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

Дисциплина «Методы оптимизации»

ЛЕКЦИЯ 12

Классический метод минимизации и основы методов спуска

Образовательная программа 6B06103 Математическое и компьютерное моделирование

Объём: 2 академических часа. **Место в рабочей программе:** лекция 12.

Цель: научиться находить и классифицировать стационарные точки функции нескольких переменных, обосновывать минимум и выбирать шаг, обеспечивающий уменьшение целевой функции.

План занятия: необходимые и достаточные условия экстремума; критерий Сильвестра; вырожденные стационарные точки; связь с выпуклостью; направление спуска; правило Армихо; условия сходимости и критерии остановки; вычислительные примеры.

Результаты обучения: студент решает систему стационарности, исследует матрицу Гессе, различает локальный и глобальный выводы, проверяет направление спуска и выполняет несколько итераций с дроблением шага.

В лекции 11 были получены градиент и квадратичная модель Тейлора. Теперь используем их для ответа на два вопроса: является ли найденная точка минимумом и как приблизиться к стационарной точке, когда систему уравнений нельзя решить явно. Основное рассмотрение относится к гладкой **безусловной** минимизации. Ограничения требуют отдельного учёта допустимых направлений [1–3].

1 Постановка задачи и значение слова минимум

Рассматривается задача

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

Точка $\mathbf{x}^*$ называется точкой локального минимума, если существует $r > 0$, для которого

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) \quad \text{при} \quad \| \mathbf{x} - \mathbf{x}^* \|_2 < r.$$

Для строгого локального минимума неравенство строгое при $\mathbf{x} \neq \mathbf{x}^*$. Глобальный минимум требует того же сравнения со всеми точками области определения. Поэтому исследование в малой окрестности само по себе не решает вопрос о глобальном минимуме.

Например, $f(x, y) = x^2 + y^2$ имеет единственный глобальный минимум в начале координат: оба слагаемых неотрицательны и одновременно равны нулю только там. Функция $f(x, y) = e^x + y^2$ имеет нижнюю грань 0, но нигде её не достигает. Следовательно, ограниченность снизу не гарантирует существования точки минимума.

Условие существования. Непрерывная функция достигает минимума на непустом компактном множестве. Для непрерывной функции на всём $\mathbb{R}^n$ достаточным условием является коэрцитивность:

$$f(\mathbf{x}) \to +\infty \quad \text{при} \quad \| \mathbf{x} \|_2 \to \infty.$$

Действительно, за пределами достаточно большого шара значения больше, чем в фиксированной пробной точке. Поиск минимума можно ограничить замкнутым шаром, на котором применима теорема Вейерштрасса.

2 Необходимое условие первого порядка

Пусть f дифференцируема во внутренней точке локального экстремума $\mathbf{x}^*$. Тогда

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

Обоснование. Зафиксируем любое направление $\mathbf{p}$ и рассмотрим функцию одной переменной $\varphi(t) = f(\mathbf{x}^* + t\mathbf{p})$. При достаточно малых положительных и отрицательных t точка остаётся в области определения. Из одномерного условия экстремума следует

$$\varphi'(0) = \nabla f(\mathbf{x}^*)^{\mathsf{T}}\mathbf{p} = 0.$$

Равенство выполняется для всех $\mathbf{p}$, в частности для координатных векторов. Значит, каждая частная производная равна нулю.

Точки, удовлетворяющие системе

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = 0, \quad i = 1, \dots, n,$$

называются **стационарными**. Это кандидаты на экстремум. В стационарной точке может быть минимум, максимум или седло. Например, у $x^2 - y^2$ градиент в начале координат нулевой, но по оси x функция возрастает, а по оси y убывает.

Граница и негладкость. Для $f(x) = x$ на $[0, +\infty)$ минимум достигается при $x = 0$, хотя $f'(0) = 1$. Для $|x|$ минимум также находится в нуле, но обычной производной там нет. Формула $\nabla f = 0$ применяется при выполнении её предпосылок.

3 Условия второго порядка

Пусть $f \in C^2$ в окрестности стационарной точки. Разложение из предыдущей лекции принимает вид

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) = \frac{1}{2}\mathbf{h}^{\mathsf{T}} H(\mathbf{x}^*)\mathbf{h} + o(\| \mathbf{h} \|_2^2).$$

Линейное слагаемое исчезло. Знак изменения функции в первую очередь определяется квадратичной формой Гессиана.

