы углубляться не будем. МСГ постепенно уступает позиции более трудоёмким, но зато и более быстрым ньютоновским и квазиньютоновским методам, а также гибридным процессам, сочетающим конструкции МСГ и МН. Смысл используемых в МСГ расчётных формул удобнее всего пояснить *на классе конечномерных квадратичных задач безусловной минимизации сильно выпуклых функций*:

$$J(u) = \frac{1}{2}\,\langle Au,\, u\rangle - \langle f,\, u\rangle \to \inf, \quad u \in H = R^n\,, \tag{137}$$

где $A = A^\top = A^* > 0$ — квадратная симметричная положительно определённая матрица размера $n \times n$, $f \in R^n$ — заданный вектор. Поскольку в этой задаче ограничения отсутствуют, она эквивалентна уловию Фермá $J'(u) = 0$, принимающему в случае (137) вид СЛАУ

$$Au = f, \tag{138}$$

для решения которой существует много различных конечношаговых методов. МСГ можно расссматривать как один из таких итерационных процессов, который для задач вида (137) или (138) вряд ли обладает какими-то заметными преимуществами перед другими методами, но зато после незначительных модификаций превращается в достаточно мощный инструмент численного решения широкого класса задач безусловной минимизации:

$$J(u) \to \inf, \quad u \in H\,,$$

с гладкими функциями $J(u)$ общего вида, а также и в бесконечномерных пространствах H. Именно благодаря таким успешным применениям МСГ и приобрёл свою популярность.

Опишем итерационную процедуру МСГ для конечномерной квадратичной задачи (137) или, что то же самое, для СЛАУ (138). Обозначим единственное решение этих задач через u_* и предположим, что в нашем распоряжении

имеется базис $\{p_k\}_{k=0}^{n-1}$ пространства R^n, обладающий свойствами попарной ортогональности относительно оператора A :

$$\langle Ap_k,\, p_m\rangle = 0 \quad \forall\, k \neq m. \tag{139}$$

В своё время, когда МСГ только разрабатывался, условие (139) было названо «условием сопряжённости» и впоследствии закрепилось в названии метода. Несколько странная нумерация базисных векторов выбрана для того, чтобы без последующих переобозначений придерживаться общепринятой формы описания итерационного процесса МСГ, в котором на каждом шаге наряду с очередным приближением к решению u_{k+1} будет подбираться и очередной базисный вектор p_{k+1}. Если же предположить, что базис $\{p_k\}_{k=0}^{n-1}$ уже нам известен, то для любого элемента $u_0 \in R^n$, выбранного нами в качестве начального приближения, мы могли бы разность $u_* - u_0$ (как и любой другой вектор из R^n) разложить по этому базису:

$$u_* - u_0 = \alpha_0\, p_0 + \alpha_1\, p_1 + \cdots + \alpha_{n-1}\, p_{n-1}\,, \tag{140}$$

а затем найти коэффициенты разложения в (125) последовательным скалярным домножением на Ap_k :

$$\begin{aligned}\langle u_* - u_0,\, Ap_k\rangle &= \alpha_0\,\langle p_0, Ap_k\rangle + \alpha_1\,\langle p_1,\, Ap_k\rangle + \cdots + \alpha_{n-1}\,\langle p_{n-1},\, Ap_k\rangle \overset{(139)}{=} \\ &= \alpha_k\,\langle p_k,\, Ap_k\rangle \quad \overset{Au_*=f}{\Longrightarrow} \quad \alpha_k = \frac{\langle f - Au_0,\, p_k\rangle}{\langle Ap_k,\, p_k\rangle}, \quad k = 0, 1, \ldots, n-1. \end{aligned} \tag{141}$$

В самом же МСГ предлагается строить следующие приближения и очередные базисные векторы, ориентируясь на (139) – (141), следующим образом:

$$u_0 \in R^n - \text{любая}\,, \quad p_0 = -J'(u_0)\,, \tag{142}$$

$$u_{k+1} = u_k + \alpha_k\, p_k, \quad p_{k+1} = -J'(u_{k+1}) + \beta_k\, p_k\,, \tag{143}$$

$$\alpha_k = \frac{\langle f - Au_0,\, p_k\rangle}{\langle Ap_k,\, p_k\rangle}, \quad \beta_k = \frac{\langle J'(u_{k+1}),\, Ap_k\rangle}{\langle Ap_k,\, p_k\rangle}, \quad k = 0, 1, \ldots \tag{144}$$

Замечание 23. *В* (144) *значения шагов* α_k *взяты непосредственно из формул* (141), *верных в случае, когда базис* $\{p_k\}_{k=0}^{n-1}$ *известен нам заранее. Что касается шагов* β_k, *то в* (144) *они выбираются с таким расчётом, чтобы новое базисное направление* p_{k+1} *получилось «сопряжённым» по отношению к* p_k *в смысле условий* (139).

Из конструкций (142) – (144) пока неясно, будет ли описываемый итерационный процесс конечным или бесконечным и будет ли он сходиться к искомому решению u_*. Чтобы ответить на эти вопросы, сначала дакажем вспомогательное утверждение.

Лемма. *Пусть* $\{p_k\}_{k=0}^{n-1}$ *— базис в* R^n, *состоящий из элементов, удовлетворяющих условиям попарной ортогональности* (139) *относительно матрицы* $A = A^\top > 0$, *точка* $u_0 \in R^n$ *выбрана произвольно, а элементы* u_k *и коэффициенты* α_k *определяются по правилам* (143) *и* (141). *Тогда для всех номеров* $k = 0, 1, \ldots, n-1$ *выполняются условия*

$$\alpha_k = \frac{\langle -J'(u_k),\, p_k\rangle}{\langle Ap_k,\, p_k\rangle} = \arg\min_{\alpha\in R^1} J(u_k + \alpha\, p_k)\,, \tag{145}$$

$$J'(u_{k+1}) - J'(u_k) = \alpha_k\, Ap_k\,, \tag{146}$$

$$\langle J'(u_{k+1}),\, p_k\rangle = 0\,. \tag{147}$$

Доказательство. Для рассматриваемой квадратичной функции (137) градиент равен

$$J'(u) = Au - f\,, \tag{148}$$

поэтому

$$J'(u_{k+1}) - J'(u_k) \overset{(148)}{=} Au_{k+1} - Au_k \overset{(143)}{=} \alpha_k\, Ap_k\,,$$

т. е. соотношение (146) доказано. Первое равенство в (145) верно, поскольку

$$\langle -J'(u_k),\, p_k\rangle \overset{(148)}{=} \langle f - Au_k,\, p_k\rangle \overset{(143)}{=} \langle f - A(u_0 + \alpha_0 p_0 + \cdots + \alpha_{k-1}p_{k-1}),\, p_k\rangle \overset{(139)}{=}$$
$$= \langle f - Au_0,\, p_k\rangle\,.$$

Рассматривая значения функии $f_k(\alpha) = J(u_k + \alpha p_k)$ и решая точно задачу одномерного поиска $f_k(\alpha) \to \inf$ с помощью условия Фермá $f_k'(\alpha) = 0$, получим второе равенство в (145):

$$f_k'(\alpha) = \langle Ap_k,\, p_k\rangle\,\alpha + \langle Au_k - f,\, p_k\rangle \overset{(148)}{=} \langle Ap_k,\, p_k\rangle\,\alpha + \langle J'(u_k),\, p_k\rangle = 0\,,$$
$$\Longrightarrow \quad \alpha = \arg\min_{\alpha\in R^1} J(u_k + \alpha p_k) = \frac{\langle -J'(u_k),\, p_k\rangle}{\langle Ap_k,\, p_k\rangle}\,.$$

Из (145) и (146) следует свойство (147):

$$\langle J'(u_{k+1}),\, p_k\rangle \overset{(146)}{=} \langle J'(u_k) + \alpha_k\, Ap_k,\, p_k\rangle \overset{(145)}{=} 0\,.$$

Лемма доказана.

Перед тем как формулировать теорему сходимости МСГ, рассмотрим «нештатные» ситуации, которые могут возникнуть при реализации алгоритма (142) – (144). Речь идёт о зацикливании $u_{k+1} = u_k$ или равенстве нулю значенателей $\langle Ap_k,\, p_k\rangle$ в выражениях (144) для шагов α_k и β_k. Оказывается, все такие «нештатные» ситуации сигнализируют о том, что оптимальное решение u_* уже найдено и вычисления следует остановить. Действительно, если $\langle Ap_k,\, p_k\rangle = 0$, то в силу положительной определённости матрицы A имеем

$$p_k \overset{(143)}{=} -J'(u_k) + \beta_{k-1}\, p_{k-1} = 0 \quad \Longrightarrow \quad \|J'(u_k)\|^2 = \langle J'(u_k),\, J'(u_k)\rangle =$$
$$= \langle J'(u_k),\, \beta_{k-1}\, p_{k-1}\rangle \overset{(147)}{=} 0 \quad \Longrightarrow \quad J'(u_k) = 0 \quad \Longrightarrow \quad u_k = u_*\,.$$

В случае зацикливания в силу (143) будет равно нулю произведение $\alpha_k\, p_k$. Случай $p_k = 0$ мы только что рассмотрели. Если же $\alpha_k = 0$, то в соответствии с (145) имеем

$$\langle -J'(u_k),\, p_k\rangle = 0 \overset{(143)}{=} \langle -J'(u_k),\, -J'(u_k) + \beta_{k-1}\, p_{k-1}\rangle \overset{(147)}{=} \|J'(u_k)\|^2$$
$$\Longrightarrow \quad J'(u_k) = 0 \quad \Longrightarrow \quad u_k = u_*\,.$$

Теорема 21. `(о конечной сходимости МСГ)`
Для квадратичной задачи безусловной минимизации (137) *с матрицей* $A = A^\top > 0$ *итерационный процесс* (142) – (144) *обладает свойствами*

$$\langle Ap_k,\, p_m\rangle = 0 \qquad \forall\, k \neq m\,, \tag{149}$$
$$\langle J'(u_k),\, J'(u_m)\rangle = 0 \qquad \forall\, k \neq m\,, \tag{150}$$
$$\langle J'(u_k),\, p_m\rangle = 0 \qquad \forall\, m \leqslant k-1\,, \tag{151}$$

и, как следует из (149), *сходится к оптимальному решению* u_* *за* ***конечное число шагов****, не превосходящее* $n = \dim R^n$.

Доказательство. Утверждения теоремы будем доказывать по индукции. Сначала убеждаемся в том, что все они действительно выполняются для начальных номеров $k = 1$ и $m = 0$. Затем делаем индуктивное предположение о том, что все три свойства (149) – (151) выполняются для номеров, не превосходящих k. Покажем, что в таком случае эти свойства останутся верными и для всех номеров, включая $k+1$. «Нештатные» ситуации мы уже проанализировали, поэтому снова их рассматривать не будем.

Начнём с утверждения (150), взяв максимальные номера k и $k+1$:

$$\langle J'(u_{k+1}),\, J'(u_k)\rangle \overset{(146)}{=} \langle J'(u_k),\, J'(u_k)\rangle + \alpha_k \langle Ap_k,\, J'(u_k)\rangle =$$
$$= \big[\, \langle Ap_k,\, J'(u_k)\rangle \overset{(143)}{=} \langle Ap_k,\, -p_k + \beta_{k-1}\, p_{k-1}\rangle \overset{(149)}{=} -\langle Ap_k,\, p_k\rangle \,\big] =$$
$$= \langle J'(u_k),\, J'(u_k)\rangle - \alpha_k \langle Ap_k,\, p_k\rangle = 0\,, \quad (152)$$

поскольку

$$\alpha_k \overset{(145)}{=} \frac{\langle -J'(u_k),\, p_k\rangle}{\langle Ap_k,\, p_k\rangle} \overset{(143)}{=} \frac{\langle -J'(u_k),\, -J'(u_k) + \beta_{k-1}\, p_{k-1}\rangle}{\langle Ap_k,\, p_k\rangle} \overset{(151)}{=} \frac{\langle J'(u_k),\, J'(u_k)\rangle}{\langle Ap_k,\, p_k\rangle}\,.$$

Ортогональность градиентов для номеров $k+1$ и $m \leqslant k-1$ устанавливается проще:

$$\langle J'(u_{k+1}),\, J'(u_m)\rangle \overset{(146)}{=} \langle J'(u_k),\, J'(u_m)\rangle + \alpha_k \langle Ap_k,\, J'(u_m)\rangle \overset{(150),(143)}{=}$$
$$= \alpha_k \langle Ap_k,\, -p_m + \beta_{m-1}\, p_{m-1}\rangle \overset{(149)}{=} 0\,.$$

Теперь обратимся к свойству (151). Для соседних номеров $k+1$ и k оно уже доказано в лемме (см. (147)), а для номеров $k+1$ и $m \leqslant k-1$ имеем

$$\langle J'(u_{k+1}),\, p_m\rangle \overset{(146)}{=} \langle J'(u_k),\, p_m\rangle + \alpha_k \langle Ap_k,\, p_m\rangle \overset{(149),(151)}{=} 0\,.$$

Остаётся проверить главное свойство (149), гарантирующее конечную сходимость МСГ. Для соседних номеров $k+1$ и k оно верно в силу соответствующего выбора коэффициента β_k в (144), а для номеров $k+1$ и $m \leqslant k-1$ будем иметь

$$\langle Ap_m,\, p_{k+1}\rangle \overset{(143)}{=} \langle Ap_m,\, -J'(u_{k+1}) + \beta_k\, p_k\rangle \overset{(149)}{=} \langle Ap_m,\, -J'(u_{k+1})\rangle \overset{(146)}{=}$$
$$= -\frac{1}{\alpha_m}\langle J'(u_{m+1}) - J'(u_m),\, J'(u_{k+1})\rangle \overset{(150)}{=} 0\,.$$

Теорема 21 доказана.

