НАО «Восточно-Казахстанский технический университет имени Д. Серикбаева»

Дисциплина «Методы оптимизации»

ЛЕКЦИЯ 10

Многомерная оптимизация и геометрия целевой функции

Образовательная программа 6B06103 Математическое и компьютерное моделирование

Объём: 2 академических часа. **Место в рабочей программе:** лекция 10.

Цель: научиться формулировать задачу оптимизации нескольких переменных, читать поверхности и линии уровня, различать виды экстремума и выбирать класс численных методов по свойствам задачи.

План занятия: постановка задачи; допустимое множество; график и линии уровня; локальный и глобальный минимум; выпуклость; геометрия квадратичной функции; ограничения и масштаб переменных; классификация методов; связь с одномерным поиском; вычислительный эксперимент.

Результаты обучения: студент объясняет геометрический смысл решения, проверяет его допустимость, строит линии уровня, различает седло и минимум, определяет используемую методом информацию и выполняет сеточный эксперимент с оценкой его ограничений.

1 От одной переменной к вектору параметров

В предыдущих лекциях положение точки задавалось одним числом. Во многих приложениях требуется одновременно выбрать несколько параметров: размеры конструкции, коэффициенты математической модели, объёмы выпуска продукции или параметры управления. Их удобно объединить в вектор

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^{\mathsf{T}} \in \mathbb{R}^n.$$

Целевая функция $f\colon \mathbb{R}^n \to \mathbb{R}$ каждому допустимому набору параметров сопоставляет одно число — затраты, ошибку, расход ресурсов или другой показатель качества. Задача минимизации записывается в виде [1]

$$\min_{\mathbf{x}\in D} f(\mathbf{x}), \qquad D \subseteq \mathbb{R}^n.$$

Здесь D — множество допустимых решений. Требуется найти точку $\mathbf{x}^* \in D$ и значение $f^* = f(\mathbf{x}^*)$. Это разные объекты: координаты описывают выбранные параметры, а значение функции — качество результата. В отчёте следует указывать оба объекта.

При $D = \mathbb{R}^n$ рассматривается безусловная задача. Если D задано ограничениями, задача условная. Область определения функции также должна учитываться: выражение с логарифмом или делением нельзя вычислять в произвольной точке пространства.

Типичная форма допустимого множества имеет вид

$$D = \{\mathbf{x}\colon g_i(\mathbf{x}) \le 0,\ i = 1, \dots, m;\ h_j(\mathbf{x}) = 0,\ j = 1, \dots, r\}.$$

Задача максимизации сводится к минимизации функции $-f$. При этом точка решения сохраняется, а оптимальное значение меняет знак. Число переменных n называют размерностью задачи; число ограничений не является её размерностью.

Пример постановки. Пусть нужно подобрать два параметра модели так, чтобы уменьшить взвешенную сумму квадратов отклонений:

$$f(x_1, x_2) = \sum_{i=1}^{m} w_i \, (a_i x_1 + b_i x_2 - y_i)^2, \qquad w_i > 0.$$

Переменные здесь x_1, x_2, а величины a_i, b_i, y_i, w_i заданы. Если параметры должны быть неотрицательными, условия $x_1 \geq 0$, $x_2 \geq 0$ включаются в D. Наличие суммы квадратов гарантирует неотрицательность функции, но само по себе не гарантирует единственность решения: разные наборы параметров могут давать одинаковые предсказания.

2 Поверхность и линии уровня

Для двух переменных график функции — поверхность в трёхмерном пространстве:

$$S = \{(x_1, x_2, z) : z = f(x_1, x_2)\}.$$

Высота поверхности показывает значение целевой функции. Поиск минимума можно представить как поиск самой низкой допустимой точки поверхности. Если ограничено множество D, рассматривается только часть поверхности над этим множеством.

Для фиксированного числа γ множество уровня определяется равенством

$$L_\gamma = \{(x_1, x_2) : f(x_1, x_2) = \gamma\}.$$

Геометрически поверхность пересекают горизонтальной плоскостью $z = \gamma$, после чего проецируют полученное пересечение на плоскость переменных. В обычном невырожденном случае получаются линии уровня. При особом значении γ множество может быть одной точкой; оно также может быть пустым или состоять из нескольких несвязных частей.

Рассмотрим функцию, которая будет использоваться в нескольких примерах лекции:

$$f(x_1, x_2) = 2(x_1 - 1)^2 + 2(x_1 - 1)(x_2 + 1) + 2(x_2 + 1)^2 + 3.$$

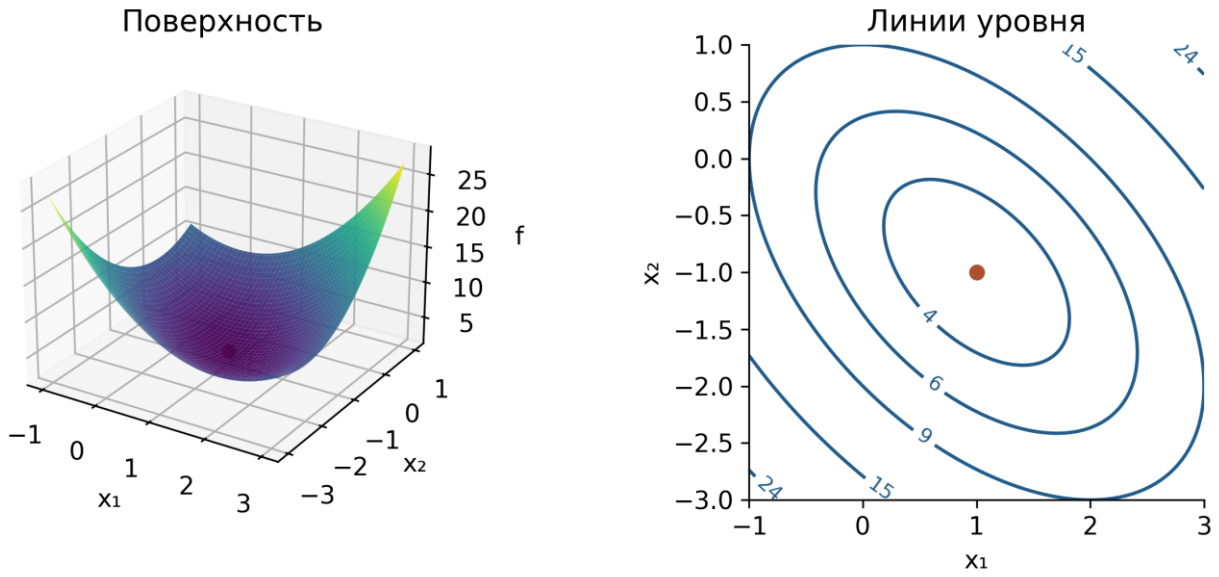


Рисунок 1 — Два способа изображения одной целевой функции

На поверхности минимум виден как нижняя точка чаши. На плоскости он расположен в центре вложенных эллипсов. Числа у линий обозначают значения f, а не координаты. При приближении к центру значения уменьшаются до 3.

