

Методы оптимизации

Ответы к дифференцированному зачёту

Конспект покрывает все 24 пункта из списка вопросов: восемь теоретических тем, восемь методов и восемь типовых задач. У каждого пункта есть короткая фраза, которой можно начать ответ, нормальный разбор, объяснение «на пальцах» и частая ошибка.

Источник списка: двухстраничный PDF «Вопросы к дифференцированному зачёту по дисциплине “Методы оптимизации”», SHA-256 b88955ebefe882413fda0b8ec672f6940a6d868c28fa5155de9abba74c0d3565.

Как пользоваться конспектом

1. Сначала выучи блок **«За 5 секунд»** — это первая фраза ответа.
2. Затем пойми блок **«На пальцах»**, чтобы не заучивать формулы вслепую.
3. Формулы и алгоритм используй, когда преподаватель попросит подробнее.
4. В задачах сначала называй предпосылки метода, потом выполняй шаги.

Во всём конспекте задача с неравенствами записывается в соглашении

$$g_i(x) \leq 0.$$

При другом направлении неравенства меняется знак соответствующего множителя ККТ.

Содержание

- Первый вопрос: теория
- Второй вопрос: методы
- Третий вопрос: типовые задачи

Первый вопрос: теория

1.1. Глобальный и локальный экстремум. Теорема Вейерштрасса

За 5 секунд

Локальный экстремум — лучший результат только в некоторой окрестности точки, глобальный — на всём допустимом множестве. Непрерывная функция на непустом компактном множестве обязательно достигает и минимума, и максимума.

Нормальный ответ

Пусть функция f задана на допустимом множестве D .

- $x^* \in D$ — точка **локального минимума**, если существует окрестность $U(x^*)$, в которой $f(x^*) \leq f(x)$ для всех $x \in D \cap U(x^*)$.
- $x^* \in D$ — точка **глобального минимума**, если $f(x^*) \leq f(x)$ для всех $x \in D$.
- Для максимума знаки неравенств меняются на противоположные.

Глобальный минимум автоматически будет и локальным, но локальный минимум не обязан быть глобальным.

Теорема Вейерштрасса: если $D \subset \mathbb{R}^n$ непусто и компактно, то то есть замкнуто и ограничено, а f непрерывна на D , то существуют точки $x_{\min}, x_{\max} \in D$, для которых

$$f(x_{\min}) = \min_{x \in D} f(x), \quad f(x_{\max}) = \max_{x \in D} f(x).$$

Теорема гарантирует **существование** экстремумов, но не говорит, где они находятся и единственны ли они. Если множество открыто или неограниченно, экстремум всё равно может существовать, но одной теоремы Вейерштрасса уже недостаточно.

На пальцах

Локальный минимум — самая низкая точка в одной яме. Глобальный минимум — дно самой глубокой ямы на всём рельефе. Вейерштрасс говорит: если рельеф непрерывен, а территория закрыта и конечна, то самая высокая и самая низкая точки где-то на ней обязательно есть.

Как отвечать или решать

1. Назвать допустимое множество D .
2. Уточнить, локальный или глобальный экстремум ищется.
3. Для существования проверить непрерывность f и компактность D .
4. После этого отдельно искать и сравнивать кандидатов.

Частая ошибка: говорить, что непрерывность сама по себе гарантирует минимум. Например, $f(x) = e^x$ непрерывна на всей прямой, но глобального минимума на $\mathbb{R}$ не достигает.

1.2. Необходимое и достаточное условия локального экстремума

За 5 секунд

Для дифференцируемой функции во внутренней точке экстремума градиент равен нулю. Положительно определённый Гессиан даёт строгий минимум, отрицательно определённый — строгий максимум.

Нормальный ответ

Необходимое условие Ферма: если x^* — внутренняя точка локального экстремума и f дифференцируема в x^* , то

$$\nabla f(x^*) = 0.$$

В одномерном случае это $f'(x^*) = 0$. Условие необходимое, но не достаточное: у функции $f(x) = x^3$ точка 0 стационарная, однако экстремума в ней нет.

Если f дважды непрерывно дифференцируема и $\nabla f(x^*) = 0$, то по Гессиану $H(x^*)$:

- $H(x^*)$ положительно определён — x^* — строгий локальный минимум;
- $H(x^*)$ отрицательно определён — строгим локальным максимумом;
- $H(x^*)$ неопределён — это седловая точка;
- $H(x^*)$ полуопределён или вырожден — тест ответа не даёт.

Для одной переменной тест упрощается:

$$f'(x^*) = 0, \quad f''(x^*) > 0 \Rightarrow \text{локальный минимум,}$$

$$f'(x^*) = 0, \quad f''(x^*) < 0 \Rightarrow \text{локальный максимум.}$$

Есть и более общее достаточное условие: если при переходе через x^* производная меняет знак с минуса на плюс, то это минимум; с плюса на минус — максимум.

На пальцах

Градиент — стрелка спуска или подъёма. На гладком дне ямы стрелка исчезает, поэтому градиент равен нулю. Но нулевая стрелка бывает и на плоском перевале. Гессиан показывает форму поверхности: «чаша вверх» — минимум, «чаша вниз» — максимум, изгиб в разные стороны — седло.

Как отвечать или решать

1. Найти все точки, где $\nabla f = 0$.
2. Не забыть точки недифференцируемости и границы области.
3. В стационарных точках исследовать Гессиан или знаки производной.
4. Для глобального ответа сравнить значения во всех кандидатах.

Частая ошибка: после $\nabla f = 0$ сразу писать «найден минимум». Это только список кандидатов.

1.3. Градиент и матрица Гессе

За 5 секунд

Градиент состоит из первых частных производных и показывает направление самого быстрого роста. Гессиан состоит из вторых производных и описывает кривизну функции.

Нормальный ответ

Для $f(x_1, \dots, x_n)$ градиент — вектор

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

Направленная производная по единичному вектору d равна

$$D_d f(x) = \nabla f(x)^T d.$$

Поэтому ∇f направлен в сторону наибольшего локального роста, а $-\nabla f$ — наибольшего локального убывания. Градиент перпендикулярен линии или поверхности уровня $f(x) = c$.