3.1 Необходимое условие

В точке локального минимума

$$\mathbf{p}^{\mathsf{T}} H(\mathbf{x}^*)\mathbf{p} \geq 0 \quad \text{для всех } \mathbf{p},$$

то есть $H(\mathbf{x}^*)$ положительно полуопределена. В этой лекции полуопределённость понимается включительно: положительно определённая матрица также удовлетворяет нестрогому неравенству.

Для доказательства снова рассматриваем $\varphi(t) = f(\mathbf{x}^* + t\mathbf{p})$. В точке минимума $\varphi''(0) \geq 0$, а

$$\varphi''(0) = \mathbf{p}^{\mathsf{T}} H(\mathbf{x}^*)\mathbf{p}.$$

3.2 Достаточное условие

Если $\nabla f(\mathbf{x}^*) = \mathbf{0}$ и $H(\mathbf{x}^*)$ положительно определена, то $\mathbf{x}^*$ — строгий локальный минимум [1].

Действительно, для симметричной положительно определённой матрицы существует $m > 0$ такое, что

$$\mathbf{h}^{\mathsf{T}} H(\mathbf{x}^*)\mathbf{h} \geq m \parallel \mathbf{h} \parallel_2^2.$$

По определению малого остатка можно выбрать окрестность, где его абсолютная величина не превосходит $m \parallel \mathbf{h} \parallel_2^2/4$. Тогда

$$f(\mathbf{x}^* + \mathbf{h}) - f(\mathbf{x}^*) \geq \frac{m}{4} \parallel \mathbf{h} \parallel_2^2 > 0 \quad (\mathbf{h} \neq \mathbf{0}).$$

Аналогично отрицательная определённость даёт строгий локальный максимум. Если у Гессиана есть как положительные, так и отрицательные собственные значения, то вдоль соответствующих направлений функция изменяется с разными знаками: стационарная точка является седловой.

Гессиан в стационарной точке	Что можно заключить
Все собственные значения положительны	Строгий локальный минимум
Все собственные значения отрицательны	Строгий локальный максимум
Есть положительное и отрицательное значения	Седловая точка
Есть нулевые значения, остальные одного знака	Квадратичного приближения недостаточно
Все собственные значения нулевые	Нужны другие способы исследования

4 Критерий Сильвестра и случай двух переменных

Для симметричной матрицы положительная определённость эквивалентна положительности всех **угловых главных миноров**:

$$\Delta_1 > 0, \quad \Delta_2 > 0, \quad \dots, \quad \Delta_n > 0.$$

Здесь Δ_k — определитель левого верхнего блока размера $k \times k$. Для отрицательной определённости требуется чередование знаков: $(-1)^k \Delta_k > 0$.

Для двух переменных

$$H = \begin{pmatrix} A & B \\ B & C \end{pmatrix}, \quad D = AC - B^2.$$

В стационарной точке получаем простой порядок проверки:

1. Если $D > 0$ и $A > 0$, имеется строгий локальный минимум.
2. Если $D > 0$ и $A < 0$, имеется строгий локальный максимум.
3. Если $D < 0$, имеется седловая точка.
4. Если $D = 0$, этот тест не даёт окончательного ответа.

Почему возникают именно такие условия. При $A \neq 0$ квадратичную форму можно представить как

$$Ah_1^2 + 2Bh_1h_2 + Ch_2^2 = A\left(h_1 + \frac{B}{A}h_2\right)^2 + \frac{D}{A}h_2^2.$$

При $A > 0$ и $D > 0$ оба коэффициента положительны. При $D < 0$ форма принимает значения обоих знаков.

Типичная ошибка. Для положительной полуопределённости нельзя просто заменить $>$ на $\geq$ в проверке угловых миноров. Матрица $\mathrm{diag}(0, -1)$ имеет $\Delta_1 = \Delta_2 = 0$, но её квадратичная форма отрицательна на векторе $(0; 1)$. Для полуопределённости проверяются все главные миноры либо все собственные значения. В двумерном случае нужны одновременно $A \geq 0$, $C \geq 0$, $AC - B^2 \geq 0$.