Подуровневое множество включает все точки со значением функции не выше заданного порога:

$$S_\gamma = \{\mathbf{x}: f(\mathbf{x}) \le \gamma\}.$$

Для рассматриваемой функции при $\gamma > 3$ это заполненный эллипс, а L_γ — его граница. При $\gamma = 3$ оба множества состоят из точки $(1; -1)$; при $\gamma < 3$ они пусты. Не следует смешивать равенство уровня и неравенство подуровня.

Как читать рисунок. Сначала проверьте подписи осей и значения уровней. Затем определите, в какую сторону значения убывают, где расположено допустимое множество и какие линии его пересекают. Для сравнения формы эллипсов нужен одинаковый масштаб единицы по обеим осям. Изображение, растянутое по одной оси, искажает геометрическое представление о задаче.

Если уровни выбраны с одинаковым приращением значения функции, близко расположенные линии обычно соответствуют более быстрому изменению функции в поперечном направлении. При неравномерном выборе уровней делать такой вывод по одним расстояниям нельзя.

3 Локальный и глобальный минимум

Точка $\mathbf{x}^* \in D$ является **глобальным минимумом**, если

$$f(\mathbf{x}^*) \le f(\mathbf{x}) \quad \text{для всех } \mathbf{x} \in D.$$

Она является **локальным минимумом**, если существует $\delta > 0$, для которого

$$f(\mathbf{x}^*) \le f(\mathbf{x}) \quad \text{при } \mathbf{x} \in D,\ \| \mathbf{x} - \mathbf{x}^* \|_2 < \delta.$$

Евклидова норма задаёт расстояние между точками:

$$\| \mathbf{x} - \mathbf{y} \|_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

В локальном определении сравниваются только близкие допустимые точки. Глобальный минимум всегда локальный, но локальный минимум может уступать другой, удалённой точке. Если неравенство строго для всех отличных от $\mathbf{x}^*$ точек соответствующей области сравнения, минимум называют строгим.

Несколько решений. Для функции

$$h(x_1, x_2) = (x_1^2 - 1)^2 + x_2^2$$

каждое слагаемое неотрицательно. Значение ноль достигается ровно в точках $(1; 0)$ и $(-1; 0)$. Поэтому обе точки — глобальные минимумы, и каждая является строгим локальным минимумом. Ни одна из них не является строгим глобальным минимумом: существует другая точка с тем же значением.

В точке $(0; 0)$ значение равно 1. Вдоль оси x_2 имеем $h(0, t) = 1 + t^2 > 1$ при $t \neq 0$, а вдоль оси x_1

$$h(t, 0) = 1 - 2t^2 + t^4 < 1 \quad \text{при } 0 < |t| < \sqrt{2}.$$

Следовательно, в любой достаточно малой окрестности есть как большие, так и меньшие значения. Точка $(0; 0)$ седловая и не является экстремумом.

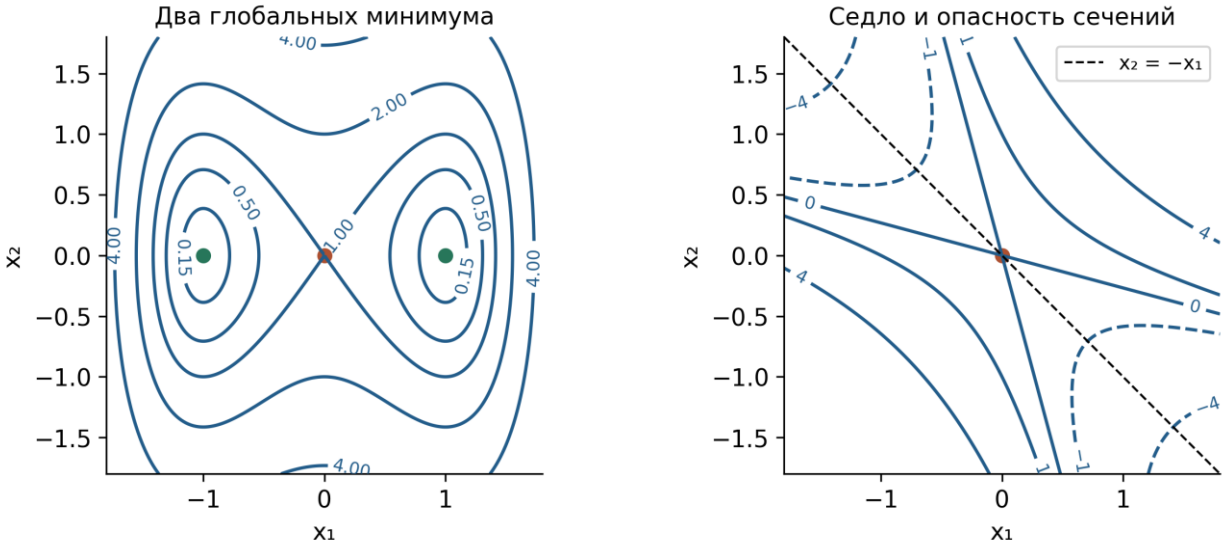


Рисунок 2 — Несколько минимумов и седловая геометрия

Почему недостаточно проверить координатные сечения. Рассмотрим другую функцию

$$q(x_1, x_2) = x_1^2 + 4x_1x_2 + x_2^2.$$

Вдоль каждой координатной оси функция положительна вне начала координат: $q(t, 0) = q(0, t) = t^2$. Однако на диагонали $x_2 = -x_1$ получаем $q(t, -t) = -2t^2 < 0$. Значит, начало координат — седло, хотя оба координатных сечения имеют в нём минимум. Проверка нескольких направлений не заменяет анализ всей окрестности.

