Методы оптимизации — коллоквиум 2026

Всё в одном файле: формат и порядок подготовки, объяснение курса с нуля, шаблон устного ответа, разбор трёх тестовых вариантов, карта типов задач и полная теория по 80 вопросам списка Kollok-26_v1.

Оглавление

НАЧАЛО

С чего начать: формат и порядок

2. КАК ОТВЕЧАТЬ И БИЛЕТЫ

Шаблон ответа

Вариант 1 разобран

Вариант 2 разобран

Вариант 3 разобран

Карта задач

СПРАВОЧНО

Где лежит теория: записи пар, книги

1. ПОНЯТЬ С НУЛЯ

Одна идея на весь курс

Пять разделов — одна идея

3. ТЕОРИЯ

Вводная часть В1–В30

О1–О8. Без ограничений, матрицы

О9–О15. Выпуклость

О16–О18. Равенства

О19–О22. Неравенства

О23–О32. Численные методы

О33–О35. Выпуклое программирование

О36. Динамическое программирование

О37–О40. Функционалы, вариации

О41–О46. Простейшая задача, Больца

О47–О50. Условия 2-го порядка

Шпаргалка на одну страницу

С чего начать: формат и порядок действий

Как будет на коллоквиуме (так преподаватель объяснила в чате 26.11):

- Билет — это **три задачи A, B, C по 2 балла**, как в тестовых вариантах. Иногда одна из трёх — просто вопрос из списка.

- Решаешь **вслух**: называешь определение или теорему, говоришь, что по ней надо проверить, и проверяешь на числах. Ей важно услышать ход мысли.
- Отдельных вопросов из списка почти не будет, но без них задачу не обосновать. Поэтому теория ниже всё равно нужна: это то, что ты называешь на шаге 2.
- В старом описании (Kollok-26_v1) перед билетом ещё были 2 вопроса из вводной части без подготовки. Если это останется, вводную часть B1–B30 надо знать на 100 %.

Порядок действий — идёшь по странице сверху вниз:

1. Одна идея на весь курс и пять разделов — прочитать, понять картинку. 30–40 минут.
2. Шаблон ответа — четыре фразы, по ним отвечается любая задача.
3. Варианты 1–3 — читаешь задачу, закрываешь решение, говоришь вслух по шаблону, сверяешь. Под каждой задачей ссылки на нужную теорию.
4. Карта задач — 15 типов задач. Для каждого: какую теорему назвать и что проверить.
5. Вводная часть B1–B30 — короткие определения. Кнопка «Скрыть ответы» наверху: читаешь вопрос, отвечаешь вслух, нажимаешь на ответ и сверяешь.
6. Основная часть O1–O50 — полные формулировки, доказательства и примеры. Открывай по ссылкам из карты задач, читать подряд не обязательно.
7. Перед сдачей: варианты 1–3 вслух на время (2–3 минуты на задачу) и кнопка «Случайный билет».

Одна идея на весь курс

Оптимизация — это поиск самой низкой (или самой высокой) точки. Представь, что ты стоишь на местности с холмами и ямами и ищешь дно ямы.

Главная мысль всего курса: если ты на дне ямы, то **ни один шаг, который тебе разрешён, не ведёт вниз**. Всё остальное в курсе — это та же фраза, переведённая на язык формул для разных ситуаций:

- где можно ходить (везде, по линии, по области),
- по чему ходим (по точкам $x \in \mathbb{R}^n$ или по целым функциям $x(t)$),
- насколько точно смотрим на «наклон» (первая производная — наклон, вторая — изгиб).

Три вопроса, которые задают про любую точку

1. Необходимое условие — «если это минимум, то обязательно выполнено вот это». Например, производная равна нулю. Это фильтр: точки, где условие не выполнено, сразу выбрасываем. Но прошедшие фильтр — ещё не обязательно минимумы. Это просто **кандидаты**.

2. Достаточное условие — «если выполнено вот это, то точно минимум». Например, производная ноль и вторая производная положительна (яма выгнута вверх).

3. Существует ли решение вообще — может, функция уходит в минус бесконечность и дна нет.

Пример на пальцах: $f(x) = x^3$. В нуле $f'(0) = 0$ — необходимое условие выполнено, кандидат. Но это не минимум, а перегиб: слева значения меньше, справа больше. Поэтому **необходимое ≠ достаточное**. Это различие спрашивают постоянно (например, в варианте 1).

Наклон и изгиб

Градиент ∇f — вектор из частных производных. Он показывает, куда функция растёт быстрее всего. Значит, $-\nabla f$ — куда она быстрее всего падает.

Матрица Гессе $\nabla^2 f$ — матрица вторых производных. Она описывает изгиб: чаша вверх (минимум), чаша вниз (максимум) или седло (по одной оси вверх, по другой вниз).

«Чаша вверх» на языке матриц — **положительно определённая** матрица: $h^\top A h > 0$ для любого $h \neq 0$. Проверяется **критерием Сильвестра**: все угловые миноры > 0.

Пять разделов курса — одна и та же идея

Раздел	Где можно ходить	«Наверху ямы нельзя сделать шаг вниз» превращается в
Без ограничений	куда угодно в $\mathbb{R}^n$	$\nabla f(x^*) = 0$ (теорема Ферма). Дальше по Гессе: $\succ 0$ — минимум, $\prec 0$ — максимум, знакопеременная — седло.
Ограничения-равенства $g(x) = 0$	только по поверхности (линии) $g = 0$	Градиент f не обязан быть нулём — он может смотреть «поперёк» линии, туда, куда ходить всё равно нельзя. Поэтому ∇f параллелен ∇g: $\nabla f + \lambda \nabla g = 0$. Это **метод множителей Лагранжа**.
Неравенства $g(x) \le 0$	по области, у её края нельзя выходить наружу	То же самое, но $\lambda \ge 0$, и если точка не на границе, то $\lambda = 0$ (граница «не мешает»). Это **теорема Куна–Таккера**.
Численные методы	—	Если решить $\nabla f = 0$ руками нельзя, идём к дну шагами: шаг против градиента (градиентный спуск) или прыжок в минимум параболы-приближения (метод Ньютона).
Вариационное исчисление	ищем не точку, а целую функцию $x(t)$	«Шаг» — это добавка $\alpha h(t)$ к функции. Аналог производной называется **вариация** δJ. Аналог $\nabla f = 0$ — **уравнение Эйлера–Лагранжа**. Его решения называют **экстремалями** — это кандидаты.

И отдельная тема, которая помогает везде — **выпуклость**. Выпуклая функция — это одна сплошная чаша без лишних ям. У неё любая найденная «яма» — сразу самое глубокое место (локальный минимум = глобальный). Поэтому в выпуклых задачах **необходимые условия становятся достаточными**.

Что такое регулярность (в варианте 1 спрашивают)

В методе Лагранжа есть тонкость. Полная функция Лагранжа: $\mathcal{L} = \lambda_0 f + \lambda_1 g_1 + \dots$ Если $\lambda_0 = 0$, то f вообще пропадает из условий, и они ничего не говорят про нашу функцию. Чтобы такого не было, нужна **регулярность**: градиенты ограничений $\nabla g_i(x^*)$ линейно независимы. Тогда можно взять $\lambda_0 = 1$. При одном ограничении это просто $\nabla g(x^*) \ne 0$.

Что такое вариация (в варианте 1 и 2 спрашивают)

Берём функцию $x(t)$, чуть сдвигаем её в направлении $h(t)$: $x + \alpha h$. Смотрим, как меняется J, когда α маленькое:

$$\delta J(x, h) = \lim_{\alpha \to 0} \frac{J(x + \alpha h) - J(x)}{\alpha} = \frac{d}{d\alpha} J(x + \alpha h)\Big|_{\alpha = 0}.$$

Это обычная производная по числу α. Практически: подставь $x + \alpha h$, раскрой скобки, возьми коэффициент при α в первой степени.

Как отвечать: шаблон из четырёх фраз

Преподаватель сказала прямо: видите задачу → называете теорему → говорите, что по ней надо проверить → проверяете на конкретных числах. Она хочет слышать ход мысли. Поэтому любой ответ строится одинаково:

1. **Что это за задача.** «Это задача с ограничением-равенством» / «это простейшая задача вариационного исчисления» / «это задача Больца».
2. **Какое определение или теорему применяю.** Назвать и сформулировать.
3. **Значит, надо проверить вот это.** Перевести теорему в конкретные действия.
4. **Проверяю и делаю вывод.** Посчитать и ответить на вопрос задачи одной фразой.

Если забыл точную формулировку — проговори идею своими словами («в точке минимума нельзя сдвинуться вниз, значит производная по любому направлению ноль»). Это лучше молчания: видно, что ты понимаешь, о чём речь.

Вариант 1

1A. Регулярность и достаточность в задаче $x^2 + y^2 \to \text{extr}$, $3x + 4y = 12$

Теория: B16, B13, B14, B12, O14

ГОВОРИШЬ ВСЛУХ

Это задача с одним ограничением-равенством: $f = x^2 + y^2$, $g = 3x + 4y - 12 = 0$.

Определение регулярности: точка регулярна, если градиенты ограничений в ней линейно независимы. У нас одно ограничение, значит, надо проверить, что $\nabla g \neq 0$.

Проверяю: $\nabla g = (3, 4) \neq 0$ в любой точке. Значит, условие регулярности выполнено везде, и в функции Лагранжа можно взять $\lambda_0 = 1$:

$$\mathcal{L} = x^2 + y^2 + \lambda(3x + 4y - 12).$$

Необходимые условия: $\mathcal{L}_x = 2x + 3\lambda = 0$, $\mathcal{L}_y = 2y + 4\lambda = 0$, $3x + 4y = 12$. Отсюда $x = -\frac{3}{2}\lambda$, $y = -2\lambda$, подставляю: $-\frac{25}{2}\lambda = 12$, $\lambda = -\frac{24}{25}$, точка $\left(\frac{36}{25}, \frac{48}{25}\right)$.

Достаточны ли условия. Применяю теорему о выпуклой задаче: если функция выпуклая, а допустимое множество выпуклое, то точка, удовлетворяющая необходимым условиям, — глобальный минимум.

Проверяю: гессиан f равен $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succ 0$ (по Сильвестру $\Delta_1 = 2 > 0$, $\Delta_2 = 4 > 0$), значит f строго выпукла.

Ограничение линейное, прямая — выпуклое множество. Значит, задача на минимум выпуклая, и для минимума необходимые условия являются и достаточными: найденная точка — единственный глобальный минимум.

Для максимума — нет: условия дали ту же точку, а максимума не существует, потому что на прямой $x^2 + y^2 \to +\infty$. Так что в общем случае необходимые условия не достаточны; здесь достаточны только для минимума благодаря выпуклости.

Смысл на картинке: ищем ближайшую к началу координат точку прямой. Это основание перпендикуляра, и он один.

1B. Вариация $J(x) = \int_0^1 (\dot{x}^2 + x^2)\,dt - 2x(1)\sinh 1$

Теория: B26, O38

ГОВОРИШЬ ВСЛУХ

Определение: вариация по Лагранжу — это $\delta J(x, h) = \dfrac{d}{d\alpha} J(x + \alpha h)\Big|_{\alpha=0}$, где $h \in C^1[0, 1]$ — направление.

Подставляю $x + \alpha h$:

$$J(x+\alpha h)=\int_0^1\big((\dot x+\alpha\dot h)^2+(x+\alpha h)^2\big)dt-2\big(x(1)+\alpha h(1)\big)\sinh 1.$$

Раскрываю скобки:

$$=J(x)+\alpha\Big[\int_0^1(2\dot x\dot h+2xh)\,dt-2h(1)\sinh 1\Big]+\alpha^2\int_0^1(\dot h^2+h^2)\,dt.$$

Производная по α в нуле — это коэффициент при α:

$$\delta J(x,h)=\int_0^1\big(2\dot x\dot h+2xh\big)\,dt-2h(1)\sinh 1.$$

Если спросят дальше: это задача Больца. Уравнение Эйлера $\ddot x=x$, условия трансверсальности $\dot x(0)=0$, $\dot x(1)=\sinh 1$, экстремаль $x=\cosh t$.

1C. Направление убывания $f=x^2+xy-2y^2+50$ в точке $(2,2)$

Теория: B20, B21, O24

ГОВОРИШЬ ВСЛУХ

Определение: вектор h — направление убывания в точке x^0, если $f(x^0+\alpha h)<f(x^0)$ при всех достаточно малых $\alpha>0$.

Достаточное условие: если $\langle\nabla f(x^0),h\rangle<0$, то h — направление убывания. Значит, надо найти градиент и взять вектор с отрицательным скалярным произведением.

Считаю: $\nabla f=(2x+y,\ x-4y)$, в точке $(2,2)$: $\nabla f=(6,-6)$.

Беру $h=-\nabla f=(-6,6)$, или проще $h=(-1,1)$: $\langle(6,-6),(-1,1)\rangle=-12<0$. Значит, $h=(-1,1)$ — направление убывания. Это направление наискорейшего убывания (антиградиент).

Вариант 2

2A. g, h выпуклы на выпуклом D. Что можно сказать про $f = 2g - h$?

Теория: B8, B9, O11

ГОВОРИШЬ ВСЛУХ

Определение выпуклой функции: $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$ для всех $x, y \in D$, $\lambda \in [0, 1]$.

Свойства: выпуклая функция, умноженная на неотрицательное число, остаётся выпуклой. Сумма выпуклых функций выпукла. А при умножении на минус выпуклая функция становится вогнутой.

Значит, $2g$ выпукла, $-h$ вогнута, и f — это сумма выпуклой и вогнутой функций. Про такую сумму в общем случае ничего сказать нельзя. Показываю на примерах:

- $g = x^2$, $h = x^2$: $f = x^2$ — выпуклая;
- $g = 0$, $h = x^2$: $f = -x^2$ — вогнутая, не выпуклая;
- $g = x^2$, $h = x^4$: $f = 2x^2 - x^4$ — ни выпуклая, ни вогнутая ($f'' = 4 - 12x^2$ меняет знак).

Вывод: f может быть какой угодно. Выпуклость гарантирована, только если h линейная (аффинная): тогда $-h$ тоже выпукла.

2B. Верно ли, что метод Ньютона находит минимум за одну итерацию?

Теория: O32

ГОВОРИШЬ ВСЛУХ

Метод Ньютона: $x^{k+1} = x^k - \left[\nabla^2 f(x^k)\right]^{-1} \nabla f(x^k)$. Идея: заменить функцию её квадратичным приближением по Тейлору и прыгнуть в минимум этой параболы.

В общем случае — **неверно**. Если функция сама квадратичная, приближение совпадает с ней, и прыжок попадает точно в минимум. Для квадратичной $f = \frac{1}{2} x^\top A x - b^\top x$ с $A \succ 0$: $\nabla f = Ax - b$, $\nabla^2 f = A$, и $x^1 = x^0 - A^{-1}(Ax^0 - b) = A^{-1} b$ — это и есть точка минимума, за один шаг из любой начальной точки.

Контрпример для неквадратичной: $f = x^4$, минимум в нуле. Шаг: $x^{k+1} = x^k - \dfrac{4(x^k)^3}{12(x^k)^2} = \frac{2}{3} x^k$. Из $x^0 = 1$ получаем $\frac{2}{3}, \frac{4}{9}, \ldots$ — к нулю сходится, но за конечное число шагов не приходит.

Ответ: неверно в общем случае; верно для квадратичной функции с положительно определённой матрицей.

2C. Является ли $x(t) = t + 1$ экстремалью в задаче $\int_0^1 \dot{x}^2 dt \to \text{extr}$, $x(0) = 1$, $x(1) = 0$?

Теория: B27, O44

ГОВОРИШЬ ВСЛУХ

Это простейшая задача вариационного исчисления, $L = \dot{x}^2$.

Определение: экстремаль — решение уравнения Эйлера–Лагранжа $-\dfrac{d}{dt}L_{\dot{x}} + L_x = 0$. **Допустимая экстремаль** — экстремаль, которая ещё и удовлетворяет граничным условиям.

Считаю: $L_x = 0$, $L_{\dot{x}} = 2\dot{x}$, уравнение: $-2\ddot{x} = 0$, то есть $\ddot{x} = 0$.

Для $x = t + 1$: $\ddot{x} = 0$ — уравнение выполнено, значит это **экстремаль**.

Граничные условия: $x(0) = 1$ — да, а $x(1) = 2 \neq 0$ — нет. Значит, это **не допустимая** экстремаль. Допустимая здесь $x = 1 - t$.

Вариант 3

3A. Функция Лагранжа и условия 1-го порядка для $f_0 \to \max$, $f_1 = 10$, $f_2 \geq 0$

Теория: B17, B18, O19

ГОВОРИШЬ ВСЛУХ

Это задача с равенством и неравенством. Теорема Куна–Таккера сформулирована для минимума с неравенствами вида ≤ 0, поэтому сначала привожу задачу к этому виду:

$$-f_0(x) \to \min, \qquad f_1(x) - 10 = 0, \qquad -f_2(x) \leq 0.$$

Функция Лагранжа:

$$\mathcal{L} = \lambda_0\big(-f_0(x)\big) + \mu\big(f_1(x) - 10\big) + \lambda\big(-f_2(x)\big).$$

Теорема (Кун–Таккер / Джон): пусть f_0, f_1, f_2 непрерывно дифференцируемы около x^* и x^* — точка локального максимума. Тогда найдутся λ_0, μ, λ, не все равные нулю, такие что:

1. стационарность: $-\lambda_0 \nabla f_0(x^*) + \mu \nabla f_1(x^*) - \lambda \nabla f_2(x^*) = 0$;
2. знаки: $\lambda_0 \geq 0$, $\lambda \geq 0$; μ — любого знака (у равенства знак не ограничен);
3. дополняющая нежёсткость: $\lambda f_2(x^*) = 0$ — если ограничение не активно ($f_2 > 0$), то $\lambda = 0$;
4. допустимость: $f_1(x^*) = 10$, $f_2(x^*) \geq 0$.

Если точка регулярна, то $\lambda_0 \neq 0$ и можно взять $\lambda_0 = 1$.