В заключение остановимся кратко на описании таких модификаций расчётных формул (142) – (144), которые были бы пригодны для применения к задачам минимизации с функциями $J(u)$ общего вида. Понятно, что в обязательной коррекции нуждаются лишь формулы (144), содержащие оператор A, не привязанный к функции $J(u)$. Для α_k мы уже получили в лемме более универсальную версию, подобную той, что используется в МСС:

$$\alpha_k = \arg\min_{\alpha \in R^1} J(u_k + \alpha p_k)\,. \quad (153)$$

Теперь обратимся к β_k. Заметим, что в квадратичном случае

$$\beta_k = \frac{\langle J'(u_{k+1}),\, Ap_k\rangle}{\langle Ap_k,\, p_k\rangle} = \frac{\alpha_k \langle J'(u_{k+1}),\, Ap_k\rangle}{\alpha_k \langle Ap_k,\, p_k\rangle} \overset{(146)}{=}$$

$$= \frac{\langle J'(u_{k+1}),\, J'(u_{k+1}) - J'(u_k)\rangle}{\langle J'(u_{k+1}) - J'(u_k),\, p_k\rangle} =$$

$$= \left[\langle J'(u_{k+1}),\, p_k\rangle \overset{(151)}{=} 0, \quad p_k \overset{(143)}{=} -J'(u_k) + \beta_{k-1}\, p_{k-1}, \quad \langle J'(u_k),\, p_{k-1}\rangle \overset{(151)}{=} 0\right] =$$

$$= \frac{\langle J'(u_{k+1}),\, J'(u_{k+1}) - J'(u_k)\rangle}{\|J'(u_k)\|^2}. \quad (154)$$

Формулами (142), (143), (153) и (154) можно пользоваться при решении задач безусловной минимизации функций $J(u)$ общего вида, в том числе и в бесконечномерных пространствах. Сверхлинейной скорости сходимости МСГ не достигает, но обычно превосходит другие градиентные методы и по скорости и по точности и по устойчивости к различного рода погрешностям, округлениям и пр. Отметим также, что для квадратичных функций можно выписать много других внешне отличающихся от (154) выражений для одного и того же значения β_k, которые будут существенно отличаться друг от друга для функций, не являющихся квадратичными, причём научные статьи с исследованиями соответствующих версий МСГ до сих пор пишутся и публикуются.

Задачи линейного программирования. Симплекс-метод

Симплекс-метод ориентирован на решение *конечномерных* задач минимизации *линейных функций* на множествах, заданных *конечным* числом *линейных ограничений* типа равенств и неравенств:

$$J(u) = \langle c, u\rangle \to \inf, \quad u \in U = \{u \in R^n \mid Au = b,\ Du \leqslant f\}. \tag{155}$$

Здесь заданы матрица A размера $m \times n$, матрица D размера $k \times n$, и векторы $c \in R^n,\ b \in R^m,\ f \in R^k$. Задачи вида (155) называют *общими задачами линейного программирования (ЛП),* а допустимое множество U — *многогранным множеством* или *многогранником.*

Процедуру симплекс-метода удобнее описывать на подклассе задач ЛП, которые называют *каноническими*:

$$J(u) = \langle c, u\rangle \to \inf, \quad u \in U_0 = \{u \in R^n \mid Au = b,\ u \geqslant 0\}\,. \tag{156}$$

Допустимое множество U_0 из канонической постановки (156) часто называют *каноническим многогранником* или *каноническим симплексом.* Заметим, что любая общая задача ЛП (155) может быть приведена к каноническому виду (156), правда, к сожалению, за счёт заметного повышения размерности. Один из возможных способов такого сведéния основан на том, что любое число x представи́мо в виде разности двух неотрицательных чисел, например,

$$x = \max\{x, 0\} - \max\{-x, 0\}\,.$$

Каждая из переменных u_i общей задачи (155) заменяется разностью

$$u_i = v_i - w_i, \quad v_i \geqslant 0, \quad w_i \geqslant 0\,,$$

и вводятся дополнительные переменные

$$y = f - D(v - w), \quad y \geqslant 0\,.$$

В результате от n-мерной общей задачи ЛП (155) мы перейдём к $(2n + k)$-мерной канонической задаче ЛП относительно переменных $z = (v, w, y)$ с функцией

$$I(z) = \langle c, v - w\rangle\,,$$

ограничениями типа равенств

$$A(v - w) = b, \quad D(v - w) + y = f$$

и ограничениями-неравенствами канонического вида:

$$v \geqslant 0, \quad w \geqslant 0, \quad y \geqslant 0\,.$$

Далее мы будем иметь дело только с каноническими задачами (156).

На простых примерах легко понять, что в задачах ЛП, как канонических, так и общих, особую роль играют *вершины* многогранных множеств — их *угловые* точки.

Определение 16. *Пусть L — линейное пространство и $U \subset L$ — некоторое множество. Точка $v \in U$ этого множества называется* ***угловой*** *или* ***крайней****, если из представления*

$$v = \alpha\, x + (1-\alpha)\, y, \quad \text{где} \quad x,\, y \in U, \quad \alpha \in (0,1),$$

следует, что $v = x = y$.

Замечание 24. *Для многогранных множеств угловыми точками являются именно вершины многогранников . Что касается других множеств, которые не являются многогранными, то их точки, относящиеся согласно определению 16 к категории угловых, далеко не всегда похожи на* ***угловые****, как, к примеру, граничные точки плоского круга. Для них более подходящим будет термин* ***«крайние»****.*

Одна из основополагающих идей симплекс-метода состоит в организации целенаправленного перебора угловых точек допустимого многогранника. Правило распознавания угловых точек (вершин) канонического многогранника U_0 содержит

Теорема 22. (критерий угловой точки)
Пусть U_0 – канонический многогранник вида (156), A_j, $j = 1, 2, ..., n$, – *столбцы матрицы A и* $\operatorname{rank} A = r \geqslant 1$. *Тогда для того, чтобы точка $v = (v_1, v_2, \ldots, v_n) \in U$ была угловой точкой множества U_0, необходимо и достаточно, чтобы для некоторого набора* ***базисных*** *столбцов A_{j_i}, $i = 1, 2, ..., r$, матрицы A равенство $Au = b$ из определения множества U_0 выполнялось в виде*

$$A_{j_1} v_{j_1} + A_{j_2} v_{j_2} + \cdots + A_{j_r} v_{j_r} = b\,, \tag{157}$$

а все не представленные в (157) *координаты точки v обязательно были* ***нулевыми***.

Доказательство. *Необходимость.* Пусть v — угловая точка множества U_0. Если $v = 0$, то и $b = 0$ и тогда соотношение (157) будет выполняться для любого набора базисных столбцов матрицы A. Если же $v \neq 0$ ($v \geqslant 0$), то некоторые её координаты будут положительными, а все остальные будут равны нулю:

$$\exists\, v_{j_1} > 0,\ v_{j_2} > 0,\ \ldots,\ v_{j_k} > 0; \qquad v_j = 0\ \ \forall\, j \notin \{j_1, j_2, \ldots, j_k\}\,. \tag{158}$$

Согласно (158) ограничения-равенства для точки v будут выполняться в виде

$$A_{j_1} v_{j_1} + A_{j_2} v_{j_2} + \cdots + A_{j_k} v_{j_k} = b\,.$$

Покажем, что столбцы A_{j_i}, $i = 1, 2, \ldots, k$, *линейно независимы*. Рассмотрим их линейную комбинацию

$$A_{j_1} \gamma_{j_1} + A_{j_2} \gamma_{j_2} + \cdots + A_{j_k} \gamma_{j_k} = 0 \tag{159}$$

и докажем, что в (159) все коэффициенты равны нулю: $\gamma_{j_i} = 0\ \forall\, i = 1, 2, \ldots, k$. Дополним имеющийся k-мерный набор γ_{j_i}, $i = 1, 2, \ldots, k$, *нулями* до n-мерного вектора $\gamma \in R^n$. В терминах γ равенство (159) записывается в виде

$$A\gamma = 0\,. \tag{160}$$

Рассмотрим пары точек

$$v_{\pm\varepsilon} = v \pm \varepsilon\, \gamma\,, \quad \varepsilon > 0\,.$$

Поскольку $Av = b$, то в силу (160)

$$Av_{\pm\varepsilon} = b \quad \forall\, \varepsilon > 0\,,$$

а в силу (158)

$$v_{\pm\varepsilon} \geqslant 0 \quad \forall\, \varepsilon > 0 \text{ (достаточно малых)}\,,$$

поэтому для всех достаточно малых $\varepsilon > 0$ обе точки $v_{\pm\varepsilon}$ будут принадлежать каноническому многограннику U_0 и при этом

$$v = \frac{1}{2}\, v_{\varepsilon} + \frac{1}{2}\, v_{-\varepsilon}\,,$$

а тогда по определению угловой точки имеем $v = v_{\varepsilon} = v_{-\varepsilon}$, следовательно, $\gamma = 0$ и, тем самым, доказана *линейная независимость* столбцов $A_{j_i},\ i = 1, 2, \ldots, k$. По условию $\operatorname{rank} A = r$, значит, $k \leqslant r$ и, дополняя набор $A_{j_i},\ i = 1, 2, \ldots, k$, до базисного, мы получим соотношение (157). Необходимость доказана.

Достаточность. Пусть $v \in U_0$ — некоторая допустимая точка, для которой выполняется условие (157) и все отсутствующие в записи (157) координаты точки v равны нулю. Покажем, что тогда v является *угловой* точкой многогранника U_0. Пусть $x,\ y \in U_0$ и при некотором $\alpha \in (0, 1)$

$$v = \alpha\, x + (1 - \alpha)\, y\,.$$

Для всех j, для которых $v_j = 0$, в силу того, что $x_j \geqslant 0,\ y_j \geqslant 0$ и $\alpha \in (0, 1)$, будем иметь $v_j = x_j = y_j = 0$, после чего совпадение всех остальных (неотрицательных) координат $v_{j_i} = x_{j_i} = y_{j_i},\ i = 1, 2, \ldots, r$, будет следовать из *линейной независимости* присутствующих в (157) столбцов матрицы A и того, что $Av = Ax = Ay = b$. Теорема доказана.

При описании процедуры симплекс-метода будем предполагать, что из СЛАУ $Au = b$ *исключены все линейно зависимые уравнения* и что канонический многогранник U_0 является непустым множеством и не вырождается в точку:

$$1 \leqslant m = r = \operatorname{rank} A < n. \tag{161}$$

Для запуска симплекс-метода нужно задать стартовую угловую точку (вершину) v^0 канонического многогранника U_0. Тогда все следующие приближения v^k также будут угловыми точками и значения функции не будут возрастать: $J(v^{k+1}) \leqslant J(v^k),\ k = 0, 1, \ldots$ На k-ой итерации симплекс-метода мы имеем угловую точку v^k и соответствующий ей по теореме 22 *базисный* набор столбцов матрицы A с *базисными* номерами

$$J_b^k = \{\, j_1,\, j_2, \ldots, j_r\}.$$

Все переменные $u = (u_1, u_2, \ldots, u_n)$ разделяются на две группы: *базисную* и *свободную*. В базисную группу u_b входят переменные u_j с номерами $j \in J_b^k$, а в свободную группу u_f — все остальные переменные с номерами

$$j \in J_f^k = \{1, 2, \ldots, n\} \setminus J_b^k.$$

Этим разбиениям будет соответствовать выделение из матрицы A квадратного $r \times r$ *невырожденного* блока

$$B^k = \left(A_{j_1} \middle| A_{j_2} \middle| \ldots \middle| A_{j_r} \right).$$

Из остальных столбцов матрицы A, не вошедших в блок B^k, сформируем прямоугольный $r \times (n-r)$ блок F^k. При этом СЛАУ $Au = b$ можно будет записать в более подробной форме

$$B^k u_b + F^k u_f = b\,.$$

Заметим, что для угловой точки v^k, у которой все небазисные координаты равны нулю, данное ограничение примет укороченный вид (157):

$$B^k v_b^k = b\,.$$

Если теперь базисные переменные u_b явно выразить через свободные:

$$u_b = u_b(u_f) = (B^k)^{-1}(b - F^k u_f) = v_b^k - (B^k)^{-1} F^k u_f, \quad \dim u_f = n - r, \quad (162)$$

то n-мерная каноническая задача (156) превратится в *эквивалентную* ей $(n-r)$-мерную *неканоническую* задачу ЛП:

$$g^k(u_f) = J(v^k) - \sum_{j \in J_f^k} \Delta_j^k\, u_j \to \inf, \quad u_f \geqslant 0, \quad u_b(u_f) \geqslant 0\,. \qquad (163)$$