4 Существование решения и роль выпуклости

Прежде чем запускать алгоритм, нужно понять, существует ли искомый минимум. Непрерывная функция достигает минимума на непустом компактном множестве. В конечномерном пространстве компактность означает замкнутость и ограниченность.

Ограниченность функции снизу ещё не означает достижение нижней грани. Например, $f(x_1, x_2) = e^{x_1} + x_2^2$ положительна на $\mathbb{R}^2$. При $x_1 \to -\infty$ и $x_2 = 0$ значения стремятся к нулю, но ни в одной конечной точке не равны нулю. Нижняя грань существует, точки минимума нет.

Другой достаточный случай — непрерывная функция на $\mathbb{R}^n$, для которой $f(\mathbf{x}) \to +\infty$ при $\|\mathbf{x}\|_2 \to \infty$. Такое свойство называют коэрцитивностью. Оно не позволяет минимизирующим точкам уходить бесконечно далеко; поиск можно ограничить подходящим компактным подуровневым множеством.

Множество D выпукло, если вместе с любыми двумя точками содержит соединяющий их отрезок:

$$\mathbf{x}, \mathbf{y} \in D, \quad 0 \le \theta \le 1 \;\Rightarrow\; \theta\mathbf{x} + (1-\theta)\mathbf{y} \in D.$$

Функция выпукла на выпуклом множестве D, если

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \le \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}).$$

Геометрически график вдоль любого отрезка лежит не выше хорды, соединяющей его концы. Для строго выпуклой функции неравенство строго при разных концах и $0 < \theta < 1$.

Следствие. Для выпуклой функции на выпуклом множестве любой локальный минимум глобален. Действительно, если нашлась бы допустимая точка $\mathbf{y}$ с меньшим значением, то точки $\mathbf{x}_\theta = (1-\theta)\mathbf{x}^* + \theta\mathbf{y}$ при малом положительном θ лежали бы сколь угодно близко к $\mathbf{x}^*$ и удовлетворяли бы

$$f(\mathbf{x}_\theta) \le (1-\theta) f(\mathbf{x}^*) + \theta f(\mathbf{y}) < f(\mathbf{x}^*).$$

Это противоречит локальной минимальности. У строго выпуклой функции на выпуклом множестве может быть не более одной точки минимума. Существование решения при этом нужно обосновать отдельно.

Выпуклость не следует устанавливать только по рисунку на ограниченном окне. Он может не показывать удалённые особенности функции. В следующих лекциях появятся дифференциальные признаки; в квадратичном примере ниже достаточно алгебраических преобразований.

5 Подробный разбор квадратичной функции

Для функции из раздела 2 введём смещённые координаты $u = x_1 - 1$, $v = x_2 + 1$. Тогда

$$f = 2u^2 + 2uv + 2v^2 + 3 = (u+v)^2 + u^2 + v^2 + 3.$$

Все три квадрата неотрицательны и одновременно равны нулю только при $u = v = 0$. Поэтому минимум единственный и глобальный:

$$\mathbf{x}^* = (1; -1)^{\mathsf{T}}, \qquad f^* = 3.$$

В матричной записи

$$f(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^*)^{\mathsf{T}} Q (\mathbf{x} - \mathbf{x}^*) + 3, \qquad Q = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Ортогональная замена координат позволяет определить форму линий уровня. Введём

$$w = \frac{u+v}{\sqrt{2}}, \qquad z = \frac{u-v}{\sqrt{2}}.$$

Обратные формулы имеют вид $u = (w+z)/\sqrt{2}$, $v = (w-z)/\sqrt{2}$. Подстановка даёт

$$f = 3w^2 + z^2 + 3.$$

Поэтому при уровне $\gamma > 3$

$$\frac{w^2}{(\gamma-3)/3} + \frac{z^2}{\gamma-3} = 1.$$

Короткая полуось равна $\sqrt{(\gamma-3)/3}$ и направлена вдоль вектора $(1;1)$. Длинная полуось равна $\sqrt{\gamma-3}$ и направлена вдоль $(1;-1)$. На уровне $\gamma = 6$ их длины равны соответственно 1 и $\sqrt{3}$. Перекрёстный член $2uv$ поворачивает эллипсы относительно исходных координатных осей.