3B. Докажите, что $\{x \in \mathbb{R}^3 : \ 2x_1 + 3x_2 + 4x_3 = 26\}$ выпукло

Теория: B7, O9

ГОВОРИШЬ ВСЛУХ

Определение: множество выпукло, если вместе с любыми двумя точками содержит весь отрезок между ними: для любых x, y из множества и $\lambda \in [0, 1]$ точка $z = \lambda x + (1 - \lambda) y$ тоже в нём.

Беру x, y из множества: $2x_1 + 3x_2 + 4x_3 = 26$ и $2y_1 + 3y_2 + 4y_3 = 26$. Беру $z = \lambda x + (1 - \lambda) y$ и проверяю:

$$2z_1 + 3z_2 + 4z_3 = \lambda(2x_1 + 3x_2 + 4x_3) + (1 - \lambda)(2y_1 + 3y_2 + 4y_3) = 26\lambda + 26(1 - \lambda) = 26.$$

Значит, z лежит в множестве, и множество выпукло. Всё держится на том, что левая часть линейна.

3C. Является ли $x(t) = 0$ допустимой экстремалью в задаче $\int_0^2 \dot{x}^2 dt + 2x^2(0) - 4x^2(2) \to \text{extr}$?

Теория: B28, O46

ГОВОРИШЬ ВСЛУХ

Концы не закреплены, а есть внеинтегральная часть — это **задача Больца**: $L = \dot{x}^2$, $l = 2x_0^2 - 4x_1^2$, где $x_0 = x(0)$, $x_1 = x(2)$.

Необходимые условия для задачи Больца: уравнение Эйлера–Лагранжа $-\dfrac{d}{dt}L_{\dot{x}} + L_x = 0$ и условия трансверсальности $L_{\dot{x}}(t_0) = l_{x_0}$, $L_{\dot{x}}(t_1) = -l_{x_1}$. Допустимая экстремаль должна удовлетворять всем им.

Считаю: $L_{\dot{x}} = 2\dot{x}$, $l_{x_0} = 4x_0$, $l_{x_1} = -8x_1$.

- Уравнение Эйлера: $\ddot{x} = 0$. Для $x \equiv 0$ выполнено.
- Левый конец: $2\dot{x}(0) = 4x(0)$, то есть $0 = 0$ — выполнено.
- Правый конец: $2\dot{x}(2) = 8x(2)$, то есть $0 = 0$ — выполнено.

Ответ: да, $x \equiv 0$ — допустимая экстремаль.

Если спросят, экстремум ли это: нет. На константе $x \equiv c$ значение $2c^2 - 4c^2 = -2c^2 < 0 = J(0)$, значит не минимум. На $x = c\sin\frac{\pi t}{2}$ (ноль на обоих концах) значение $\int \dot{x}^2 > 0$, значит не максимум. Экстремаль — только кандидат.

Карта задач: увидел → применяю → проверяю

Билеты собраны из небольшого набора типов. Ниже все основные. Учить надо именно эти связки, а не 80 вопросов подряд.

В задаче видишь	Применяешь	Проверяешь	Теория
Найти экстремумы $f(x, y)$ без ограничений	Необходимое условие (Ферма) и достаточное по Гессе	Решить $\nabla f = 0$. В каждой точке Гессе и Сильвестр: все $\Delta_k > 0$ — min; знаки $-, +, -$ — max; $\Delta_2 < 0$ — седло	B4, B5, B6, O8
Одномерная, $f'(x^*) = 0$ и $f''(x^*) = 0$	Теорема о первой ненулевой производной	Ищешь первую ненулевую $f^{(m)}(x^*)$. m чётное: > 0 — min, < 0 — max. m нечётное — экстремума нет (как у x^3)	O4
Знакоопределённость матрицы	Критерий Сильвестра	Угловые миноры. Если какой-то равен 0 — смотреть все главные миноры	B2, B3, O7
Существует ли минимум	Вейерштрасс / бесконечно растущая функция	Непрерывна на компакте или $f \to +\infty$ при $\|x\| \to \infty$	O2, O3
Выпукло ли множество	Определение выпуклого множества	Взять x, y, $z = \lambda x + (1-\lambda)y$, показать, что z в множестве	B7, O9
Выпукла ли функция	Критерий: $\nabla^2 f \succeq 0$ везде; или свойства (сумма, умножение на $\alpha \geq 0$)	Посчитать Гессе, Сильвестр. Для «что можно сказать» — контрпримеры	B8, B10, O11, O13
Ограничения-равенства	Функция Лагранжа, регулярность	∇g_i линейно независимы → $\lambda_0 = 1$; решить $\nabla \mathcal{L} = 0$, $g = 0$	B13, B14, B16, O16
Достаточны ли необходимые условия	Выпуклая задача или достаточные условия 2-го порядка	f выпукла и множество выпукло → достаточны. Иначе: $d^\top \nabla^2 \mathcal{L}\, d > 0$ на касательных d	B12, B15, O14, O18

Неравенства	Кун–Таккер	Привести к $\min$, $g \le 0$. Стационарность, $\lambda \ge 0$, $\lambda_i g_i = 0$, допустимость	B17, B18, O19
Направление убывания	Достаточное условие $\langle \nabla f, h \rangle < 0$	Посчитать градиент в точке, взять $h = -\nabla f$	B20, B21, O24
Шаг градиентного спуска / Ньютона	Формулы шага	$x^{k+1} = x^k - \alpha \nabla f(x^k)$; Ньютон: $x^{k+1} = x^k - [\nabla^2 f]^{-1} \nabla f$	B22, B23, O26, O32
Вычислить вариацию	Определение вариации по Лагранжу	Подставить $x + \alpha h$, взять коэффициент при α	B25, B26, O38
$\int L\,dt$ с закреплёнными концами	Простейшая задача, уравнение Эйлера–Лагранжа	Решить $-\frac{d}{dt} L_{\dot{x}} + L_x = 0$, подставить граничные условия	B27, B29, O44
$\int L\,dt + l(x(t_0), x(t_1))$, концы свободны	Задача Больца	Эйлер–Лагранж + трансверсальность: $L_{\dot{x}}(t_0) = l_{x_0}$, $L_{\dot{x}}(t_1) = -l_{x_1}$ («на правом конце минус»)	B28, O46
Минимум или максимум на экстремали	Условие Лежандра	$L_{\dot{x}\dot{x}}$ на экстремали: ≥ 0 необходимо для min, ≤ 0 для max	B30, O48

Вводная часть — знать наизусть

Отвечать надо коротко и точно. Ниже текст, который можно говорить почти дословно.

B1 Локальное и глобальное решение, строгие экстремумы, отрицания

Задача: $f(x) \to \min$, $x \in X \subseteq \mathbb{R}^n$ (X — допустимое множество; без ограничений $X = \mathbb{R}^n$).

- **Глобальный минимум:** $x^* \in X$ и $f(x^*) \le f(x)$ для всех $x \in X$. **Строгий глобальный:** $f(x^*) < f(x)$ для всех $x \in X$, $x \ne x^*$.
- **Локальный минимум:** $x^* \in X$ и $\exists \varepsilon > 0$: $f(x^*) \le f(x)$ для всех $x \in X \cap U_\varepsilon(x^*)$. **Строгий локальный:** $f(x^*) < f(x)$ для всех $x \in X \cap U_\varepsilon(x^*)$, $x \ne x^*$.
- Максимум — те же определения с обратными знаками неравенств. Экстремум = минимум или максимум.

Отрицания (кванторы меняются местами, неравенство переворачивается):

- x^* **не локальный минимум**: $\forall \varepsilon > 0\ \exists x \in X \cap U_\varepsilon(x^*)$: $f(x) < f(x^*)$.
- x^* **не точка локального экстремума** (ни min, ни max): $\forall \varepsilon > 0\ \exists x_1, x_2 \in X \cap U_\varepsilon(x^*)$: $f(x_1) < f(x^*) < f(x_2)$.
- x^* **не глобальный минимум**: $\exists x \in X$: $f(x) < f(x^*)$. **Не глобальный экстремум:** $\exists x_1, x_2 \in X$: $f(x_1) < f(x^*) < f(x_2)$.

В задаче с ограничениями всё то же самое, только сравниваем лишь с допустимыми точками $x \in X$.

B2 Знакоопределённость симметрической матрицы

$A = A^\top \in \mathbb{R}^{n \times n}$, квадратичная форма $Q(x) = x^\top A x$.

- положительно определена ($A \succ 0$): $x^\top A x > 0$ для всех $x \ne 0$;
- отрицательно определена ($A \prec 0$): $x^\top A x < 0$ для всех $x \ne 0$;
- неотрицательно определена ($A \succeq 0$): $x^\top A x \ge 0$ для всех x;
- неположительно определена ($A \preceq 0$): $x^\top A x \le 0$ для всех x;
- знакопеременная: $\exists x, y$: $x^\top A x > 0$, $y^\top A y < 0$.

B3 Критерий Сильвестра (расширенный)

Δ_k — **угловые** (ведущие) миноры: определители левых верхних $k \times k$ подматриц.

- $A \succ 0 \iff \Delta_1 > 0,\ \Delta_2 > 0,\ \ldots,\ \Delta_n > 0.$
- $A \prec 0 \iff (-1)^k \Delta_k > 0$, т.е. знаки чередуются: $\Delta_1 < 0,\ \Delta_2 > 0,\ \Delta_3 < 0, \ldots$
- $A \succeq 0 \iff$ **все главные** миноры (не только угловые!) ≥ 0.
- $A \preceq 0 \iff$ все главные миноры порядка k удовлетворяют $(-1)^k M_k \geq 0$.
- Иначе знакопеременная. Достаточный признак: какой-то угловой минор чётного порядка $\Delta_{2k} < 0$, или все $\Delta_k \neq 0$, но знаки не по одной из двух схем выше.

Главный минор — определитель подматрицы на строках и столбцах с одними и теми же номерами $i_1 < \cdots < i_k$.

ЛОВУШКА
$A = \mathrm{diag}(0, -1)$: $\Delta_1 = 0,\ \Delta_2 = 0$ — «все угловые ≥ 0», но A не неотрицательно определена (главный минор $-1 < 0$). Для полуопределённости смотрим все главные миноры.

B4 Необходимое условие 1-го порядка (без ограничений)

Если f дифференцируема в точке x^* и x^* — точка локального экстремума, то $\nabla f(x^*) = 0$ (теорема Ферма). Такие точки называются стационарными.

B5 Необходимые условия 2-го порядка (без ограничений)

Пусть f дважды дифференцируема в x^*.

- x^* — локальный минимум $\Rightarrow \nabla f(x^*) = 0$ и $\nabla^2 f(x^*) \succeq 0$.
- x^* — локальный максимум $\Rightarrow \nabla f(x^*) = 0$ и $\nabla^2 f(x^*) \preceq 0$.

B6 Достаточные условия (без ограничений)

Пусть f дважды дифференцируема в x^* и $\nabla f(x^*) = 0$.

- $\nabla^2 f(x^*) \succ 0 \Rightarrow x^*$ — точка строгого локального минимума;
- $\nabla^2 f(x^*) \prec 0 \Rightarrow$ строгий локальный максимум;
- $\nabla^2 f(x^*)$ знакопеременная $\Rightarrow$ экстремума нет (седло);
- полуопределённая и вырожденная — нужны члены высших порядков.

B7 Выпуклое множество

$X \subseteq \mathbb{R}^n$ выпукло, если $\forall x, y \in X,\ \forall \lambda \in [0,1]$: $\lambda x + (1-\lambda) y \in X$ (отрезок между любыми двумя точками лежит в X).

B8 Выпуклая функция (неравенство Йенсена)

f задана на выпуклом X. f выпукла, если $\forall x, y \in X,\ \forall \lambda \in [0,1]$:

$$f(\lambda x + (1-\lambda) y) \le \lambda f(x) + (1-\lambda) f(y).$$

Строго выпукла, если при $x \neq y$, $\lambda \in (0,1)$ неравенство строгое. Вогнутая: $-f$ выпукла. Общий вид неравенства Йенсена: $f\left(\sum \lambda_i x_i\right) \le \sum \lambda_i f(x_i)$ при $\lambda_i \ge 0$, $\sum \lambda_i = 1$.

B9 Сумма выпуклых функций

Если $f_1, \dots, f_m$ выпуклы на выпуклом X и $\alpha_i \ge 0$, то $\sum \alpha_i f_i$ выпукла на X. (Складываем неравенства Йенсена, домноженные на $\alpha_i \ge 0$.) Если хотя бы одна f_i строго выпукла с $\alpha_i > 0$ — сумма строго выпукла.

B10 Критерий выпуклости дважды дифференцируемой функции

$f \in C^2(X)$, X — открытое выпуклое множество. Тогда f выпукла на $X \iff \nabla^2 f(x) \succeq 0$ для всех $x \in X$. Если $\nabla^2 f(x) \succ 0$ для всех x, то f строго выпукла (обратное неверно: x^4 строго выпукла, но $f''(0) = 0$).

B11 Выпуклая задача

$f(x) \to \min$, $x \in X$, где X — выпуклое множество и f — выпуклая на X функция. (Для максимума: f вогнутая.) Типичный вид: $X = \{x :\ g_i(x) \le 0,\ g_i \text{ выпуклы};\ \langle a_j, x \rangle = b_j\}$.

B12 Локальный и глобальный минимум в выпуклой задаче

Теорема. В выпуклой задаче любой локальный минимум является глобальным. Множество точек минимума выпукло. Если f строго выпукла, точка минимума единственна (если существует).

B13 Задача с равенствами, функция Лагранжа

$f(x) \to \text{extr}$, $g_i(x) = 0$, $i = 1, \dots, m$, $x \in \mathbb{R}^n$ (обычно $m < n$).

$$\mathcal{L}(x, \lambda_0, \lambda) = \lambda_0 f(x) + \sum_{i=1}^{m} \lambda_i g_i(x).$$

В регулярном случае $\lambda_0 = 1$: $\mathcal{L} = f + \sum \lambda_i g_i$.

B14 Необходимые условия 1-го порядка с равенствами (правило множителей Лагранжа)

f, g_i непрерывно дифференцируемы в окрестности x^*, x^* — точка локального экстремума. Тогда $\exists (\lambda_0, \lambda_1, \dots, \lambda_m) \neq 0$:

$$\nabla_x \mathcal{L}(x^*, \lambda_0, \lambda) = \lambda_0 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) = 0, \qquad g_i(x^*) = 0.$$

Если x^* регулярна (градиенты $\nabla g_i(x^*)$ линейно независимы), то $\lambda_0 \neq 0$ и можно взять $\lambda_0 = 1$. Знак λ_i не ограничен.

B15 Достаточные условия с равенствами

$f, g_i \in C^2$, x^* допустима, $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$ при $\lambda_0 = 1$. Пусть

$$d^\top \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*)\, d > 0 \quad \forall d \neq 0 : \ \langle \nabla g_i(x^*), d \rangle = 0, \ i = 1..m.$$

Тогда x^* — строгий локальный минимум. Если < 0 на том же подпространстве — строгий локальный максимум. (Гессиан функции Лагранжа проверяется только на касательном подпространстве, а не на всём $\mathbb{R}^n$.)

B16 Условие регулярности (равенства)

Точка x^* регулярна, если градиенты ограничений $\nabla g_1(x^*), \dots, \nabla g_m(x^*)$ линейно независимы, т.е. ранг матрицы Якоби $\left(\partial g_i / \partial x_j\right)(x^*)$ равен m.

B17 Задача с равенствами и неравенствами, функция Лагранжа

$f(x) \to \min$, $g_i(x) \le 0$, $i = 1..m$; $h_j(x) = 0$, $j = 1..k$.

$$\mathcal{L}(x, \lambda_0, \lambda, \mu) = \lambda_0 f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{k} \mu_j h_j(x).$$

Неравенства всегда сначала приводим к виду $g_i \le 0$ (для задачи на минимум).

B18 Необходимые условия 1-го порядка (теорема Куна–Таккера / Джона)

$f, g_i, h_j \in C^1$, x^* — локальный минимум. Тогда $\exists \lambda_0, \lambda, \mu$, не все нули:

1. стационарность: $\lambda_0 \nabla f(x^*) + \sum \lambda_i \nabla g_i(x^*) + \sum \mu_j \nabla h_j(x^*) = 0$;
2. неотрицательность: $\lambda_0 \ge 0$, $\lambda_i \ge 0$ (μ_j любого знака);
3. дополняющая нежёсткость: $\lambda_i g_i(x^*) = 0$ для всех i (неактивное ограничение получает $\lambda_i = 0$);
4. допустимость: $g_i(x^*) \le 0$, $h_j(x^*) = 0$.

Если точка регулярна, то $\lambda_0 = 1$. **Для максимума** (при $\lambda_0 = 1$ и тех же $g_i \le 0$) знаки меняются: $\lambda_i \le 0$. Проще всего: заменить f на $-f$.

B19 Множество обобщённых допустимых направлений

$I(x^0) = \{i : \ g_i(x^0) = 0\}$ — активные ограничения-неравенства. Конус (линеаризованных) допустимых направлений:

$$K(x^0) = \Big\{ d \in \mathbb{R}^n : \ \langle \nabla h_j(x^0), d \rangle = 0 \ \ \forall j; \ \ \langle \nabla g_i(x^0), d \rangle \le 0 \ \ \forall i \in I(x^0) \Big\}.$$

Неактивные неравенства не накладывают условий. Для условий 2-го порядка берут подмножество (критический конус): дополнительно $\langle \nabla g_i(x^0), d \rangle = 0$ для активных i с $\lambda_i > 0$.

B20 Направление убывания

Вектор $h \ne 0$ задаёт направление убывания f в точке x^0, если $\exists \alpha_0 > 0$: $f(x^0 + \alpha h) < f(x^0)$ для всех $\alpha \in (0, \alpha_0)$.

B21 Достаточное условие направления убывания

f дифференцируема в x^0 и $\langle \nabla f(x^0), h \rangle < 0 \Rightarrow h$ — направление убывания. В частности, $h = -\nabla f(x^0)$ (если $\nabla f(x^0) \ne 0$) — направление наискорейшего убывания.