Матрица Гессе — матрица вторых частных производных:

$$H(x) = \nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]_{i,j=1}^n.$$

Для дважды непрерывно дифференцируемой функции она симметрична. Вблизи точки x при малом p работает квадратичная модель

$$f(x+p) \approx f(x) + \nabla f(x)^T p + \frac{1}{2} p^T H(x) p.$$

Градиент используют для поиска стационарных точек и построения направлений спуска. Гессиан классифицирует стационарные точки и задаёт ньютоновский шаг. Положительную определённость можно проверить по собственным значениям или по критерию Сильвестра: все угловые главные миноры положительны.

На пальцах

Градиент отвечает на вопрос «куда поверхность сейчас сильнее всего идёт вверх?». Гессиан отвечает «как быстро и в каких направлениях меняется наклон?». Первый похож на стрелку, второй — на описание формы чаши.

Мини-пример

Для $f(x, y) = x^2 + 4y^2$:

$$\nabla f = (2x, 8y)^T, \quad H = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}.$$

В точке $(0, 0)$ градиент нулевой, а оба собственных значения Гессиана положительны. Значит, это строгий локальный, а здесь и глобальный, минимум.

Частая ошибка: проверять только $\det H > 0$. В размерности выше двух этого недостаточно, а даже в двумерном случае нужен ещё знак первого диагонального элемента.

1.4. Исключение интервалов

За 5 секунд

Методы исключения интервалов ищут минимум унимодальной функции на $[a, b]$: сравнивают значения во внутренних точках и на каждой итерации отбрасывают часть интервала, где минимум находиться не может.

Нормальный ответ

Пусть f унимодальна на $[a, b]$: сначала убывает до единственной точки минимума, затем возрастает. Возьмём $a < x_1 < x_2 < b$.

- Если $f(x_1) < f(x_2)$, минимум не лежит правее x_2 , поэтому новый интервал равен $[a, x_2]$.
- Если $f(x_1) > f(x_2)$, отбрасывается левая часть и остаётся $[x_1, b]$.
- При равенстве можно оставить $[x_1, x_2]$.

На этом принципе построены дихотомия, метод золотого сечения, метод Фибоначчи, тернарный поиск и другие интервальные методы. Они не требуют производных и дают понятную оценку ошибки через длину оставшегося интервала:

$$|x_{\text{answer}} - x^*| \leq \frac{b_k - a_k}{2}$$

для ответа в середине интервала.

Главная предпосылка — унимодальность на исходном интервале. Если минимумов несколько, алгоритм может отбросить область с лучшим из них.

На пальцах

Мы знаем, что на отрезке только одна яма. Смотрим высоту в двух местах. Если справа уже выше, дальше вправо спускаться к единственному дну невозможно — эту часть выкидываем. Так зажимаем минимум всё более узкими границами.

Как применять

1. Найти или обосновать интервал унимодальности $[a, b]$.
2. Выбрать правило внутренних точек.
3. Сравнить значения и обновить только одну из границ.
4. Повторять до $b - a \leq 2\varepsilon$ или до заданного числа шагов.

Частая ошибка: запускать метод на многомодальной функции и считать найденный локальный минимум глобальным.

1.5. Аппроксимация в оптимизации

За 5 секунд

Аппроксимация заменяет сложную функцию простой локальной моделью — например, прямой или параболой. Экстремум модели даёт следующий кандидат, после чего модель строят заново.

Нормальный ответ

Аппроксимация — приближённое представление функции другой, более простой функцией. В оптимизации часто используют разложение Тейлора:

$$f(x + p) \approx f(x) + \nabla f(x)^T p$$

или квадратичную модель

$$m(p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T B p,$$

где B — точный Гессиан или его приближение.

Примеры применения:

- через три точки одномерной функции проводят параболу и берут её вершину;
- метод Ньютона минимизирует локальную квадратичную модель с $B = H(x)$;
- квазиньютоновские методы постепенно строят приближение Гессiana;
- линейная аппроксимация ограничений используется в последовательных методах.

Модель точна только локально. Поэтому кандидат надо проверить: не вышел ли он слишком далеко, уменьшил ли реальную функцию, имеет ли модель правильную кривизну. Для этого применяют поиск длины шага или доверительную область.

На пальцах

Сложную дорогу прямо под ногами заменяем простой дугой. По дуге легко угадать, где низ. Делаем шаг туда, снова смотрим на настоящую дорогу и строим новую дугу. Ошибка начинается, когда принимаем маленький локальный эскиз за карту всей местности.

Как применять

1. Выбрать точки или производные, по которым строится модель.
2. Найти минимум простой модели.
3. Проверить допустимость и качество предложенного шага.
4. Принять шаг либо уменьшить область доверия и перестроить модель.

Частая ошибка: без проверки принимать вершину параболы, оказавшуюся далеко за исходным интервалом.

1.6. Условная и безусловная оптимизация. Теорема Лагранжа

За 5 секунд

В безусловной задаче ограничений нет, а в условной решение обязано лежать на допустимом множестве. При регулярных ограничениях-равенствах в условном экстремуме градиент цели выражается линейной комбинацией градиентов ограничений.

Нормальный ответ

Безусловная задача имеет вид

$$\min_{x \in \mathbb{R}^n} f(x).$$

Условная задача с m равенствами:

$$\min f(x), \quad h_j(x) = 0, \quad j = 1, \dots, m.$$

Составим функцию Лагранжа

$$L(x, \lambda) = f(x) + \sum_{j=1}^m \lambda_j h_j(x).$$

Если x^* — локальный условный экстремум, функции дифференцируемы, а градиенты $\nabla h_j(x^*)$ линейно независимы, то существуют множители λ_j , для которых

$$\nabla_x L(x^*, \lambda) = \nabla f(x^*) + \sum_{j=1}^m \lambda_j \nabla h_j(x^*) = 0,$$

$$h_j(x^*) = 0, \quad j = 1, \dots, m.$$

Это необходимое условие: система даёт кандидатов, которые требуют классификации. Геометрически градиент цели ортогонален допустимым касательным направлениям и лежит в линейной оболочке нормалей к ограничениям.

В общей форме теоремы используют множитель λ_0 при цели и требуют, чтобы не все множители были нулевыми. При регулярности можно нормировать $\lambda_0 = 1$; без регулярности возможен аномальный случай $\lambda_0 = 0$.