5 Классический метод на квадратичном примере

Классический метод состоит в аналитическом решении системы стационарности и последующей классификации её решений. Разберём функцию из лекций 10 и 11:

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

Введём $u = x_1 - 1$, $v = x_2 + 1$. Тогда

$$\nabla f = \begin{pmatrix} 4u + 2v \\ 2u + 4v \end{pmatrix}, \quad H = \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}.$$

Шаг 1. Найдём стационарную точку. Система

$$4u + 2v = 0, \quad 2u + 4v = 0$$

имеет единственное решение $u = v = 0$. Следовательно, $\mathbf{x}^* = (1; -1)^\mathsf{T}$.

Шаг 2. Классифицируем точку. Угловые миноры равны $\Delta_1 = 4 > 0$ и $\Delta_2 = 16 - 4 = 12 > 0$. Гессиан положительно определён, поэтому найден строгий локальный минимум.

Шаг 3. Проверим глобальность. Здесь можно выполнить точное преобразование:

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

Правая часть положительна при $(u, v) \neq (0,0)$. Поэтому минимум единственный и глобальный, $f_{\min} = 3$. Такое доказательство сильнее локального теста: сравнение выполнено со всеми точками плоскости.

Общий квадратичный случай. Для симметричной положительно определённой матрицы A

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T} A \mathbf{x} - \mathbf{b}^\mathsf{T}\mathbf{x} + c$$

имеем $\nabla f = A\mathbf{x} - \mathbf{b}$, $H = A$. Стационарная точка — решение системы $A\mathbf{x}^* = \mathbf{b}$. При вычислениях решают линейную систему; явное построение обратной матрицы обычно не требуется. Точное равенство

$$f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\mathsf{T} A (\mathbf{x} - \mathbf{x}^*)$$

сразу доказывает единственность глобального минимума.

6 Несколько стационарных точек и вырожденный Гессиан

6.1 Невыпуклый пример

Пусть

$$F(x, y) = x^4 - 2x^2 + y^2.$$

Система стационарности имеет вид

$$4x(x^2 - 1) = 0, \quad 2y = 0.$$

Получаем три точки: $(0; 0)$, $(1; 0)$, $(-1; 0)$. Гессиан равен

$$H_F(x, y) = \begin{pmatrix} 12x^2 - 4 & 0 \\ 0 & 2 \end{pmatrix}.$$

В начале координат собственные значения -4 и 2, поэтому это седло. В точках $(\pm 1; 0)$ собственные значения 8 и 2: обе точки являются строгими локальными минимумами. Более того,

$$F(x, y) = (x^2 - 1)^2 + y^2 - 1 \geq -1,$$

и равенство достигается ровно в этих двух точках. Следовательно, оба минимума глобальные. Несколько глобальных минимумов допустимы для невыпуклой функции.