B22 Шаг градиентного спуска с постоянным шагом

$$x^{k+1} = x^k - \alpha\,\nabla f(x^k), \qquad \alpha > 0 \text{ фиксировано.}$$

B23 Шаг наискорейшего градиентного спуска

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k), \qquad \alpha_k = \arg\min_{\alpha \geq 0} f\big(x^k - \alpha \nabla f(x^k)\big).$$

Для квадратичной $f = \frac{1}{2}x^\top A x - b^\top x$: $\alpha_k = \dfrac{\langle g_k, g_k\rangle}{\langle Ag_k, g_k\rangle}$, $g_k = \nabla f(x^k)$.

B24 Пространства C, C^1, C_0^1 и их нормы

- $C[t_0, t_1]$ — непрерывные на отрезке функции, $\|x\|_C = \max_{t\in[t_0,t_1]} |x(t)|$.
- $C^1[t_0, t_1]$ — непрерывно дифференцируемые функции, $\|x\|_{C^1} = \max\left(\max|x(t)|,\ \max|\dot{x}(t)|\right)$ (эквивалентный вариант: $\max|x| + \max|\dot{x}|$).
- $C_0^1[t_0, t_1] = \{h \in C^1[t_0, t_1] :\ h(t_0) = h(t_1) = 0\}$.

B25 Вариация функционала по направлению

$J : X \to \mathbb{R}$, X — линейное нормированное пространство. Вариация (производная) по направлению h в точке x:

$$\delta^+ J(x, h) = \lim_{\alpha \to +0} \frac{J(x + \alpha h) - J(x)}{\alpha} \quad \text{(односторонний предел)}.$$

B26 Вариация по Лагранжу (первая вариация)

Если для каждого $h \in X$ существует двусторонний предел

$$\delta J(x, h) = \lim_{\alpha \to 0} \frac{J(x + \alpha h) - J(x)}{\alpha} = \frac{d}{d\alpha} J(x + \alpha h)\Big|_{\alpha = 0},$$

то $\delta J(x, h)$ называется первой вариацией (вариацией по Лагранжу) J в точке x. Линейность по h не требуется (если она линейна и непрерывна — это производная Гато).

B27 Простейшая задача КВИ и уравнение Эйлера–Лагранжа

$$J(x) = \int_{t_0}^{t_1} L\big(t, x(t), \dot{x}(t)\big)\, dt \to \text{extr}, \qquad x \in C^1[t_0, t_1],\ x(t_0) = x_0,\ x(t_1) = x_1.$$

Уравнение Эйлера–Лагранжа:

$$-\frac{d}{dt} L_{\dot{x}}\big(t, x, \dot{x}\big) + L_x\big(t, x, \dot{x}\big) = 0.$$

Его решения называются экстремалями.

B28 Задача Больца, уравнение Э–Л и трансверсальность

$$B(x) = \int_{t_0}^{t_1} L(t, x, \dot{x})\, dt + l\big(x(t_0), x(t_1)\big) \to \text{extr}, \qquad x \in C^1[t_0, t_1]$$

(концы свободны).

Уравнение Э–Л: $-\frac{d}{dt} L_{\dot{x}} + L_x = 0$. Условия трансверсальности:

$$L_{\dot{x}}\big|_{t=t_0} = l_{x_0}, \qquad L_{\dot{x}}\big|_{t=t_1} = -l_{x_1},$$

где $l_{x_0} = \partial l / \partial x(t_0)$, $l_{x_1} = \partial l / \partial x(t_1)$. Запоминалка: «на правом конце минус».

B29 Допустимые направления в простейшей задаче и в задаче Больца

Простейшая задача: $h \in C_0^1[t_0, t_1]$, т.е. $h(t_0) = h(t_1) = 0$ — тогда $\hat{x} + \alpha h$ удовлетворяет граничным условиям при любом α. Задача Больца: любые $h \in C^1[t_0, t_1]$ (ограничений на концы нет).

B30 Условие Лежандра и усиленное условие Лежандра

Для экстремали $\hat{x}$: $P(t) = L_{\dot{x}\dot{x}}\big(t, \hat{x}(t), \dot{\hat{x}}(t)\big)$.

- Условие Лежандра (необходимое для минимума): $P(t) \geq 0$ при всех $t \in [t_0, t_1]$; для максимума $P(t) \leq 0$.
- Усиленное условие Лежандра: $P(t) > 0$ на $[t_0, t_1]$ (для максимума $P(t) < 0$).

Основная часть. О1–О8: задачи без ограничений, матрицы

О1 Постановка задачи оптимизации

Задача оптимизации: найти $f(x) \to \min$ (или $\max$) при $x \in X$.

- $f: X \to \mathbb{R}$ — **целевая функция**;
- $X \subseteq \mathbb{R}^n$ — **допустимое множество**, его элементы — **допустимые решения** (планы). X обычно задаётся ограничениями: $g_i(x) \le 0$, $h_j(x) = 0$. Если $X = \mathbb{R}^n$ — задача без ограничений (безусловная);
- **решение** — точка глобального минимума x^*; значение задачи $f^* = \inf_X f$.

Определения глобального/локального (строгого) минимума — как в В1. Связь: $\max_X f = -\min_X(-f)$, точки те же. Глобальный минимум всегда локальный, обратное неверно. Возможные исходы: решение есть; f^* конечно, но не достигается (e^x на $\mathbb{R}$); $f^* = -\infty$; $X = \varnothing$.

О2 Существование решения. Теорема Вейерштрасса и следствие

Теорема Вейерштрасса. Если $X \subset \mathbb{R}^n$ — непустой компакт (замкнутое ограниченное множество), а f непрерывна на X, то f достигает на X своих наибольшего и наименьшего значений.

Следствие. Пусть X замкнуто, f непрерывна на X, и для некоторого c множество Лебега $M_c = \{x \in X : f(x) \le c\}$ непусто и ограничено. Тогда глобальный минимум f на X существует.

> ***Доказательство.***
> M_c замкнуто (прообраз замкнутого луча $(-\infty, c]$ при непрерывном отображении, пересечённый с замкнутым X) и ограничено, значит компакт. По Вейерштрассу на M_c есть точка минимума x^*. Для $x \in X \setminus M_c$: $f(x) > c \ge f(x^*)$. Значит x^* — глобальный минимум на всём X.

Для максимума то же с множеством $\{f \ge c\}$.

О3 Бесконечно растущая функция. Существование минимума

Определение. Непрерывная f на $X \subseteq \mathbb{R}^n$ — бесконечно растущая (коэрцитивная), если $f(x^k) \to +\infty$ для любой последовательности $x^k \in X$ с $\|x^k\| \to \infty$. Если X открыто (например, интервал $(-\pi, \pi)$), то ещё и для любой последовательности, сходящейся к границе X.

Теорема. Непрерывная бесконечно растущая на X функция имеет на X глобальный минимум.

> ***Доказательство.***
> Возьмём любую $x^0 \in X$, $c = f(x^0)$, $M = \{x \in X : \ f(x) \le c\} \ni x^0$. **Ограничено:** иначе есть $x^k \in M$, $\|x^k\| \to \infty$, тогда $f(x^k) \to +\infty$, но $f(x^k) \le c$ — противоречие. **Замкнуто:** если $x^k \in M$, $x^k \to \bar{x}$, то $\bar{x} \in X$ (иначе $\bar{x}$ на границе и $f(x^k) \to +\infty$), и по непрерывности $f(\bar{x}) \le c$. Значит M — непустой компакт, дальше как в следствии теоремы Вейерштрасса.

ПРИМЕР

$f = y_1^4 + y_2^4 - (y_1 + y_2)^2$ бесконечно растущая на $\mathbb{R}^2$: $(y_1 + y_2)^2 \le 2(y_1^2 + y_2^2) = 2\|y\|^2$, а $y_1^4 + y_2^4 \ge \frac{1}{2}\|y\|^4$, поэтому $f \ge \frac{1}{2}\|y\|^4 - 2\|y\|^2 \to \infty$. Значит минимум есть, и его ищем среди стационарных точек: $(1, 1)$, $(-1, -1)$ с $f = -2$.

$f = \operatorname{tg}\frac{x}{2}$ на $(-\pi, \pi)$ *не* бесконечно растущая: при $x \to -\pi + 0$ она уходит в $-\infty$, а не в $+\infty$ (у неё и минимума нет). А вот $\operatorname{tg}^2\frac{x}{2}$ на том же интервале бесконечно растущая.

О4 Условия оптимальности в одномерной задаче док-во

$f : \mathbb{R} \to \mathbb{R}$, x^* — внутренняя точка.

1) Необходимое 1-го порядка (Ферма). f дифференцируема в x^*, x^* — локальный экстремум $\Rightarrow f'(x^*) = 0$.

> ***Доказательство.***
> Пусть x^* — локальный минимум: $f(x^* + h) - f(x^*) \geq 0$ при малых $|h|$. При $h > 0$: $\frac{f(x^*+h)-f(x^*)}{h} \geq 0$, в пределе $f'(x^*) \geq 0$. При $h < 0$ дробь ≤ 0, в пределе $f'(x^*) \leq 0$. Итого $f'(x^*) = 0$.

2) Необходимое 2-го порядка. f дважды дифференцируема, x^* — локальный минимум $\Rightarrow f'(x^*) = 0$, $f''(x^*) \geq 0$ (для максимума ≤ 0).

> ***Доказательство.***
> По Тейлору (форма Пеано) с $f'(x^*) = 0$: $0 \leq f(x^* + h) - f(x^*) = \frac{1}{2} f''(x^*) h^2 + o(h^2)$. Делим на h^2: $\frac{1}{2} f''(x^*) + o(1) \geq 0$, при $h \to 0$ получаем $f''(x^*) \geq 0$.

3) Достаточное. $f'(x^*) = 0$, $f''(x^*) > 0 \Rightarrow$ строгий локальный минимум ($f'' < 0$ — максимум).

> ***Доказательство.***
> $f(x^* + h) - f(x^*) = h^2 \left(\frac{1}{2} f''(x^*) + o(1) \right)$. При малых $h \neq 0$ скобка $> \frac{1}{4} f''(x^*) > 0$, значит разность > 0.

4) Общий случай. Пусть $f'(x^*) = \cdots = f^{(m-1)}(x^*) = 0$, $f^{(m)}(x^*) \neq 0$. Если m чётное: $f^{(m)} > 0$ — строгий минимум, < 0 — строгий максимум. Если m нечётное — экстремума нет. (Доказательство то же: $f(x^* + h) - f(x^*) = \frac{f^{(m)}(x^*)}{m!} h^m + o(h^m)$; при нечётном m знак h^m меняется.)

ПРИМЕР
$f = x^2 \sin x$: $f(x) = x^3 - \ldots$, первая ненулевая производная $f'''(0) = 6$, $m = 3$ нечётное — в нуле экстремума нет.
$f = x^4 \cos x = x^4 - \ldots$: $f^{(4)}(0) = 24 > 0$, $m = 4$ — строгий минимум.

О5 Знакоопределённость симметрической матрицы

Определения — В2. Дополнительно стоит сказать:

- Любую квадратичную форму можно записать с симметрической матрицей: $x^\top B x = x^\top \frac{B + B^\top}{2} x$.
- $A \succ 0 \Rightarrow$ все диагональные элементы $a_{ii} = e_i^\top A e_i > 0$; $A \succ 0 \Rightarrow \det A > 0$, A обратима и $A^{-1} \succ 0$.
- $A \succ 0 \iff -A \prec 0$; $A \succeq 0$ и $\det A \neq 0 \iff A \succ 0$.
- Способы проверки: собственные числа (О6), критерий Сильвестра (О7), выделение полных квадратов (метод Лагранжа).

ПРИМЕР
$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$: $x^\top A x = 2x_1^2 + 2x_1 x_2 + 2x_2^2 = (x_1 + x_2)^2 + x_1^2 + x_2^2 > 0$ при $x \neq 0$ — положительно определена. $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$: на $x = (1, -1)$ форма равна -2, на $(1, 1)$ равна 6 — знакопеременная.

О6 Собственные числа и знакоопределённость

У симметрической A все собственные числа $\lambda_1, \dots, \lambda_n$ вещественны и есть ортонормированный базис из собственных векторов: $A = Q\Lambda Q^\top$, $Q^\top Q = I$.

Теорема. $A \succ 0 \iff$ все $\lambda_i > 0$; $A \succeq 0 \iff$ все $\lambda_i \geq 0$; $A \prec 0 \iff$ все $\lambda_i < 0$; $A \preceq 0 \iff$ все $\lambda_i \leq 0$; знакопеременная $\iff$ есть собственные числа разных знаков.

> ***Доказательство.***
> Замена $y = Q^\top x$ (взаимно однозначная, $x = 0 \iff y = 0$): $x^\top A x = y^\top \Lambda y = \sum \lambda_i y_i^2$. Если все $\lambda_i > 0$, сумма положительна при $y \neq 0$. Обратно, подставив собственный вектор $x = q_i$: $q_i^\top A q_i = \lambda_i \|q_i\|^2 = \lambda_i$, значит знак формы на q_i — это знак λ_i.

Следствие (отношение Рэлея): $\lambda_{\min}\|x\|^2 \leq x^\top A x \leq \lambda_{\max}\|x\|^2$. Это используется в доказательстве достаточных условий (О8).

О7 Критерий Сильвестра (расширенный)

Формулировка — В3. Как применять:

1. Считаем угловые миноры $\Delta_1, \dots, \Delta_n$.
2. Все > 0 — $A \succ 0$. Знаки $-, +, -, \dots$ — $A \prec 0$.
3. Все $\Delta_k \neq 0$, но схема другая — знакопеременная. Если $\Delta_{2k} < 0$ для какого-то чётного порядка — знакопеременная сразу.
4. Есть нулевые Δ_k — переходим ко всем главным минорам: все ≥ 0 — $\succeq 0$; все $(-1)^k M_k \geq 0$ — $\preceq 0$; иначе знакопеременная.

Связь $A \prec 0 \iff -A \succ 0$: у $-A$ миноры порядка k равны $(-1)^k \Delta_k$, отсюда чередование знаков.

ПРИМЕР

Листок 1, №3: $f = -x_1^2 + 2x_1x_2 - 3x_2^2 - 4x_3^2$, точка 0. $\nabla^2 f = \begin{pmatrix} -2 & 2 & 0 \\ 2 & -6 & 0 \\ 0 & 0 & -8 \end{pmatrix}$, $\Delta_1 = -2 < 0$, $\Delta_2 = 12 - 4 = 8 > 0$, $\Delta_3 = -8 \cdot 8 = -64 < 0$. Знаки чередуются — $\prec 0$, строгий максимум.
Листок 3, №5: $f = -x_1^2 + 2x_1x_2 - x_2^2 - 4x_3^2$: $\Delta_1 = -2$, $\Delta_2 = 4 - 4 = 0$ — Сильвестр в строгой форме не работает. Главные миноры 1-го порядка $-2, -2, -8 \leq 0$, 2-го $0, 16, 16 \geq 0$, 3-го $0 \leq 0$ — матрица $\preceq 0$. А сама функция $f = -(x_1 - x_2)^2 - 4x_3^2 \leq 0 = f(t, t, 0)$ — нестрогий глобальный максимум на прямой $x_1 = x_2$, $x_3 = 0$.

О8 n-мерная задача без ограничений: условия 1-го и 2-го порядка, достаточные условия

Формулировки: В4, В5, В6. Доказательства сводятся к одномерному случаю.

> ***Доказательство.***
> **Необходимые.** Пусть x^* — локальный минимум. Для любого $h \in \mathbb{R}^n$ функция $\varphi(t) = f(x^* + th)$ имеет локальный минимум в $t = 0$. $\varphi'(0) = \langle \nabla f(x^*), h \rangle = 0$ для всех h, значит $\nabla f(x^*) = 0$. $\varphi''(0) = h^\top \nabla^2 f(x^*) h \geq 0$ для всех h, значит $\nabla^2 f(x^*) \succeq 0$.
> **Достаточные.** Тейлор: $f(x^* + h) - f(x^*) = \frac{1}{2} h^\top \nabla^2 f(x^*) h + o(\|h\|^2) \geq \frac{1}{2}\lambda_{\min}\|h\|^2 + o(\|h\|^2) = \|h\|^2\left(\frac{\lambda_{\min}}{2} + o(1)\right) > 0$ при малых $h \neq 0$, т.к. $\lambda_{\min} > 0$ при $\nabla^2 f \succ 0$.
> **Седло.** Если $\nabla^2 f$ знакопеременна: есть h_1 с $h_1^\top \nabla^2 f h_1 > 0$ и h_2 с < 0. Вдоль h_1 функция растёт, вдоль h_2 убывает — экстремума нет.

Алгоритм: 1) решить $\nabla f = 0$; 2) в каждой стационарной точке посчитать $\nabla^2 f$ и проверить по Сильвестру; 3) в вырожденных случаях исследовать f напрямую (подставить кривые, выделить квадраты).

ПРИМЕР

Листок 2, №1: $f = 3xy - x^2 y - xy^2$. $f_x = y(3 - 2x - y) = 0$, $f_y = x(3 - x - 2y) = 0$. Стационарные: $(0,0), (3,0), (0,3), (1,1)$. $\nabla^2 f = \begin{pmatrix} -2y & 3 - 2x - 2y \\ 3 - 2x - 2y & -2x \end{pmatrix}$. В $(1,1)$: $\begin{pmatrix} -2 & -1 \\ -1 & -2 \end{pmatrix}$, $\Delta_1 = -2$, $\Delta_2 = 3 -$ максимум, $f = 1$. В $(0,0)$: $\begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}$, $\Delta_2 = -9 < 0$ — седло. В $(3,0)$ и $(0,3)$ тоже $\Delta_2 = -9$ — сёдла.

О9–О15: выпуклость

О9 Выпуклые множества, пересечение, примеры, полиэдры

док-во

Определение — В7.

Теорема. Пересечение любого (даже бесконечного) семейства выпуклых множеств выпукло.

> ***Доказательство.***
> $X = \bigcap_\alpha X_\alpha$. Возьмём $x, y \in X$, $\lambda \in [0,1]$. Для каждого α: $x, y \in X_\alpha$, X_α выпукло, значит $\lambda x + (1-\lambda)y \in X_\alpha$. Это верно для всех α, значит точка лежит в пересечении.