На пальцах

Без ограничений можно идти куда угодно. С равенством ты обязан идти по рельсе. В условном минимуме по рельсе ниже уже не пройти, хотя в сторону от рельса склон может продолжаться. Множитель Лагранжа показывает, какой «силой» рельса уравнивается градиент цели.

Как решать

1. Перенести каждое равенство в форму $h_j(x) = 0$.
2. Составить $L = f + \sum \lambda_j h_j$.
3. Решить $\nabla_x L = 0$ вместе с $h_j = 0$.
4. Проверить допустимость и тип каждого кандидата.

Частая ошибка: решить только $\nabla_x L = 0$ и забыть добавить сами ограничения $h_j(x) = 0$.

1.7. Ограничения-равенства и неравенства. Условия ККТ

За 5 секунд

ККТ — это четыре условия: стационарность, допустимость исходной точки, неотрицательность множителей неравенств и дополняющая нежёсткость $\mu_i g_i = 0$.

Нормальный ответ

Рассмотрим задачу минимизации

$$\min f(x)$$

при

$$h_j(x) = 0, \quad g_i(x) \leq 0.$$

Функция Лагранжа:

$$L(x, \lambda, \mu) = f(x) + \sum_j \lambda_j h_j(x) + \sum_i \mu_i g_i(x).$$

Если x^* — локальный минимум и выполнено условие регулярности ограничений, то существуют λ_j и μ_i , удовлетворяющие условиям Каруша—Куна—Таккера:

1. Стационарность

$$\nabla f(x^*) + \sum_j \lambda_j \nabla h_j(x^*) + \sum_i \mu_i \nabla g_i(x^*) = 0.$$

2. Прямая допустимость

$$h_j(x^*) = 0, \quad g_i(x^*) \leq 0.$$

3. Двойственная допустимость

$$\mu_i \geq 0.$$

4. Дополняющая нежёсткость

$$\mu_i g_i(x^*) = 0.$$

Если $g_i(x^*) < 0$, ограничение неактивно и обязательно $\mu_i = 0$. Если $\mu_i > 0$, ограничение активно: $g_i(x^*) = 0$. Множители равенств λ_j могут иметь любой знак.

В общем случае ККТ необходимы, но не достаточны. Для выпуклой задачи минимизации с выпуклыми f, g_i , аффинными h_j и подходящей регулярностью, например условием Слейтера, они также достаточны для глобального минимума.

На пальцах

Неравенство — это стена, но касаться её необязательно. Если точка далеко от стены, стена на решение не давит и её множитель равен нулю. Если стена мешает идти дальше, мы стоим прямо на ней, и её множитель может быть положительным.

Как решать

1. Привести все неравенства к одному соглашению $g_i \leq 0$.
2. Перебрать согласованные наборы активных ограничений.
3. Для активных положить $g_i = 0$, для неактивных — $\mu_i = 0$.
4. Решить стационарную систему.
5. Отфильтровать кандидатов по $g_i \leq 0$, $\mu_i \geq 0$ и сравнить значения.

Частая ошибка: поменять $g_i \leq 0$ на $g_i \geq 0$, но оставить прежнее условие $\mu_i \geq 0$.

1.8. Линейное программирование и каноническая форма

За 5 секунд

В линейном программировании и цель, и ограничения линейны. Допустимое множество образует многогранник, поэтому при существовании конечного оптимума хотя бы одно оптимальное решение находится в его вершине.

Нормальный ответ

Общая задача линейного программирования имеет линейную цель

$$\max c^T x$$

и линейные равенства или неравенства. Её допустимое множество — выпуклый многогранник. Возможны четыре основных исхода: единственный оптимум, множество оптимумов на ребре или грани, неограниченная целевая функция либо пустое допустимое множество.

За каноническую форму здесь примем форму, удобную для симплекс-метода:

$$\max c^T x, \quad Ax = b, \quad x \geq 0.$$

В разных учебниках слова «каноническая» и «стандартная» могут меняться местами, поэтому на ответе полезно сразу назвать принятую запись.

Переход к ней:

- ограничение $a_i^T x \leq b_i$ превращается в равенство добавлением запасной переменной $s_i \geq 0$: $a_i^T x + s_i = b_i$;
- для $a_i^T x \geq b_i$ вычитается избыточная переменная и при необходимости вводится искусственная переменная для начального базиса;
- свободная по знаку переменная заменяется как $x_j = x_j^+ - x_j^-$, где $x_j^+, x_j^- \geq 0$;
- минимум превращается в максимум сменой знака целевой функции;
- отрицательную правую часть можно умножить вместе со всей строкой на -1 .

Графический метод перебирает вершины в двумерной задаче. Симплекс-метод переходит между соседними базисными допустимыми вершинами, пока улучшение цели невозможно.

На пальцах

Линейная функция на многоугольнике улучшается, пока мы двигаем её линию уровня к краю. Последняя точка касания обычно вершина; если линия совпала с ребром, лучшим становится всё ребро.

Как отвечать или решать

1. Привести цель к выбранному направлению, например максимуму.
2. Превратить ограничения в равенства и все переменные — в неотрицательные.
3. Найти начальный допустимый базис.
4. Решить графически при двух переменных или симплекс-методом в общем случае.

Частая ошибка: добавить запасную переменную к ограничению $\geq$ с плюсом. Для такого ограничения избыточная переменная вычитается.

Второй вопрос: методы

2.1. Метод Ферма

За 5 секунд

Метод Ферма сводит поиск внутреннего экстремума гладкой функции к поиску стационарных точек: $f'(x) = 0$ в одномерном случае или $\nabla f(x) = 0$ в многомерном. Затем кандидатов обязательно классифицируют.

Идея и формулы

Если дифференцируемая функция имеет локальный экстремум во внутренней точке, то её первая производная по любому направлению равна нулю. Отсюда

$$f'(x^*) = 0$$

для одной переменной и

$$\nabla f(x^*) = 0$$

для нескольких переменных.

Это аналитический метод получения кандидатов, а не готовое доказательство экстремума. Полный алгоритм:

1. Определить область допустимых значений.
2. Найти точки, где производная равна нулю.
3. Добавить точки, где производная не существует.
4. Если область имеет границу, добавить граничные кандидаты.
5. Классифицировать точки по знаку производной, второй производной, Гессиану или сравнению значений.

Для глобального экстремума на отрезке $[a, b]$ сравнивают значения во всех критических точках внутри отрезка и на концах a, b .