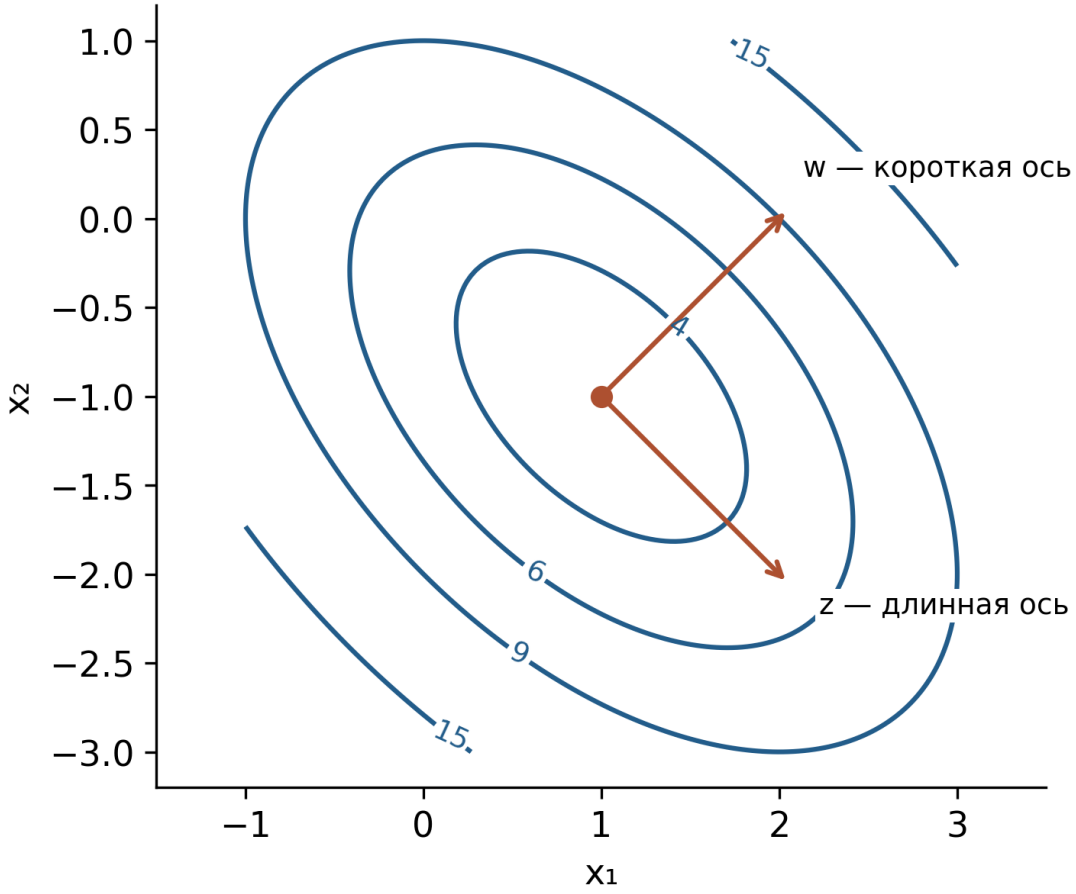


Рисунок 3 — Главные направления линий уровня

Матрица Q имеет собственные значения 1 и 3. Здесь не требуется заранее знать спектральную теорему: эти коэффициенты уже получены прямой заменой координат. Они показывают, насколько быстро функция увеличивается вдоль двух взаимно перпендикулярных направлений.

Поскольку эта замена сохраняет расстояние, $w^2 + z^2 = u^2 + v^2$. Получаем полезные оценки:

$$\| \mathbf{x} - \mathbf{x}^* \|_2^2 \leq f(\mathbf{x}) - 3 \leq 3 \| \mathbf{x} - \mathbf{x}^* \|_2^2.$$

Например, если известно, что $f(\mathbf{x}) - 3 \le 10^{-4}$, то расстояние до минимума не превышает 10^{-2}. Малый разрыв по функции и малое расстояние — разные показатели, связанные здесь конкретным неравенством. Нельзя переносить эту оценку на произвольную функцию без дополнительных условий.

6 Как ограничения изменяют решение

Рассмотрим задачу

$$F(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 1)^2 \to \min,$$

$$D = \{(x_1, x_2): x_1 \ge 0,\ x_2 \ge 0,\ x_1 + x_2 \le 2\}.$$

Без ограничений минимум достигается в точке $(2; 1)$ и равен нулю. Эта точка недопустима, поскольку $2 + 1 > 2$. Допустимое множество — заполненный треугольник с вершинами $(0; 0)$, $(2; 0)$, $(0; 2)$.

Целевая функция равна квадрату расстояния до точки $(2; 1)$. Следовательно, нужно найти ближайшую к ней точку треугольника. Рассмотрим наклонную сторону: $x_2 = 2 - x_1$, $0 \le x_1 \le 2$. На ней

$$F(x_1, 2 - x_1) = (x_1 - 2)^2 + (1 - x_1)^2 = 2\left(x_1 - \frac{3}{2}\right)^2 + \frac{1}{2}.$$

Кандидат на минимум — $(3/2; 1/2)$ со значением $1/2$. Чтобы доказать глобальность на всём треугольнике, положим $\mathbf{p} = (3/2; 1/2)$, $\mathbf{a} = (2; 1)$. Для любой $\mathbf{x} \in D$

$$\| \mathbf{x} - \mathbf{a} \|_2^2 = \| \mathbf{x} - \mathbf{p} \|_2^2 + \| \mathbf{p} - \mathbf{a} \|_2^2 + 2(\mathbf{x} - \mathbf{p})^{\mathsf{T}}(\mathbf{p} - \mathbf{a}).$$

Последний член равен $2 - x_1 - x_2 \ge 0$, а $\| \mathbf{p} - \mathbf{a} \|_2^2 = 1/2$. Поэтому $F(\mathbf{x}) \ge 1/2$, причём равенство возможно только при $\mathbf{x} = \mathbf{p}$. Решение доказано для всего допустимого множества.