Примеры: $\varnothing$, точка, $\mathbb{R}^n$; прямая, отрезок, луч; гиперплоскость $\{x : \langle a, x\rangle = b\}$; полупространство $\{x : \langle a, x\rangle \le b\}$; шар $\{x : \|x - a\| \le r\}$; аффинное подпространство $\{x : Ax = b\}$; неотрицательный ортант $\mathbb{R}^n_+$; множество неотрицательно определённых матриц.

> ***Доказательство.***
> Шар: $\|\lambda x + (1-\lambda)y - a\| = \|\lambda(x-a) + (1-\lambda)(y-a)\| \le \lambda\|x-a\| + (1-\lambda)\|y-a\| \le r$.
> Полупространство: $\langle a, \lambda x + (1-\lambda)y\rangle = \lambda\langle a, x\rangle + (1-\lambda)\langle a, y\rangle \le \lambda b + (1-\lambda)b = b$.

Полиэдр — множество решений конечной системы линейных неравенств $P = \{x : \ Ax \le b\} = \bigcap_{i=1}^m \{x : \langle a_i, x\rangle \le b_i\}$. Это пересечение конечного числа полупространств, поэтому выпукло. Ограниченный полиэдр — многогранник (выпуклая оболочка своих вершин). Пример — допустимое множество задачи линейного программирования.

Не выпуклы: окружность (сфера), объединение двух непересекающихся отрезков, $\{x : x_1 x_2 \ge 1\}$ целиком (две ветви).

О10 Выпуклая функция. Определение через надграфик

Определение — В8.

Надграфик (эпиграф): $\operatorname{epi} f = \{(x, t) \in X \times \mathbb{R} : \ t \ge f(x)\}$.

Теорема. f выпукла на выпуклом $X \iff \operatorname{epi} f$ — выпуклое множество в $\mathbb{R}^{n+1}$.

> ***Доказательство.***
> $(\Rightarrow)$ $(x,s), (y,t) \in \operatorname{epi} f$: $s \ge f(x)$, $t \ge f(y)$. Тогда $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) \le \lambda s + (1-\lambda)t$, т.е. $\lambda(x,s) + (1-\lambda)(y,t) \in \operatorname{epi} f$.
> $(\Leftarrow)$ $(x, f(x))$ и $(y, f(y))$ лежат в $\operatorname{epi} f$, по выпуклости там и $\big(\lambda x + (1-\lambda)y,\ \lambda f(x) + (1-\lambda)f(y)\big)$, т.е. $\lambda f(x) + (1-\lambda)f(y) \ge f(\lambda x + (1-\lambda)y)$.

Геометрически: хорда графика лежит не ниже графика. Ещё полезно: множества Лебега $\{x : f(x) \le c\}$ выпуклой функции выпуклы (обратное неверно).

О11 Сумма, максимум, композиция выпуклых функций

док-во

1. Сумма. f_1, f_2 выпуклы, $\alpha_1, \alpha_2 \geq 0 \Rightarrow \alpha_1 f_1 + \alpha_2 f_2$ выпукла.

> ***Доказательство.***
> $\alpha_i f_i(\lambda x + (1-\lambda)y) \leq \lambda \alpha_i f_i(x) + (1-\lambda)\alpha_i f_i(y)$ (умножили неравенство Йенсена на $\alpha_i \geq 0$). Складываем по i.

2. Максимум. f_i выпуклы ($i \in I$, можно бесконечное семейство, если супремум конечен) $\Rightarrow f(x) = \max_i f_i(x)$ выпукла.

> ***Доказательство.***
> Для каждого i: $f_i(\lambda x + (1-\lambda)y) \leq \lambda f_i(x) + (1-\lambda) f_i(y) \leq \lambda f(x) + (1-\lambda) f(y)$. Правая часть не зависит от i, берём максимум по i слева: $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$. (Иначе: $\operatorname{epi} \max f_i = \bigcap \operatorname{epi} f_i$ — пересечение выпуклых.)

3. Композиция. $g : X \to \mathbb{R}$ выпукла, $\varphi : \mathbb{R} \to \mathbb{R}$ выпукла и **неубывающая** $\Rightarrow \varphi(g(x))$ выпукла.

> ***Доказательство.***
> $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda) g(y)$. Так как φ не убывает: $\varphi\big(g(\lambda x + (1-\lambda)y)\big) \leq \varphi\big(\lambda g(x) + (1-\lambda) g(y)\big)$, а по выпуклости φ это $\leq \lambda \varphi(g(x)) + (1-\lambda)\varphi(g(y))$.

Аффинная замена: f выпукла $\Rightarrow f(Ax + b)$ выпукла (подставляем $A(\lambda x + (1-\lambda)y) + b = \lambda(Ax+b) + (1-\lambda)(Ay+b)$).

ЛОВУШКА

Без монотонности композиция может быть невыпуклой: $\varphi(u) = u^2$ (не монотонна) и $g(x) = x^2 - 1$ дают $(x^2-1)^2$ — не выпукла. А $e^{g(x)}$ при выпуклой g выпукла (e^u выпукла и возрастает).

О12 Критерий выпуклости дифференцируемой функции

f дифференцируема на открытом выпуклом X. Тогда

$$f \text{ выпукла} \iff f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \forall x, y \in X.$$

Геометрически: график лежит не ниже любой касательной гиперплоскости. Строгая выпуклость $\iff$ неравенство строгое при $y \neq x$. Эквивалентно: монотонность градиента $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0$.

> ***Доказательство.***
> (Если спросят.) $(\Rightarrow)$ $f(x + \lambda(y - x)) \leq f(x) + \lambda(f(y) - f(x))$, отсюда $\frac{f(x+\lambda(y-x))-f(x)}{\lambda} \leq f(y) - f(x)$; при $\lambda \to +0$ слева $\langle \nabla f(x), y - x \rangle$.
> $(\Leftarrow)$ $z = \lambda x + (1 - \lambda)y$. Записываем неравенство в точке z для x и для y, умножаем на λ и $1 - \lambda$, складываем: $\lambda f(x) + (1 - \lambda) f(y) \geq f(z) + \langle \nabla f(z), \lambda x + (1 - \lambda)y - z \rangle = f(z)$.

ПРИМЕР
Листок 3, №8 (дивергенция Кульбака–Лейблера): $f(p) = \sum p_i \ln p_i$ строго выпукла на $\mathbb{R}^n_{++}$ ($f'' = 1/p_i > 0$), и $D(p, q) = f(p) - f(q) - \langle \nabla f(q), p - q \rangle \geq 0$ ровно по этому критерию, с равенством только при $p = q$.

О13 Критерий выпуклости дважды дифференцируемой функции

Формулировка — В10. Для $n = 1$: $f''(x) \geq 0$ на интервале.

> ***Доказательство.***
> (Идея.) Выпуклость f на X равносильна выпуклости всех её сужений на прямые $\varphi(t) = f(x + th)$. $\varphi''(t) = h^\top \nabla^2 f(x + th) h$. Одномерный критерий: φ выпукла $\iff$ $\varphi'' \geq 0$ (производная φ' не убывает). Для всех x, h это ровно $\nabla^2 f \succeq 0$.

ПРИМЕР
Листок 3, №4: $f = ae^{2x} + be^x + c$, $a, b > 0$: $f'' = 4ae^{2x} + be^x > 0$ — строго выпукла.
№6: $f = \frac{1}{x_1 x_2}$ на $\mathbb{R}^2_{++}$: $\nabla^2 f = \frac{1}{x_1 x_2} \begin{pmatrix} 2/x_1^2 & 1/(x_1 x_2) \\ 1/(x_1 x_2) & 2/x_2^2 \end{pmatrix}$, $\Delta_1 > 0$, $\Delta_2 \propto \frac{4-1}{x_1^2 x_2^2} > 0$ — выпукла.
№9: $f = x_1 x_2$: $\nabla^2 f = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\Delta_2 = -1$ — знакопеременна, не выпукла.

О14 Выпуклая задача. Локальный минимум = глобальный

док-во

Определение — В11. **Теорема:** в выпуклой задаче любая точка локального минимума — точка глобального минимума.

> ***Доказательство.***
> Пусть x^* — локальный минимум: $f(x^*) \le f(x)$ для $x \in X \cap U_\varepsilon(x^*)$. Предположим, что есть $y \in X$ с $f(y) < f(x^*)$. Рассмотрим $z_\lambda = \lambda y + (1-\lambda)x^*$, $\lambda \in (0,1]$. По выпуклости X: $z_\lambda \in X$. По выпуклости f:
>
> $$f(z_\lambda) \le \lambda f(y) + (1-\lambda) f(x^*) < \lambda f(x^*) + (1-\lambda) f(x^*) = f(x^*).$$
>
> При $\lambda < \varepsilon / \|y - x^*\|$ имеем $\|z_\lambda - x^*\| = \lambda \|y - x^*\| < \varepsilon$, т.е. $z_\lambda \in X \cap U_\varepsilon(x^*)$ и $f(z_\lambda) < f(x^*)$ — противоречие с локальной минимальностью.

Дополнения. (1) Множество точек минимума X^* выпукло: если $x, y \in X^*$, то $f(\lambda x + (1-\lambda)y) \le \lambda f^* + (1-\lambda) f^* = f^*$. (2) Если f строго выпукла, минимум единственный: если бы $x \ne y$ оба минимумы, то $f(\frac{x+y}{2}) < f^*$.

О15 Критерий оптимальности в выпуклой задаче (дифференцируемая f)

док-во

X выпукло, f выпукла и дифференцируема (на открытом множестве, содержащем X). Тогда

$$x^* \in X \text{ — точка минимума } \iff \langle \nabla f(x^*), x - x^* \rangle \ge 0 \quad \forall x \in X.$$

Если x^* — внутренняя точка X (в частности, $X = \mathbb{R}^n$), условие равносильно $\nabla f(x^*) = 0$. То есть в выпуклой задаче без ограничений стационарная точка — сразу глобальный минимум.

> ***Доказательство.***
> ($\Leftarrow$) По критерию О12: $f(x) \ge f(x^*) + \langle \nabla f(x^*), x - x^* \rangle \ge f(x^*)$ для всех $x \in X$. (Здесь нужна выпуклость f.)
> ($\Rightarrow$) Пусть для некоторого $x \in X$: $\langle \nabla f(x^*), x - x^* \rangle < 0$. Положим $\varphi(\lambda) = f(x^* + \lambda(x - x^*))$, $\lambda \in [0,1]$; точки лежат в X по выпуклости. $\varphi'(0) = \langle \nabla f(x^*), x - x^* \rangle < 0$, значит при малых $\lambda > 0$ $\varphi(\lambda) < \varphi(0) = f(x^*)$ — противоречие. (Здесь выпуклость f не нужна: это необходимое условие для любой гладкой f на выпуклом X.)
> Внутренняя точка: можно взять $x = x^* \pm \varepsilon e_i$, получаем $\pm \partial_i f(x^*) \ge 0$, т.е. $\nabla f(x^*) = 0$.

Геометрически: антиградиент в точке минимума «смотрит наружу» от X, ни одно допустимое направление не является направлением убывания.

О16–О18: ограничения-равенства

О16 Задача с равенствами: необходимые условия 1-го порядка, регулярность

док-во

Постановка и $\mathcal{L}$ — В13. Теорема (правило множителей Лагранжа) — В14.

Регулярность: $\nabla g_1(x^*), \dots, \nabla g_m(x^*)$ линейно независимы (В16).

Утверждение. Если x^* регулярна, то в правиле множителей $\lambda_0 \neq 0$ (и можно считать $\lambda_0 = 1$), а множители λ_i определены однозначно.

> ***Доказательство.***
> Пусть $\lambda_0 = 0$. Тогда $\sum_{i=1}^m \lambda_i \nabla g_i(x^*) = 0$, причём не все λ_i равны нулю (весь набор $(\lambda_0, \lambda) \neq 0$). Это нетривиальная линейная комбинация градиентов ограничений, равная нулю, — противоречие с линейной независимостью. Значит $\lambda_0 \neq 0$, делим всё на λ_0.
> Единственность: если $\nabla f + \sum \lambda_i \nabla g_i = 0$ и $\nabla f + \sum \lambda'_i \nabla g_i = 0$, то $\sum (\lambda_i - \lambda'_i) \nabla g_i = 0$, откуда $\lambda = \lambda'$.

> ***Доказательство.***
> (Схема доказательства самой теоремы в регулярном случае — если попросят.) Матрица Якоби $\partial g / \partial x$ ($m \times n$) имеет ранг m, перенумеруем переменные: $x = (y, z)$, $y \in \mathbb{R}^m$, $\det g_y(x^*) \neq 0$. По теореме о неявной функции вблизи x^* множество $\{g = 0\}$ задаётся как $y = \varphi(z)$, $\varphi_z = -g_y^{-1} g_z$. Тогда z^* — локальный экстремум функции без ограничений $F(z) = f(\varphi(z), z)$, и $0 = F_z = f_z + f_y \varphi_z = f_z - f_y g_y^{-1} g_z$. Положим $\lambda = -f_y g_y^{-1}$ (строка). Тогда $f_y + \lambda g_y = 0$ и $f_z + \lambda g_z = 0$, т.е. $\nabla f + \sum \lambda_i \nabla g_i = 0$.

Нерегулярный случай: пример $f = x \to \min$, $g = x^2 + y^2 = 0$ (или $y^2 - x^3 = 0$): $x^* = (0, 0)$, $\nabla g = 0$ — нужно брать $\lambda_0 = 0$, с $\lambda_0 = 1$ система несовместна.

ПРИМЕР

Листок 5, №1: $xy \to \text{extr}$, $2x^2 + y^2 = 1$. $\mathcal{L} = xy + \lambda(2x^2 + y^2 - 1)$: $y + 4\lambda x = 0$, $x + 2\lambda y = 0$. Отсюда $y^2 = 2x^2$, с ограничением $x^2 = \frac{1}{4}$, $y^2 = \frac{1}{2}$. Точки $(\pm\frac{1}{2}, \pm\frac{1}{\sqrt{2}})$: при одинаковых знаках $xy = \frac{1}{2\sqrt{2}}$ — глобальный максимум, при разных $-\frac{1}{2\sqrt{2}}$ — глобальный минимум (эллипс компакт, Вейерштрасс).

О17 Необходимые условия 2-го порядка с равенствами

$f, g_i \in C^2$, x^* — регулярная точка локального минимума, λ^* — её множители ($\lambda_0 = 1$). Тогда

$$d^\top \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*)\, d \geq 0 \quad \forall d:\ \langle \nabla g_i(x^*), d\rangle = 0,\ i = 1..m.$$

Для максимума: ≤ 0 на том же подпространстве $T(x^*) = \{d : J(x^*)d = 0\}$ (касательное подпространство к поверхности ограничений).

> ***Доказательство.***
> (Идея.) Для любого $d \in T(x^*)$ в регулярном случае есть гладкая кривая $x(t)$ на поверхности $g = 0$ с $x(0) = x^*$, $\dot{x}(0) = d$. На ней $g_i(x(t)) \equiv 0$, поэтому $f(x(t)) = \mathcal{L}(x(t), \lambda^*)$. Функция $\psi(t) = \mathcal{L}(x(t), \lambda^*)$ имеет минимум в 0; $\psi'(0) = \langle \nabla_x \mathcal{L}, d\rangle = 0$, $\psi''(0) = d^\top \nabla^2_{xx}\mathcal{L}\, d + \langle \nabla_x \mathcal{L}, \ddot{x}(0)\rangle = d^\top \nabla^2_{xx}\mathcal{L}\, d \geq 0$.

Почему гессиан именно $\mathcal{L}$, а не f: кривизна самой поверхности ограничений входит через $\lambda_i \nabla^2 g_i$.

О18 Достаточные условия строгого локального экстремума с равенствами

Формулировка — В15: x^* допустима, $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$, и $d^\top \nabla^2_{xx}\mathcal{L}\, d > 0$ (< 0) для всех $d \neq 0$ из $T(x^*)$ — строгий минимум (максимум). Регулярность здесь не требуется.

Как проверять на практике:

1. Выразить d из линейных условий $J(x^*)d = 0$ (у подпространства размерность $n - m$) и подставить в квадратичную форму — получится форма от $n - m$ переменных, её проверяем Сильвестром.
2. Если $\nabla^2_{xx}\mathcal{L} \succ 0$ на всём $\mathbb{R}^n$ — тем более минимум.
3. Окаймлённый гессиан $\bar{H} = \begin{pmatrix} 0 & J \\ J^\top & \nabla^2_{xx}\mathcal{L} \end{pmatrix}$: смотрим его последние $n - m$ угловых миноров $D_{2m+1}, \ldots, D_{m+n}$. Минимум: все имеют знак $(-1)^m$. Максимум: знаки чередуются, D_{m+k} имеет знак $(-1)^k$. Для $n = 2, m = 1$: $\det \bar{H} < 0$ — минимум, > 0 — максимум.

ПРИМЕР
Листок 5, №3: $x^2 - y^2 \to \text{extr}$, $x^2 + y^2 = 1$. $\mathcal{L} = x^2 - y^2 + \lambda(x^2 + y^2 - 1)$: $2x(1 + \lambda) = 0$, $2y(\lambda - 1) = 0$.
Точки $(0, \pm 1)$ с $\lambda = 1$ и $(\pm 1, 0)$ с $\lambda = -1$.
$(\pm 1, 0)$, $\lambda = -1$: $\nabla^2 \mathcal{L} = \text{diag}(0, -4)$; касательное направление $d = (0, 1)$ ($\nabla g = (\pm 2, 0) \perp d$), $d^\top \nabla^2 \mathcal{L} d = -4 < 0$ — максимум, $f = 1$.
$(0, \pm 1)$, $\lambda = 1$: $\nabla^2 \mathcal{L} = \text{diag}(4, 0)$, $d = (1, 0)$: $4 > 0$ — минимум, $f = -1$.

О19–О22: равенства и неравенства

О19 Необходимые условия 1-го порядка. Регулярность

Задача $f \to \min$, $g_i(x) \le 0$ ($i = 1..m$), $h_j(x) = 0$ ($j = 1..k$), $\mathcal{L} = \lambda_0 f + \sum \lambda_i g_i + \sum \mu_j h_j$.