Здесь

$$\begin{aligned} g^k(u_f) = J(u_b(u_f), u_f) &\overset{c=(c_b,c_f)}{=} \langle c_b,\, u_b(u_f)\rangle_{R^r} + \langle c_f,\, u_f\rangle_{R^{n-r}} \overset{(162)}{=} \\ &= \langle c_b,\, v_b^k - (B^k)^{-1} F^k u_f\rangle_{R^r} + \langle c_f,\, u_f\rangle_{R^{n-r}} = \\ &= \langle c_b,\, v_b^k\rangle_{R^r} - \langle c_b,\, (B^k)^{-1} F^k u_f\rangle_{R^r} + \langle c_f,\, u_f\rangle_{R^{n-r}} \overset{v_f=0}{=} \\ &= J(v^k) - \langle c_b,\, (B^k)^{-1} F^k u_f\rangle_{R^r} + \langle c_f,\, u_f\rangle_{R^{n-r}} = \\ &= J(v^k) - \langle \Delta^k,\, u_f\rangle_{R^{n-r}}, \quad \text{где} \quad \Delta^k = \left((B^k)^{-1} F^k\right)^\top c_b - c_f \in R^{n-r}\,. \end{aligned}$$

Неканоническими в задаче (163) являются r ограничений $u_b(u_f) \geqslant 0$, которым должны были удовлетворять исключенные базисные переменные:

$$(B^k)^{-1} F^k u_f \overset{(162)}{\leqslant} v_b^k\,. \qquad (164)$$

Как видно из (163), шансы уменьшить уже достигнутое значение $J(v^k)$ вариацией свободных переменных u_f имеются лишь в том случае, когда среди коэффициентов Δ_j^k есть хотя бы один положительный. Другими словами, перспективными в плане минимизации являются свободные переменные, номера которых попадают во множество

$$J_f^{k+} = \{\, j \in J_f^k \mid \ \Delta_j^k > 0\}. \tag{165}$$

Если оказалось, что $J_f^{k+} = \varnothing$, то никакой возможности уменьшить значение функции $g^k(u_f)$ нет, процесс останавливается, а искомое решение найдено: $v^k \in U_*, \ \ J_* = J(v^k)$.

Если $J_f^{k+} \neq \varnothing$, то выбирается некоторый номер $j \in J_f^+$ и принимается решение варьировать только *одну* свободную переменную u_j, а остальным свободным переменным присваиваются нулевые значения. Во избежание зацикливания симплекс-метода можно использовать известное *правило Блэнда*, упорядочивающее выбор номеров в случае их неединственности:

$$j_* = \min_{j \in J_f^{k+}} j\,. \tag{166}$$

Итак, $u_f = (0, \ldots, 0, u_{j_*}, 0, \ldots, 0)$ и задача (163) превращается в *одномерную* задачу минимизации линейной функции:

$$g^k(u_{j_*}) = J(v^k) - \Delta_{j_*}^k\, u_{j_*} \to \inf, \quad u_{j_*} \geqslant 0, \quad \big((B^k)^{-1}F^k\big)_{j_*} u_{j_*} \overset{(164)}{\leqslant} v_b^k\,. \tag{167}$$

Здесь $\big((B^k)^{-1}F^k\big)_{j_*}$ — j_*-ый столбец произведения матриц $(B^k)^{-1}F^k$, который равен произведению $(B^k)^{-1}A_{j_*}$ матрицы $(B^k)^{-1}$ на j_*-ый столбец A_{j_*} матрицы A. Введём для этого r-мерного вектора обозначение

$$(B^k)^{-1}A_{j_*} = \gamma_*^k = \big(\gamma_{*1}^k, \gamma_{*2}^k, \ldots, \gamma_{*r}^k\big) \in R^r\,, \tag{168}$$

и запишем ограничения задачи (167) более подробно:

$$u_{j_*} \geqslant 0; \quad \gamma_{*1}^k\, u_{j_*} \leqslant v_{j_1}^k,\ \gamma_{*2}^k\, u_{j_*} \leqslant v_{j_2}^k, \ldots, \gamma_{*r}^k\, u_{j_*} \leqslant v_{j_r}^k\,. \tag{169}$$

Если оказалось, что в условиях (169) все r компонент вектора γ_*^k неположительны: $\gamma_{*i}^k \leqslant 0$, т. е. на самом деле *нет никаких реальных ограничений сверху* на переменную u_{j_*}, то итерации прекращаются и делается вывод: решения у задачи (156) не существует, $J_* = -\infty$, $U_* = \varnothing$.

Остаётся рассмотреть возможность, когда $J_f^{k+} \neq \varnothing$ и для выбранного номера $j_* \in J_f^{k+}$ условия (169) содержат *реальные ограничения сверху* на переменную u_{j_*}. Данная возможность характеризуется условием

$$I^{k+} = \left\{ i \in \{1, 2, \ldots, r\} \;\middle|\; \gamma_{*i}^k > 0 \right\} \neq \varnothing \,, \tag{170}$$

с учётом которого находится решение одномерной задачи (167):

$$u_{j_*} = \min_{i \in I^{k+}} \frac{v_{j_i}^k}{\gamma_{*i}^k} \,. \tag{171}$$

После этого, используя найденное в (171) значение u_{j_*}, осуществляется переход от имеющейся угловой точки v^k к следующей точке

$$v^{k+1} = \big(\underbrace{u_b(u_{f*})}_{r}, \; u_{f*} = (\underbrace{0, \ldots, 0, u_{j_*}, 0, \ldots, 0}_{n-r}) \big). \tag{172}$$

Эта точка по построению принадлежит каноническому многограннику U_0 и $J(v^{k+1}) \leqslant J(v^k)$. Для доказательства того, что она является *угловой* точкой, воспользуемся теоремой 22. Подходящий набор *базисных* номеров J_b^{k+1} этой точки сформируем следующим образом. Составим список номеров, на которых реализуется минимум в (171):

$$I_*^{k+} = \left\{ s \in I^{k+} \;\middle|\; \frac{v_{j_s}^k}{\gamma_{*s}^k} = \min_{i \in I^{k+}} \frac{v_{j_i}^k}{\gamma_{*i}^k} \right\}, \tag{173}$$

и из этого списка выберем некоторый конкретный номер $s \in I_*^{k+}$. Чтобы избежать зацикливания симплекс-метода, номер s предлагается выбирать по *правилу Блэнда*, а именно:

$$s_* = \min_{s \in I_*^{k+}} s \,. \tag{174}$$

После этого из базисного набора номеров J_b^k угловой точки v^k исключается номер j_{s_*} и добавляется номер j_* :

$$J_b^{k+1} = J_b^k \setminus \{j_{s_*}\} \cup \{j_*\}. \tag{175}$$

Соответствующие изменения происходят и в матричных блоках B^k и F^k. Из старого блока B^k изымается столбец $A_{j_{s_*}}$ и перемещается в новый свободный блок F^{k+1}, а из старого свободного блока F^k столбец A_{j_*} переносится в новый базисный блок B^{k+1} :

$$B^{k+1} = \underbrace{\left(A_{j_1} \middle| A_{j_2} \middle| \ldots \middle| A_{j_r} \right)}_{B^k} \setminus \{A_{j_{s_*}}\} \cup \{A_{j_*}\} \,. \tag{176}$$

Процедура симплекс-метода в сочетании с антициклином Блэнда полностью описана. Доказательство того, что следующее приближение v^{k+1} будет угловой точкой множества U_0, будет дано на следующей лекции.

Симплекс-метод (продолжение)

На предыдущей лекции была описана процедура перехода от текущей *угловой* точки v^k канонического многогранника U_0 к следующей *допустимой* точке v^{k+1} и для доказательства того, что v^{k+1} обязательно будет *угловой* точкой, были предъявлены в (175) и (176) соответствующие наборы *базисных* номеров J_b^{k+1} и *базисных* столбцов B^{k+1} матрицы A. Проверим, что эти наборы полностью отвечают требованиям теоремы 22. Для этого убедимся в том, что столбцы матрицы A, составляющие блок B^{k+1}, *линейно независимы* и *равна нулю* j_{s_*}-ая координата точки v^{k+1}, которая на k-ой итерации была выведена из разряда базисных. Равенство $v_{j_{s_*}}^{k+1} = 0$ следует из того, что в соответствии с (162), (168) $u_b = \gamma_*^k\, u_{j_*}$, а также правил определения значения u_{j_*} в (171) и выбора номера s_* в (173), (174). Линейную независимость столбцов матрицы B^{k+1} докажем по определению. Приравняем к нулю их линейную комбинацию:

$$\sum_{i=1(i\neq s_*)}^{r} \alpha_i\, A_{j_i} + \alpha_{j_*}\, A_{j_*} = 0 \tag{177}$$

и заместим в (177) добавленный столбец A_{j_*} его выражением через столбцы предыдущего базисного блока B^k :

$$A_{j_*} = B^k \big(B^k\big)^{-1} A_{j_*} \overset{(168)}{=} B^k \gamma_*^k = \sum_{i=1}^{r} \gamma_{*i}^k\, A_{j_i}\,.$$

В результате равенство (177) примет вид

$$\sum_{i=1(i\neq s_*)}^{r} \big(\alpha_i + \alpha_{j_*}\gamma_{*i}^k\big)\, A_{j_i} + \alpha_{j_*}\gamma_{*s_*}^k\, A_{j_{s*}} = 0\,,$$

в котором присутствуют *линейно независимые* столбцы матрицы A, следовательно,

$$\alpha_i + \alpha_{j_*}\gamma_{*i}^k = 0 \quad \forall\, i \neq s_*; \quad \alpha_{j_*}\gamma_{*s_*}^k = 0\,.$$

В последнем равенстве номер s_* выбирался из множества $I_*^{k+} \subset I^{k+}$, значит,

$$\gamma_{*s_*}^k > 0 \quad \Longrightarrow \quad \alpha_{j_*} = 0 \quad \Longrightarrow \quad \alpha_i = 0 \quad \forall\, i \neq s_*$$

и, тем самым, линейная независимость столбцов матрицы B^{k+1} доказана, а по теореме 22 точка v^{k+1} является *угловой* точкой канонического многогранника U_0.

Остаётся обсудить вопрос о выборе *стартовой угловой* точки $v^0 \in U_0$. Оказывается, одну из таких точек можно найти с помощью самого симплекс-метода, решив с его помощью специальную вспомогательную каноническую задачу ЛП, в которой стартовая угловая точка находится просто. Данный приём получил название ***метода искусственного базиса.*** Для применения этого метода удобно считать, что выполняется условие $b \geqslant 0$ (если какая-то координата вектора b отрицательна, умножаем на (-1) соответствующее линейное уравнение в СЛАУ $Au = b$). Никаких предварительных прореживаний СЛАУ $Au = b$ с целью удаления из неё линейно зависимых уравнений при этом производить не требуется. Предлагается рассмотреть следующую *вспомогательную каноническую* задачу ЛП в пространстве переменных $z = (x, u) \in R^{m+n}$, где m – количество строк матрицы A :

$$g(z) = x_1 + x_2 + \cdots + x_m \to \inf, \quad z \in Z_0 = \{z \geqslant 0, \;\; x + Au = b\}. \qquad (178)$$

Заметим, что у матрицы $(I|A)$ размера $m \times (m + n)$ из СЛАУ (178) ранг равен m, т. е. её строки *линейно независимы* и эта СЛАУ в прореживании не нуждается. Тогда по теореме 22 точка $z^0 = (x = b, u = 0)$ является *угловой* точкой канонического многогранника Z_0. Из этой точки z^0 можно запустить симплекс-метод (с антициклином Блэнда) и он *за конечное число шагов* найдет решение $z_* = (x_*, v_*)$ задачи (178), которое существует, поскольку $g_* = \inf\limits_{z \in Z_0} g(z) \geqslant 0 > -\infty$. При этом

$$g_* = 0 \quad \Longleftrightarrow \quad U_0 \neq \varnothing,$$

значит, если вдруг симплекс-метод в задаче (178) выдал результат $g_* > 0$, то $U_0 = \varnothing$ и исходная каноническая задача (156) нуждается в коррекции. Предположим, что $g_* = 0$ и симплекс-метод остановился в оптимальной точке $z_* = (x_* = 0, v_*)$ задачи (178), являющейся *угловой* точкой многогранника Z_0. При этом компонента $v_* \geqslant 0$ и $Av_* = b$, т. е. $v_* \in U_0$. Покажем, что v_* — *угловая* точка многогранника U_0. В соответствии с определением представим её в виде выпуклой линейной комбинации двух других точек из U_0 :

$$v_* = \alpha\, v_1 + (1 - \alpha)\, v_2, \quad v_1,\, v_2 \in U_0, \quad \alpha \in (0, 1)\,.$$

Тогда в пространстве R^{m+n} переменных $z = (x, u)$ будет выполняться равенство

$$z_* = (x_* = 0, v_*) = \alpha\,(0, v_1) + (1 - \alpha)\,(0, v_2), \quad (0, v_1),\, (0, v_2) \in Z_0, \quad \alpha \in (0, 1)\,,$$

и так как точка $z_* = (x_* = 0, v_*)$ — угловая точка множества Z_0, то $z_* = (x_* = 0, v_*) = (0, v_1) = (0, v_2)$, следовательно, $v_* = v_1 = v_2$, ч.т.д.