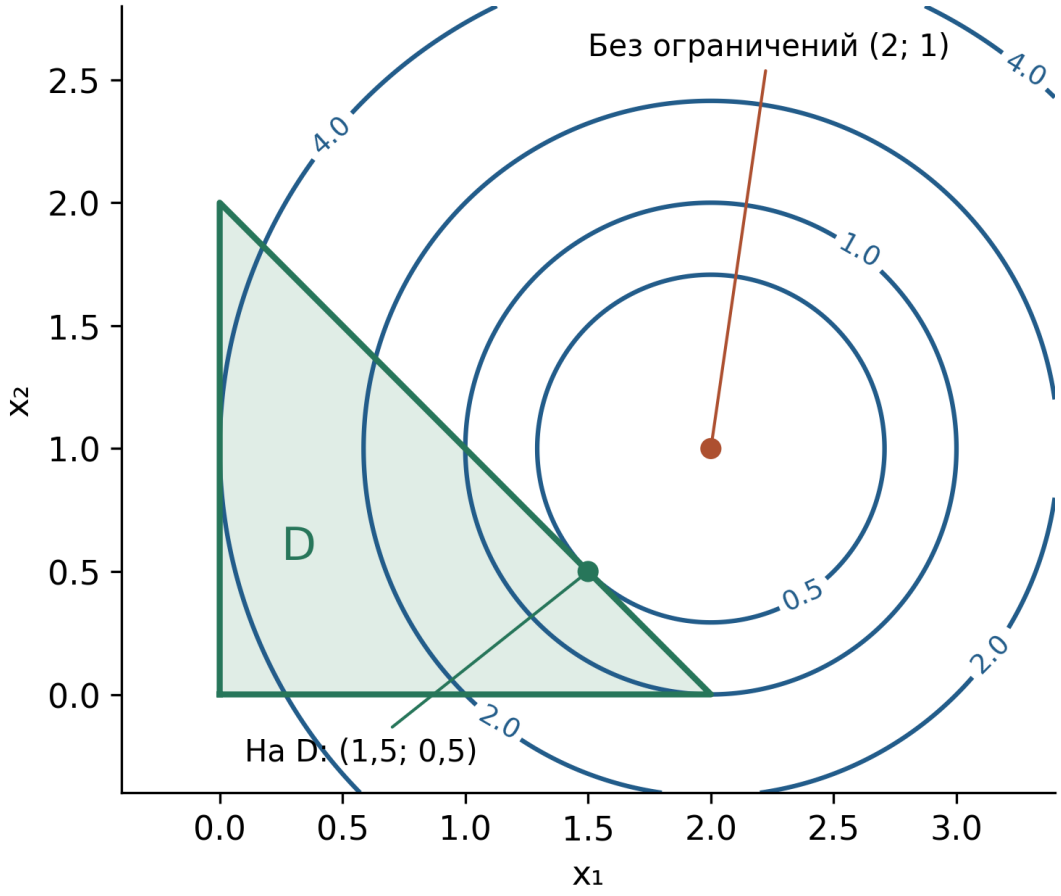


Рисунок 4 — Ограничения переносят минимум на границу допустимой области

Самая низкая линия уровня, имеющая общую точку с треугольником, касается его наклонной стороны. Значения функции меньше 1/2 существуют на плоскости, но соответствуют недопустимым точкам. Поэтому условие допустимости проверяют вместе со значением функции, а не после выбора лучшего результата.

7 Масштаб переменных и вытянутость уровней

Рассмотрим функцию $f(u, v) = u^2 + 25v^2$. Линии уровня — эллипсы с отношением длин полуосей 5. Одинаковое изменение u и v по-разному влияет на значение функции: во втором направлении рост существенно быстрее.

После замены $z_1 = u, z_2 = 5v$ получаем $f = z_1^2 + z_2^2$. В новых координатах линии уровня — окружности. Физическая задача сохраняется при условии, что вместе с переменными правильно преобразованы ограничения и затем выполнен обратный переход к исходным параметрам.

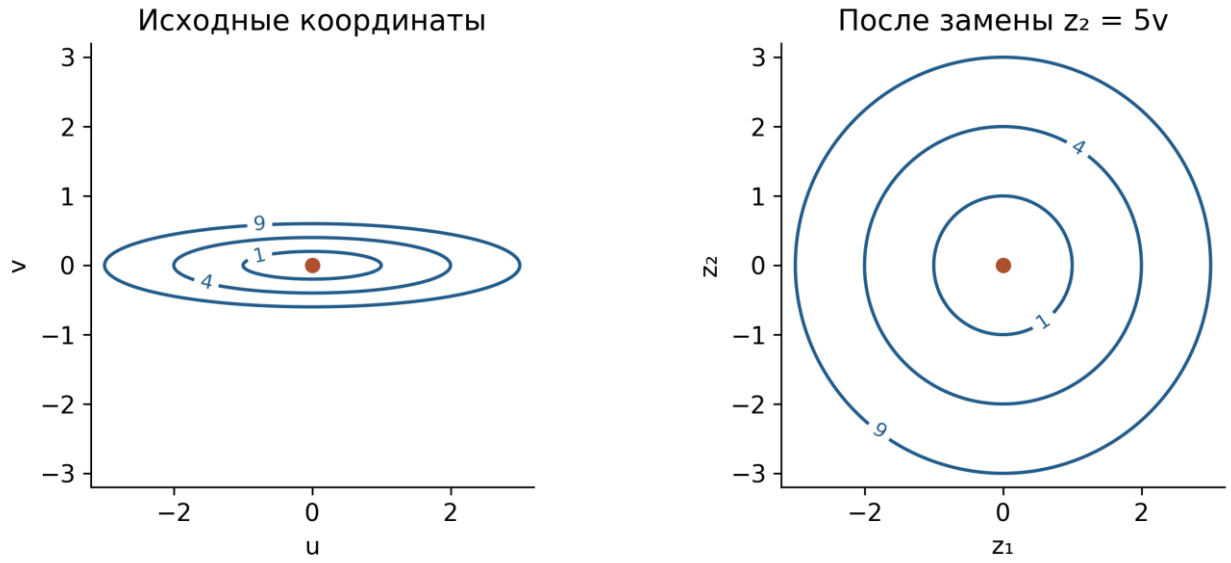


Рисунок 5 — Изменение масштаба переменных меняет геометрию численного поиска

Для положительно определённой квадратичной формы отношение наибольшего собственного значения матрицы к наименьшему характеризует неодинаковую кривизну. Здесь оно равно 25, а отношение полуосей — $\sqrt{25} = 5$. Эти величины нельзя смешивать.

Вытянутые уровни могут затруднять работу простых методов спуска: направление улучшения и подходящая длина шага существенно зависят от положения точки. Это не означает, что любой алгоритм обязательно медленен; поведение определяется способом выбора направления, шага и масштабирования.

Если одна переменная измеряется в метрах, другая в тысячах тенге, необработанная евклидова длина шага может плохо отражать значимость изменения параметров. Вводят характерные масштабы $s_i > 0$ и безразмерные переменные $z_i = (x_i - x_i^{\text{ref}})/s_i$. Выбранные масштабы следует обосновывать смыслом задачи.

8 Классификация основных методов

Классификация нужна, чтобы сопоставить свойства задачи и доступную вычислительную информацию. Один метод может одновременно быть локальным, детерминированным и использующим производные. Эти признаки относятся к разным основаниям классификации.