Теорема (Джон; Каруш–Кун–Таккер). Если $f, g_i, h_j \in C^1$ в окрестности x^* и x^* — локальный минимум, то существуют $\lambda_0 \ge 0$, $\lambda_i \ge 0$, $\mu_j \in \mathbb{R}$, не все равные нулю, такие что:

1. $\nabla_x \mathcal{L}(x^*) = \lambda_0 \nabla f(x^*) + \sum_i \lambda_i \nabla g_i(x^*) + \sum_j \mu_j \nabla h_j(x^*) = 0$;
2. $\lambda_i g_i(x^*) = 0$ для всех i (дополняющая нежёсткость);
3. $g_i(x^*) \le 0$, $h_j(x^*) = 0$.

Условие регулярности (LICQ): градиенты всех равенств $\nabla h_j(x^*)$ и всех активных неравенств $\nabla g_i(x^*)$, $i \in I(x^*)$, линейно независимы. Тогда $\lambda_0 \ne 0$, можно взять $\lambda_0 = 1$, множители единственны. Доказательство $\lambda_0 \ne 0$ то же, что в О16: при $\lambda_0 = 0$ получаем нулевую нетривиальную комбинацию градиентов активных ограничений (у неактивных $\lambda_i = 0$). Более слабое условие — Мангасарьяна–Фромовица: ∇h_j линейно независимы и $\exists d$: $\langle \nabla h_j, d\rangle = 0$, $\langle \nabla g_i, d\rangle < 0$ для активных i. Для выпуклых задач — условие Слейтера (О34).

Максимум: $f \to \max$ эквивалентно $-f \to \min$. При $\lambda_0 = 1$ и тех же $g_i \le 0$ получаем $\nabla f + \sum \lambda_i \nabla g_i + \dots = 0$ с $\lambda_i \le 0$.

Смысл: в точке минимума антиградиент $-\nabla f$ является неотрицательной комбинацией внешних нормалей активных ограничений (плюс любая комбинация нормалей равенств): «сдвинуться в сторону убывания нельзя, не нарушив ограничения».

Алгоритм решения: перебираем варианты, какие неравенства активны ($\lambda_i > 0$ или $= 0$); в каждом решаем систему, проверяем допустимость и знаки множителей; отдельно проверяем нерегулярные точки ($\lambda_0 = 0$); затем выбираем лучшее или проверяем условия 2-го порядка.

ПРИМЕР

Листок 7, №1: $x^2 + y^2 - 6x - 10y \to \min$, $x^2 + y^2 \le 17$. $\mathcal{L} = x^2 + y^2 - 6x - 10y + \lambda(x^2 + y^2 - 17)$. Если $\lambda = 0$: $(3, 5)$, но $9 + 25 = 34 > 17$ — недопустима. Значит ограничение активно: $(1+\lambda)x = 3$, $(1+\lambda)y = 5$, $x^2 + y^2 = 17 \Rightarrow (1+\lambda)^2 = 34/17 = 2$, $1 + \lambda = \sqrt{2}$, $\lambda = \sqrt{2} - 1 > 0$ ✓. $x^* = \frac{3}{\sqrt{2}}$, $y^* = \frac{5}{\sqrt{2}}$ — минимум (задача выпуклая, ККТ достаточны). Максимум: $1 + \lambda = -\sqrt{2}$, $\lambda < 0$ (знак для максимума), точка $-(\frac{3}{\sqrt{2}}, \frac{5}{\sqrt{2}})$.

О20 Достаточные условия 1-го порядка (равенства и неравенства)

Пусть в допустимой точке x^* выполнены условия ККТ с $\lambda_0 = 1$, $\lambda^* \geq 0$, μ^*. Пусть $I^+ = \{i \in I(x^*) : \ \lambda_i^* > 0\}$ (активные с положительным множителем).

Теорема. Если градиенты $\{\nabla h_j(x^*)\}_{j=1}^k \cup \{\nabla g_i(x^*)\}_{i \in I^+}$ содержат n линейно независимых векторов (в частности, $k + |I^+| = n$ при LICQ), то x^* — точка строгого локального минимума.

Эквивалентная форма: если критический конус $K^* = \{d : \ \langle \nabla h_j, d\rangle = 0;\ \langle \nabla g_i, d\rangle = 0,\ i \in I^+;\ \langle \nabla g_i, d\rangle \leq 0,\ i \in I \setminus I^+\} = \{0\}$, то строгий минимум. Это ситуация «угловой точки»: ограничений с положительными множителями столько, что они фиксируют точку, и гессиан смотреть не нужно.

> ***Доказательство.***
> (Идея.) Для $d \neq 0$ из конуса допустимых направлений либо $\langle \nabla g_i, d\rangle < 0$ для какого-то $i \in I^+$, тогда $\langle \nabla f, d\rangle = -\sum \lambda_i \langle \nabla g_i, d\rangle - \sum \mu_j \langle \nabla h_j, d\rangle > 0$ — функция растёт с ненулевой первой производной; либо $d \in K^* = \{0\}$. Значит вдоль любого допустимого направления f строго растёт к первому порядку, равномерно по сфере направлений (компакт), отсюда строгий минимум.

Максимум: то же с $\lambda_i^* < 0$ (при соглашении $\lambda_0 = 1$, $g_i \leq 0$).

ПРИМЕР
$f = x + y \to \min$, $x \geq 0$, $y \geq 0$ ($g_1 = -x \leq 0$, $g_2 = -y \leq 0$). В $(0,0)$: $(1,1) - \lambda_1(1,0) - \lambda_2(0,1) = 0$, $\lambda_1 = \lambda_2 = 1 > 0$, два активных при $n = 2$ — строгий минимум без вторых производных.

О21 Достаточные условия 2-го порядка (равенства и неравенства)

$f, g_i, h_j \in C^2$; в допустимой x^* выполнены ККТ с $\lambda_0 = 1$, $\lambda^* \geq 0$, μ^*. Если

$$d^\top \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*)\, d > 0 \quad \forall d \in K^* \setminus \{0\},$$

$$K^* = \Big\{ d : \ \langle \nabla h_j(x^*), d\rangle = 0 \ \forall j;\ \langle \nabla g_i(x^*), d\rangle = 0,\ i \in I^+;\ \langle \nabla g_i(x^*), d\rangle \leq 0,\ i \in I(x^*) \setminus I^+ \Big\},$$

то x^* — строгий локальный минимум. Для максимума: множители $\lambda_i \leq 0$ и форма < 0 на K^*.

Если строгая дополняющая нежёсткость (все активные имеют $\lambda_i > 0$), то K^* — подпространство $\{d : \langle \nabla h_j, d\rangle = 0,\ \langle \nabla g_i, d\rangle = 0,\ i \in I\}$ — то же, что в задаче с равенствами, где активные неравенства считаются равенствами. Неактивные ограничения в проверке не участвуют.

Как связано с О20: если $K^* = \{0\}$, условие О21 выполнено автоматически.

О22 Необходимые условия 2-го порядка (равенства и неравенства)

$f, g_i, h_j \in C^2$, x^* — **регулярная** (LICQ) точка локального минимума, (λ^*, μ^*) — её (единственные) множители ККТ. Тогда

$$d^\top \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*)\, d \geq 0 \quad \forall d \in K^*,$$

с тем же критическим конусом K^*, что в О21. (Часто формулируют проще, на подпространстве $\{d : \langle \nabla h_j, d\rangle = 0,\ \langle \nabla g_i, d\rangle = 0,\ i \in I(x^*)\}$, — оно содержится в K^*, так что это следствие.) Для максимума — ≤ 0.

Разница с достаточными: нестрогое ≥ 0 против строгого > 0; в необходимых нужна регулярность, в достаточных — нет.

ПРИМЕР

Листок 7, №4 при $a > 0$: $x + y^2 + z^2 \to \min$, $x + y + z = a$, $x \geq 0$. Без активного $x \geq 0$: $\mathcal{L} = x + y^2 + z^2 + \mu(x + y + z - a) - \lambda x$: $1 + \mu - \lambda = 0$, $2y + \mu = 0$, $2z + \mu = 0$. Если $\lambda = 0$: $\mu = -1$, $y = z = \frac{1}{2}$, $x = a - 1$ — допустимо при $a \geq 1$. Если $x = 0$: $y = z = \frac{a}{2}$, $\mu = -a$, $\lambda = 1 - a \geq 0$ при $a \leq 1$. Задача выпуклая, поэтому найденные ККТ-точки — глобальные минимумы: $a \geq 1 - (a - 1, \frac{1}{2}, \frac{1}{2})$, $a \leq 1 - (0, \frac{a}{2}, \frac{a}{2})$. Максимума нет ($x \to +\infty$).

О23–О32: численные методы безусловной оптимизации

О23 Методы спуска (общая схема)

Итерация:

$$x^{k+1} = x^k + \alpha_k h^k,$$

где h^k — направление убывания (В20), $\alpha_k > 0$ — шаг, выбранный так, что $f(x^{k+1}) < f(x^k)$ (релаксационная последовательность).

- **Порядок метода:** нулевого (только значения f: покоординатный, случайный поиск, Нелдер–Мид), первого (ещё ∇f: градиентные, сопряжённые градиенты), второго ($\nabla^2 f$: Ньютон).
- **Выбор шага:** постоянный; дробление (начинаем с β, умножаем на $\lambda \in (0,1)$, пока не выполнится условие убывания, например Армихо $f(x^k + \alpha h^k) \le f(x^k) + \varepsilon\alpha\langle\nabla f(x^k), h^k\rangle$, $\varepsilon \in (0,1)$); наискорейший ($\alpha_k = \arg\min_{\alpha \ge 0} f(x^k + \alpha h^k)$ — одномерная минимизация, например методом золотого сечения).
- **Критерии остановки:** $\|\nabla f(x^k)\| < \varepsilon$; $\|x^{k+1} - x^k\| < \varepsilon$; $|f(x^{k+1}) - f(x^k)| < \varepsilon$ (часто несколько сразу).
- **Сходимость:** обычно гарантируется только к стационарной точке; к глобальному минимуму — для выпуклых f. Скорость: линейная $\|x^{k+1} - x^*\| \le q\|x^k - x^*\|$, $q < 1$; сверхлинейная; квадратичная $\le C\|x^k - x^*\|^2$.

О24 Направление убывания: необходимое и достаточное условия (1-й порядок)

док-во

f дифференцируема в x^0, $h \ne 0$.

- **Достаточное:** $\langle\nabla f(x^0), h\rangle < 0 \Rightarrow h$ — направление убывания.
- **Необходимое:** h — направление убывания $\Rightarrow \langle\nabla f(x^0), h\rangle \le 0$.

> ***Доказательство.***
> По определению дифференцируемости $f(x^0 + \alpha h) = f(x^0) + \alpha\langle\nabla f(x^0), h\rangle + o(\alpha)$, т.е.
>
> $$\frac{f(x^0 + \alpha h) - f(x^0)}{\alpha} = \langle\nabla f(x^0), h\rangle + o(1), \quad \alpha \to +0.$$
>
> Достаточность: правая часть стремится к отрицательному числу, значит при малых $\alpha > 0$ она отрицательна, т.е. $f(x^0 + \alpha h) < f(x^0)$.
> Необходимость: если бы $\langle\nabla f(x^0), h\rangle > 0$, то при малых $\alpha > 0$ разность была бы положительна — h не направление убывания. Значит ≤ 0.

При $\langle\nabla f, h\rangle = 0$ вопрос не решается первым порядком (см. О25). Среди $\|h\| = 1$ минимум $\langle\nabla f, h\rangle$ (по Коши–Буняковскому) даёт $h = -\nabla f / \|\nabla f\|$ — антиградиент, направление наискорейшего убывания.

О25 Направление убывания для дважды непрерывно дифференцируемой f

док-во

$f \in C^2$ в окрестности x^0, $h \neq 0$.

- **Необходимое:** если h — направление убывания, то $\langle \nabla f(x^0), h \rangle \leq 0$, и если $\langle \nabla f(x^0), h \rangle = 0$, то $h^\top \nabla^2 f(x^0) h \leq 0$.
- **Достаточное:** если $\langle \nabla f(x^0), h \rangle < 0$, либо $\langle \nabla f(x^0), h \rangle = 0$ и $h^\top \nabla^2 f(x^0) h < 0$, то h — направление убывания.

> ***Доказательство.***
> Тейлор: $f(x^0 + \alpha h) - f(x^0) = \alpha \langle \nabla f(x^0), h \rangle + \frac{\alpha^2}{2} h^\top \nabla^2 f(x^0) h + o(\alpha^2)$.
> Случай $\langle \nabla f, h \rangle \neq 0$ — как в О24. Пусть $\langle \nabla f, h \rangle = 0$. Тогда $\frac{f(x^0+\alpha h) - f(x^0)}{\alpha^2} = \frac{1}{2} h^\top \nabla^2 f h + o(1)$. Достаточность: при $h^\top \nabla^2 f h < 0$ правая часть отрицательна при малых α. Необходимость: если $h^\top \nabla^2 f h > 0$, разность положительна при малых α — противоречие.

ПРИМЕР

В седловой точке $\nabla f = 0$: направления убывания — те h, где $h^\top \nabla^2 f h < 0$. Для $f = x^2 - y^2$ в нуле это $h = (0, 1)$.

О26 Покоординатный спуск. Градиентный спуск. Наискорейший спуск

Покоординатный спуск (0-й порядок). Направления — по очереди координатные орты $e_1, \dots, e_n$ (цикл). На шаге по e_i: делаем одномерную минимизацию $\min_\alpha f(x + \alpha e_i)$ (или пробный шаг $\pm\alpha$: если f уменьшилась — принимаем, иначе пробуем в другую сторону; после неудачного цикла уменьшаем α). Простой, но медленный для «овражных» функций (сильная связь переменных). Для квадратичной функции с диагональной матрицей сходится за один цикл.

Градиентный спуск. $x^{k+1} = x^k - \alpha_k \nabla f(x^k)$.

- Постоянный шаг: сходится, если ∇f липшицев с константой L и $0 < \alpha < 2/L$.
- Дробление шага: берём $\alpha = \beta$; пока $f(x^k - \alpha \nabla f(x^k)) \geq f(x^k)$ (или не выполнено условие Армихо) — $\alpha := \lambda\alpha$.

Наискорейший спуск. $\alpha_k = \arg\min_{\alpha \geq 0} f(x^k - \alpha \nabla f(x^k))$. Свойство: **соседние градиенты ортогональны**, $\langle \nabla f(x^{k+1}), \nabla f(x^k) \rangle = 0$ (из $\varphi'(\alpha_k) = -\langle \nabla f(x^{k+1}), \nabla f(x^k) \rangle = 0$), отсюда зигзаг. Для квадратичной f с $A \succ 0$: $\alpha_k = \frac{\|g_k\|^2}{g_k^\top A g_k}$, сходимость линейная, $f(x^k) - f^* \leq \left(\frac{\kappa-1}{\kappa+1}\right)^{2k} (f(x^0) - f^*)$, $\kappa = \lambda_{\max}/\lambda_{\min}$ — чем хуже обусловленность, тем медленнее.

ПРИМЕР

Листок 4, №1: $f = x_1^2 + 4x_2^2$, $x^0 = (1, 1)$, $\beta = 1$, $\lambda = \frac{1}{2}$. $\nabla f = (2x_1, 8x_2)$, $\nabla f(x^0) = (2, 8)$, $f(x^0) = 5$.
$\alpha = 1$: $(-1, -7)$, $f = 197$ — нет. $\alpha = \frac{1}{2}$: $(0, -3)$, $f = 36$ — нет. $\alpha = \frac{1}{4}$: $(\frac{1}{2}, -1)$, $f = 4.25 < 5$ — принимаем, $x^1 = (\frac{1}{2}, -1)$.
$\nabla f(x^1) = (1, -8)$. $\alpha = 1$: $(-\frac{1}{2}, 7)$ — нет; $\frac{1}{2}$: $(0, 3)$, $f = 36$ — нет; $\frac{1}{4}$: $(\frac{1}{4}, 1)$, $f = 4.0625 < 4.25$ — принимаем, $x^2 = (\frac{1}{4}, 1)$. Видно, как мешает плохая обусловленность (4:1).

ПРИМЕР

Листок 4, №2: $f = x_1^2 + 2x_2^2$, $x^0 = (2, 1)$, $A = \mathrm{diag}(2, 4)$. $g_0 = (4, 4)$, $\alpha_0 = \frac{32}{2 \cdot 16 + 4 \cdot 16} = \frac{1}{3}$, $x^1 = (\frac{2}{3}, -\frac{1}{3})$, f: $6 \to \frac{2}{3}$. $g_1 = (\frac{4}{3}, -\frac{4}{3})$ ($\perp g_0$ ✓), $\alpha_1 = \frac{32/9}{(2+4) \cdot 16/9} = \frac{1}{3}$, $x^2 = (\frac{2}{9}, \frac{1}{9})$, $f = \frac{2}{27}$; $|f(x^2) - f(x^1)| \approx 0.59$. Дальше каждый шаг делит f на 9: $x^3 = (\frac{2}{27}, -\frac{1}{27})$, $f \approx 0.0082$, разность $0.066 < 0.1$ — стоп.

О27 Случайный поиск, наилучшая проба, статистический градиент, имитация отжига

Методы нулевого порядка: используют только значения f, работают для негладких и «шумных» функций. ξ — случайный вектор, равномерно распределённый на единичной сфере.