На пальцах

На гладком дне или вершине касательная горизонтальна. Метод Ферма находит все места с горизонтальной касательной. Среди них могут быть дно, вершина и плоский перегиб, поэтому потом надо посмотреть на форму функции.

Мини-пример

Для $f(x) = x^3 - 3x$:

$$f'(x) = 3x^2 - 3 = 0 \Rightarrow x = \pm 1.$$

Так как $f''(x) = 6x$, в $x = -1$ имеем максимум, а в $x = 1$ — минимум.

Частая ошибка: забыть концы отрезка. На $[0, 1]$ функция $f(x) = x$ не имеет стационарных точек, но имеет минимум в 0 и максимум в 1.

2.2. Метод золотого сечения

За 5 секунд

Золотое сечение — метод поиска минимума унимодальной функции на отрезке без производных. Интервал каждый раз сокращается в постоянной пропорции, а одно старое значение функции переиспользуется.

Алгоритм

Пусть минимум ищется на $[a, b]$, а

$$\tau = \frac{\sqrt{5} - 1}{2} \approx 0.618.$$

Внутренние точки:

$$x_1 = b - \tau(b - a), \quad x_2 = a + \tau(b - a).$$

Далее:

- если $f(x_1) \leq f(x_2)$, оставить $[a, x_2]$;
- если $f(x_1) > f(x_2)$, оставить $[x_1, b]$.

После обновления одна из внутренних точек уже находится в правильном месте нового интервала. Поэтому после двух начальных вычислений на каждую итерацию нужно только одно новое вычисление f .

После k сокращений длина интервала примерно равна

$$\ell_k = \tau^k(b - a).$$

Чтобы заранее оценить число итераций для длины не больше 2ε , можно взять

$$k \geq \frac{\ln(2\varepsilon/(b - a))}{\ln \tau}.$$

Ответом служит середина последнего интервала или лучшая из вычисленных точек.

На пальцах

Мы ставим две метки так хитро, что после отбрасывания части отрезка одна метка снова подходит для следующего шага. Поэтому не приходится каждый раз заново считать функцию в двух местах.

Что важно назвать

- функция должна быть унимодальной на исходном отрезке;
- производные не нужны;
- сходимость надёжная, но линейная;
- естественная оценка ошибки — длина оставшегося интервала.

Частая ошибка: после сокращения пересчитать оба значения функции и потерять главное преимущество метода.

2.3. Метод парабол

За 5 секунд

Метод парабол строит по трём точкам квадратичную аппроксимацию функции и берёт вершину параболы как новое приближение минимума.

Алгоритм

Выбирают три различные точки $x_1 < x_2 < x_3$, желательно так, чтобы $f(x_2) < f(x_1)$ и $f(x_2) < f(x_3)$. Через точки $(x_i, f(x_i))$ проводят параболу

$$q(x) = ax^2 + bx + c.$$

Если $a > 0$, её вершина

$$\bar{x} = -\frac{b}{2a}$$

даёт кандидата на минимум. Без явного решения для a, b, c вершину можно вычислить как

$$\bar{x} = x_2 - \frac{1}{2} \frac{(x_2 - x_1)^2(f_2 - f_3) - (x_2 - x_3)^2(f_2 - f_1)}{(x_2 - x_1)(f_2 - f_3) - (x_2 - x_3)(f_2 - f_1)},$$

где $f_i = f(x_i)$.

Затем вычисляют $f(\bar{x})$ и из четырёх точек оставляют три, которые снова окружают лучшую точку. Процесс повторяют до малости изменения x , функции или интервала.

Метод может сходиться быстрее интервальных методов, если функция рядом с минимумом хорошо похожа на параболу. Но он не гарантирует сохранение интервала: знаменатель может быть почти нулевым, парабола может открываться вниз, а её вершина — оказаться далеко от текущих точек. Практические гибриды, например метод Брента, в таких случаях переходят к золотому сечению.

На пальцах

По трём замерам рисуем простую чашу. Дно нарисованной чаши подсказывает, куда смотреть дальше. После нового замера перерисовываем чашу точнее.

Частая ошибка: принимать $\bar{x}$ без проверки, даже если парабола вырождена, направлена ветвями вниз или вершина вышла из рабочего интервала.

2.4. Метод Ньютона и квазиньютоновские методы

За 5 секунд

Ньютон минимизирует локальную квадратичную модель с точным Гессианом. Квазиньютоновские методы, например BFGS, восстанавливают кривизну по изменениям градиента и не требуют вычислять полный Гессиан.

Метод Ньютона

Для одной переменной Ньютон применяется к уравнению $f'(x) = 0$:

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

В нескольких переменных ньютоновское направление p_k находят из системы

$$H(x_k)p_k = -\nabla f(x_k),$$

после чего

$$x_{k+1} = x_k + \alpha_k p_k.$$

При хорошем старте, невырожденном положительно определённом Гессиане в минимуме и достаточной гладкости метод локально имеет квадратичную сходимость. Матрицу H^{-1} явно не вычисляют — решают линейную систему.

Вдали от минимума Ньютон может дать слишком большой шаг или направление не на спуск, если Гессиан не положительно определён. Поэтому применяют демпфирование, поиск длины шага или доверительную область.

Квазиньютоновская идея

Вместо точного Гессиана строят матрицу B_k по векторам

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

так, чтобы выполнялось секущее условие

$$B_{k+1}s_k = y_k.$$

BFGS при условии $s_k^T y_k > 0$ сохраняет положительную определённость аппроксимации и обычно имеет сверхлинейную локальную сходимость. L-BFGS хранит только несколько последних пар (s_k, y_k) и подходит для большой размерности.

На пальцах

Ньютон каждый раз строит настоящую чашу по наклону и кривизне и прыгает в её дно. BFGS начинает с грубой чаши и поправляет её по тому, как реально менялся наклон после прошлых шагов.

Частая ошибка: считать, что ньютоновский шаг автоматически ведёт к минимуму. При отрицательной кривизне он может вести к максимуму или вообще не быть направлением спуска.

2.5. Метод покоординатного спуска

За 5 секунд

Покоординатный спуск по очереди оптимизирует функцию по одной координате, фиксируя остальные, и повторяет циклы до прекращения заметного улучшения.