В заключение сформулируем следующее утверждение, содержащее обоснование намеченного нами в самом начале плана поиска оптимального решения задачи ЛП среди угловых точек допустимого многогранника.

Теорема 23. `(о свойствах канонических задач ЛП)`
В канонических задачах линейного программирования (156)

1) если допустимое множество U_0 непусто, то оно содержит хотя бы одну угловую точку;

2) если нижняя грань функционала J_* конечна, то множество U_* оптимальных решений задачи (156) *непусто и содержит по крайней мере одну угловую точку множества U_0.*

Замечание 25.

- *Теоретически нельзя исключить, что при неудачном выборе стартовой угловой точки симплекс-метод по пути к оптимальному решению пройдёт* ***по всем без исключения вершинам*** *многогранника, число которых может быть астрономическим. К счастью, подобные «катастрофы» симплекс-метода в реальных задачах пока не зафиксированы.*

- *В утверждении 2) теоремы 23 проявляется специфика задач ЛП. Для нелинейных функций легко привести пример, в котором из того, что $J_* > -\infty$,* ***не следует****, что $U_* \neq \varnothing$:*
$$J(u) = e^u \to \inf, \quad u \in U = R^1 .$$

- *Практические возможности симплекс-метода (в сочетании с антициклином) не следует переоценивать: реальные данные A и b, как правило, известны* ***неточно****, а тогда при буквальной реализации описанной здесь процедуры могут возникнуть настолько существенные перекосы, что итоговые результаты окажутся непригодными для использования. Для устойчивой работы симплекс-метода с зашумлёнными данными не обойтись без тех или иных вспомогательных регуляризирующих процедур.*

Метод покоординатного спуска

На предыдущих лекциях мы, в основном, рассматривали итерационные методы минимизации, которые для своей реализации требуют вычисления первых или вторых производных целевой функции. Однако в практических задачах нередко либо минимизируемая функция не обладает нужной гладкостью, либо вычисление ее производных с нужной точностью оказывается слишком трудоемким. Для подобных ситуаций желательно иметь в запасе методы минимизации, при реализации которых требуется вычислять лишь значения функции. Одним из таких методов является *метод покоординатного спуска.*

Рассмотрим этот метод применительно к задаче минимизации *без ограничений в конечномерном пространстве*:

$$J(u) \to \inf, \quad u \in R^n. \tag{179}$$

Выберем в пространстве R^n некоторый базис, например, стандартный ОНБ $\{e_i\}_{i=1}^n$. При реализации метода будет производиться циклический перебор этих базисных векторов, поэтому для удобства описания итерационного процесса (теоретически бесконечного) выстроим их в единую бесконечную цепочку:

$$p_0 = e_1,\, p_1 = e_2, \ldots, p_{n-1} = e_n, \\ p_n = e_1,\, p_{n+1} = e_2, \ldots, p_{2n-1} = e_n, \quad p_{2n} = e_1, \ldots \tag{180}$$

Перед запуском метода выбирается некоторая стартовая точка $u_0 \in R^n$, стартовый шаг $\alpha_0 > 0$ и коэффициент дробления шага $\lambda \in (0, 1)$. Допустим, что на k-ой итерации найдено k-ое приближение u_k и текущее значение шага равно $\alpha_k > 0$. Для определения очередного приближения выполняются следующие действия. Берётся базисный вектор p_k и вычисляется значение функции в точке $u = u_k + \alpha_k p_k$.

$$\text{Если} \quad J(u_k + \alpha_k p_k) < J(u_k), \quad \text{то} \quad u_{k+1} = u_k + \alpha_k p_k, \quad \alpha_{k+1} = \alpha_k \tag{181}$$

и процесс продолжается из точки u_{k+1} со следующим по порядку базисным направлением p_{k+1}. Если $J(u_k + \alpha_k p_k) \geq J(u_k)$, то вычисляем значение функции в противоположной точке $u = u_k - \alpha_k p_k$.

$$\text{Если} \quad J(u_k - \alpha_k p_k) < J(u_k), \quad \text{то} \quad u_{k+1} = u_k - \alpha_k p_k, \quad \alpha_{k+1} = \alpha_k \tag{182}$$

и переходим к следующей итерации с очередным базисным вектором p_{k+1}. Можно также сразу вычислять оба значения $J(u_k \pm \alpha_k p_k)$ и выбирать из них

наименьшее. Будем называть $(k+1)$-ую итерацию *удачной*, если переход от u_k к u_{k+1} произошел по одному из условий (181) или (182) и сопровождался строгим убыванием значения функции. Если же нарушены оба условия (181) и (182), назовем $(k+1)$-ую итерацию *неудачной*. Дальнейшие действия зависят от предыстории. В процессе вычислений ведется подсчет числа *неудачных* итераций, *случившихся подряд*. Если их общее количество вместе с неудачей на текущем $(k+1)$-ом шаге еще не достигло n, то полагают

$$u_{k+1} = u_k, \quad \alpha_{k+1} = \alpha_k \tag{183}$$

и переходят к очередному базисному направлению p_{k+1}. Если же все $n-1$ итерации, предшествующие неудаче, зафиксированной на $(k+1)$-ом шаге, также были неудачными, то производится дробление шага α_k с априорно выбранным коэффициентом $\lambda \in (0,1)$:

$$u_{k+1} = u_k, \quad \alpha_{k+1} = \lambda\,\alpha_k \tag{184}$$

и проверка основных условий (181) и (182) продолжается на более мелкой сетке. Достаточные условия сходимости описанного варианта метода покоординатного спуска содержит

Теорема 24. *Пусть функция $J(u)$ выпукла на R^n и принадлежит классу $C^1(R^n)$, а начальное приближение u_0 таково, что соответствующее ему множество Лебега $M_0 = \{u \in R^n \mid J(u) \leqslant J(u_0)\}$ ограничено. Тогда последовательность u_k, вырабатываемая методом* (180) – (184), *сходится и по функции и по аргументу:*

$$\lim_{k\to\infty} J(u_k) = J_*, \quad \lim_{k\to\infty} \rho(u_k, U_*) = 0. \tag{185}$$

Доказательство. По условию множество M_0 ограничено, а из непрерывности функции $J(u)$ следует его замкнутость. Кроме того, это множество по построению содержит в себе все точки, перспективные в плане минимизации $J(u)$ на всём пространстве R^n, поэтому в силу классической (конечномерной) теоремы Вейерштрасса будем иметь

$$J_* > -\infty, \quad U_* \neq \varnothing.$$

Из описания метода (180) – (184) следует монотонность:

$$J(u_{k+1}) \leq J(u_k), \quad k = 0, 1, \ldots,$$

поэтому $u_k \in M_0$ и существует $\lim\limits_{k\to\infty} J(u_k) \geq J_*$. Покажем, что $\lim\limits_{k\to\infty} \alpha_k = 0$, т. е. моментов дробления шага метода будет бесконечно много. Допустим

противное: пусть последнее дробление состоялось на N-ом шаге, после чего процесс вычислений (бесконечный) продолжается с фиксированным шагом $\alpha_k = \alpha_N = \alpha > 0,\ k = N, N+1, \ldots$ Рассмотрим в пространстве R^n дискретную решетку (сетку) M_α с одним и тем же равномерным шагом $\alpha > 0$ по всем координатным направлениям и поместим точку u_N в один из ее узлов. Из описания метода покоординатного спуска следует, что все следующие приближения u_k являются узлами этой самой решетки, т. е. $u_k \in M_\alpha\ \forall\ k \geq N$ и остаются в пределах *ограниченного* множества M_0, в котором может находиться лишь *конечное число узлов* сетки M_α. Но при перемещении по *конечному* множеству узлов невозможно *бесконечное* число раз наблюдать строгое убывание значений функции, без которого обязательно произошло бы дробление шага. Полученное противоречие показывает, что процесс дробления α_k бесконечен и $\lim\limits_{k\to\infty} \alpha_k = 0$.

Пусть $k_1 < k_2 < \ldots < k_m < \ldots$ – номера тех итераций, на которых длина шага α_k дробится. В соответствии с описанной процедурой (180) – (184) этим дроблениям предшествуют серии ровно из n *неудачных* итераций по всем базисным направлениям:

$$J(u_{k_m} + \alpha_{k_m} e_i) \geq J(u_{k_m}), \quad J(u_{k_m} - \alpha_{k_m} e_i) \geq J(u_{k_m}), \quad i = 1, 2, \ldots \qquad (186)$$

Из последовательности точек u_{k_m}, принадлежащих *ограниченному* множеству M_0, можно выбрать сходящуюся подпоследовательность. Без ограничения общности можем считать, что сама последовательность u_{k_m} сходится к некоторой точке $w \in M_0$. С помощью формулы конечных приращений перепишем неравенства (186) в виде

$$\langle J'(u_{k_m} + \theta_m^+ \alpha_{k_m} e_i), e_i\rangle\, \alpha_{k_m} \geq 0, \quad \langle J'(u_{k_m} - \theta_m^- \alpha_{k_m} e_i), e_i\rangle\, (-\alpha_{k_m}) \geq 0, \quad (187)$$

при некоторых $\theta_m^+,\ \theta_m^- \in [0, 1],\ m = 1, 2, \ldots$ Поделим неравенства (187) на $\alpha_{k_m} > 0$, после чего перейдем в них к пределу при $m \to \infty$. С учетом условия $J'(x) \in C(R^n)$ непрерывности градиента, ограниченности значений $\theta_m^\pm$ и сходимости $\alpha_{k_m} \to 0$ получим следующие соотношения в предельной для элементов u_{k_m} точке w :

$$\langle J'(w), e_i\rangle \geq 0, \quad \langle J'(w), e_i\rangle \leq 0, \quad i = 1, 2, \ldots, n.$$

Так как $\{e_i\}_{i=1}^n$ – базис в R^n, то отсюда следует, что $J'(w) = 0$, а поскольку функция $J(u)$ выпукла, то точка w является одной из точек минимума: $w \in U_*$. Получается, что подпоследовательность u_{k_m} является минимизирующей: $\lim\limits_{m\to\infty} J(u_{k_m}) = J(w) = J_*$. Отсюда и из монотонности последовательности $J(u_k)$ следует, что минимизирующей является и вся последовательность

u_k : $\lim\limits_{k\to\infty} J(u_k) = J_*$, т. е. утверждение (185) теоремы доказано. Второе утверждение о сходимости по аргументу выводится из сходимости по функции и компактности в R^n множества M_0 (см. классическую теорему Вейерштрасса). Теорема 24 доказана.

Заметим, что при реализации описанного варианта метода покоординатного спуска не используются значения градиентов минимизируемой функции, однако в условии теоремы о сходимости метода присутствует требование гладкости этой функции. Приведем пример, показывающий, что при отсутствии гладкости без усиления остальных условий теоремы сходимость метода покоординатного спуска гарантировать нельзя.

Пример 4. *Рассмотрим двумерную задачу минимизации*

$$J(u) = (x-1)^2 + (y-1)^2 + 2|x-y| \to \inf, \quad u = (x,y) \in R^2.$$

Функция J непрерывна, сильно выпукла, ограничена снизу на всем пространстве R^2, достигает своей нижней грани $J_* = 0$ в единственной точке $u_* = (x_* = 1, y_* = 1)$ и на прямой $y = x$ не является дифференцируемой. Нетрудно проверить, что при выборе в R^2 стандартных базисных направлений $e_1 = (1,0)$, $e_2 = (0,1)$ в случае запуска метода покоординатного спуска из начала координат $u_0 = (0,0)$ независимо от выбора стартового шага $\alpha_0 > 0$ процесс зациклится в начальной точке: $u_k = u_0 = (0,0)$, $k = 1,2,\ldots$ и сходимости к решению задачи минимизации не будет ни по функции, ни по аргументу:

$$J(u_k) - J_* = J(u_0) - J_* = 2, \quad |u_k - u_*| = |u_0 - u_*| = \sqrt{2}, \quad k = 1,2,\ldots$$

Существуют и другие варианты метода покоординатного спуска. Так, вместо стандартного базиса из единичных координатных векторов, который использовался в (180), можно взять в R^n произвольный базис $\{e_i\}_{i=1}^n$, не обязательно ортонормированный. При этом утверждения теоремы останутся в силе, правда, «покоординатным спуском» соответствующий итерационный процесс можно будет назвать лишь с некоторой натяжкой. Кстати, если в приведённом с негладкой функцией $J(u)$ в качестве базисных направлений вместо стандартных координатных взять векторы $e_1 = (1,1)$, $e_2 = (-1,1)$, то из того же самого начального приближения $u_0 = (0,0)$ для любого стартового шага $\alpha_0 > 0$ последовательность u_k будет сходиться к решению задачи.

Достоинства метода: простота, скромные требования к гладкости минимизируемой функции.

Недостатки метода: невысокая скорость сходимости, неспособность к распознаванию принадлежности вычисляемых приближений u_k оптимальному множеству U_*.

Метод штрафных функций

Наличие в оптимизационных задачах ограничений довольно часто вызывает дополнительные затруднения при их численном решении. Метод *штрафных функций* или, короче, *метод штрафов* позволяет при определенных условиях отказаться от строгого соблюдения всех без исключения ограничений и разрешить нарушение некоторых из них, например, тех, учёт которых вызывает наибольшие неудобства при численном решении. Но речь идет не о полном снятии таких ограничений, а об их учёте в более мягкой форме, предполагающей штрафные санкции за их нарушения.