- **Случайный поиск (с возвратом при неудачном шаге).** $y = x^k + \alpha\xi^k$. Если $f(y) < f(x^k) - x^{k+1} = y$, иначе остаёмся ($x^{k+1} = x^k$) и генерируем новое ξ. После M неудач подряд уменьшаем α; остановка, когда $\alpha < \varepsilon$. Вариант: при неудаче пробовать $x^k - \alpha\xi$.
- **Метод наилучшей пробы.** Генерируем m случайных векторов $\xi_1, \dots, \xi_m$, вычисляем $f(x^k + \alpha\xi_j)$, выбираем лучшее j^*. Если $f(x^k + \alpha\xi_{j^*}) < f(x^k)$ — шагаем туда, иначе уменьшаем α (или повторяем пробы).
- **Метод статистического градиента.** Оцениваем градиент по случайным пробам: делаем m пробных шагов малой длины g: $\Delta_j = f(x^k + g\xi_j) - f(x^k)$. Оценка градиента: $\tilde{\nabla} = \sum_j \Delta_j \xi_j$ (при $m \to \infty$ направление стремится к ∇f). Шаг: $x^{k+1} = x^k - \alpha \frac{\tilde{\nabla}}{\|\tilde{\nabla}\|}$. Удобен, когда градиент нельзя посчитать аналитически.
- **Имитация отжига** (глобальный поиск). Из x^k генерируем кандидата y в окрестности. $\Delta = f(y) - f(x^k)$. Если $\Delta < 0$ — принимаем; если $\Delta \geq 0$ — принимаем с вероятностью $P = e^{-\Delta/T_k}$ (иначе остаёмся). Температура $T_k \to 0$: $T_{k+1} = \gamma T_k$, $\gamma \in (0,1)$ (или $T_k = T_0/\ln(1+k)$ — с гарантией сходимости к глобальному минимуму по вероятности). Благодаря иногда принимаемым ухудшениям метод выбирается из локальных минимумов; при малой T превращается в случайный спуск. Запоминаем лучшую точку за всё время.

О28 Сопряжённые направления: определение, линейная независимость

док-во

$A = A^\top \succ 0$. Ненулевые векторы $d_0, \dots, d_k$ называются **A-сопряжёнными** (взаимно сопряжёнными относительно A), если $d_i^\top A d_j = 0$ при $i \neq j$. При $A = I$ — обычная ортогональность.

Теорема. Ненулевые взаимно A-сопряжённые векторы линейно независимы. Следовательно, их не больше n, а n таких векторов образуют базис.

> ***Доказательство.***
> Пусть $\sum_i c_i d_i = 0$. Умножим слева на $d_j^\top A$: $\sum_i c_i\, d_j^\top A d_i = c_j\, d_j^\top A d_j = 0$ (остальные слагаемые нули по сопряжённости). Так как $d_j \neq 0$ и $A \succ 0$, $d_j^\top A d_j > 0$, значит $c_j = 0$. Это для всех j.

Примеры: собственные векторы A (ортогональные) — A-сопряжены: $q_i^\top A q_j = \lambda_j q_i^\top q_j = 0$.

О29 Метод сопряжённых направлений

Для квадратичной функции $f(x) = \frac{1}{2}x^\top Ax - b^\top x + c$, $A \succ 0$ (минимум в $x^* = A^{-1}b$).

Пусть $d_0, \dots, d_{n-1}$ — A-сопряжённые. Из любой x^0:

$$x^{k+1} = x^k + \alpha_k d_k, \qquad \alpha_k = \arg\min_\alpha f(x^k + \alpha d_k) = -\frac{\langle \nabla f(x^k), d_k\rangle}{d_k^\top A d_k}.$$

Теорема. Метод находит минимум квадратичной функции не более чем за n шагов: $x^n = x^*$. Более того, x^k минимизирует f на $x^0 + \operatorname{span}(d_0, \dots, d_{k-1})$ и $\langle \nabla f(x^k), d_i\rangle = 0$ при $i < k$.

> ***Доказательство.***
> (Идея.) Разложим $x^* - x^0 = \sum_i \sigma_i d_i$ (базис по О28). Умножая на $d_k^\top A$: $\sigma_k = \frac{d_k^\top A(x^* - x^0)}{d_k^\top A d_k}$. Так как $x^k - x^0 \in \operatorname{span}(d_0..d_{k-1})$ и $d_k^\top A d_i = 0$, имеем $d_k^\top A(x^* - x^0) = d_k^\top A(x^* - x^k) = -d_k^\top \nabla f(x^k)$ (т.к. $\nabla f(x^k) = A(x^k - x^*)$). Значит $\sigma_k = \alpha_k$, и после n шагов $x^n = x^0 + \sum \alpha_k d_k = x^*$.

Для неквадратичных функций метод применяют итерационно, с обновлением направлений каждые n шагов (рестарт). Варианты: построение направлений через ортогонализацию (О30), из градиентов (О31), без производных — метод Пауэлла.

О30 Алгоритм построения сопряжённых направлений

A-ортогонализация Грама–Шмидта. Берём любые линейно независимые $p_0, \dots, p_{n-1}$ (например, орты $e_1, \dots, e_n$). Строим:

$$d_0 = p_0, \qquad d_k = p_k - \sum_{i=0}^{k-1} \frac{p_k^\top A d_i}{d_i^\top A d_i}\, d_i, \quad k = 1, \dots, n-1.$$

> ***Доказательство.***
> Индукция: пусть $d_0, \dots, d_{k-1}$ взаимно сопряжены. Для $j < k$: $d_j^\top A d_k = d_j^\top A p_k - \sum_i \frac{p_k^\top A d_i}{d_i^\top A d_i} d_j^\top A d_i = d_j^\top A p_k - p_k^\top A d_j = 0$ (в сумме остаётся только $i = j$, A симметрична). $d_k \neq 0$, так как p_k не лежит в $\operatorname{span}(p_0..p_{k-1}) = \operatorname{span}(d_0..d_{k-1})$.

Другие способы: собственные векторы A; через градиенты по ходу спуска — метод сопряжённых градиентов (О31, не нужно хранить все предыдущие d_i: достаточно одного); без знания A — свойство параллельных подпространств (Пауэлл): если y_1, y_2 — минимумы f на двух параллельных прямых с направлением d, то $y_2 - y_1$ сопряжено с d.

ПРИМЕР

$A = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$, $p_0 = e_1$, $p_1 = e_2$. $d_0 = (1, 0)$, $d_0^\top A d_0 = 2$, $p_1^\top A d_0 = -1$. $d_1 = (0, 1) + \frac{1}{2}(1, 0) = (\frac{1}{2}, 1)$.
Проверка: $d_0^\top A d_1 = (2, -1) \cdot (\frac{1}{2}, 1) = 0$ ✓.

О31 Метод сопряжённых градиентов

Обозначим $g_k = \nabla f(x^k)$.

1. $d_0 = -g_0$.
2. $x^{k+1} = x^k + \alpha_k d_k$, $\alpha_k = \arg\min_{\alpha \geq 0} f(x^k + \alpha d_k)$ (для квадратичной: $\alpha_k = \frac{\|g_k\|^2}{d_k^\top A d_k}$).
3. $d_{k+1} = -g_{k+1} + \beta_k d_k$, где $\beta_k = \dfrac{\|g_{k+1}\|^2}{\|g_k\|^2}$ (Флетчер–Ривс) или $\beta_k = \dfrac{\langle g_{k+1}, g_{k+1} - g_k \rangle}{\|g_k\|^2}$ (Полак–Рибьер).
4. Остановка: $\|g_k\| < \varepsilon$. Для неквадратичных — рестарт $d = -g$ каждые n шагов.

Свойства на квадратичной функции: направления d_k взаимно A-сопряжены, градиенты g_k взаимно ортогональны, минимум достигается не более чем за n шагов. Обе формулы β на квадратичной функции совпадают (т.к. $\langle g_{k+1}, g_k \rangle = 0$). Формула β_k получается из требования $d_{k+1}^\top A d_k = 0$: $\beta_k = \frac{g_{k+1}^\top A d_k}{d_k^\top A d_k}$, и это преобразуется к виду ФР. Плюсы: память $O(n)$, не нужен гессиан, намного быстрее наискорейшего спуска.

ПРИМЕР

Листок 4, №3: $f = x_1^2 + \frac{1}{2}x_2^2 - x_1 x_2$, $x^0 = (1, 1)$. $A = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$, $\nabla f = (2x_1 - x_2,\ x_2 - x_1)$.
$g_0 = (1, 0)$, $d_0 = (-1, 0)$, $d_0^\top A d_0 = 2$, $\alpha_0 = \frac{1}{2}$, $x^1 = (\frac{1}{2}, 1)$.
$g_1 = (0, \frac{1}{2})$ ($\perp g_0$ ✓), $\beta_0 = \frac{1/4}{1} = \frac{1}{4}$, $d_1 = (0, -\frac{1}{2}) + \frac{1}{4}(-1, 0) = (-\frac{1}{4}, -\frac{1}{2})$.
$Ad_1 = (0, -\frac{1}{4})$, $d_1^\top A d_1 = \frac{1}{8}$, $\alpha_1 = \frac{1/4}{1/8} = 2$, $x^2 = (\frac{1}{2}, 1) + 2(-\frac{1}{4}, -\frac{1}{2}) = (0, 0)$. Минимум за $n = 2$ шага, $f^* = 0$.

О32 Метод Ньютона. Квазиньютоновские методы

Метод Ньютона. Приближаем f квадратичной моделью $f(x^k) + \langle g_k, x - x^k\rangle + \frac{1}{2}(x-x^k)^\top \nabla^2 f(x^k)(x-x^k)$ и берём её минимум:

$$x^{k+1} = x^k - \left[\nabla^2 f(x^k)\right]^{-1} \nabla f(x^k)$$

(на практике решаем систему $\nabla^2 f(x^k)\, p = -g_k$). Эквивалентно — метод Ньютона для системы $\nabla f = 0$.

- Квадратичную функцию с $A \succ 0$ минимизирует за 1 шаг.
- Если $\nabla^2 f(x^*) \succ 0$ и гессиан липшицев, то вблизи x^* сходимость **квадратичная**.
- Минусы: нужен гессиан и решение системы ($O(n^3)$); вдали от минимума может расходиться; если $\nabla^2 f \not\succ 0$, направление может не быть направлением убывания, метод сходится к любой стационарной точке, в том числе к седлу или максимуму.
- Модификации: демпфированный Ньютон $x^{k+1} = x^k - \alpha_k [\nabla^2 f]^{-1} g_k$ с подбором α_k; Левенберг–Марквардт $[\nabla^2 f + \mu I]^{-1}$.

ПРИМЕР

Листок 4–5, №4: $f = x^3 + 2y^2 - 3x - 4y$, старт $(-\frac{1}{2}, -1)$. $\nabla f = (3x^2 - 3,\ 4y - 4)$, $\nabla^2 f = \mathrm{diag}(6x, 4)$. $x^1 = -\frac{1}{2} - \frac{-9/4}{-3} = -\frac{5}{4}$, $y^1 = -1 - \frac{-8}{4} = 1$. Дальше $x \to -1$, $y = 1$ уже точно. Но $(-1, 1)$ — седло ($6x = -6 < 0$)! Ньютон сошёлся к стационарной точке, которая не минимум. Минимум — $(1, 1)$, $f = -4$; к нему метод пошёл бы из точки с $x > 0$.

Квазиньютоновские методы. $x^{k+1} = x^k - \alpha_k H_k g_k$, где $H_k \approx [\nabla^2 f]^{-1}$ строится по градиентам, без вторых производных. $H_0 = I$ (первый шаг — градиентный). Обозначим $s_k = x^{k+1} - x^k$, $y_k = g_{k+1} - g_k$. Требование — **квазиньютоновское (секущее) условие** $H_{k+1} y_k = s_k$.

- **DFP** (Дэвидон–Флетчер–Пауэлл): $H_{k+1} = H_k + \dfrac{s_k s_k^\top}{s_k^\top y_k} - \dfrac{H_k y_k y_k^\top H_k}{y_k^\top H_k y_k}$.
- **BFGS** (Бройден–Флетчер–Гольдфарб–Шанно): $H_{k+1} = (I - \rho_k s_k y_k^\top) H_k (I - \rho_k y_k s_k^\top) + \rho_k s_k s_k^\top$, $\rho_k = 1/(y_k^\top s_k)$. На практике лучший.
- Свойства: $H_k \succ 0$ сохраняется (если $y_k^\top s_k > 0$), поэтому направление всегда направление убывания; на квадратичной функции с точным одномерным поиском — минимум за n шагов, $H_n = A^{-1}$, направления сопряжены; сходимость сверхлинейная.

О33–О35: выпуклое программирование

О33 ЗВП. Выпуклость допустимого множества док-во

Задача выпуклого программирования:

$$f(x) \to \min, \quad g_i(x) \le 0,\ i = 1..m, \quad x \in P,$$

где f, g_i — выпуклые функции, P — выпуклое множество (например $\mathbb{R}^n$ или $\mathbb{R}^n_+$). Допускаются также линейные (аффинные) равенства $\langle a_j, x\rangle = b_j$. Нелинейные равенства не допускаются — их множество в общем случае невыпукло (окружность).

Теорема. Допустимое множество $X = \{x \in P :\ g_i(x) \le 0\}$ ЗВП выпукло.

> ***Доказательство.***
> 1) Каждое $X_i = \{x : g_i(x) \le 0\}$ выпукло: для $x, y \in X_i$, $\lambda \in [0,1]$: $g_i(\lambda x + (1-\lambda)y) \le \lambda g_i(x) + (1-\lambda) g_i(y) \le \lambda \cdot 0 + (1-\lambda)\cdot 0 = 0$. 2) Аффинное равенство задаёт гиперплоскость — выпукло (О9). 3) $X = P \cap \bigcap_i X_i \cap \dots$ — пересечение выпуклых множеств, выпукло по О9.

Значит ЗВП — выпуклая задача, и к ней применимо всё из О14, О15: локальный минимум = глобальный.

О34 Необходимые и достаточные условия оптимальности в ЗВП

док-во

$\mathcal{L}(x, \lambda) = f(x) + \sum \lambda_i g_i(x)$ (плюс $\sum \mu_j h_j$ для аффинных равенств).

Условие Слейтера: существует $\bar{x} \in P$ с $g_i(\bar{x}) < 0$ для всех i (строго допустимая точка; равенства выполнены).

Необходимые условия. Если x^* — решение ЗВП и выполнено условие Слейтера, то существует $\lambda^* \geq 0$, такое что: (а) x^* минимизирует $\mathcal{L}(x, \lambda^*)$ на P; в гладком случае с $P = \mathbb{R}^n$ это $\nabla f(x^*) + \sum \lambda_i^* \nabla g_i(x^*) = 0$; (б) $\lambda_i^* g_i(x^*) = 0$. Здесь $\lambda_0 = 1$ гарантируется именно условием Слейтера.

Достаточные условия. Если x^* допустима и существуют $\lambda^* \geq 0$ (μ^* любые) с

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0, \qquad \lambda_i^* g_i(x^*) = 0,$$

то x^* — **глобальный** минимум ЗВП. (Слейтер и регулярность не нужны.)

> ***Доказательство.***
> При $\lambda^* \geq 0$ функция $x \mapsto \mathcal{L}(x, \lambda^*, \mu^*) = f + \sum \lambda_i^* g_i + \sum \mu_j^* h_j$ выпукла (неотрицательная комбинация выпуклых плюс аффинные). Градиент в x^* равен нулю, по О15 x^* — глобальный минимум $\mathcal{L}(\cdot, \lambda^*, \mu^*)$. Для любой допустимой x ($g_i(x) \leq 0$, $h_j(x) = 0$, $\lambda_i^* \geq 0$):
>
> $$f(x) \geq f(x) + \sum \lambda_i^* g_i(x) + \sum \mu_j^* h_j(x) = \mathcal{L}(x) \geq \mathcal{L}(x^*) = f(x^*) + \underbrace{\sum \lambda_i^* g_i(x^*)}_{=0} + \underbrace{\sum \mu_j^* h_j(x^*)}_{=0} = f(x^*).$$

Итог: в ЗВП со Слейтером условия ККТ — **необходимые и достаточные** для глобального минимума.

ЛОВУШКА

Без Слейтера необходимость может нарушиться: $x \to \min$, $x^2 \leq 0$. Единственная допустимая точка 0, но $1 + \lambda \cdot 2 \cdot 0 = 0$ невозможно — ККТ с $\lambda_0 = 1$ не выполняется.

О35 Теорема Куна–Таккера о седловой точке

ЗВП: $f(x) \to \min$, $g_i(x) \le 0$, $x \in P$; $\mathcal{L}(x, \lambda) = f(x) + \sum \lambda_i g_i(x)$, $\lambda \in \mathbb{R}^m_+$.

Седловая точка (x^*, λ^*), $x^* \in P$, $\lambda^* \ge 0$:

$$\mathcal{L}(x^*, \lambda) \le \mathcal{L}(x^*, \lambda^*) \le \mathcal{L}(x, \lambda^*) \qquad \forall x \in P,\ \forall \lambda \ge 0$$

(минимум по x, максимум по λ).

Теорема Куна–Таккера. Пусть f, g_i выпуклы, P выпукло и выполнено условие Слейтера. Тогда x^* — решение ЗВП $\iff$ существует $\lambda^* \ge 0$, при котором (x^*, λ^*) — седловая точка функции Лагранжа.

> ***Доказательство.***
> **Достаточность** (верна без выпуклости и Слейтера). Левое неравенство: $\sum \lambda_i g_i(x^*) \le \sum \lambda_i^* g_i(x^*)$ для всех $\lambda \ge 0$. Если бы $g_i(x^*) > 0$, то при $\lambda_i \to \infty$ левая часть неограниченна — значит $g_i(x^*) \le 0$, x^* допустима. При $\lambda = 0$: $\sum \lambda_i^* g_i(x^*) \ge 0$, но каждое слагаемое ≤ 0, значит $\lambda_i^* g_i(x^*) = 0$ (дополняющая нежёсткость). Правое неравенство: для допустимой x
>
> $$f(x^*) = \mathcal{L}(x^*, \lambda^*) \le \mathcal{L}(x, \lambda^*) = f(x) + \sum \lambda_i^* g_i(x) \le f(x).$$
>
> **Необходимость** (идея). Множество $S = \{(u_0, u) \in \mathbb{R}^{m+1} :\ \exists x \in P,\ f(x) < u_0,\ g_i(x) \le u_i\}$ выпукло (по выпуклости f, g_i, P) и не содержит точку $(f(x^*), 0)$ (иначе нашлась бы допустимая x с $f(x) < f(x^*)$). Отделяем их гиперплоскостью; нормаль даёт $(\lambda_0, \lambda) \ge 0$, Слейтер гарантирует $\lambda_0 > 0$. Получаем $\mathcal{L}(x, \lambda^*) \ge f(x^*)$ для всех $x \in P$ и $\lambda_i^* g_i(x^*) = 0$, откуда обе части седлового неравенства.