Алгоритм

Пусть $x = (x_1, \dots, x_n)$ и дано начальное приближение $x^{(0)}$. На одном цикле последовательно решаются одномерные задачи:

$$x_i^{(k+1)} = \arg \min_t f(x_1^{(k+1)}, \dots, x_{i-1}^{(k+1)}, t, x_{i+1}^{(k)}, \dots, x_n^{(k)}).$$

В derivative-free варианте вместо точного одномерного минимума пробуют шаги $\pm \alpha e_i$. Если улучшения нет по всем координатам, α уменьшают.

Метод прост, почти не требует памяти и хорош, когда координатные подзадачи решаются дёшево или функция разделяется по переменным. Для гладкой выпуклой задачи при корректном выборе шагов он сходится к глобальному минимуму.

Слабое место — повернутые узкие овраги: оси координат не совпадают с удобными направлениями, поэтому траектория долго идёт зигзагом. Также важны масштабы переменных и порядок обхода координат.

На пальцах

В комнате можно двигаться только сначала вдоль стены по x , потом поперёк по y , затем снова по x . До нужной точки дойдём, но если коридор расположен по диагонали, придётся много раз переставлять ноги зигзагом.

Как выполнять итерацию на рисунке

1. Из текущей точки провести линию параллельно первой координатной оси.
2. Найти на ней минимум функции и перейти туда.
3. Повторить для следующей оси.
4. Полный обход всех осей считать одним циклом.

Частая ошибка: назвать один шаг по одной координате полной итерацией, не оговорив принятую нумерацию.

2.6. Метод штрафных функций

За 5 секунд

Штрафной метод превращает условную задачу в последовательность безусловных: к цели добавляют большую цену за нарушение ограничений и постепенно усиливают эту цену.

Внешний штраф

Для задачи

$$\min f(x), \quad h_j(x) = 0, \quad g_i(x) \leq 0$$

можно построить

$$F(x, r) = f(x) + r \left(\sum_j h_j(x)^2 + \sum_i \max(0, g_i(x))^2 \right), \quad r > 0.$$

Для последовательности $r_1 < r_2 < \dots, r_k \rightarrow \infty$ минимизируют $F(x, r_k)$, каждый раз начиная с предыдущего решения. Нарушение ограничений становится всё дороже, и при подходящих условиях решения приближаются к решению исходной задачи.

Квадратичный штраф гладкий, но для высокой точности требует большого r , из-за чего задача становится плохо обусловленной. Негладкий точный штраф вида

$$f(x) + r \left(\sum_j |h_j(x)| + \sum_i \max(0, g_i(x)) \right)$$

при достаточно большом конечном r может иметь то же решение, но требует методов для негладких функций.

Есть и **внутренние барьерные методы**: например,

$$F(x, \mu) = f(x) - \mu \sum_i \ln(-g_i(x)),$$

где стартовая точка должна строго удовлетворять $g_i(x) < 0$, а $\mu \rightarrow 0$. Барьер не позволяет пересечь границу допустимой области.

На пальцах

Представь запрещённую зону. Внешний штраф разрешает временно туда зайти, но начисляет всё больший штраф. Барьер ставит у границы стену, которая становится тем круче, чем ближе мы подходим.

Алгоритм

1. Выбрать штраф, который равен нулю для допустимой точки.
2. Решить безусловную задачу при умеренном параметре.
3. Увеличить r или уменьшить μ и стартовать с прошлого решения.
4. Контролировать одновременно стационарность и нарушение ограничений.

Частая ошибка: взять сразу огромный штраф. Это не обязательно ускоряет решение и часто делает численную задачу почти вырожденной.

2.7. Simple Genetic Algorithm (SGA)

За 5 секунд

SGA хранит популяцию бинарных особей и повторяет отбор, скрещивание и мутацию: хорошие решения чаще дают потомков, скрещивание смешивает их признаки, а мутация поддерживает разнообразие.

Представление и цикл

В классическом SGA решение кодируется бинарной строкой. Если вещественная переменная $x \in [a, b]$ кодируется строкой длины L со значением z , то

$$x = a + \frac{z}{2^L - 1} (b - a).$$

Одно поколение:

1. **Инициализация:** создать N случайных особей.

2. **Оценивание:** декодировать каждую особь и вычислить целевую функцию и fitness.
3. **Отбор:** выбрать родителей с преимуществом более приспособленных, например рулеткой или турнирным отбором.
4. **Скращивание:** с вероятностью p_c обменивать частями двух строк.
5. **Мутация:** с малой вероятностью p_m инвертировать отдельные биты.
6. **Замещение:** сформировать новое поколение; элитизм сохраняет лучшее найденное решение.

Для минимизации нельзя бездумно использовать f как fitness, ведь в отборе большее fitness обычно означает лучшее решение. Нужна монотонная трансформация или ранговый отбор, причём вероятности должны быть неотрицательными.

SGA не доказывает нахождение глобального оптимума за конечный бюджет. Его сила — одновременное исследование разных областей; слабости — преждевременное схлопывание популяции, потеря лучшего решения без элитизма и зависимость от кодирования и параметров.

На пальцах

Есть много черновиков решения. Более удачные чаще становятся «родителями». Части двух черновиков смешиваются, а случайные мутации иногда добавляют новую идею. Так популяция постепенно сдвигается к хорошим областям, но гарантий, что она нашла самую лучшую, нет.

Частая ошибка: убрать мутацию или сделать отбор слишком жёстким. Тогда все особи быстро становятся похожими и поиск застревает.

2.8. Симплекс-метод для линейного программирования

За 5 секунд

Симплекс-метод идёт по базисным допустимым вершинам многогранника. На каждом шаге одна переменная входит в базис, другая выходит, а целевая функция улучшается до выполнения критерия оптимальности.

Подготовка

Задачу приводят к форме

$$\max c^T x, \quad Ax = b, \quad x \geq 0.$$

Нужно начальное базисное допустимое решение: выбрать m базисных переменных для m независимых равенств, положить небазисные переменные равными нулю и получить неотрицательные значения базисных. Если очевидного базиса нет, используют первую фазу или метод искусственного базиса.