Опишем одну из возможных схем применения метода штрафов к задачам минимизации с ограничениями в гильбертовом пространстве $H,\ \dim H \leqslant \infty$:

$$J(u) \to \inf, \quad u \in U \subset H,$$
$$U = \Big\{ u \in U_0 \;\Big|\; g_1(u) \leqslant 0, \ldots, g_m(u) \leqslant 0,\ g_{m+1}(u) = 0, \ldots, g_{m+s}(u) = 0 \Big\}. \tag{188}$$

Штраф будет «выписываться» за нарушение ограничений типа равенств и неравенств, задаваемым функциями $g_i(u) : H \to R^1$. Неструктурированные ограничения, задаваемые в задаче (188) множеством U_0, мы считаем «терпимыми», и обязуемся их не нарушать. Случаи, когда $U_0 = H$ или отсутствуют ограничения типа равенств ($s = 0$) или неравенств ($m = 0$), не исключаются, но хотя бы одно из ограничений типа равенств или неравенств все же должно присутствовать, иначе исчезнет объект штрафования.

Рассмотрим один из наиболее популярных способов штрафования. За нарушения ограничений типа неравенств будем выписывать индивидуальные штрафы типа срезки:

$$g_i^+(u) = \max\{\, g_i(u),\ 0 \,\}, \quad i = 1, 2, \ldots, m. \tag{189}$$

При нарушении ограничений типа равенств будем использовать модули:

$$g_i^+(u) = |\, g_i(u) \,|, \quad i = m+1, m+2, \ldots, m+s. \tag{190}$$

Из индивидуальных штрафов (189), (190) составляется *общий* или *суммарный* штраф

$$P(u) = \sum_{i=1}^{m+s} \Big(g_i^+(u) \Big)^{p_i}, \quad p_i \geqslant 1, \quad i = 1, 2, \ldots, m+s. \tag{191}$$

Показатели степеней p_i в (191) можно задавать по-разному; на практике довольно часто берут одинаковые значения $p_i = 2,\ i = 1, 2, \ldots, m + s$.

Отметим следующие важные свойства введенных штрафных функций. Во-первых, $P(u) \geqslant 0\ \forall u \in H$, а, во-вторых,

$$u \in U \iff \left\{ \begin{array}{c} u \in U_0, \\ P(u) = 0, \end{array} \right. \iff \left\{ \begin{array}{c} u \in U_0, \\ g_i^+(u) = 0\ \ \forall i = 1, 2, \ldots, m + s. \end{array} \right. \tag{192}$$

Общий штраф с коэффициентами $A_k > 0,\ k = 1, 2, \ldots$, добавляется к исходной функции $J(u)$ и осуществляется переход от исходной постановки (188) на множестве U к последовательности задач минимизации оштрафованной функции на «терпимом» множестве U_0 :

$$\Phi_k(u) = J(u) + A_k P(u) \to \inf, \quad u \in U_0, \quad k = 1, 2, \ldots \tag{193}$$

Пусть u_k – приближенные решения задач (193):

$$u_k \in U_0, \ \ \Phi_{k*} = \inf_{u \in U_0} \Phi_k(u) \leqslant \Phi_k(u_k) \leqslant \Phi_{k*} + \varepsilon_k, \ \ \varepsilon_k > 0, \ \ k = 1, 2, \ldots \tag{194}$$

Метод отыскания таких точек $u_k \in U_0$ не конкретизируется. Можно привлечь одну из уже рассмотренных нами вычислительных процедур, например, один из градиентных методов.

Теорема 25. (о сходимости метода штрафных функций)
Пусть H – гильбертово пространство, множество $U_0 \subset H$ слабо замкнуто в H, исходная функция $J(u)$ и все индивидуальные штрафы $g_i^+(u)$ слабо полунепрерывны снизу на U_0. Пусть также нижняя грань функции $J(u)$ на «терпимом» множестве U_0 конечна:

$$J_0 = \inf_{u \in U_0} J(u) > -\infty,$$

а δ-расширение

$$U(\delta) = \{\, u \in U_0 \mid g_i^+(u) \leqslant \delta,\ i = 1, 2, \ldots, m + s\}$$

допустимого множества U ограничено в H при некотором $\delta > 0$. Тогда, если $A_k \to +\infty$ и $\varepsilon_k \to 0$, то для элементов u_k, определенных из условий (194), *имеет место схоимость по функции*

$$J(u_k) \to J_* \quad \textit{при} \quad k \to \infty, \tag{195}$$

а у самой последовательности $\{u_k\}_{k=1}^{\infty}$ имеются слабые в H предельные точки, причем каждая из них принадлежит множеству оптимальных решений U_* задачи (188).

Доказательство. Из ограниченности снизу функции $J(u)$ на U_0, положительности A_k и неотрицательности $P(u)$ имеем

$$\Phi_k(u) = J(u) + A_k P(u) \geqslant J_0 \quad \forall\, u \in U_0 \quad \Longrightarrow \quad \Phi_{k*} \geqslant J_0 > -\infty,$$

откуда следует, по крайней мере, существование элементов u_k, удовлетворяющих условиям (194). Справедливы неравенства:

$$J(u_k) \leqslant \Phi_k(u_k) \overset{(194)}{\leqslant} \Phi_{k*} + \varepsilon_k \overset{\forall\, u \in U_0}{\leqslant} \Phi_k(u) + \varepsilon_k = J(u) + A_k P(u) + \varepsilon_k.$$

В эти неравенства можно подставлять любые элементы $u \in U \subset U_0$ и если взять $\inf\limits_{u \in U}$ и перейти к верхнему пределу при $k \to \infty$, то

$$\overline{\lim_{k\to\infty}}\, J(u_k) \leqslant \overline{\lim_{k\to\infty}}\, \Phi_k(u_k) \leqslant \overline{\lim_{k\to\infty}}\, \Phi_{k*} \leqslant J_*. \tag{196}$$

Конечность появившейся в (196) нижней грани J_* будет следовать из вложения $U \subset U_0$ и конечности J_0. Непосредственно из (194) извлекается оценка для значений суммарного штрафа:

$$0 \leqslant P(u_k) \leqslant \frac{\Phi_{k*} - J_0 + \varepsilon_k}{A_k}.$$

По условию $A_k \to \infty$, поэтому, учитывая (196), получаем сходимость

$$\lim_{k\to\infty} P(u_k) = 0 \quad \Longleftrightarrow \quad \lim_{k\to\infty} g_i^+(u_k) = 0 \quad \forall\, i = 1, \ldots, m+s. \tag{197}$$

Из (197) следует, что для $\delta > 0$ из условия теоремы найдется такой достаточно большой номер $k_0(\delta)$, что

$$\forall\, k \geqslant k_0(\delta)\ \ g_i^+(u_k) \leqslant \delta,\ i = 1, \ldots, m+s, \quad \Longleftrightarrow \quad \forall\, k \geqslant k_0(\delta)\ \ u_k \in U(\delta).$$

Так как по условию множество $U(\delta)$ ограничено, то у последовательности $\{u_k\}_{k=1}^{\infty}$ имеются слабые в H предельные точки. Пусть u_0 — одна из них и подпоследовательность $\{u_{k_m}\}_{m=1}^{\infty} \subset \{u_k\}_{k=1}^{\infty}$ слабо в H сходится к u_0. Так как $u_{k_m} \in U_0$ и множество U_0 по условию слабо замкнуто, то $u_0 \in U_0$.

По условию все индивидуальные штрафы $g_i^+(u)$ слабо полунепрерывны снизу, поэтому

$$0 \leqslant g_i^+(u_0) \leqslant \underline{\lim_{m\to\infty}}\, g_i^+(u_{k_m}) \overset{(197)}{=} 0.$$

Это означает, что $u_0 \in U$, т. е. слабая предельная точка u_0 является допустимой в исходной задаче (188). Функция $J(u)$ также слабо полунепрервна снизу, поэтому

$$J_* \leqslant J(u_0) \leqslant \underline{\lim_{m\to\infty}}\, J(u_{k_m}) \leqslant \overline{\lim_{m\to\infty}}\, J(u_{k_m}) \overset{(196)}{\leqslant} J_*.$$

Поскольку и предельная точка u_0 и слабо сходящаяся к ней подпоследовательность $\{u_{k_m}\}_{m=1}^{\infty}$ выбирались произвольно, то утверждение (195) о сходимости по функции доказано. Из последних неравенств и установленного включения $u_0 \in U$ следует, что $u_0 \in U_*$, т. е. точка u_0 является одним из оптимальных решений исходной задачи (188). Отсюда и из того, что точка u_0 выбиралась произвольно, следует и второе утверждение теоремы о слабой сходимости по аргументу. ▼

Замечание 26.

- *Требование слабой полунепрерывности снизу по отношению к индивидуальным штрафам $g_i^+(u)$, $i = m+1, \ldots, m+s$, отвечающим за ограничения типа равенств, в бесконечномерных пространствах является весьма жестким. Так, например, далеко не самое сложное ограничение $g(u) = \|u\| - 1 = 0$, задающее в гильбертовом пространстве H единичную сферу с центром в нуле, будучи оштрафованным по стандартному «тарифу» (190), приведет к функции $g^+(u) = \big|\|u\| - 1\big|$, которая не является слабо полунепрерывной снизу. Например, на элементах ОНБ $\{e_k\}_{k=1}^{\infty}$, которые, как известно, слабо в H сходятся к нулю, будем иметь:*

$$\lim_{k\to\infty} g^+(e_k) = 0 < g^+(0) = 1.$$

  Таким образом, в бесконечномерных пространствах данное требование означает наличие усиленных свойств непрерывности функций, задающих ограничения типа равенств (если эти ограничения линейны, то всё в порядке). В пространствах конечной размерности подобных неудобств не возникает, поскольку для любых непрерывных функций $g(u)$ их индивидуальные штрафы $g^+(u)$ вида (190) также непрерывны.

- *Нельзя утверждать, что описанный выше переход от **исходной одной** оптимизационной задачи (188) с ограничениями **к бесконечной последовательности** задач (194), но зато с ослабленными ограничениями, целесообразен в любом случае, но, по крайней мере, его полезно иметь в виду и применять, исходя из имеющихся технологических возможностей, соображений трудоемкости и т. п.*

Правило множителей Лагранжа для выпуклых задач

Речь пойдёт о важном в идеологическом плане приёме «снятия ограничений», который был предложен Лагранжем и который часто в его честь называют Принципом Лагранжа (ПЛ). Как будет видно, в конструкции правила множителей Лагранжа просматриваются родственные черты с методом штрафов, который мы уже обсудили.

Рассмотрим следующий (не самый общий) класс задач:

$$\begin{gathered} J(u) \to \inf, \quad u \in U \subset L, \\ U = \Big\{ u \in U_0 \;\Big|\; g_1(u) \leqslant 0, \ldots, g_m(u) \leqslant 0 \Big\}. \end{gathered} \tag{198}$$

Здесь L — произвольное линейное пространство любой размерности, даже не обязательно топологическое. Предполагается выполненным следующее **основное предположение выпуклости**:

$$\begin{gathered} J(u),\ g_1(u), \ldots, g_m(u) \ \text{— выпуклые функции}, \\ U_0 \ \text{— выпуклое множество}. \end{gathered} \tag{199}$$

Предложение Лагранжа — ввести функцию (её называют функцией Лагранжа) переменных $(u, \lambda),\ \lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in R^{m+1}$ следующего вида:

$$L(u, \lambda) = \lambda_0 J(u) + \sum_{i=1}^{m} \lambda_i g_i(u), \quad u \in U_0,\ \lambda \in R^{m+1}_+ = \{\lambda \in R^{m+1} \mid \lambda \geqslant 0\}.$$

Сформулирует теорему, содержащую необходимые и достаточные условия оптимальности элемента u_* в задаче (198).

Теорема 26. `(ПЛ для выпуклых задач)`
Пусть задача (198) *выпукла в смысле* (199). *Тогда если* $u_* \in U_*$, *то необходимо существует набор множителей Лагранжа* $\lambda^* \neq 0$, *для которого*

$$\min_{u \in U_0} L(u, \lambda^*) = L(u_*, \lambda^*), \quad \text{(принцип минимума)} \tag{a}$$

$$\lambda_i^* \geqslant 0, \quad i = 0, 1, \ldots, m, \quad \text{(неотрицательность множителей)} \tag{b}$$

$$\lambda_i^*\, g_i(u_*) = 0, \quad i = 1, 2, \ldots, m. \quad \text{(усл-я дополняющей нежесткости)} \tag{c}$$

Обратно, если для некоторой пары (u_*, λ^*) *выполняются условия* (a),(b),(c), *причём* $u_* \in U$, *(элемент* u_* *является допустимым), а* $\lambda_0^* \neq 0$, *то* $u_* \in U_*$, *т.е. элемент* u_* *является оптимальным решением задачи* (198).

Доказательство. Базовым инструментом доказательства будет простейшая *теорема отделимости* точки от непустого выпуклого множества в конечномерном пространстве. Само утверждение будет сформулировано ниже. Заинтересованные студенты могут доказать эту теорему отделимости самостоятельно или заглянуть в книгу [1, кн. 1, гл. 4, § 5].