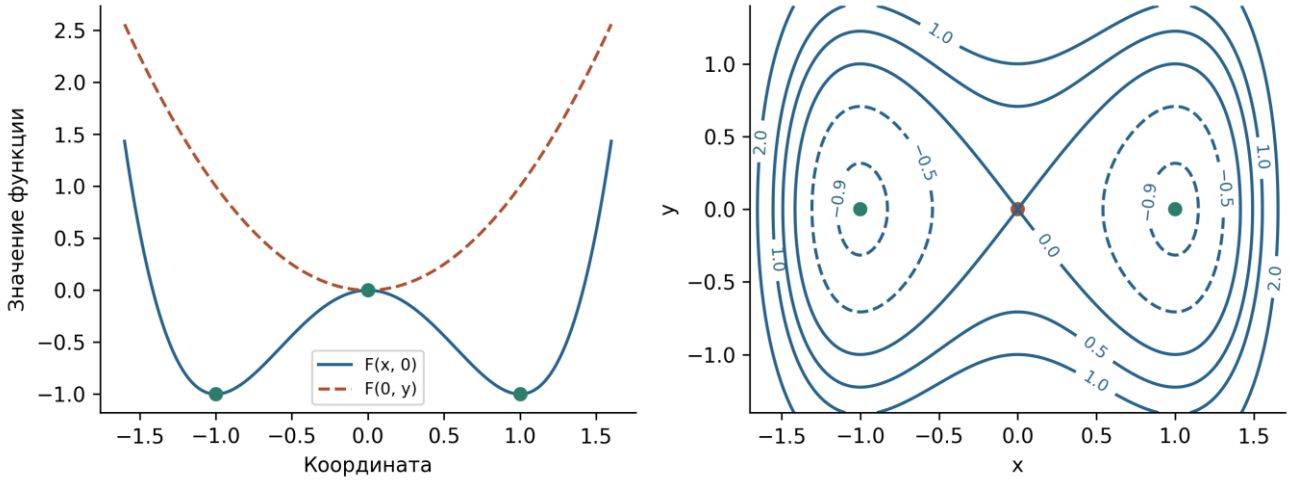


Рисунок 1 — Разные типы стационарных точек функции $x^4 - 2x^2 + y^2$

На левом графике показано сечение $y = 0$. Точка $x = 0$ выглядит максимумом вдоль этого сечения, но вдоль оси y функция равна y^2 и имеет минимум. Сравнение направлений объясняет седловой характер точки на плоскости.

6.2 Когда вторых производных недостаточно

Рассмотрим две функции

$$f_1(x, y) = x^4 + y^4, \quad f_2(x, y) = x^4 - y^4.$$

В начале координат у обеих функций нулевые градиент и Гессиан. Однако $f_1 > 0$ вне начала координат, поэтому там строгий глобальный минимум. У f_2 значения положительны на оси x и отрицательны на оси y, поэтому там седло.

Из нулевого определителя Гессиана нельзя заключать ни наличие, ни отсутствие минимума. Следует исследовать следующие члены разложения, выделить неотрицательные слагаемые или сравнить значения на специально выбранных кривых. Даже проверка конечного числа направлений в общем случае не доказывает минимум во всей окрестности.

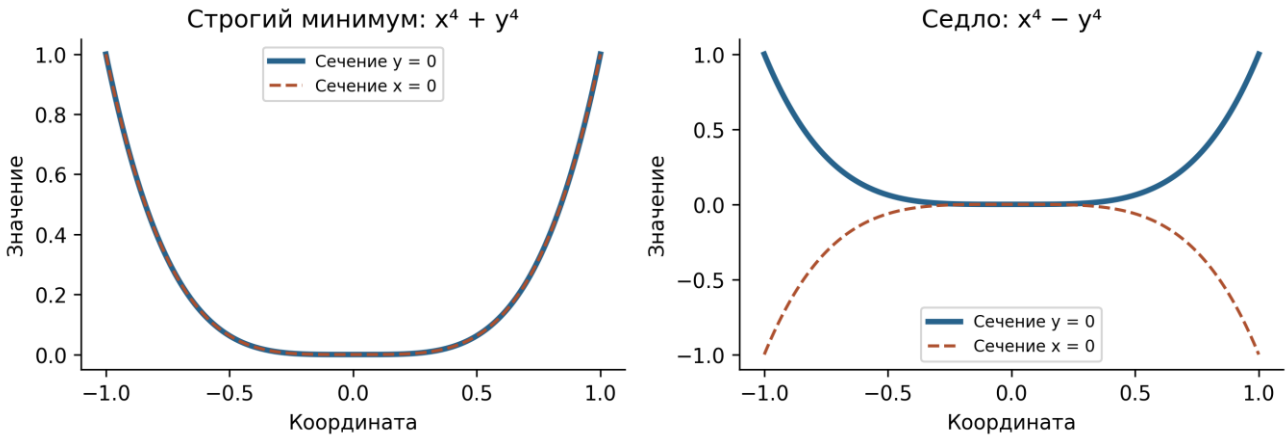


Рисунок 2 — Одинаковый нулевой Гессиан допускает разное поведение функции

На рисунке показаны сечения по координатным осям. Для f_1 оба сечения совпадают, поэтому дополнительно используется неотрицательность всей суммы. Для f_2 два сечения уже достаточны, чтобы опровергнуть минимум и максимум.