8.1 По используемой информации

Класс	Что используется	Примеры	Что учитывать
Нулевой порядок	Значения $f(\mathbf{x})$	Прямой поиск, метод Нелдера — Мида	Подходит при отсутствии надёжных производных; число вычислений функции может быть большим
Первый порядок	Значения функции и первые производные	Градиентный метод, наискорейший спуск	Нужны корректные производные и правило выбора шага
Второй порядок	Первые и вторые производные	Метод Ньютона	Требуется решить линейную систему; вне окрестности минимума нужны дополнительные меры

Здесь «порядок» обозначает порядок используемых производных. Он не совпадает с порядком сходимости алгоритма. Например, метод второго порядка не имеет квадратичную сходимость без соответствующих предположений.

Квазиньютоновские методы строят приближение матрицы вторых производных по изменениям точек и градиентов; точную матрицу вторых производных они обычно не требуют. Координатный спуск последовательно изменяет отдельные переменные, но его отнесение к классу по информации зависит от реализации одномерного поиска: она может использовать только значения функции или также производные.

Метод Нелдера — Мида называют симплексным методом прямого поиска. Его не следует смешивать с симплекс-методом линейного программирования: это разные алгоритмы и разные конструкции.

8.2 По структуре задачи и характеру поиска

Безусловные и условные методы. Первые предназначены для задач без явно заданных ограничений; вторые учитывают допустимую область, например через проекцию, штрафы, барьеры или работу с активными ограничениями. Обычный шаг безусловного метода может оказаться недопустимым в условной задаче.

Локальные и глобальные методы. Локальный метод использует информацию около текущих приближений и при соответствующих условиях стремится к локальному решению. В невыпуклой задаче результат может зависеть от начальной точки. Глобальные подходы исследуют несколько областей; строгая гарантия глобальности требует дополнительных свойств и оценок. Несколько запусков из разных точек повышают охват, но сами по себе не являются доказательством глобального минимума.

Детерминированные и стохастические методы. Первые при одинаковых данных и правилах дают одну и ту же последовательность. Вторые используют случайность при выборе направлений, выборок или пробных точек. Для воспроизводимости эксперимента фиксируют генератор случайных чисел, но это не превращает вероятностный результат в строгую гарантию оптимальности.

Гладкие и негладкие задачи. Наличие изломов меняет допустимые методы анализа. Например, функция $|x_1| + |x_2|$ имеет минимум в начале координат, но обычные частные производные там не существуют. Отсутствие гладкости не означает отсутствия решения.

8.3 Что выяснить перед выбором метода

Сначала определяют размерность и допустимое множество. Затем выясняют, можно ли вычислять производные, насколько дороги значения функции, есть ли шум, известна ли выпуклость и требуется ли глобальное решение. Для гладкой задачи небольшой размерности при доступных производных естественно рассматривать методы спуска. Если функция является результатом сложного моделирования и производные ненадёжны, уместно изучить методы прямого поиска. При ограничениях выбирают способ их явного соблюдения или обоснованного учёта.

Число итераций само по себе не измеряет стоимость решения. Одна итерация может включать десятки пробных вычислений, расчёт производных или решение большой системы. В эксперименте раздельно фиксируют затраты на функцию, производные и линейную алгебру, а также время и достигнутую точность.

9 Связь многомерного поиска с одномерной оптимизацией

Многие методы формируют последовательность

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{d}^k, \qquad \alpha_k \geq 0.$$

Вектор $\mathbf{d}^k$ задаёт направление, а число α_k — длину шага в единицах этого вектора. Изменение масштаба направления требует согласованного изменения шага. После выбора направления вводят функцию одной переменной

$$\varphi_k(\alpha) = f(\mathbf{x}^k + \alpha \mathbf{d}^k).$$

Именно здесь применяются ранее изученные методы одномерной оптимизации. Для использования золотого сечения нужно выбрать допустимый отрезок шагов и обосновать унимодальность на нём. В условной задаче дополнительно требуется $\mathbf{x}^k + \alpha \mathbf{d}^k \in D$.

Подробный пример. Возьмём функцию раздела 5, начальную точку $\mathbf{x}^0 = (3; 1)$ и направление $\mathbf{d} = (-1; -2)$. Тогда

$$x_1 = 3 - \alpha, \quad x_2 = 1 - 2\alpha, \quad u = 2 - \alpha, \quad v = 2 - 2\alpha.$$

Подставляя эти выражения в функцию, получаем

$$\varphi(\alpha) = 27 - 36\alpha + 14\alpha^2 = 14\left(\alpha - \frac{9}{7}\right)^2 + \frac{27}{7}.$$

Минимум на луче $\alpha \geq 0$ достигается при $\alpha^* = 9/7$. Новая точка и её значение равны

$$\mathbf{x}^1 = \left(\frac{12}{7}; -\frac{11}{7}\right), \qquad f(\mathbf{x}^1) = \frac{27}{7} \approx 3{,}857143.$$

Значение уменьшилось с 27, но глобальный минимум 3 ещё не достигнут. Луч не проходит через точку $(1; -1)$, поэтому точное решение одномерной подзадачи не решает автоматически всю многомерную задачу.