Связь с двойственностью: $\min_x \max_{\lambda \ge 0} \mathcal{L} = \max_{\lambda \ge 0} \min_x \mathcal{L}$ — значения прямой и двойственной задач совпадают. В гладком случае с $P = \mathbb{R}^n$ седловое условие равносильно условиям ККТ (О34).

О36: динамическое программирование

О36 Задача ДП. Принцип оптимальности и уравнения Беллмана

Модель. Управляемая система проходит N шагов: $X_0 \xrightarrow{u_1} X_1 \xrightarrow{u_2} \cdots \xrightarrow{u_N} X_N$.

- X_k — состояние после k-го шага, $u_k \in V_k$ — управление на k-м шаге (V_k — допустимые управления, могут зависеть от X_{k-1}).
- Переход $X_k = \varphi_k(X_{k-1}, u_k)$ — **нет последействия**: новое состояние зависит только от текущего состояния и управления, но не от предыстории.
- Выигрыш на шаге $W_k(X_{k-1}, u_k)$, критерий **аддитивный**: $W = \sum_{k=1}^{N} W_k(X_{k-1}, u_k) \to \max$.

Принцип оптимальности Беллмана. Каковы бы ни были начальное состояние и решение на первом шаге, решения на оставшихся шагах должны составлять оптимальное управление относительно состояния, полученного после первого шага. («Хвост оптимальной траектории оптимален».)

Функция Беллмана $F_k(X_{k-1})$ — максимальный выигрыш за шаги $k, \ldots, N$, если перед k-м шагом система в состоянии X_{k-1}.

Функциональные уравнения Беллмана:

$$F_N(X_{N-1}) = \max_{u_N \in V_N} W_N(X_{N-1}, u_N),$$

$$F_k(X_{k-1}) = \max_{u_k \in V_k} \Big\{ W_k(X_{k-1}, u_k) + F_{k+1}\big(\varphi_k(X_{k-1}, u_k)\big) \Big\}, \quad k = N-1, \ldots, 1.$$

Смысл: выигрыш на этом шаге плюс лучшее, что можно получить потом из того состояния, куда попали.

Алгоритм. 1) Обратный ход: для $k = N, \ldots, 1$ и всех возможных X_{k-1} считаем F_k и запоминаем условно-оптимальное управление $\hat{u}_k(X_{k-1})$. 2) $F_1(X_0)$ — оптимальное значение. 3) Прямой ход: $u_1^* = \hat{u}_1(X_0)$, $X_1 = \varphi_1(X_0, u_1^*)$, $u_2^* = \hat{u}_2(X_1)$ и т.д. Если из состояния нельзя выполнить условия — $F_k = -\infty$; если максимум при нескольких u — запоминаем все (оптимальных планов несколько).

Почему это работает: вместо перебора всех комбинаций управлений (экспонента от N) решаем N маленьких задач. Обоснование — принцип оптимальности (следует из аддитивности и отсутствия последействия).

ПРИМЕР

Листок 10, №3: 80 единиц ресурса, 3 предприятия, порции по 20. Состояние s — остаток ресурса. $F_3(s) = z_3(s)$: $0, 33, 40, 50, 80$ при $s = 0, 20, 40, 60, 80$.
$F_2(s) = \max_x\{z_2(x) + F_3(s-x)\}$: $F_2(0) = 0$, $F_2(20) = 33$, $F_2(40) = \max(40, 21+33, 23) = 54$,
$F_2(60) = \max(50, 21+40, 23+33, 34) = 61$, $F_2(80) = \max(80, 21+50, 23+40, 34+33, 90) = 90$.
$F_1(80) = \max\{0+90,\ 34+61,\ 47+54,\ 67+33,\ 70+0\} = \max\{90, 95, 101, 100, 70\} = 101$ при $x_1 = 40$.
Прямой ход: осталось 40, $F_2(40) = 54$ при $x_2 = 20$, $x_3 = 20$. Ответ: $(40, 20, 20)$, прибыль $47 + 21 + 33 = 101$.

О37–О40: функционалы и вариации

О37 Нормы в $C[t_0, t_1]$ и $C^1[t_0, t_1]$

$\|x\|_C = \max_{[t_0,t_1]} |x(t)|$ (равномерная норма; максимум существует по Вейерштрассу). $\|x\|_{C^1} = \max\left(\|x\|_C, \|\dot{x}\|_C\right)$ (или эквивалентная $\|x\|_C + \|\dot{x}\|_C$).

Аксиомы нормы: $\|x\| \geq 0$, $\|x\| = 0 \iff x \equiv 0$; $\|\alpha x\| = |\alpha|\|x\|$; $\|x + y\| \leq \|x\| + \|y\|$ (так как $|x(t) + y(t)| \leq |x(t)| + |y(t)| \leq \|x\| + \|y\|$, берём максимум по t). Оба пространства полные (банаховы).

Сильный и слабый экстремум. Локальный экстремум относительно нормы C называют сильным, относительно нормы C^1 — слабым. Окрестность в C^1 меньше (близки и функции, и производные), поэтому сильный экстремум всегда слабый, обратное неверно. В задачах КВИ на C^1 обычно ищут слабый экстремум.

ПРИМЕР

Листок 13, №1: расстояние между x^4 и $3x + 2$ в $C^1[1, 3]$. $d(t) = t^4 - 3t - 2$, $d'(t) = 4t^3 - 3 > 0$ на $[1, 3]$, так что d монотонна: $d(1) = -4$, $d(3) = 70$, $\max|d| = 70$. $|d'|$: от 1 до 105, $\max|d'| = 105$. С нормой-максимумом $\rho = 105$, с нормой-суммой $\rho = 175$ (уточни, какая норма была на лекции).

О38 Вариация по направлению, вариация по Лагранжу, производная Гато

X — линейное нормированное пространство, $J : X \to \mathbb{R}$, $x \in X$, $h \in X$.

- **Вариация (производная) по направлению** h: $\delta^+ J(x, h) = \lim_{\alpha \to +0} \frac{J(x+\alpha h) - J(x)}{\alpha}$ — односторонний предел.
- **Первая вариация (по Лагранжу):** если для каждого h существует двусторонний предел $\delta J(x, h) = \lim_{\alpha \to 0} \frac{J(x+\alpha h) - J(x)}{\alpha} = \frac{d}{d\alpha} J(x + \alpha h)\big|_{\alpha=0}$. Это однородная функция h: $\delta J(x, ch) = c\,\delta J(x, h)$, но не обязательно линейная.
- **Производная Гато:** если первая вариация существует и $h \mapsto \delta J(x, h)$ — линейный непрерывный функционал, то J дифференцируема по Гато, $J'_\Gamma(x)[h] = \delta J(x, h)$.
- (Для полноты) **Производная Фреше:** $J(x + h) = J(x) + \Lambda h + o(\|h\|)$, Λ линейный непрерывный.

Цепочка: Фреше $\Rightarrow$ Гато $\Rightarrow$ вариация по Лагранжу $\Rightarrow$ производные по всем направлениям. Обратные импликации неверны.

ПРИМЕР

Листок 11: $|x|$ в нуле: $\delta^+ J(0, h) = |h|$ для любых h, но двустороннего предела нет ($\alpha < 0$ даёт $-|h|$) — вариации по Лагранжу нет. $r\cos 3\varphi$ в нуле: $\delta J(0, h) = |h| \cos 3\varphi_h$ — существует (однородна и нечётна), но не линейна по h — производной Гато нет.

Подсчёт вариации: $J(f) = \left(\int_0^1 f\,dt\right)^3$ на $C[0, 1]$: $\frac{d}{d\alpha}\left(\int (f + \alpha h)\right)^3\Big|_0 = 3\left(\int_0^1 f\right)^2 \int_0^1 h\,dt$.

О39 Необходимое условие экстремума функционала без ограничений

док-во

Теорема. Если $\hat{x}$ — точка локального экстремума функционала J на линейном нормированном пространстве X и в $\hat{x}$ существует первая вариация $\delta J(\hat{x}, h)$, то $\delta J(\hat{x}, h) = 0$ для всех $h \in X$.

> ***Доказательство.***
> Пусть $\hat{x}$ — локальный минимум: $J(x) \geq J(\hat{x})$ при $\|x - \hat{x}\| < \varepsilon$. Зафиксируем $h \neq 0$ и рассмотрим функцию одной переменной $\varphi(\alpha) = J(\hat{x} + \alpha h)$. При $|\alpha| < \varepsilon / \|h\|$ точка $\hat{x} + \alpha h$ лежит в ε-окрестности, поэтому $\varphi(\alpha) \geq \varphi(0)$ — у φ локальный минимум в 0. По условию φ дифференцируема в 0, $\varphi'(0) = \delta J(\hat{x}, h)$. По теореме Ферма $\varphi'(0) = 0$. Для $h = 0$ вариация равна нулю тривиально.

Если существует только односторонняя вариация по направлению, то для минимума получаем лишь $\delta^{+} J(\hat{x}, h) \geq 0$ для всех h (так у $|x|$ в нуле).

О40 Допустимые направления. Необходимое условие при ограничениях

док-во

Задача $J(x) \to \text{extr}$, $x \in M \subset X$. Вектор h — **допустимое направление** в точке $\hat{x} \in M$, если $\exists \alpha_0 > 0$: $\hat{x} + \alpha h \in M$ для всех $|\alpha| < \alpha_0$ (двусторонняя допустимость).

Пример: в простейшей задаче $M = \{x \in C^1 : x(t_0) = x_0,\ x(t_1) = x_1\}$ допустимые направления — ровно $h \in C_0^1$, и $\hat{x} + \alpha h \in M$ при любом α. Множество допустимых направлений — линейное подпространство, когда M — аффинное многообразие.

Теорема. Если $\hat{x}$ — точка локального экстремума J на M и для допустимого направления h существует $\delta J(\hat{x}, h)$, то $\delta J(\hat{x}, h) = 0$.

> ***Доказательство.***
> $\varphi(\alpha) = J(\hat{x} + \alpha h)$ определена на $(-\alpha_0, \alpha_0)$ и все точки $\hat{x} + \alpha h$ допустимы. При малых $|\alpha|$ они в окрестности $\hat{x}$, поэтому $\varphi(\alpha) \geq \varphi(0)$ (для минимума). $\varphi'(0) = \delta J(\hat{x}, h)$ существует, по Ферма $= 0$.

Если направление допустимо лишь в одну сторону ($\alpha \in [0, \alpha_0)$), получаем для минимума $\delta J(\hat{x}, h) \geq 0$ — аналог условия $\langle \nabla f, x - x^* \rangle \geq 0$ из О15.

О41–О46: простейшая задача КВИ и задача Больца

О41 Простейшая задача КВИ. Локальный экстремум. Вывод первой вариации

Постановка — В27. Интегрант $L(t, x, \dot{x})$ (лагранжиан); предполагаем $L, L_x, L_{\dot{x}}$ непрерывны. Допустимые функции — $x \in C^1[t_0, t_1]$ с $x(t_0) = x_0$, $x(t_1) = x_1$.

Локальный (слабый) минимум: допустимая $\hat{x}$ такая, что $\exists \varepsilon > 0$: $J(x) \geq J(\hat{x})$ для всех допустимых x с $\|x - \hat{x}\|_{C^1} < \varepsilon$. Сильный — то же с нормой C.

Первая вариация. Для $h \in C_0^1$:

$$\varphi(\alpha) = J(\hat{x} + \alpha h) = \int_{t_0}^{t_1} L\big(t, \hat{x} + \alpha h, \dot{\hat{x}} + \alpha \dot{h}\big)\, dt.$$

Подынтегральная функция и её производная по α непрерывны, дифференцируем под знаком интеграла:

$$\delta J(\hat{x}, h) = \varphi'(0) = \int_{t_0}^{t_1} \Big(\hat{L}_x(t)\, h(t) + \hat{L}_{\dot{x}}(t)\, \dot{h}(t)\Big)\, dt,$$

где $\hat{L}_x(t) = L_x(t, \hat{x}(t), \dot{\hat{x}}(t))$, аналогично $\hat{L}_{\dot{x}}$. Вариация линейна и непрерывна по h (функционал дифференцируем по Гато, даже по Фреше в C^1).

О42 Лемма Лагранжа док-во

Лемма. Если $a \in C[t_0, t_1]$ и $\int_{t_0}^{t_1} a(t)h(t)\, dt = 0$ для всех $h \in C_0^1[t_0, t_1]$, то $a(t) \equiv 0$.

> ***Доказательство.***
> Пусть $a(\tau) \neq 0$ в некоторой точке $\tau \in (t_0, t_1)$, например $a(\tau) > 0$. По непрерывности $a(t) > 0$ на $[\tau - \delta, \tau + \delta] \subset (t_0, t_1)$. Возьмём
>
> $$h(t) = \begin{cases} (t - \tau + \delta)^2 (\tau + \delta - t)^2, & |t - \tau| \leq \delta, \\ 0, & \text{иначе.} \end{cases}$$
>
> $h \in C^1$ (в точках склейки и функция, и производная равны нулю), $h(t_0) = h(t_1) = 0$, $h > 0$ внутри отрезка. Тогда $\int ah\, dt = \int_{\tau-\delta}^{\tau+\delta} ah\, dt > 0$ — противоречие. Значит $a = 0$ на (t_0, t_1), а по непрерывности и на концах.

О43 Лемма Дюбуа-Реймона

док-во

Лемма. Если $b \in C[t_0, t_1]$ и $\int_{t_0}^{t_1} b(t)\dot{h}(t)\,dt = 0$ для всех $h \in C_0^1[t_0, t_1]$, то $b(t) \equiv \text{const}$.

> ***Доказательство.***
> Положим $c = \frac{1}{t_1 - t_0}\int_{t_0}^{t_1} b(t)\,dt$ и $h(t) = \int_{t_0}^{t} \big(b(s) - c\big)ds$. Тогда $h \in C^1$, $h(t_0) = 0$, $h(t_1) = \int b\,dt - c(t_1 - t_0) = 0$, т.е. $h \in C_0^1$, и $\dot{h} = b - c$. Считаем:
>
> $$\int_{t_0}^{t_1} (b-c)^2 dt = \int (b-c)\dot{h}\,dt = \underbrace{\int b\dot{h}\,dt}_{=0} - c\underbrace{\int \dot{h}\,dt}_{=h(t_1)-h(t_0)=0} = 0.$$
>
> Подынтегральная функция непрерывна и неотрицательна, значит $b(t) \equiv c$.

Объединённая лемма (именно она даёт уравнение Эйлера): если $a, b \in C$ и $\int (ah + b\dot{h})\,dt = 0$ для всех $h \in C_0^1$, то $b \in C^1$ и $\dot{b} = a$.

> ***Доказательство.***
> $A(t) = \int_{t_0}^{t} a(s)ds$. Интегрируем по частям: $\int ah\,dt = [Ah]_{t_0}^{t_1} - \int A\dot{h}\,dt = -\int A\dot{h}\,dt$ (т.к. $h(t_0) = h(t_1) = 0$). Получаем $\int (b - A)\dot{h}\,dt = 0$, по лемме Дюбуа-Реймона $b - A = \text{const}$, т.е. $b = A + c \in C^1$, $\dot{b} = a$.

О44 Необходимые условия в простейшей задаче (уравнение Эйлера)

док-во

Теорема. Пусть $L, L_x, L_{\dot{x}}$ непрерывны, $\hat{x} \in C^1$ — точка слабого локального экстремума простейшей задачи. Тогда $\hat{L}_{\dot{x}}(t)$ непрерывно дифференцируема и выполнено уравнение Эйлера–Лагранжа

$$-\frac{d}{dt}\hat{L}_{\dot{x}}(t) + \hat{L}_x(t) = 0, \quad t \in [t_0, t_1].$$

> ***Доказательство.***
> Все $h \in C_0^1$ — допустимые направления (О40), вариация существует (О41), значит $\delta J(\hat{x}, h) = \int (\hat{L}_x h + \hat{L}_{\dot{x}}\dot{h})dt = 0$ для всех $h \in C_0^1$. По объединённой лемме (О43) с $a = \hat{L}_x$, $b = \hat{L}_{\dot{x}}$: $\hat{L}_{\dot{x}} \in C^1$ и $\frac{d}{dt}\hat{L}_{\dot{x}} = \hat{L}_x$. (Вариант, если $\hat{x} \in C^2$ и $L \in C^2$: проинтегрировать $\int \hat{L}_{\dot{x}}\dot{h}$ по частям, получить $\int (\hat{L}_x - \frac{d}{dt}\hat{L}_{\dot{x}})h\,dt = 0$ и применить лемму Лагранжа.)

Частные случаи (первые интегралы): L не зависит от $x \Rightarrow L_{\dot{x}} = \text{const}$ (интеграл импульса); L не зависит от t $\Rightarrow L - \dot{x}L_{\dot{x}} = \text{const}$ (интеграл энергии).

Схема решения задачи: 1) выписать уравнение Эйлера, решить; 2) константы из граничных условий; 3) проверить, что экстремаль даёт минимум/максимум: через выпуклость L по $(x, \dot{x})$ (О50) или напрямую $J(\hat{x} + h) - J(\hat{x})$, или по Лежандру–Якоби; 4) указать, достигается ли другой экстремум (обычно $\sup = +\infty$ или $\inf = -\infty$ — привести последовательность).

ПРИМЕР

Листок 12, №1: $\int_1^e t\dot{x}^2 dt$, $x(1) = 0$, $x(e) = 1$. L не зависит от x: $2t\dot{x} = C$, $\dot{x} = \frac{C}{2t}$, $x = C_1 \ln t + C_2$, по краевым условиям $\hat{x} = \ln t$. Проверка: $J(\hat{x} + h) - J(\hat{x}) = \int t(2\dot{\hat{x}}\dot{h} + \dot{h}^2) = \int 2\dot{h}\,dt + \int t\dot{h}^2 = 0 + \int t\dot{h}^2 \geq 0$ — глобальный минимум, $J_{\min} = \int_1^e \frac{1}{t}\,dt = 1$. Максимума нет ($\sup = +\infty$).