7 Выпуклость и обоснование глобального результата

Пусть f дифференцируема и выпукла на выпуклой открытой области. Тогда для любых её точек справедливо неравенство первого порядка

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathsf{T}}(\mathbf{y} - \mathbf{x}).$$

Если $\nabla f(\mathbf{x}^*) = \mathbf{0}$, правая часть равна $f(\mathbf{x}^*)$. Следовательно, любая стационарная точка выпуклой функции является глобальным минимумом. Строгая выпуклость обеспечивает не более одной точки минимума; существование надо обосновать отдельно.

Для C^2-функции на выпуклой открытой области выпуклость эквивалентна условию $H(\mathbf{x}) \succcurlyeq 0$ во всей области. Положительная определённость Гессиана только в одной точке даёт локальный результат. Если она выполнена везде, функция строго выпукла.

Если существует единая константа $m > 0$ такая, что $H(\mathbf{x}) \succcurlyeq mI$ во всей области, функция сильно выпукла с константой m. Для неё

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathsf{T}}(\mathbf{y} - \mathbf{x}) + \frac{m}{2} \| \mathbf{y} - \mathbf{x} \|_2^2.$$

У основного квадратичного примера собственные значения Гессиана равны 2 и 6, поэтому можно взять $m = 2$. В частности,

$$f(\mathbf{x}) - 3 \geq \| \mathbf{x} - (1; -1)^{\mathsf{T}} \|_2^2.$$

Это связывает ошибку по значению функции с расстоянием до решения. Для общей невыпуклой функции подобная оценка без дополнительных условий неверна.

8 Направление спуска и общая схема итераций

При сложной функции явное решение $\nabla f = 0$ может быть недоступно. Тогда строят последовательность

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha_k \mathbf{p}_k, \quad \alpha_k > 0.$$

Здесь $\mathbf{p}_k$ задаёт направление, а α_k — длину шага в данной параметризации. **Строгое направление спуска** в точке $\mathbf{x}$ удовлетворяет условию

$$\nabla f(\mathbf{x})^{\mathsf{T}}\mathbf{p} < 0.$$

Из дифференцируемости следует

$$f(\mathbf{x}+\alpha\mathbf{p}) = f(\mathbf{x}) + \alpha\nabla f(\mathbf{x})^{\mathsf{T}}\mathbf{p} + o(\alpha), \quad \alpha \to 0.$$

При достаточно малом положительном α отрицательный линейный член преобладает над остатком. Поэтому функция уменьшается. Утверждение является локальным: большой шаг может увеличить функцию.

При ненулевом градиенте направление $\mathbf{p} = -\nabla f$ всегда является строгим направлением спуска, поскольку

$$\nabla f^{\mathsf{T}}\mathbf{p} = -\| \nabla f \|_2^2 < 0.$$

Более общий выбор $\mathbf{p} = -B\nabla f$ также обеспечивает спуск, если B симметрична и положительно определена. Этот факт объясняет возможность учитывать масштаб координат. Для сходимости последовательности с меняющимися матрицами B_k нужны дополнительные равномерные оценки; одного знака скалярного произведения недостаточно.

Если $\nabla f = 0$, антиградиент равен нулю. Это не исключает седла: в начале координат невыпуклого примера алгоритм с таким направлением не начнёт движение. Необходимо отдельно исследовать характер стационарной точки.

9 Лемма о спуске и допустимая величина шага

Предположим, что градиент удовлетворяет условию Липшица с константой $L > 0$ на области, содержащей отрезок от $\mathbf{x}$ до $\mathbf{x}+\mathbf{s}$:

$$\| \nabla f(\mathbf{u}) - \nabla f(\mathbf{v}) \|_2 \le L \| \mathbf{u} - \mathbf{v} \|_2.$$

Тогда выполняется **лемма о спуске**:

$$f(\mathbf{x}+\mathbf{s}) \le f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathsf{T}}\mathbf{s} + \frac{L}{2}\| \mathbf{s} \|_2^2.$$

Вывод. По формуле Ньютона — Лейбница вдоль отрезка

$$f(\mathbf{x}+\mathbf{s}) - f(\mathbf{x}) = \int_0^1 \nabla f(\mathbf{x}+t\mathbf{s})^{\mathsf{T}}\mathbf{s}\, dt.$$

Вычтем $\nabla f(\mathbf{x})^{\mathsf{T}}\mathbf{s}$. Подынтегральная разность по неравенству Коши — Буняковского и условию Липшица не превосходит $Lt\,\| \mathbf{s} \|_2^2$. Интегрирование даёт $L\,\| \mathbf{s} \|_2^2/2$.