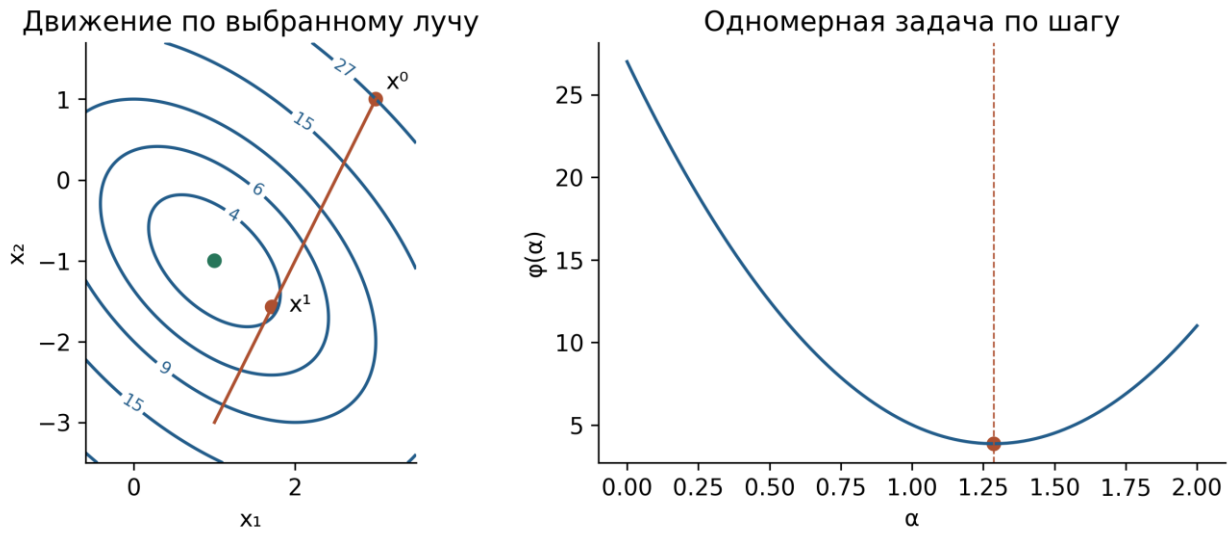


Рисунок 6 — На выбранном луче решается отдельная одномерная задача

Для самопроверки полезны следующие значения:

α	0	1/2	1	9/7	3/2	2
$\varphi(\alpha)$	27	12,5	5	27/7	4,5	11

Даже удачное направление не разрешает брать произвольно большой шаг. На выбранном луче функция после $\alpha = 9/7$ возрастает, а при достаточно большом шаге превысит исходное значение. Выбор направления и выбор шага — две связанные, но отдельные части алгоритма.

10 Сеточный эксперимент и его ограничения

Для наглядного исследования функции двух переменных можно вычислить её значения на прямоугольной сетке. Если по каждой оси выбрано N узлов, потребуется N^2 значений. Лучший узел минимизирует функцию среди узлов, а не среди всех точек прямоугольника.

В n измерениях число узлов становится N^n. При $N = 101$ в двумерном случае это 10 201 узел, в трёхмерном — 1 030 301, а в шестимерном — 1 061 520 150 601. Поэтому плотная сетка полезна для иллюстрации маломерных задач, но не является универсальным способом оптимизации.

В приложенном файле geometry_2d.py сеточный поиск реализован через массивы NumPy. Основная идея такова:

```
xs = np.linspace(xmin, xmax, N)
ys = np.linspace(ymin, ymax, N)
X, Y = np.meshgrid(xs, ys)
Z = f(X, Y)
iy, ix = np.unravel_index(np.argmin(Z), Z.shape)
x_best, y_best = xs[ix], ys[iy]
```

Первый индекс массива соответствует строке, то есть координате x_2, второй — столбцу и координате x_1. Перестановка индексов — частая причина неверных координат при правильном минимальном значении. Полная программа проверяет границы, число узлов и конечность вычисленных значений.

Для основной функции выбран прямоугольник $[-1{,}3;\,3{,}7] \times [-3{,}1;\,1{,}9]$. Он содержит точный минимум, но тот не обязан совпадать с узлом сетки. Результаты расчёта:

N	**Узлов**	x_1	x_2	f	**Ошибка координат**
21	441	0,9500	-1,1000	3,035000	0,111803
41	1681	0,9500	-0,9750	3,003750	0,055902
81	6561	1,0125	-0,9750	3,002187	0,027951

Что можно доказать для квадратичного примера. Пусть

$$f(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^*)^{\mathsf{T}} Q (\mathbf{x} - \mathbf{x}^*) + d,$$

где Q симметрична и положительно определена, а $\mathbf{x}^*$ лежит в прямоугольнике. Обозначим через $\lambda_{\min}$ и $\lambda_{\max}$ крайние собственные значения Q. Существует узел $\mathbf{y}$ с отклонением по каждой координате не больше половины шага соответствующей сетки. Поэтому

$$f(\mathbf{y}) - d \le \lambda_{\max} \left(\frac{h_1^2}{4} + \frac{h_2^2}{4} \right).$$

Если $\hat{\mathbf{x}}$ — лучший узел, то $f(\hat{\mathbf{x}}) \le f(\mathbf{y})$, а снизу разрыв оценивается через $\lambda_{\min} \| \hat{\mathbf{x}} - \mathbf{x}^* \|_2^2$. Следовательно,

$$\| \hat{\mathbf{x}} - \mathbf{x}^* \|_2 \le \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}} \left(\frac{h_1^2}{4} + \frac{h_2^2}{4} \right)}.$$

Это оценка для указанного класса функций и прямоугольной сетки, при точных вычислениях. Для произвольной функции лучший узел не обязательно ближайший к истинному минимуму. Узкий минимум может находиться между узлами; простой просмотр сетки не исключает такой ситуации.

10.1 Запуск программы

Для работы нужны Python 3.10 или новее, NumPy и Matplotlib. После распаковки архива выполните в терминале:

```
python -m pip install numpy matplotlib
python geometry_2d.py --points 41
python geometry_2d.py --points 81
python geometry_2d.py --variant 7 --points 81
```

В Windows вместо python можно использовать py. Скрипт сохраняет график линий уровня с разрешением 300 dpi, файл JSON с параметрами и результатами и таблицу CSV. Все результаты помещаются в отдельную подпапку, имя которой содержит номер варианта и число узлов по оси. Повторный запуск с теми же параметрами перезаписывает результаты этого запуска.

Функция exact_line_step отдельно вычисляет точный шаг для положительно определённой квадратичной функции на луче. Если формальный минимум на прямой соответствует отрицательному шагу, возвращается $\alpha = 0$, поскольку рассматривается условие $\alpha \ge 0$. Это вспомогательный учебный расчёт, а не реализация полного многомерного метода.

11 Самостоятельная работа

Для каждого варианта дана функция

$$f(x_1, x_2) = a(x_1 - s)^2 + 2b(x_1 - s)(x_2 - t) + c(x_2 - t)^2 + d.$$

Все параметры приведены в таблице. Буква d в этой формуле обозначает постоянное слагаемое и не является вектором направления.