1) *Необходимость.* Нам будет удобнее считать, что

$$J(u_*) = 0. \tag{d}$$

Договорённость (d) не является ограничением общности, т.к. от $J(u)$ можно перейти к функционалу $\widetilde{J}(u) = J(u) - J(u_*)$ со сдвинутыми на постоянную величину $J(u_*)$ значениями. К таким сдвигам утверждение (a) безразлично, а в (b) и в (c) функционал $J(u)$ вообще не присутствует. Рассмотрим в пространстве R^{m+1} специальное множество

$$M = \{\mu \in R^{m+1} \mid \exists\, u \in U_0 :\ J(u) < \mu_0,\ \ g_i(u) \leqslant \mu_i,\ \ i = 1, 2, \ldots, m\}.$$

Множество M *непусто*. Действительно, ему принадлежат, в частности, все точки μ с положительными координатами $\mu_i > 0,\ \ i = 0, 1, \ldots, m$, для которых в силу принятой договорённости (d) подходящей точкой из U_0 может служить оптимальное решение $u_* \in U_* \subset U_0$.

Множество M *выпукло*. Это свойство легко проверяется по определению выпуклости с помощью основного предположения (199).

Кроме того, $0 \notin M$, т.к. в противном случае нашёлся бы *допустимый* элемент $u \in U$ с *ме́ньшим* по сравнению с $J(u_*) = 0$ значением функционала J, что противоречило бы принятой договорённости (d).

С учётом перечисленных свойств множества M его и непринадлежащую ему точку 0 можно разделить проходящей через 0 гиперплоскостью с нормальным вектором $\lambda^* \neq 0$, так что

$$\langle \lambda^*,\, \mu \rangle_{R^{m+1}} \geqslant 0 = \langle \lambda^*,\, 0 \rangle_{R^{m+1}} \quad \forall\, \mu \in M. \tag{200}$$

Пользуясь имеющимся в (200) произволом в выборе μ, подставим в (200) содержащиеся во множестве M элементы

$$\mu^{\varepsilon}(i) = (\varepsilon, \ldots, \varepsilon, \underbrace{1}_{i-\text{ая коорд.}}, \varepsilon, \ldots, \varepsilon), \quad \varepsilon > 0\,.$$

Переходя к пределу при $\varepsilon \to +0$, получим неравенство $\lambda_i^* \geqslant 0$. Номер $i = 0, 1, \ldots, m$ был нами выбран произвольно, поэтому соотношения (b) *неотрицательности множителей Лагранжа* доказаны.

Если для некоторого номера $i \in \{1, 2, \ldots, m\}$ i-ое ограничение *активно*, т.е. $g_i(u_*) = 0$, то соответствующее условие $\lambda_i^* \, g_i(u_*) = 0$ из списка (c) выполняется тривиальным образом. Рассмотрим любое из *пассивных* ограничений, для которого $g_i(u_*) < 0$, и сформируем для него семейство элементов

$$\mu^\varepsilon(i) = (\varepsilon, 0, \ldots, 0, \underbrace{g_i(u_*)}_{i-\text{ая коорд.}}, 0, \ldots, 0) \in M, \quad \varepsilon > 0\,.$$

Точкой, подтверждающей принадлежность $\mu^\varepsilon(i) \in M$, будет, например, элемент $u_* \in U_* \subset U_0$, для которого $J(u_*) = 0 < \varepsilon$, $g_j(u_*) \leqslant 0$, $j \neq i$, а $g_i(u_*) = g_i(u_*)$. Подставляя такие $\mu^\varepsilon(i)$ в (200) и устремляя $\varepsilon \to +0$, получим неравенство $\lambda_i^* \, g_i(u_*) \geqslant 0$. Мы исходили из того, что $g_i(u_*) < 0$, поэтому $\lambda_i^* \leqslant 0$, а тогда в силу уже доказанного свойства (b) получаем, что $\lambda_i^* = 0$. Тем самым, справедливость *условий дополняющей нежёсткости* (c) установлена для всех номеров $i = 1, 2, \ldots, m$.

Для доказательства *принципа минимума* (a) фиксируем произвольную точку $u \in U_0$, сформируем соответствующее ей семейство элементов

$$\mu^\varepsilon(u) = (J(u) + \varepsilon, g_1(u), g_2(u), \ldots, g_m(u)) \in M, \quad \varepsilon > 0,$$

и подставим их в (200). После перехода к пределу при $\varepsilon \to +0$ получим неравенство $L(u, \lambda^*) \geqslant 0 \;\forall u \in U_0$, которое, с учётом того, что $J(u_*) \overset{(d)}{=} 0$ и $\lambda_i^* \, g_i(u_*) \overset{(c)}{=} 0$, $i = 1, 2, \ldots, m$, превращается в (a). Необходимость доказана.

2) *Достаточность.* Исходим из того, что $u_* \in U$, $\lambda_0^* > 0$ и выполнены условия (199),(a),(b),(c). Тогда, без ограничения общности считая, что $\lambda_0^* = 1$, будем иметь

$$J(u_*) \overset{\text{(c)}}{=} J(u_*) + \sum_{i=1}^m \lambda_i^* \, g_i(u_*) = L(u_*, \lambda^*) \overset{\text{(a)}}{\leqslant} L(u, \lambda^*) =$$

$$= J(u) + \sum_{i=1}^m \underbrace{\lambda_i^*}_{\geqslant 0} \underbrace{g_i(u)}_{\leqslant 0} \leqslant J(u) \quad \forall u \in U \quad \Longrightarrow \quad u_* \in U_*.$$

Достаточность доказана, доказательство теоремы 25 завершено. ▼

Замечание 27. *В теореме 26 рассматривалась задача вида* (198)*, в записи которой отсутствуют ограничения типа равенств:*

$$g_i(u) = 0, \quad i = m+1, m+2, \ldots, m+s. \tag{201}$$

Формально их можно было бы включить в состав множества U_0*, однако требование* выпуклости*, принципиальное для теоремы 25, в любом случае фактически обязывало бы функции* $g_i(u),\ i = m+1, m+2, \ldots, m+s$*, быть* ***линейными****.* Линейные *ограничения-равенства* (201) *можно было бы явно включить в постановку задачи* (198)*, приписать им дополнительные множители Лагранжа* $\lambda_i,\ i = m+1, m+2, \ldots, m+s$*, не являющиеся знакоопределёнными, и тогда теорема 25 по существу осталась бы в силе. Более точные формулировки можно найти в [1] и [2].*

Правило множителей Лагранжа для выпуклых задач
(продолжение)

Случай, когда $\lambda_0^* > 0$ ($\lambda_0^* = 1$) и функция Лагранжа принимает *классический вид*

$$L(u, \lambda) = J(u) + \sum_{i=1}^{m} \lambda_i \, g_i(u),$$

принято называть *регулярным*. К сожалению, регулярными являются далеко не все оптимизационные задачи вида (198).

Пример (нерегулярной задачи).
Рассмотрим следующую одномерную задачу минимизации:

$$J(u) = -u \to \inf, \quad u \in U = \{u \in U_0 = R^1 \mid g(u) = u^2 \leqslant 0\},$$

принадлежащую классу (199) выпуклых задач. У этой задачи единственным оптимальным решением является единственная допустимая точка $u_* = 0$. По теореме 26 для неё найдётся нетривиальный набор множителей Лагранжа $\lambda^* = (\lambda_0^*, \lambda_1^*)$, для которого пара (u_*, λ^*) удовлетворяет условиям (a),(b),(c). Пусть $\lambda^* = (\lambda_0^*, \lambda_1^*)$ — любой из таких наборов. Запишем условие (a):

$$-\lambda_0^* \, u + \lambda_1^* \, u^2 \geqslant -\lambda_0^* \, u_* + \lambda_1^* \, u_*^2 = 0 \quad \forall u \in U_0 = R^1.$$

Пользуясь произволом в выборе u, делаем вывод о том, что $\forall u > 0$ верно неравенство $-\lambda_0^* + \lambda_1^* \, u \geqslant 0$, а тогда, устремляя $u \to +0$, приходим к неравенству $-\lambda_0^* \geqslant 0$, которое с учётом нетрицательности $\lambda_0^* \geqslant 0$ приводит к равенству $\lambda_0^* = 0$ для *любого* набора множителей Лагранжа.

Поскольку с *нерегулярными* задачами, в которых $\lambda_0^* = 0$ и функция $J(u)$ остаётся «не у дел», иметь дело не очень приятно, хотелось бы иметь какие-то *достаточные условия регулярности*, гарантирующие нетривиальность главного множителя Лагранжа λ_0^*. Одним из них является так называемое

Достаточное условие регулярности Слейтера. *Пусть в задаче* (198), *выпуклой в смысле* (199),

$$\exists \, u_0 \in U_0 \; : \quad g_i(u_0) < 0 \quad \forall \, i = 1, 2, \ldots, m. \tag{S}$$

Тогда в любом наборе множителей Лагранжа λ^*, *удовлетворяющем в паре с элементом* u_* *условиям* (a),(b),(c), *обязательно* $\lambda_0^* > 0$.

Доказательство легко провести от противного: если бы в некотором наборе множителей Лагранжа $\lambda^* \neq 0$ из теоремы 26 компонента $\lambda_0^* = 0$, то в слейтеровой точке u_0 из (S) выполнялось бы строгое неравенство

$$L(u_0, \lambda^*) = 0 \cdot J(u_0) + \sum_{i=1}^{m} \lambda_i^* \, g_i(u_0) < 0,$$

поскольку среди λ_i^* есть хотя бы один *положительный* множитель, а все значения $g_i(u_0)$ *отрицательны* в силу (S). С другой стороны, из (c) и того, что $\lambda_0^* = 0$, имеем равенство $0 = L(u_*, \lambda^*)$. Получается, что $L(u_0, \lambda^*) < L(u_*, \lambda^*)$, а это противоречит принципу минимума (a).

В регулярном выпуклом случае правило множителей Лагранжа часто формулируют в другой весьма распространённой форме — в терминах седловой точки функции Лагранжа — и называют теоремой Куна-Таккера.

Определение 17. *Пусть X и Y — два произвольных множества и $f : X \times Y \to R^1$ — функция, определённая на их декартовом произведении. Точка $(x_*, y^*) \in X \times Y$ называется* седловой точкой *функции f на множестве $X \times Y$, если*

$$f(x_*, y) \leqslant f(x_*, y^*) \leqslant f(x, y^*) \quad \forall\, x \in X \;\; \forall\, y \in Y.$$

Теорема 27. (теорема Куна-Таккера) *Пусть выполнены условия теоремы* 26 *и условия регулярности Слейтера* (S). *Тогда для оптимальности элемента u_* в задаче* (198) *необходимо и достаточно, чтобы классическая функция Лагранжа с $\lambda_0 = 1$ имела седловую точку $\big(u_*, \lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*)\big)$ на множестве $U_0 \times R_+^m$.*

Доказательство. Запишем седловые неравенства для классической функции Лагранжа:

$$\begin{gathered} J(u_*) + \sum_{i=1}^{m} \lambda_i \, g_i(u_*) \leqslant J(u_*) + \sum_{i=1}^{m} \lambda_i^* \, g_i(u_*) \leqslant \\ \leqslant J(u) + \sum_{i=1}^{m} \lambda_i^* \, g_i(u) \quad \forall\, u \in U_0 \;\; \forall\, \lambda \in R_+^m. \end{gathered} \tag{202}$$

1) *Необходимость.* Пусть $u_* \in U_*$, т. е. точка u_* является оптимальным решением задачи (198). Тогда по теореме 26 и с учётом достаточного условия регулярности (S) утверждение (a) (принцип минимума) превратится в

правое неравенство (202). Условия дополняющей нежёсткости (c) приводят к равенству $\sum\limits_{i=1}^{m} \lambda_i^* \, g_i(u_*) = 0$, а поскольку $\lambda_i \geqslant 0$ и $g_i(u_*) \leqslant 0,\ i = 1, 2 \ldots, m$, то $\sum\limits_{i=1}^{m} \lambda_i \, g_i(u_*) \leqslant 0$, поэтому левое неравенство (202) тоже выполняется.

2) *Достаточность.* Пусть в некоторой точке $u_* \in U_0$ при некоторых $\lambda^* \geqslant 0$ выполнены седловые неравенства (202). Тогда из левого неравенства (202) имеем

$$\sum_{i=1}^{m} \left(\lambda_i - \lambda_i^*\right) g_i(u_*) \leqslant 0 \quad \forall\, \lambda \geqslant 0. \tag{203}$$

Фиксируем некоторое $\varepsilon > 0$ и подставим в (203) вектор

$$\lambda = \left(\lambda_1^*, \ldots, \lambda_{i-1}^*, \lambda_i^* + \varepsilon, \lambda_{i+1}^*, \ldots, \lambda_m^*\right) \geqslant 0.$$

Неравенство (203) примет вид $\varepsilon \, g_i(u_*) \leqslant 0$, значит, $g_i(u_*) \leqslant 0$, а поскольку это верно для любого номера $i = 1, 2, \ldots, m$, то $u_* \in U$, т. е. элемент u_* является *допустимым* в задаче (198). Если подставить в (203) вектор

$$\lambda = \left(\lambda_1^*, \ldots, \lambda_{i-1}^*, \lambda_i = 0, \lambda_{i+1}^*, \ldots, \lambda_m^*\right) \geqslant 0,$$

получим неравенство $-\lambda_i^* \, g_i(u_*) \leqslant 0$. Учитывая, что $\lambda_i^* \geqslant 0$ и $g_i(u_*) \leqslant 0$, приходим к равенству $\lambda_i^* \, g_i(u_*) = 0$, справедливому для любых номеров $i = 1, 2, \ldots, m$, что означает выполнение условий (c) дополняющей нежёсткости.