Одна симплекс-итерация

1. Вычислить приведённые стоимости небазисных переменных.
2. Если ни одна переменная не может улучшить цель, текущий базис оптимален.
3. Выбрать **входящую переменную**, способную улучшить цель.
4. По минимальному положительному отношению

$$\frac{b_i}{a_{ij}}, \quad a_{ij} > 0,$$

выбрать **выходящую переменную**, чтобы сохранить допустимость.

5. Выполнить поворот, то есть элементарными преобразованиями сделать ведущий элемент единицей, а остальные элементы ведущего столбца — нулями.
6. Повторять до критерия оптимальности.

Знаки критерия оптимальности зависят от записи таблицы, поэтому важен не заученный знак, а смысл: ни одна небазисная переменная не должна улучшать цель.

Особые случаи:

- в ведущем столбце нет положительных элементов — цель неограничена;
- после первой фазы искусственная цель не равна нулю — задача несовместна;
- нулевое допустимое отношение — вырожденный переход;
- несколько допустимых входящих переменных с нулевой приведённой стоимостью в оптимальной таблице могут означать альтернативные оптимумы.

На пальцах

Каждый базис — вершина допустимого многогранника. Мы выбираем ребро, по которому цель становится лучше, и идём до следующей вершины. Когда ни по одному ребру улучшения нет, пришли в оптимальную вершину.

Частая ошибка: применять тест отношений и разрешать отрицательные или нулевые элементы ведущего столбца как обычные положительные кандидаты.

Третий вопрос: типовые задачи

В исходном списке для этих пунктов нет конкретных функций, ограничений и рисунков. Поэтому ниже дан универсальный порядок решения и небольшой пример. На зачёте числа из примера надо заменить данными своего варианта.

3.1. Найти экстремум методом Ферма и определить его вид

За 5 секунд

Нахожу область определения и все критические точки, решая $f'(x) = 0$, затем по знаку f'' или изменению знака f' определяю минимум и максимум.

Универсальный алгоритм

Для функции одной переменной:

1. Найти область определения.
2. Вычислить $f'(x)$.
3. Решить $f'(x) = 0$ и добавить точки, где f' не существует, но существует f .
4. Классифицировать каждую точку:
 - $f''(x^*) > 0$ — локальный минимум;
 - $f''(x^*) < 0$ — локальный максимум;
 - $f''(x^*) = 0$ — исследовать знак f' слева и справа или старшие производные.
5. Если дан отрезок, добавить его концы и сравнить значения для глобального ответа.

Для нескольких переменных вместо $f' = 0$ решают $\nabla f = 0$, а тип точки определяют по Гессиану.

Мини-пример

Пусть

$$f(x) = x^3 - 3x.$$

Тогда

$$f'(x) = 3x^2 - 3 = 3(x - 1)(x + 1),$$

поэтому критические точки $x = -1$ и $x = 1$. Вторая производная:

$$f''(x) = 6x.$$

Получаем:

- $f''(-1) = -6 < 0$ — в $x = -1$ локальный максимум, $f(-1) = 2$;
- $f''(1) = 6 > 0$ — в $x = 1$ локальный минимум, $f(1) = -2$.

На всей прямой глобальных экстремумов нет, потому что кубическая функция неограниченна сверху и снизу.

На пальцах

Сначала ищем все горизонтальные места графика. Потом смотрим, что перед нами: вершина горки, дно ямы или просто плоский перегиб.

Частая ошибка: назвать локальные точки глобальными, не проверив область и поведение функции на её границе или на бесконечности.

3.2. Найти критическую точку при ограничении методом Лагранжа

За 5 секунд

Переношу ограничение в форму $h(x) = 0$, составляю $L = f + \lambda h$ и решаю $\nabla_x L = 0$ вместе с самим ограничением.

Универсальный алгоритм

Для задачи

$$\min f(x), \quad h(x) = 0$$

выполнить:

1. Составить $L(x, \lambda) = f(x) + \lambda h(x)$.
2. Выписать систему

$$\nabla_x L(x, \lambda) = 0, \quad h(x) = 0.$$

3. Решить систему относительно x и λ .
4. Проверить исходное ограничение у каждого кандидата.
5. Определить тип условного экстремума сравнением значений, исследованием на допустимой кривой или вторым дифференциалом на касательном пространстве.

Мини-пример

Найдём минимум

$$f(x, y) = x^2 + y^2$$

при $x + y = 1$. Запишем $h(x, y) = x + y - 1$ и

$$L = x^2 + y^2 + \lambda(x + y - 1).$$

Система:

$$2x + \lambda = 0, \quad 2y + \lambda = 0, \quad x + y - 1 = 0.$$

Из первых двух уравнений $x = y$, а из ограничения $2x = 1$. Поэтому

$$x^* = y^* = \frac{1}{2}, \quad \lambda^* = -1, \quad f(x^*, y^*) = \frac{1}{2}.$$

Это глобальный условный минимум: функция строго выпукла, а допустимое множество аффинно. Максимум на бесконечной прямой $x + y = 1$ нет.

На пальцах

Мы ищем ближайшую к началу координат точку на прямой $x + y = 1$. В точке касания окружности с прямой их нормали параллельны — именно это и записывает множитель Лагранжа.

Частая ошибка: остановиться на решении системы и не сказать, минимум это, максимум или только критическая точка.

3.3. Найти критическую точку по условиям ККТ

За 5 секунд

Привожу ограничения к $g_i \leq 0$, выписываю стационарность, прямую и двойственную допустимость и $\mu_i g_i = 0$, затем разбираю активные и неактивные ограничения.

Универсальный алгоритм

Для

$$\min f(x), \quad h_j(x) = 0, \quad g_i(x) \leq 0$$

нужно решить систему

$$\nabla f + \sum_j \lambda_j \nabla h_j + \sum_i \mu_i \nabla g_i = 0,$$

$$h_j = 0, \quad g_i \leq 0, \quad \mu_i \geq 0, \quad \mu_i g_i = 0.$$

Практически для каждого неравенства рассматривают два случая:

- оно неактивно: $g_i < 0$ и $\mu_i = 0$;
- оно активно: $g_i = 0$ и после решения требуется $\mu_i \geq 0$.

Все комбинации фильтруют по исходным ограничениям. Если задача выпуклая, точка, прошедшая ККТ, даёт глобальный минимум; иначе это только кандидат.