При $\mathbf{s} = -\alpha\mathbf{g}$, где $\mathbf{g} = \nabla f(\mathbf{x})$, получаем

$$f(\mathbf{x}-\alpha\mathbf{g}) \le f(\mathbf{x}) - \alpha\left(1 - \frac{L\alpha}{2}\right)\| \mathbf{g} \|_2^2.$$

При $\mathbf{g} \neq 0$ гарантированное уменьшение обеспечивается условием $0 < \alpha < 2/L$. Выбор $\alpha = 1/L$ даёт

$$f(\mathbf{x} - \mathbf{g}/L) \le f(\mathbf{x}) - \frac{1}{2L}\| \mathbf{g} \|_2^2.$$

В основном квадратичном примере $L = 6$, поэтому любой фиксированный шаг из $(0{,}1/3)$ гарантирует уменьшение в нестационарной точке. В конкретной точке допустимый интервал может оказаться шире этой общей оценки.

10 Выбор шага по правилу Армихо

Константа L часто неизвестна. Вместо её предварительного вычисления можно проверять фактическое уменьшение функции. Для направления спуска $\mathbf{p}$ правило Армихо требует [2]

$$f(\mathbf{x} + \alpha\mathbf{p}) \le f(\mathbf{x}) + \sigma\alpha\nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{p}, \quad 0 < \sigma < 1.$$

Правая часть меньше текущего значения. Параметр σ задаёт долю линейно предсказанного уменьшения, которую мы требуем от фактического шага.

Дробление шага. Задаём начальный шаг $\alpha_0 > 0$ и коэффициент $\beta \in (0{,}1)$. Последовательно проверяем

$$\alpha_0, \quad \beta\alpha_0, \quad \beta^2\alpha_0, \quad \dots$$

и принимаем первую величину, удовлетворяющую правилу. В примерах используем $\alpha_0 = 1$, $\beta = 1/2$, $\sigma = 0{,}1$. На каждой новой итерации снова начинаем с $\alpha_0 = 1$.

Почему дробление заканчивается. Из дифференцируемости

$$f(\mathbf{x} + \alpha\mathbf{p}) - f(\mathbf{x}) - \sigma\alpha\mathbf{g}^{\mathrm{T}}\mathbf{p} = (1 - \sigma)\alpha\mathbf{g}^{\mathrm{T}}\mathbf{p} + o(\alpha).$$

Первый член отрицателен, поэтому для всех достаточно малых положительных шагов условие выполнено. В точной арифметике геометрическая последовательность проб достигает этого диапазона за конечное число сокращений. В программе дополнительно нужны ограничение числа проб и проверка, что новая точка отличается от старой.

Для $\mathbf{p} = -\mathbf{g}$ лемма о спуске показывает, что достаточно

$$\alpha \le \frac{2(1 - \sigma)}{L}.$$

Если во всех итерациях действует одна константа L, принимаемые шаги имеют положительную нижнюю границу

$$\alpha_k \ge \underline{\alpha} = \min\left\{\alpha_0, \frac{2\beta(1 - \sigma)}{L}\right\} > 0.$$

Если первая проба принята, утверждение очевидно. Иначе предыдущая проба α_k/β была отклонена и потому превышала достаточную верхнюю границу. Отсюда следует требуемая оценка.

11 Полный ручной расчёт первых шагов

Для основной квадратичной функции возьмём $\mathbf{x}^0 = (3; 0)^{\mathrm{T}}$. В отличие от лекции 11, начальная точка здесь изменена, чтобы показать движение по двум собственным направлениям.

Имеем $u = 2$, $v = 1$, поэтому

$$f(\mathbf{x}^0) = 17, \quad \mathbf{g}_0 = (10; 8)^{\mathrm{T}}, \quad \mathbf{p}_0 = (-10; -8)^{\mathrm{T}}.$$

Проверка направления: $\mathbf{g}_0^{\mathrm{T}}\mathbf{p}_0 = -164 < 0$. Так как функция квадратична, вдоль прямой получаем точное выражение

$$\varphi(\alpha) = f(\mathbf{x}^0 + \alpha\mathbf{p}_0) = 17 - 164\alpha + 488\alpha^2.$$

Здесь $\mathbf{p}_0^{\mathsf{T}} H \mathbf{p}_0 = 976$. Граница Армихо при $\sigma = 0{,}1$ равна $17 - 16{,}4\alpha$.