№	a	b	c	s	t	d
1	2	1	2	-1	1	0
2	1	0	9	-1	-1	0
3	3	-1	2	-1	2	0
4	4	2	3	-1	0	0
5	1	0	25	-1	-2	0
6	5	-2	2	-1	1	0
7	2	1	2	0	1	3
8	1	0	9	0	-1	3

№	a	b	c	s	t	d
9	3	-1	2	0	2	3
10	4	2	3	0	0	3
11	1	0	25	0	-2	3
12	5	-2	2	0	1	3
13	2	1	2	1	1	-2
14	1	0	9	1	-1	-2
15	3	-1	2	1	2	-2
16	4	2	3	1	0	-2
17	1	0	25	1	-2	-2
18	5	-2	2	1	1	-2

Порядок выполнения.

1. Запишите функцию своего варианта. Проверьте $a > 0$ и $ac - b^2 > 0$.
2. Докажите глобальность и единственность минимума выделением квадратов. Укажите точку и значение минимума отдельно.
3. Запишите уравнения трёх линий уровня: $f = d + 1, f = d + 4, f = d + 9$. Постройте их при одинаковом масштабе осей.
4. Найдите собственные значения матрицы Q по формуле из раздела 11.1. Вычислите отношение длин полуосей и объясните вытянутость линий.
5. Из точки $\mathbf{x}^0 = (s + 2; t + 2)$ рассмотрите направление $\mathbf{p} = (-1; -2)$. Получите квадратный трёхчлен по α и найдите его минимум на луче $\alpha \geq 0$.
6. Выполните сеточный поиск при $N = 21{,}41{,}81$ в прямоугольнике $[s - 2{,}3; s + 2{,}7] \times [t - 2{,}1; t + 2{,}9]$. Сравните лучший узел с аналитическим решением.
7. Для каждой сетки укажите число вычисленных узлов, ошибку координат, разрыв по функции и теоретическую оценку из раздела 10. Объясните, почему совпадение точного минимума с узлом не доказывает универсальную точность метода.
8. Укажите, к какому классу относится сеточный поиск по используемой информации. Предложите класс метода для дальнейшего решения той же гладкой выпуклой задачи и объясните выбор.

Содержание отчёта: постановка; аналитическое решение; вывод линий уровня; рисунок; вычисление шага вдоль луча; таблица экспериментов; проверка оценки; вывод о геометрии задачи и ограничениях сеточного поиска.

11.1 Формулы для самопроверки

При $u = x_1 - s, v = x_2 - t$ выделение квадратов даёт

$$f - d = a\left(u + \frac{b}{a}v\right)^2 + \left(c - \frac{b^2}{a}\right)v^2.$$

Если $a > 0$ и $ac - b^2 > 0$, оба коэффициента положительны. Единственный минимум находится в точке $(s; t)$ и равен d.

Для $Q = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ собственные значения равны

$$\lambda_{\pm} = \frac{a + c \pm \sqrt{(a - c)^2 + 4b^2}}{2}.$$

Отношение длин большой и малой полуосей составляет $\sqrt{\lambda_+ / \lambda_-}$. При $b = 0$ оси эллипсов совпадают с координатными, а при $a = c$ и $b = 0$ получаются окружности.

Для заданных в работе начальной точки и направления

$$\varphi(\alpha) = A\alpha^2 - B\alpha + C,$$

$$A = a + 4b + 4c, \quad B = 4a + 12b + 8c, \quad C = 4a + 8b + 4c + d.$$

Положительная определённость гарантирует $A > 0$. Минимальный шаг на луче равен $\alpha^* = \max\{0; B/(2A)\}$. После его вычисления нужно подставить шаг в обе координаты, а затем проверить значение исходной функции.

12 Контрольные вопросы и задачи

1. Чем отличаются точка минимума и минимальное значение?
2. Как ограничения входят в определение локального минимума?
3. Чем поверхность функции отличается от её линий уровня?
4. Может ли множество уровня состоять из одной точки или быть пустым?
5. Почему минимум в двух координатных сечениях не гарантирует минимум функции?
6. Какие условия обеспечивают достижение минимума?
7. Почему локальный минимум выпуклой задачи является глобальным?
8. Как перекрёстный член квадратичной функции влияет на линии уровня?
9. Чем отношение полуосей отличается от отношения собственных значений?
10. Чем порядок используемых производных отличается от порядка сходимости?
11. Почему точный одномерный поиск по шагу не всегда даёт многомерный минимум?
12. Как меняется стоимость равномерной сетки с ростом размерности?

Задача 1. Найдите минимум $f(x_1, x_2) = (x_1 + 2)^2 + 4(x_2 - 1)^2 - 5$ и полуоси линии уровня $f = -1$.

Ответ: точка $(-2; 1)$, значение -5. Уравнение уровня $(x_1 + 2)^2 + 4(x_2 - 1)^2 = 4$, полуоси 2 и 1.

Задача 2. Для $q = x_1^2 + 4x_1x_2 + x_2^2$ сравните значения в точках $(0{,}1; 0{,}1)$ и $(0{,}1; -0{,}1)$ с $q(0; 0)$.

Ответ: соответственно $0{,}06$ и $-0{,}02$, тогда как в начале координат значение ноль. Начало координат является седлом.

Задача 3. Лучший узел сетки дал значение $f = 3{,}0001$ для функции раздела 5. Какое расстояние до точного минимума гарантирует аналитическая оценка?

Ответ: $\| \mathbf{x} - \mathbf{x}^* \|_2 \le \sqrt{0{,}0001} = 0{,}01$. Оценка опирается на специальную структуру данной функции.

Задача 4. Доказывает ли совпадение результатов десяти запусков локального алгоритма глобальность найденного минимума?

Ответ: само по себе не доказывает. Для вывода нужны дополнительные основания, например доказанная выпуклость задачи или корректная глобальная нижняя оценка, совпавшая с полученным значением.

13 Литература

[1] Пантелеев А. В., Летова Т. А. Методы оптимизации в примерах и задачах: учебное пособие. 4-е изд., испр. СПб.: Лань, 2015. 512 с. Глава 1, § 1 «Общая постановка задачи оптимизации и основные положения»; глава 2 «Численные методы поиска безусловного экстремума».

[2] Рабочая программа дисциплины «Методы оптимизации», 2024. Образовательная программа 6B06103 «Математическое и компьютерное моделирование». Тема 10 «Методы многомерной оптимизации. Графическая интерпретация целевой функции. Классификация основных методов многомерной оптимизации».