О45 Задача Больца. Локальный экстремум. Вывод первой вариации

Постановка — В28: $B(x) = \int_{t_0}^{t_1} L(t,x,\dot{x})\,dt + l\big(x(t_0), x(t_1)\big) \to \text{extr}$, $x \in C^1[t_0,t_1]$, концы свободны. l — терминант, непрерывно дифференцируемая функция двух переменных.

Локальный (слабый) экстремум: $\hat{x} \in C^1$, $\exists \varepsilon$: $B(x) \geq B(\hat{x})$ при $\|x - \hat{x}\|_{C^1} < \varepsilon$. Все $x \in C^1$ допустимы, допустимые направления — все $h \in C^1$.

Вариация:

$$\delta B(\hat{x}, h) = \frac{d}{d\alpha} B(\hat{x} + \alpha h)\Big|_0 = \int_{t_0}^{t_1} \big(\hat{L}_x h + \hat{L}_{\dot{x}} \dot{h}\big) dt + \hat{l}_{x_0} h(t_0) + \hat{l}_{x_1} h(t_1),$$

где $\hat{l}_{x_0} = \frac{\partial l}{\partial x_0}\big(\hat{x}(t_0), \hat{x}(t_1)\big)$, $\hat{l}_{x_1}$ аналогично (цепное правило для $l(\hat{x}(t_0) + \alpha h(t_0), \hat{x}(t_1) + \alpha h(t_1))$).

О46 Необходимые условия в задаче Больца

док-во

Теорема. Если $\hat{x}$ — слабый локальный экстремум задачи Больца ($L, L_x, L_{\dot{x}}$ непрерывны, $l \in C^1$), то $\hat{L}_{\dot{x}} \in C^1$ и

$$-\frac{d}{dt}\hat{L}_{\dot{x}} + \hat{L}_x = 0 \;\; (\text{Эйлер}), \qquad \hat{L}_{\dot{x}}(t_0) = \hat{l}_{x_0}, \quad \hat{L}_{\dot{x}}(t_1) = -\hat{l}_{x_1} \;\; (\text{трансверсальность}).$$

> ***Доказательство.***
> По О39 $\delta B(\hat{x}, h) = 0$ для всех $h \in C^1$. 1) Берём $h \in C_0^1$: внеинтегральные члены исчезают, остаётся $\int (\hat{L}_x h + \hat{L}_{\dot{x}} \dot{h}) = 0$ — по объединённой лемме уравнение Эйлера и $\hat{L}_{\dot{x}} \in C^1$. 2) Теперь произвольное $h \in C^1$. Интегрируем по частям (можно, так как $\hat{L}_{\dot{x}} \in C^1$):
>
> $$\int \hat{L}_{\dot{x}} \dot{h}\, dt = \big[\hat{L}_{\dot{x}} h\big]_{t_0}^{t_1} - \int \frac{d}{dt}\hat{L}_{\dot{x}}\, h\, dt.$$
>
> С учётом Эйлера интеграл исчезает:
>
> $$0 = \delta B = \big(\hat{L}_{\dot{x}}(t_1) + \hat{l}_{x_1}\big) h(t_1) + \big(\hat{l}_{x_0} - \hat{L}_{\dot{x}}(t_0)\big) h(t_0).$$
>
> 3) Возьмём h с $h(t_0) = 1$, $h(t_1) = 0$ (например, линейную): получаем $\hat{L}_{\dot{x}}(t_0) = \hat{l}_{x_0}$. С $h(t_0) = 0$, $h(t_1) = 1$: $\hat{L}_{\dot{x}}(t_1) = -\hat{l}_{x_1}$.

Если $l \equiv 0$ (свободный конец без терминанта) — естественные граничные условия $\hat{L}_{\dot{x}} = 0$ на концах. Если один конец закреплён, на нём условие трансверсальности заменяется граничным.

ПРИМЕР

Листок 14, №1: $\int_0^1 (\dot{x}^2 - x)dt + x^2(1) \to \text{extr}$. Эйлер: $-2\ddot{x} - 1 = 0$, $x = -\frac{t^2}{4} + C_1 t + C_2$. $l = x_1^2$: $l_{x_0} = 0$, $l_{x_1} = 2x(1)$. Трансверсальность: $2\dot{x}(0) = 0 \Rightarrow C_1 = 0$; $2\dot{x}(1) = -2x(1)$: $-\frac{1}{2} = -(-\frac{1}{4} + C_2) \Rightarrow C_2 = \frac{3}{4}$. $\hat{x} = \frac{3}{4} - \frac{t^2}{4}$.
Проверка: L выпукла по $(x, \dot{x})$, l выпукла — функционал выпуклый, экстремаль даёт глобальный минимум. $\sup = +\infty$ (например $x = nt$).

О47–О50: условия 2-го порядка в КВИ

О47 Вторая вариация в простейшей задаче

Пусть $L \in C^2$ (по $x, \dot{x}$), $\hat{x}$ — экстремаль. Вторая вариация:

$$\delta^2 J(\hat{x}, h) = \frac{d^2}{d\alpha^2} J(\hat{x} + \alpha h)\Big|_{\alpha=0} = \int_{t_0}^{t_1} \left(\hat{L}_{xx} h^2 + 2\hat{L}_{x\dot{x}} h\dot{h} + \hat{L}_{\dot{x}\dot{x}} \dot{h}^2 \right) dt.$$

(Дифференцируем дважды под интегралом: $\frac{d^2}{d\alpha^2} L(t, \hat{x} + \alpha h, \dot{\hat{x}} + \alpha \dot{h})$ — квадратичная форма гессиана L по $(x, \dot{x})$ на векторе $(h, \dot{h})$.)

Преобразование. Если $\hat{L}_{x\dot{x}} \in C^1$: $\int 2\hat{L}_{x\dot{x}} h\dot{h}\, dt = \int \hat{L}_{x\dot{x}}\, d(h^2) = \left[\hat{L}_{x\dot{x}} h^2\right]_{t_0}^{t_1} - \int \frac{d}{dt}\hat{L}_{x\dot{x}}\, h^2 dt$, внеинтегральный член ноль при $h \in C_0^1$. Получаем

$$\delta^2 J(\hat{x}, h) = \int_{t_0}^{t_1} \left(P(t)\dot{h}^2 + Q(t) h^2 \right) dt, \quad P = \hat{L}_{\dot{x}\dot{x}}, \quad Q = \hat{L}_{xx} - \frac{d}{dt}\hat{L}_{x\dot{x}}.$$

Необходимое условие 2-го порядка: если $\hat{x}$ — слабый минимум, то $\delta^2 J(\hat{x}, h) \geq 0$ для всех $h \in C_0^1$ (максимум: ≤ 0).

> ***Доказательство.***
> $\varphi(\alpha) = J(\hat{x} + \alpha h)$ имеет минимум в 0, $\varphi'(0) = 0$, значит $\varphi''(0) \geq 0$ (одномерное необходимое условие 2-го порядка, О4).

О48 Лежандр, сопряжённая точка, Якоби

Условие Лежандра (для минимума): $P(t) = \hat{L}_{\dot{x}\dot{x}}(t) \geq 0$ на $[t_0, t_1]$. **Усиленное:** $P(t) > 0$ на $[t_0, t_1]$. Для максимума — ≤ 0 и < 0.

Уравнение Якоби — уравнение Эйлера для квадратичного функционала $\delta^2 J$:

$$-\frac{d}{dt}\big(P(t)\dot{h}\big) + Q(t)h = 0.$$

Сопряжённая точка. Точка $\tau \in (t_0, t_1]$ называется сопряжённой с t_0, если существует нетривиальное решение h уравнения Якоби с $h(t_0) = 0$, $h(\tau) = 0$. На практике: решаем уравнение Якоби с $h(t_0) = 0$, $\dot{h}(t_0) = 1$ и ищем нули h правее t_0.

Условие Якоби: в интервале (t_0, t_1) нет точек, сопряжённых с t_0. **Усиленное условие Якоби:** на полуинтервале $(t_0, t_1]$ нет сопряжённых точек.

ПРИМЕР
$J = \int_0^T (\dot{x}^2 - x^2)dt$: $P = 2$, $Q = -2$, уравнение Якоби $\ddot{h} + h = 0$, $h = \sin t$, первая сопряжённая точка $\tau = \pi$. При $T < \pi$ усиленное условие Якоби выполнено, при $T \geq \pi$ — нет.

О49 Необходимые условия 2-го порядка в простейшей задаче

Теорема. Пусть $L \in C^3$ (достаточная гладкость), $\hat{x}$ — слабый локальный минимум. Тогда:

1. $\hat{x}$ — экстремаль (уравнение Эйлера);
2. выполнено условие Лежандра $\hat{L}_{\dot{x}\dot{x}}(t) \geq 0$;
3. если выполнено усиленное условие Лежандра, то выполнено условие Якоби (нет сопряжённых точек в (t_0, t_1)).

Для максимума — ≤ 0 в Лежандре, Якоби — то же.

> ***Доказательство.***
> (Идея для Лежандра.) Пусть $P(\tau) < 0$. Возьмём «острый горб» h_ε, сосредоточенный на $[\tau - \varepsilon, \tau + \varepsilon]$, например $h_\varepsilon(t) = \sin^2 \frac{\pi(t-\tau)}{\varepsilon}$ на отрезке длины ε. Тогда $\int P\dot{h}_\varepsilon^2 \sim P(\tau) \cdot \frac{C}{\varepsilon} \to -\infty$, а $\int Q h_\varepsilon^2 = O(\varepsilon)$. Значит $\delta^2 J < 0$ при малом ε — противоречие с $\delta^2 J \geq 0$.
> (Идея для Якоби.) Если $\tau \in (t_0, t_1)$ сопряжена, возьмём $h =$ решение Якоби на $[t_0, \tau]$ и 0 дальше. Интегрированием по частям $\delta^2 J(h) = 0$, т.е. h — минимум неотрицательной квадратичной формы, значит должна удовлетворять уравнению Якоби на всём отрезке; но у неё излом в τ ($\dot{h}(\tau) \neq 0$) — противоречие (при $P > 0$).

О50 Достаточные условия в простейшей задаче

Теорема (условия Лежандра–Якоби). Пусть $L \in C^3$ и $\hat{x}$ — допустимая экстремаль (уравнение Эйлера и граничные условия), на которой выполнены **усиленное условие Лежандра** $\hat{L}_{\dot{x}\dot{x}} > 0$ на $[t_0, t_1]$ и **усиленное условие Якоби** (нет точек, сопряжённых с t_0, на $(t_0, t_1]$). Тогда $\hat{x}$ — строгий слабый локальный минимум. Для максимума: $\hat{L}_{\dot{x}\dot{x}} < 0$ и усиленное условие Якоби.

Идея доказательства. При этих условиях существует решение уравнения Риккати $\dot{w} = \frac{w^2}{P} - Q$ на всём $[t_0, t_1]$ (через решение Якоби без нулей: $w = -P\dot{u}/u$), и тогда $\delta^2 J = \int P\big(\dot{h} + \frac{w}{P}h\big)^2 dt > 0$ для $h \neq 0$ (дополнение до полного квадрата, приём Лежандра). Положительная определённость второй вариации плюс оценка остатка дают строгий минимум.

Достаточное условие через выпуклость (самое удобное в задачах). Если для каждого t функция $(x, \dot{x}) \mapsto L(t, x, \dot{x})$ выпукла, то любая допустимая экстремаль — **глобальный** минимум. (Для Больца — плюс выпуклость l.)

> ***Доказательство.***
> По выпуклости (О12): $L(t, \hat{x}+h, \dot{\hat{x}}+\dot{h}) - L(t, \hat{x}, \dot{\hat{x}}) \geq \hat{L}_x h + \hat{L}_{\dot{x}}\dot{h}$. Интегрируем: $J(\hat{x}+h) - J(\hat{x}) \geq \delta J(\hat{x}, h) = 0$ для всех $h \in C_0^1$.

ПРИМЕР

Листок 13, №2: $\int_0^{2\pi}(\dot{x}^2 - x^2)dt$, $x(0) = x(2\pi) = 0$. Эйлер $\ddot{x} + x = 0$, экстремали $\hat{x} = C\sin t$, на всех $J = 0$. Лежандр выполнен ($P = 2 > 0$), но сопряжённая точка $\pi \in (0, 2\pi)$ — условие Якоби нарушено, минимума нет. Действительно, $h = \sin\frac{t}{2}$: $J = \int_0^{2\pi}\left(\frac{1}{4}\cos^2\frac{t}{2} - \sin^2\frac{t}{2}\right)dt = \frac{\pi}{4} - \pi < 0$. А $h = \sin 2t$ даёт $J = 4\pi - \pi = 3\pi > 0$. Ответ: экстремума нет, J принимает любые значения.

Шпаргалка: всё на одной странице

Задача	Необходимые	Достаточные
Без ограничений	$\nabla f = 0$; min: $\nabla^2 f \succeq 0$	$\nabla f = 0$, $\nabla^2 f \succ 0$ — строгий min
Равенства $g = 0$	$\lambda_0 \nabla f + \sum \lambda_i \nabla g_i = 0$; регулярна $\Rightarrow$ $\lambda_0 = 1$; min: $d^\top \nabla^2 \mathcal{L} d \geq 0$ на $\{Jd = 0\}$	$\nabla \mathcal{L} = 0$, $d^\top \nabla^2 \mathcal{L} d > 0$ на $\{Jd = 0\} \setminus 0$
$g \leq 0$, $h = 0$ (min)	KKT: $\nabla \mathcal{L} = 0$, $\lambda \geq 0$, $\lambda_i g_i = 0$; LICQ $\Rightarrow$ $\lambda_0 = 1$	KKT + $\nabla^2 \mathcal{L} \succ 0$ на критическом конусе; или конус $= \{0\}$
ЗВП	Слейтер $\Rightarrow$ KKT с $\lambda_0 = 1$ / седловая точка	KKT $\Rightarrow$ глобальный min
Простейшая КВИ	Эйлер $-\frac{d}{dt} L_{\dot{x}} + L_x = 0$; Лежандр $L_{\dot{x}\dot{x}} \geq 0$; Якоби	Усил. Лежандр + усил. Якоби; или L выпукла по $(x, \dot{x})$
Больца	Эйлер + $L_{\dot{x}}(t_0) = l_{x_0}$, $L_{\dot{x}}(t_1) = -l_{x_1}$	L и l выпуклы $\Rightarrow$ глобальный min

Методы: градиентный $x - \alpha \nabla f$; наискорейший $\alpha = \frac{g^\top g}{g^\top A g}$; сопряжённые градиенты $d_{k+1} = -g_{k+1} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k$, $\alpha = \frac{\|g\|^2}{d^\top A d}$; Ньютон $x - [\nabla^2 f]^{-1} \nabla f$; BFGS/DFP $H_{k+1} y_k = s_k$.

Беллман: $F_k(X) = \max_u \{W_k(X, u) + F_{k+1}(\varphi_k(X, u))\}$, обратный ход, потом прямой.

Вариации: по направлению (одностор.) $\supset$ по Лагранжу (двустор., однородна) $\supset$ Гато (линейна, непрерывна) $\supset$ Фреше.

Леммы: Лагранжа $\int ah = 0 \; \forall h \in C_0^1 \Rightarrow a \equiv 0$; Дюбуа-Реймона $\int b\dot{h} = 0 \Rightarrow b \equiv \text{const}$; объединённая $\int (ah + b\dot{h}) = 0 \Rightarrow \dot{b} = a$.

Где лежит теория

Записи пар. Плейлист «Методы оптимизации ПМ 2026» на VK Видео: vkvideo.ru/playlist/177868828_6. 12 записей по 1–1,5 часа. В таблице видно, какая запись закрывает какие вопросы.

Запись	Тема	Вопросы (вводн. / основн.)
7 сентября	Задача оптимизации без ограничений	В1–В6 / О1, О4–О8
14 сентября	n-мерная задача без ограничений	В4–В6 / О2, О3, О8
18 сентября	Выпуклость, бесконечно растущие функции	В7–В12 / О3, О9–О15
21 сентября	Численные методы	В20–В23 / О23–О32
25 сентября, 28 сентября	Ограничения-равенства	В13–В16 / О16–О18
2 октября, 16 окт. (1), 16 окт. (2)	Ограничения-неравенства, выпуклое программирование	В17–В19 / О19–О22, О33–О35
6 ноября	Динамическое программирование	— / О36
9 ноября (1)	Дифференцирование в ЛНП	В24–В26 / О37–О40
9 ноября (2)	Простейшая задача вариационного исчисления	В27, В29 / О41–О44
16 ноября	Нормы, ПЗВИ, условия 2-го порядка	В24, В30 / О37, О47–О50
в плейлисте пока нет	Задача Больца, задача Лагранжа	В28, В29 / О45, О46

Листки семинаров (задачи, без теории) лежат в Загрузках: `1_задача_оптимизации_без_ограничений.pdf` … `15. Задача Лагранжа.pdf`. Примеры ниже взяты из них.

Если нужно добрать теорию по книгам: Алексеев, Тихомиров, Фомин «Оптимальное управление» (вариации по Лагранжу, задача Больца, все условия в той же терминологии); Пантелеев, Летова «Методы оптимизации в примерах и задачах» (численные методы, задачи с ограничениями, разобранные примеры); Сухарев, Тимохов, Федоров «Курс методов оптимизации»; Галеев, Тихомиров «Оптимизация: теория, примеры, задачи».

Обозначения в этом конспекте: ∇f — градиент (столбец), $\nabla^2 f$ — матрица Гессе, $\langle a, b\rangle = a^\top b$, $U_\varepsilon(x)$ — шар радиуса ε. Минимум и максимум симметричны: $\max f = -\min(-f)$, поэтому всё формулирую для минимума

и отдельно говорю, что меняется для максимума.

Конспект собран по списку вопросов Kollok-26_v1.pdf, листкам семинаров 1–15 и стандартному курсу (Алексеев–Тихомиров–Фомин, Пантелеев–Летова, Сухарев–Тимохов–Фёдоров). Формулировки могут отличаться от лекций обозначениями (например, норма в C^1 или знаки множителей для максимума) — если на лекции было иначе, говори как на лекции.