Проба α	$\varphi(\alpha)$	Граница Армихо	Решение
1	341	0,6	Отклонить
0,5	57	8,8	Отклонить
0,25	6,5	12,9	Принять

Следовательно,

$$\mathbf{x}^1 = (3;0) + 0{,}25(-10;-8) = (0{,}5;-2), \quad f(\mathbf{x}^1) = 6{,}5.$$

Вторая итерация. В этой точке $u = -0{,}5$, $v = -1$, поэтому $\mathbf{g}_1 = (-4;-5)$ и $\mathbf{p}_1 = (4;5)$. Получаем

$$f(\mathbf{x}^1 + \alpha \mathbf{p}_1) = 6{,}5 - 41\alpha + 122\alpha^2.$$

При $\alpha = 1$ значение равно 87,5 и условие не выполнено. При $\alpha = 0{,}5$ значение 16,5 также слишком велико. При $\alpha = 0{,}25$ получаем $3{,}875 \le 5{,}475$, поэтому

$$\mathbf{x}^2 = (1{,}5;-0{,}75), \quad f(\mathbf{x}^2) = 3{,}875.$$

Сравнение с точным одномерным поиском. Из $\varphi'(\alpha) = -164 + 976\alpha = 0$ следует

$$\alpha_{\text{exact}} = \frac{41}{244} \approx 0{,}168033.$$

Это лучший шаг вдоль данной прямой, но точка после него ещё не обязана быть минимумом на всей плоскости. Правило Армихо не ищет точный минимум на прямой: оно принимает достаточно хорошее уменьшение с небольшим числом проб.

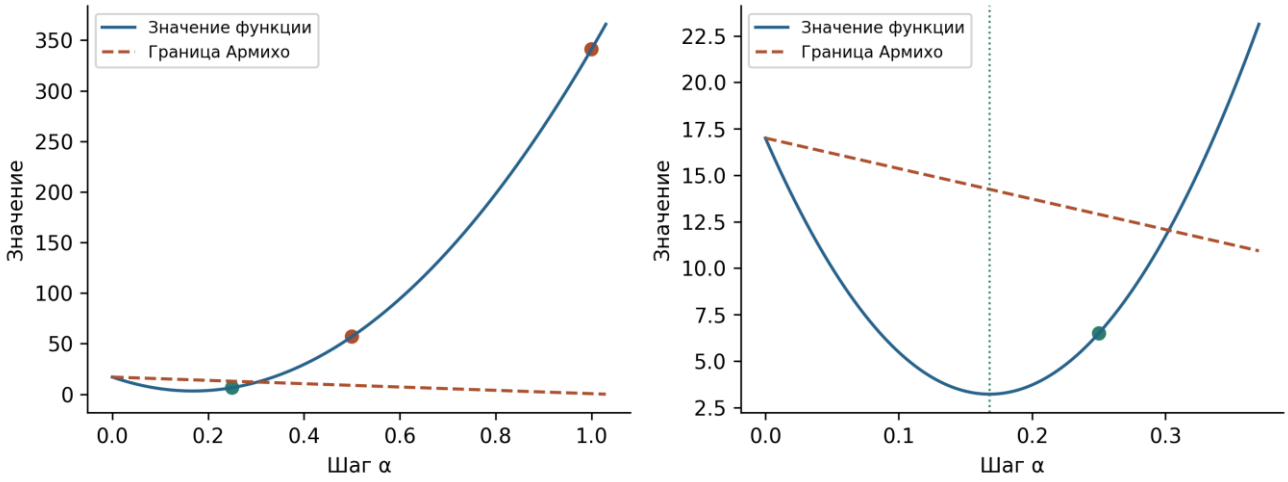


Рисунок 3 — Проверка шага по сравнению функции с границей Армихо

Левый график показывает большой рост функции при крупных пробах. Справа увеличен диапазон малых шагов. Вертикальная отметка точного минимума на прямой отличается от принятого шага 0,25.