Мини-пример

Минимизируем

$$f(x) = (x - 2)^2$$

при $x \leq 1$. Здесь $g(x) = x - 1 \leq 0$ и

$$L = (x - 2)^2 + \mu(x - 1).$$

ККТ:

$$2(x - 2) + \mu = 0, \quad x - 1 \leq 0, \quad \mu \geq 0, \quad \mu(x - 1) = 0.$$

Разберём случаи.

1. Если ограничение неактивно, то $\mu = 0$ и из стационарности $x = 2$. Но $2 \leq 1$ неверно, кандидат недопустим.
2. Если ограничение активно, то $x = 1$. Из стационарности $2(1 - 2) + \mu = 0$, значит $\mu = 2 \geq 0$.

Ответ: $x^* = 1$, $f(x^*) = 1$. Функция выпукла и ограничение линейно, поэтому это глобальный минимум.

На пальцах

Без стены функция хотела бы прийти в $x = 2$, но разрешено только $x \leq 1$. Поэтому решение прижимается к стене $x = 1$, а положительный множитель показывает, что стена действительно мешает.

Частая ошибка: объявить ограничение активным заранее и не проверить знак полученного множителя.

3.4. Решить задачу линейного программирования графически

За 5 секунд

Строю полуплоскости ограничений, нахожу допустимый многоугольник, выписываю его вершины и сравниваю в них значение линейной целевой функции.

Универсальный алгоритм

1. Каждое неравенство временно заменить равенством и построить граничную прямую.
2. Проверочной точкой, часто $(0, 0)$, определить нужную полуплоскость.
3. Пересечь все полуплоскости и получить допустимую область.
4. Найти координаты всех её вершин как пересечения пар граничных прямых.
5. Подставить вершины в цель либо двигать параллельную линию уровня $c_1x + c_2y = C$ в направлении улучшения.

6. Указать оптимальную точку и значение цели, а также проверить особые случаи: пустую область, неограниченность и целое оптимальное ребро.

Мини-пример

Решим

$$\max z = 3x + 2y$$

при

$$x + y \leq 4, \quad x \leq 2, \quad y \leq 3, \quad x, y \geq 0.$$

Допустимый многоугольник имеет вершины

$$(0, 0), (2, 0), (2, 2), (1, 3), (0, 3).$$

Значения цели соответственно равны

$$0, 6, 10, 9, 6.$$

Следовательно,

$$(x^*, y^*) = (2, 2), \quad z_{\max} = 10.$$

На пальцах

Ограничения вырезают разрешённый многоугольник. Прямая целевой функции двигается параллельно самой себе, пока в последний раз касается этого многоугольника. Точка последнего касания и есть оптимум.

Частая ошибка: проверять пересечение двух прямых, не убедившись, что оно удовлетворяет всем остальным ограничениям.

3.5. Три итерации метода дихотомии на графике

За 5 секунд

На каждой итерации беру две точки по разные стороны от середины интервала, сравниваю высоты графика и отбрасываю половину, противоположную меньшему значению.

Универсальный алгоритм

Пусть минимум унимодальной функции ищется на $[a_k, b_k]$. Выбрать малое $\delta > 0$ и построить

$$m_k = \frac{a_k + b_k}{2}, \quad x_1 = m_k - \delta, \quad x_2 = m_k + \delta.$$

Далее:

$$f(x_1) < f(x_2) \Rightarrow [a_{k+1}, b_{k+1}] = [a_k, x_2],$$

$$f(x_1) > f(x_2) \Rightarrow [a_{k+1}, b_{k+1}] = [x_1, b_k].$$

При равенстве можно оставить $[x_1, x_2]$. На рисунке после каждого сравнения надо зачеркнуть отброшенную часть оси и подписать новый интервал.

Мини-пример: ровно три итерации

Для $f(x) = (x - 2)^2$, $[a_0, b_0] = [0, 5]$ и $\delta = 0.1$:

Итерация	m	$x_1, f(x_1)$	$x_2, f(x_2)$	Новый интервал
1	2.5	2.4, 0.16	2.6, 0.36	$[0, 2.6]$
2	1.3	1.2, 0.64	1.4, 0.36	$[1.2, 2.6]$
3	1.9	1.8, 0.04	2.0, 0	$[1.8, 2.6]$

После трёх итераций минимум гарантированно остаётся в $[1.8, 2.6]$.

На пальцах

Две точки около центра показывают, в какую сторону график идёт вниз. Половину, где он уже идёт вверх от единственного дна, можно выкинуть.

Частая ошибка: поставить обе пробные точки точно в середину. Они должны быть по разные стороны от неё: $m - \delta$ и $m + \delta$.

3.6. Три итерации градиентного спуска на линиях уровня

За 5 секунд

В каждой точке градиент перпендикулярен линии уровня и смотрит вверх, поэтому для минимума делаю шаг в направлении $-\nabla f$.

Универсальный алгоритм

При начальной точке x_0 и выбранном шаге $\alpha_k > 0$:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k).$$

На рисунке:

1. В точке x_k провести нормаль к текущей линии уровня.
2. Выбрать из двух направлений то, где значения функции уменьшаются.
3. Отложить шаг заданной или выбранной длины и получить x_{k+1} .
4. Повторить три раза, подписав x_0, x_1, x_2, x_3 .

Если длина шага не дана, нужно явно назвать свой выбор. Слишком большой шаг может перескакивать через долину и расходиться, слишком маленький — сходиться медленно.

Мини-пример: ровно три итерации

Пусть

$$f(x, y) = x^2 + 2y^2, \quad \nabla f = (2x, 4y)^T, \quad \alpha = 0.2, \quad x_0 = (2, 1).$$

Тогда

$$x_1 = (2, 1) - 0.2(4, 4) = (1.2, 0.2),$$

$$x_2 = (1.2, 0.2) - 0.2(2.4, 0.8) = (0.72, 0.04),$$

$$x_3 = (0.72, 0.04) - 0.2(1.44, 0.16) = (0.432, 0.008).$$

Точки движутся внутрь эллиптических линий уровня к минимуму $(0, 0)$.

На пальцах

Линия уровня похожа на горизонталь на карте. Градиент показывает самый крутой подъём поперёк горизонтали, а минус-градиент — самый крутой спуск.

Частая ошибка: идти по $+\nabla f$ при поиске минимума или рисовать градиент вдоль линии уровня.