Теперь можно воспользоваться утверждением теоремы 25 в направлении *достаточности*, т. к. нам известно, что $\lambda_0^* = 1$, $u_* \in U$ и выполнены условия (a),(b),(c). В результате получаем, что $u_* \in U_*$.

Теорема 27 доказана. ▼

Правило множителей Лагранжа для гладких задач

Рассмотрим следующий класс задач минимизации в гильбертовом пространстве H :

$$J(u) \to \inf, \quad u \in U \subset H,$$
$$U = \Big\{ u \in H \;\Big|\; g_1(u) \leqslant 0, \ldots, g_m(u) \leqslant 0, g_{m+1}(u) = 0, \ldots, g_{m+s}(u) = 0 \Big\}. \tag{204}$$

Вид задачи почти такой же, как и в методе штрафов, только $U_0 = H$. Речь пойдёт, в основном, о *необходимых* условиях *локальной* оптимальности в форме правила множителей Лагранжа, но и о *достаточных* условиях оптимальности тоже будет кое-что сказано. Напомним, что точка u_* называется *точкой локального минимума* в задаче (204), если существует такая (достаточно малая) ε-окрестность

$$B_\varepsilon(u_*) = \big\{ u \in H \;\big|\; \|u - u_*\|_H < \varepsilon \big\}, \quad \varepsilon > 0,$$

для которой

$$J(u_*) \leqslant J(u) \quad \forall u \in B_\varepsilon(u_*) \cap U.$$

Именно для таких точек локального минимума мы приведём *необходимые* условия оптимальности в лагранжевой форме и при этом вместо требований *выпуклости* основными предположениями в данном случае будут требования *гладкости*.

Теорема 28. (`ПЛ для гладких задач`) *Пусть u_* – точка локального минимума в задаче* (204) *и все функционалы $J(u)$, $g_i(u)$, $i = 1, 2, \ldots, m+s$, непрерывно дифференцируемы по Фрешé в окрестности $B_\varepsilon(u_*)$ точки u_*. Тогда существует набор множителей Лагранжа $\lambda^* \in R^{m+s+1}$, $\lambda^* \neq 0$, для которого*

$$L'_u(u_*, \lambda^*) = \lambda_0^* J'(u_*) + \sum_{i=1}^{m+s} \lambda_i^* \, g_i'(u_*) = 0, \quad \text{(усл-е стационарности)} \tag{a}$$

$$\lambda_i^* \geqslant 0, \quad i = 0, 1, \ldots, m, \quad \text{(неотрицательность множителей)} \tag{b}$$

$$\lambda_i^* \, g_i(u_*) = 0, \quad i = 1, 2, \ldots, m. \quad \text{(усл-я дополняющей нежёсткости)} \tag{c}$$

Доказательство. Как и в выпуклых задачах, примем договорённость

$$J(u_*) = 0. \tag{d1}$$

Следующая договорённость поначалу может показаться подозрительной:

$$g_i(u_*) = 0 \quad \forall\, i = 1, 2, \ldots, m, \tag{d2}$$

поскольку она избавляет нас от необходимости *доказывать* справедливость условий дополняющей нежёсткости (c). Доводы в поддержку принятия договорённости (d2) основываются на условиях *гладкости*: если i-ое ограничение типа неравенства выполняется в точке u_* строго, т. е. $g_i(u_*) < 0$, то в мéньшей окрестности точка u_* будет точкой локального минимума в задаче (204), из постановки которой i-ое ограничение удалено. Тогда можно доказать теорему 27 для новой постановки, из которой *удалены все неактивные* ограничения типа неравенств, которые выполняются в точке u_* строго, после чего можно вспомнить о том, что раньше эти ограничения были и приписать им множители Лагранжа с нулевыми значениями: $\lambda_i^* = 0$.

Для учёта ограничений типа равенств введём отображение

$$G(u) = \big(g_{m+1}(u), g_{m+2}(u), \ldots, g_{m+s}(u)\big) : \quad H \to R^s.$$

Рассмотрим следующее ортогональное разложение коненомерного пространства R^s, связанное с производной $G'(u_*)$:

$$G'(u_*) = \big(g'_{m+1}(u_*), g'_{m+2}(u_*), \ldots, g'_{m+s}(u_*)\big) \in L(H \to R^s) :$$

$$R^s = \operatorname{Im} G'(u_*) \oplus \ker \big(G'(u_*)\big)^*. \tag{205}$$

Заметим, что в силу конечномерности пространства R^s имеем $\overline{\operatorname{Im} G'(u_*)} = \operatorname{Im} G'(u_*)$, поэтому в (205) знак замыкания над $\operatorname{Im} G'(u_*)$ отсутствует.

Замечание 28. *Напомним, что для линейных ограниченных операторов $A \in L(H \to F)$, действующих в гильбертовых пространствах H и F, справедливы ортогональные разложения*

$$F = \overline{\operatorname{Im} A} \oplus \ker A^*, \quad H = \overline{\operatorname{Im} A^*} \oplus \ker A.$$

Рассмотрим отдельно три возможности, связанные с разложенм (205) и для каждой из них завершим доказательство теоремы 28 по-своему.

1) $\operatorname{Im} G'(u_*) \neq R^s$ — «обычное» вырождение: среди ограничений типа равенств есть лишние, линейно зависимые.

2) $\operatorname{Im} G'(u_*) = R^s$ и $\ker G'(u_*) = \{0\}$ — «особое» вырождение: ограничений типа равенств слишком много и допустимое множество вырождается в точку.

3) $\operatorname{Im} G'(u_*) = R^s$ и $\dim \ker G'(u_*) \geqslant 1$ — невырожденный случай.

1) $\operatorname{Im} G'(u_*) \neq R^s$ — «обычное» вырождение. В этом случае ортогональное дополнение к $\operatorname{Im} G'(u_*)$ нетривиально:

$$\big(\operatorname{Im} G'(u_*)\big)^{\perp} = \ker \big(G'(u_*)\big)^* \neq \{0\},$$

т.е. $\exists\, \alpha = (\alpha_{m+1}, \alpha_{m+2}, \ldots, \alpha_{m+s}) \in R^s,\ \alpha \neq 0$, такой, что

$$\big(G'(u_*)\big)^* \alpha = \sum_{i=m+1}^{m+s} \alpha_i\, g_i'(u_*) = 0.$$

Данное равенство можно рассматривать как условие стационарности (a) с нетривиальным набором множителей Лагранжа

$$\lambda^* = \big(\lambda_0^* = 0, \lambda_1^* = 0, \ldots, \lambda_m^* = 0;\ \lambda_{m+1}^* = \alpha_{m+1}, \ldots, \lambda_{m+s}^* = \alpha_{m+s}\big),$$

нулевая головная часть которого обеспечивает выполнение условий неотрицательности (b). В случае 1) теорема 28 доказана.

2) В случае «особого» вырождения, когда $\operatorname{Im} G'(u_*) = R^s$ и $\ker G'(u_*) = \{0\}$, из ортогонального разложения

$$H = \operatorname{Im} \big(G'(u_*)\big)^* \oplus \ker G'(u_*)$$

следует, что

$$H = \operatorname{Im} \big(G'(u_*)\big)^* = \operatorname{span} \{g_i'(u_*)\}_{i=m+1}^{m+s},$$

т.е. $\dim H \leqslant s$. Понятно, что, тем самым, если H бесконечномерно, то никакого «особого» вырождения просто не бывает, а если оно конечномерно, то в силу условия $\operatorname{Im} G'(u_*) = R^s$ оказывается, что $\dim H = s$, и тогда s линейно независимых ограничений типа равенств фактически сжимают допустимое множество U в точку — в этом и есть смысл рассматриваемого «особого» вырождения. Разумеется, это весьма редкий, но, в принципе, возможный случай

и для него рассуждения следующие. Градиент $J'(u_*) \in H$ — это некоторый вектор из H и его можно разложить по базису $\{g_i'(u_*)\}_{i=m+1}^{m+s}$ этого пространства:

$$J'(u_*) = \sum_{i=m+1}^{m+s} \alpha_i \, g_i'(u_*), \quad \alpha_i \in R^1.$$

Данное равенство можно рассматривать как условие стационарности (a) с нетривиальным набором множителей Лагранжа

$$\lambda^* = \big(\lambda_0^* = 1, \lambda_1^* = 0, \ldots, \lambda_m^* = 0;\ \lambda_{m+1}^* = -\alpha_{m+1}, \ldots, \lambda_{m+s}^* = -\alpha_{m+s}\big),$$

головная часть которого обеспечивает выполнение условий неотрицательности (b). В случае 2) теорема 28 также доказана.

Лекция 17 (24 февраля 2022)

Правило множителей Лагранжа для гладких задач (продолжение)

3) В *невырожденном* случае, когда $\operatorname{Im} G'(u_*) = R^s$ и $\dim \ker G'(u_*) \geqslant 1$, для доказательства нам понадобится более тонкий математический аппарат, а именно, теорема Люстерника, которая даёт описание множества касательных векторов к многообразию, заданному ограничениями типа равенств.

Определение 18. *Пусть X — нормированное пространство, $M \subset X$ — некоторое множество из X и $x_0 \in M$ — некоторая точка из M. Вектор $h \in X$ называется* ***касательным*** *ко множеству M в точке x_0, если найдётся такое отображение $\varphi : R^1 \to X$, такое, что*

$$x_0 + t\,h + \varphi(t) \in M \quad \forall\, t \in (-\varepsilon, \varepsilon) \quad \text{при некотором } \varepsilon > 0,$$
$$\tfrac{\|\varphi(t)\|_X}{t} \to 0 \quad \text{при} \quad t \to 0, \quad \text{т. е.} \quad \varphi(t) = \overline{o}(t).$$

Заметим, что если h — касательный вектор, то вместе с ним касательными будут и все векторы из его линейной оболочки $\operatorname{span}\{h\}$. Множество всех касательных векторов ко множеству M в точке $x_0 \in M$ будем обозначать через $T_{x_0}M$.

Теорема Люстерника [АТФ] `(без доказательства)`
Пусть X, Y — банаховы пространства, $F : X \to Y$ — непрерывно дифференцируемое отображение, множество M задано уравнением (операторным ограничением типа равенства)

$$M = \big\{\, x \in X \;\big|\; F(x) = 0 \,\big\}.$$

Пусть $x_0 \in M$, т. е. $F(x_0) = 0$, и, кроме того, выполняется условие невырожденности (регулярности)

$$\operatorname{Im} F'(x_0) = Y.$$

Тогда

$$T_{x_0}M = \ker F'(x_0).$$

Вернёмся к доказательству теоремы 28 и дополним принятые нами договорённости (d1) и (d2) ещё одной, причём исключительно ради сокращения записей:

$$J(u) = g_0(u). \tag{d3}$$

Заметим, что (d3) согласуется и с (d1) и с (d2). Введём множества вида

$$V_k = \big\{u \in H \;\big|\; \langle g_i'(u_*), u\rangle < 0,\ i = k, k{+}1, \ldots, m;\ G'(u_*)u = 0\big\},\ k = 0, 1, \ldots, m,$$

и заметим, что

$$V_0 \subset V_1 \subset \cdots \subset V_m.$$

Покажем, что в этой цепочке $V_0 = \varnothing$, рассуждая «от противного». Если бы $\exists\, v \in V_0$, то по определению множества V_0 мы имели бы соотношения

$$\langle g_i'(u_*), v\rangle < 0,\ i = 0, 1, \ldots, m; \quad G'(u_*)v = 0.$$

Последнее означает, что $v \in \ker G'(u_*)$, а поскольку в невырожденном случае $\operatorname{Im} G'(u_*) = R^s$, то по теореме Люстерника

$$T_{u_*}\{G(u) = 0\} = \ker G'(u_*),$$

т. е. элемент v является *касательным* вектором ко множеству $\{G(u) = 0\}$ в точке u_*. Тогда по определению касательного вектора найдётся такая функция $\varphi(t) : (-\varepsilon, \varepsilon) \to H$, для которой

$$G\big(u(t)\big) = 0, \quad u(t) = u_* + t\,v + \varphi(t), \quad \forall\, t \in (-\varepsilon, \varepsilon), \quad \varphi(t) = \overline{o}(t). \qquad (206)$$

Соотношение (206) означает, что элементы $u(t)$ при всех $t \in (-\varepsilon, \varepsilon)$ удовлетворяют ограничениям типа равенств из постановки (204). Посмотрим на значения функций, задающих ограничения типа неравенств, и на значения функции $J(u) = g_0(u)$. В силу их дифференцируемости будем иметь

$$\begin{aligned} g_i(u(t)) &= g_i(u_*) + \langle g_i'(u_*),\, t\,v + \varphi(t)\rangle + \overline{o}(t) = \big[g_i(u_*) = 0\big] = \\ &= t\,\langle g_i'(u_*),\, v\rangle + \overline{o}(t) < \big[v \in V_0\big] < 0 \quad \forall\, t > 0 \text{ (достаточно малых)} \end{aligned} \qquad (207)$$

В результате для всех достаточно малых $t > 0$ элементы $u(t)$ будут *допустимыми* в задаче (204), при $t \to 0$ $\|u(t) - u_*\| \to 0$ и при этом $J(u(t)) = g_0(u(t)) < 0 = J(u_*)$, что вступает в противоречие с принятой нами договорённостью (d1). Таким образом, установлено, что $V_0 = \varnothing$.