3.7. Три итерации метода Нелдера—Мида на линиях уровня

За 5 секунд

В двумерной задаче Нелдер—Мид двигает треугольник: сортирует вершины, отражает худшую относительно середины остальных, а затем при необходимости растягивает, сжимает или уменьшает весь треугольник.

Операции одной итерации

В $\mathbb{R}^2$ нужны три вершины. Отсортируем их по значениям функции: x_l — лучшая, x_g — средняя, x_h — худшая. Центр двух лучших:

$$c = \frac{x_l + x_g}{2}.$$

Стандартные коэффициенты: отражение $\alpha = 1$, растяжение $\beta = 2$, сжатие $\gamma = 1/2$, редукция $\delta = 1/2$.

- **Отражение:** $x_r = c + \alpha(c - x_h)$.
- Если x_r лучше лучшей, проверить **растяжение** $x_e = c + \beta(x_r - c)$ и оставить лучшую из x_e, x_r .
- Если отражение недостаточно хорошо, выполнить внешнее или внутреннее **сжатие** к центру.
- Если сжатие не улучшило ситуацию, сделать **редукцию**: каждую нелучшую вершину сдвинуть к лучшей в δ раз.

На графике после каждой операции надо подписывать новую тройку вершин, а не рисовать сразу конечный треугольник.

Мини-пример: ровно три итерации

Пусть $f(x, y) = x^2 + y^2$ и начальные вершины

$$A = (1, 0), \quad f(A) = 1; \quad B = (0, 2), \quad f(B) = 4; \quad C = (3, 0), \quad f(C) = 9.$$

1. **Итерация 1.** Худшая точка C . Центр A, B равен $c = (0.5, 1)$. Отражение даёт $R = 2c - C = (-2, 2)$, $f(R) = 8$. Это лучше худшей точки, поэтому заменить C на R .
2. **Итерация 2.** Для тройки A, B, R худшая точка R . Отражение возвращает $(3, 0)$ со значением 9, то есть не помогает. Внутреннее сжатие даёт $D = c + \frac{1}{2}(R - c) = (-0.75, 1.5)$, $f(D) = 2.8125$. Заменить R на D .
3. **Итерация 3.** Для тройки A, D, B худшая точка B . Центр A, D равен $c = (0.125, 0.75)$. Отражение даёт $R_3 = (0.25, -0.5)$, $f(R_3) = 0.3125$, что лучше текущей лучшей точки. Растяжение даёт $E = (0.375, -1.75)$, $f(E) = 3.203125$, поэтому оставить отражённую точку R_3 , а не E .

После трёх итераций треугольник состоит из A, D, R_3 , а лучшая вершина — R_3 .

На пальцах

Треугольник лежит на рельефе. Мы каждый раз перекидываем его худший угол через середину двух лучших. Если там очень хорошо — тянемся дальше; если плохо — сжимаемся.

Частая ошибка: считать центр по всем трём вершинам. В операции отражения центр строится без худшей вершины.

3.8. Две итерации Bees Algorithm на графике функции

За 5 секунд

Пчелиный алгоритм совмещает глобальный и локальный поиск: разведчики находят участки, лучшие участки получают больше пчёл для поиска по соседству, а часть разведчиков снова случайно исследует всю область.

Одна итерация

Выберем параметры:

- S — всего разведчиков;
- B — число элитных участков;
- G — число остальных хороших участков;
- b и g — число завербованных пчёл вокруг элитного и хорошего участка;
- r — радиус локальной окрестности.

Далее:

1. Оценить всех S разведчиков и отсортировать их.
2. Выбрать B элитных и G других хороших участков.
3. Сгенерировать b соседей каждого элитного участка и g соседей каждого хорошего.
4. Из каждого участка оставить только одну лучшую точку.
5. Остальные $S - B - G$ места заполнить новыми глобальными разведчиками.
6. При необходимости уменьшить r и повторить.

Мини-пример: ровно две итерации

Минимизируем $f(x) = (x - 2)^2$ на $[0, 6]$. Возьмём

$$S = 5, \quad B = 1, \quad G = 1, \quad b = 2, \quad g = 1, \quad r_1 = 0.5.$$

Начальные разведчики стоят в точках 0, 1.5, 3, 4.5, 6. Их значения: 4, 0.25, 1, 6.25, 16.

- Итерация 1.** Элитный участок — 1.5, хороший — 3. Вокруг 1.5 пусть проверены точки 1.8 и 2.0; лучшая на участке — 2.0 со значением 0. Вокруг 3 проверена точка 2.6 со значением 0.36, она лучше центра участка. Три новых глобальных разведчика, например, попали в 0.5, 4, 5.5. Новая популяция: 2.0, 2.6, 0.5, 4, 5.5.
- Итерация 2.** Элитный участок — 2.0, хороший — 2.6. Уменьшим радиус до $r_2 = 0.25$. Вокруг 2.0 проверим 1.9 и 2.1, но сохраним центр 2.0, потому что его значение 0 лучше. Вокруг 2.6 точка 2.35 со значением 0.1225 становится новым представителем участка. Остальные три места снова заполняются глобальными разведчиками.

После двух итераций лучшее найденное решение — $x = 2$, $f(x) = 0$.

На пальцах

Сначала пчёлы широко осматривают поле. Там, где найдено много нектара, вокруг ищут плотнее; но несколько пчёл всё равно летят далеко, чтобы вся группа не застряла на одном неплохом цветке.

Частая ошибка: отправить всех пчёл к текущей лучшей точке. Тогда исчезает глобальная разведка и алгоритм легко застревает в локальном минимуме.

Формулы, которые стоит помнить без подсказки

$$\nabla f(x^*) = 0,$$

$$H \succ 0 \Rightarrow \min, \quad H \prec 0 \Rightarrow \max,$$

$$\nabla f + \sum_j \lambda_j \nabla h_j = 0, \quad h_j = 0,$$

$$\nabla f + \sum_j \lambda_j \nabla h_j + \sum_i \mu_i \nabla g_i = 0, \quad g_i \leq 0, \quad \mu_i \geq 0, \quad \mu_i g_i = 0,$$

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k),$$

$$H(x_k)p_k = -\nabla f(x_k), \quad x_{k+1} = x_k + p_k.$$

Главная логика курса: сначала получить полный набор кандидатов, затем проверить допустимость и только потом доказать вид и глобальность экстремума.