Проверим, возможен ли случай, когда

$$V_m = \big\{u \in H \;\big|\; \langle g_m'(u_*), u\rangle < 0,\ G'(u_*)u = 0\big\} = \varnothing,$$

в то время как в условиях *невырожденности* $\dim\ker G'(u_*) \geqslant 1$. Если $V_m = \varnothing$, то линейный функционал $\langle g_m'(u_*), u\rangle$ *не может* принимать *отрицательных* значений на подпространстве $\ker G'(u_*)$. Для *линейного* функционала это может означать только одно: этот функционал *тождественно равен нулю* на $\ker G'(u_*)$ или, другими словами,

$$g_m'(u_*) \in \big(\ker G'(u_*)\big)^{\perp} = \operatorname{Im}\big(G'(u_*)\big)^*,$$

т. е. найдётся набор $\alpha = (\alpha_{m+1}, \alpha_{m+2}, \ldots, \alpha_{m+s}) \in R^s$, для которого

$$g'_m(u_*) = \sum_{i=m+1}^{m+s} \alpha_i \, g'_i(u_*).$$

Данное равенство можно интерпретировать как утверждение (a) теоремы 28, взяв нетривиальный набор множителей Лагранжа вида

$$\lambda^* = \big(\lambda^*_0 = 0, \lambda^*_1 = 0, \ldots, \lambda^*_{m-1} = 0, \lambda^*_m = 1;\ \lambda^*_{m+1} = -\alpha_{m+1}, \ldots, \lambda^*_{m+s} = -\alpha_{m+s}\big),$$

конструкция головной части которого обеспечивает выполнение условий неотрицательности (b).

В рамках *невырожденного* случая нам остаётся рассмотреть последнюю возможность, когда

$$V_0 = V_1 = \cdots = V_k = \varnothing, \quad V_{k+1} \neq \varnothing.$$

Поставим следующую вспомогательную задачу мимнимизации с *линейными* данными:

$$J_k(u) = \langle g'_k(u_*),\, u\rangle \to \inf, \quad u \in W_k,$$
$$W_k = \big\{\, \langle g'_i(u_*),\, u\rangle \leqslant 0,\ i = k+1, \ldots, m;\ \ G'(u_*)\, u = 0\big\}, \quad (208)$$

обратим внимание на то, что она относится к классу *выпуклых* задач минимизации, которые рассматривались выше в теореме 26 и в которых в роли «терпимого» множества выступает линейное подпространство $U_0 = \ker G'(u_*)$. Покажем, что точка $u = 0$ является оптимальным решением задачи (208). Эта точка допустима в задаче (208) и $J_k(0) = 0$, поэтому $J_{k*} = \inf\limits_{u \in W_k} J_k(u) \leqslant 0$. Покажем, что случай $J_{k*} < 0$ невозможен, рассуждая от противного. Если бы $J_{k*} < 0$, то в задаче (208) нашёлся бы такой элемент v, для которого

$$G'(u_*)\, v = 0, \quad \langle g'_k(u_*),\, v\rangle < 0, \quad \langle g'_i(u_*),\, v\rangle \leqslant 0, \quad i = k+1, \ldots, m. \quad (209)$$

В рассматриваемом случае $V_{k+1} \neq \varnothing$, поэтому найдётся такой элемент $w \in V_{k+1}$, для которого знак значения $\langle g'_k(u_*),\, w\rangle$ нам неизвестен, но мы знаем, что

$$G'(u_*)\, w = 0, \quad \langle g'_i(u_*),\, w\rangle < 0, \quad i = k+1, \ldots, m. \quad (210)$$

Линейность конструкций в (209) и (210) позволяет заключить, что для линейной комбинации вида $v + \varepsilon w$ при всех достаточно малых $\varepsilon > 0$ будут выполняться соотношения

$$G'(u_*)\,(v+\varepsilon w) = 0,\ \ \langle g'_k(u_*),\, v+\varepsilon w\rangle < 0,\ \ \langle g'_i(u_*),\, v+\varepsilon w\rangle < 0,\ i = k+1, \ldots, m,$$

противоречащие тому, что $V_k = \varnothing$. Теперь можно применить к выпуклой задаче (208) теорему 25 в направлении необходимости, зная, что точка $u = 0$ является её оптимальным решением, а также зная, что у этой задачи есть слейтерова точка w со свойствами (210). Получаем, что существует *классический* набор множителей Лагранжа $\lambda_k^* = 1, \lambda_{k+1}^* \geqslant 0, \ldots, \lambda_m^* \geqslant 0$, для которого принцип минимума (a) превращается в неравенство

$$1 \cdot \langle g_k'(u_*),\, u\rangle + \sum_{i=k+1}^{m} \langle \lambda_i^*\, g_i'(u_*),\, u\rangle \geqslant 0 \quad \forall\, u \in U_0 = \ker G'(u_*). \tag{211}$$

Поскольку это неравенство *линейное*, а вариации u в (211) ведутся по *подпространству*, то *неравенство* превращается в *равенство*:

$$1 \cdot \langle g_k'(u_*),\, u\rangle + \sum_{i=k+1}^{m} \langle \lambda_i^*\, g_i'(u_*),\, u\rangle = 0 \quad \forall\, u \in \ker G'(u_*),$$

которое означает, что

$$1 \cdot g_k'(u_*) + \sum_{i=k+1}^{m} \lambda_i^*\, g_i'(u_*) \in \big(\ker G'(u_*)\big)^{\perp} = \operatorname{Im}\big(G'(u_*)\big)^*,$$

т. е. найдётся набор $\alpha = (\alpha_{m+1}, \alpha_{m+2}, \ldots, \alpha_{m+s}) \in R^s$, для которого

$$1 \cdot g_k'(u_*) + \sum_{i=k+1}^{m} \lambda_i^*\, g_i'(u_*) = \sum_{i=m+1}^{m+s} \alpha_i\, g_i'(u_*).$$

Данное равенство можно интерпретировать как утверждение (a) теоремы 28, взяв нетривиальный набор множителей Лагранжа вида

$$\lambda^* = \big(\lambda_0^* = 0, \ldots, \lambda_{k-1}^* = 0, \lambda_k^* = 1, \lambda_{k+1}^*, \ldots, \lambda_m^*, \\ \lambda_{m+1}^* = -\alpha_{m+1}, \ldots, \lambda_{m+s}^* = -\alpha_{m+s}\big),$$

конструкция головной части которого обеспечивает выполнение условий неотрицательности (b). Доказательство теоремы 28 завершено. ▼

Приведём достаточные условия регулярности задачи (204), гарантирующие, что в наборе множителей Лагранжа обязательно $\lambda_0^* = 1$. Из доказательства финальной части теоремы 28 видно, что они могум быть сформулированы в виде условий

$$V_0 = \varnothing, \quad V_1 \neq \varnothing.$$

В развёрнутой форме эти условия можно сформулировать в следующем виде.

Утверждение. *Если в дополнение к условиям теоремы* 28

$$\operatorname{Im} G'(u_*) = R^s$$

и, кроме того,

$$\exists\, h \in \ker G'(u_*): \quad \langle g_i'(u_*),\, h\rangle < 0 \quad \forall\, i = 1, 2, \ldots, m,$$

то в ***любом*** *наборе множителей Лагранжа* $\lambda_0^* = 1$.

Для доказательства используем рассуждения от противного. Предположим, что в некотором наборе множителей Лагранжа $\lambda_0^* = 0$. Тогда условие стационарности (а) примет вид $\sum\limits_{i=1}^{m+s} \lambda_i^*\, g_i'(u_*) = 0$, в частности, для элемента h будем иметь

$$0 = \Big\langle \sum_{i=1}^{m+s} \lambda_i^*\, g_i'(u_*),\, h\Big\rangle = \sum_{i=1}^{m} \lambda_i^* \left\langle g_i'(u_*),\, h\right\rangle + \sum_{i=m+1}^{m+s} \lambda_i^* \left\langle g_i'(u_*),\, h\right\rangle. \qquad (212)$$

Вторая сумма в правой части (212) будет равна нулю, т. к. $h \in \ker G'(u_*)$, а в первой сумме *все без исключения* коэффициенты при $\lambda_i^* \geqslant 0$ *отрицательны*, следовательно, нулевой итоговый результат в (212) может получиться только при $\lambda_1^* = \lambda_2^* = \cdots = \lambda_m^* = 0$. Тогда условие стационарности примет укороченный вид:

$$\sum_{i=m+1}^{m+s} \lambda_i^* \left\langle g_i'(u_*),\, u\right\rangle_H = \left\langle (\lambda_{m+1}^*, \lambda_{m+2}^*, \ldots, \lambda_{m+s}^*),\, G'(u_*)\, u\right\rangle_{R^s} = 0 \quad \forall\, u \in H,$$

и поскольку по условию $\operatorname{Im} G'(u_*) = R^s$, то остаётся единственная возможность $\lambda_{m+1}^* = \lambda_{m+2}^* = \cdots = \lambda_{m+s}^* = 0$, которая противоречит тому, что в целом набор множителей Лагранжа $\lambda^* \neq 0$, поэтому обязательно $\lambda_0^* = 1$.

Замечание 29. *Если в постановке задачи* (204) *ограничения типа неравенств отсутствуют, то* достаточным *условием регулярности, при котором* $\lambda_0^* = 1$, *будет условие невырожденности ограничений типа равенств:* $\operatorname{Im} G'(u_*) = R^s$.

Теперь приведём обещанные *достаточные условия локальной оптимальности* для гладких задач минимизации вида (204).

Теорема 29. *(*`достат. усл-я оптимальности для гладких задач`*)*
Пусть в задаче (204) *все функционалы дважды непрерывно дифференцируемы по Фреше в окрестности* $B_\varepsilon(u_*)$ *точки* u_*, *которая является допустимой:* $u_* \in U$ *и существует классический набор множителей Лагранжа* $\lambda^* = (\lambda_0^* = 1, \lambda_1^*, \ldots, \lambda_{m+s}^*)$, *такой, что для пары* (u_*, λ^*) *выполняются утверждения* (a),(b),(c) *теоремы* 27 *и, кроме того, вторая производная функции Лагранжа* $L''_{uu}(u_*, \lambda^*) \in L(H \to H)$ *является положительно определённым оператором:*

$$\langle L''_{uu}(u_*, \lambda^*)\, h,\, h\rangle \geqslant \varkappa \|h\|^2 \quad \forall\, h \in H \quad (\varkappa = \text{const} > 0).$$

Тогда точка u_* *является точкой локального минимума в задаче* (204).

Доказательство. Для любой допустимой и близрасположенной к u_* точки $u \in U \cap B_\varepsilon(u_*)$ будем иметь:

$$\begin{aligned}
J(u) &\geqslant \\
&\geqslant \left[\ \lambda_0^* = 1,\ \lambda_i^* \geqslant 0,\ g_i(u) \leqslant 0,\ i = \overline{1,m};\ g_i(u) = 0,\ i = \overline{m+1, m+s}\ \right] \geqslant \\
&\qquad \geqslant L(u, \lambda^*) = L(u, \lambda^*) \mp L(u_*, \lambda^*) = \langle L'_u(u_*, \lambda^*),\, u - u_*\rangle + \\
&\quad + \frac{1}{2}\langle L''_{uu}(u_*, \lambda^*)(u - u_*),\, u - u_*\rangle + \overline{o}(\|u - u_*\|^2) + L(u_*, \lambda^*) \geqslant \\
&\quad \geqslant \left[\ L'_u(u_*, \lambda^*) \overset{\text{(a)}}{=} 0, \quad L''_{uu}(u_*, \lambda^*) > 0, \quad L(u_*, \lambda^*) \overset{\text{(c)}}{=} J(u_*)\ \right] \geqslant \\
&\qquad\qquad \geqslant \frac{\varkappa}{2}\|u - u_*\|^2 + \overline{o}(\|u - u_*\|^2) + J(u_*) \geqslant J(u_*).
\end{aligned}$$

Теорема 29 доказана. ▼

STOP 24 февраля 2022

Список литературы

[1] *Васильев Ф.П.* Методы оптимизации: В 2-х кн. М., МЦНМО, 2011 (Факториал Пресс, 2002).

[2] *Сухарев А.Г., Тимохов А.В., Федоров В.В.* Курс методов оптимизации. М., Физматлит, 2005 (Наука, 1986).

[3] *Колмогоров А.Н., Фомин С.В.* Элементы теории функций и функционального анализа. М., Наука, 1976.

[4] *Алексеев В.М., Тихомиров В.М., Фомин С.В.* Оптимальное управление. М., Физматлит, 2005 (Наука, 1979).

[5] *Nocedal J., Wright S.J.* Numerical Optimization. Springer, NY, 2006.