k	x_1^k	x_2^k	f_k	$\| \mathbf{g}_k \|_2$	α_k
0	3	0	17	12,80625	0,25
1	0,5	-2	6,5	6,403124	0,25
2	1,5	-0,75	3,875	3,201562	0,25
3	0,875	-1,25	3,21875	1,600781	0,25
4	1,125	-0,9375	3,054688	0,8003905	0,25
5	0,96875	-1,0625	3,013672	0,4001953	—

В таблице α_k переводит точку строки k в следующую строку. Для последней показанной точки следующий шаг не приводится. Норма градиента нужна для проверки стационарности, а значение функции — для контроля уменьшения.

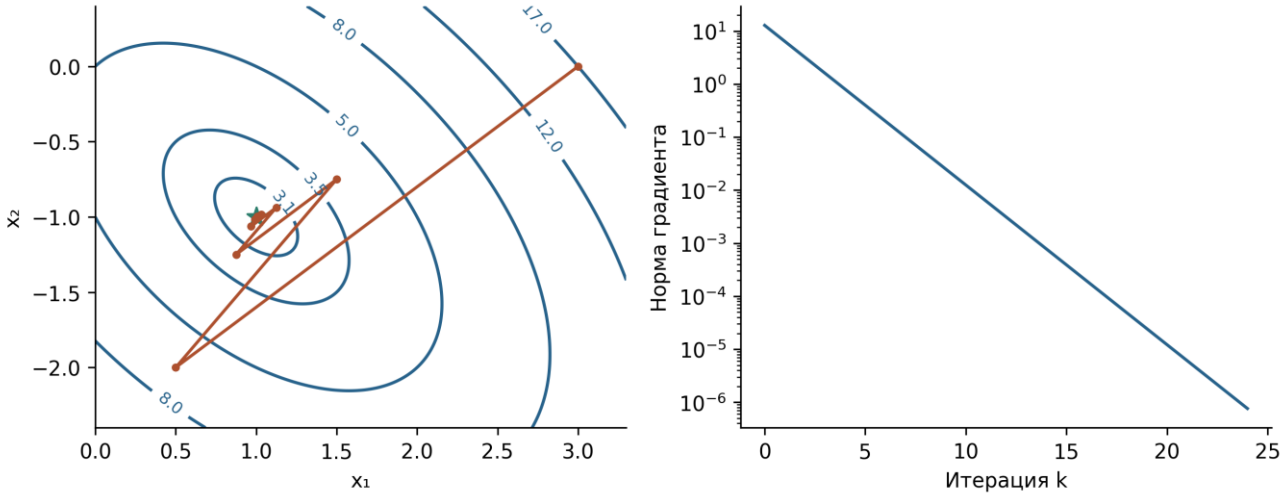


Рисунок 4 — Принятые точки и приближение к точному минимуму квадратичной функции

Траектория пересекает направление к минимуму и возвращается к нему с другой стороны. Такое поведение не является ошибкой: контролируется убывание функции, а не монотонность каждой координаты.

12 Что действительно следует из убывания функции

Если $f(\mathbf{x}^{k+1}) \le f(\mathbf{x}^k)$ и функция ограничена снизу, последовательность значений сходится. Из этого отдельно не следует ни сходимость точек, ни стационарность предельной точки.

Контрпример. Для $f(x) = x^2$ положим $x^k = 1 + 1/(k+1)$. Значения строго убывают и стремятся к 1, но точки сходятся к 1, где $f'(1) = 2 \neq 0$. Эти переходы можно записать с направлением $p_k = -1$ и шагом

$$\alpha_k = \frac{1}{(k+1)(k+2)}.$$

Направление действительно является направлением спуска, но шаги слишком быстро уменьшаются.

Достаточные условия приближения к стационарности. Пусть f ограничена снизу, её градиент глобально липшицев с константой L, используется $\mathbf{p}_k = -\mathbf{g}_k$ и дробление Армихо с фиксированными α_0, β, σ. Тогда

$$f(\mathbf{x}^k) - f(\mathbf{x}^{k+1}) \geq \sigma\alpha_k \parallel \mathbf{g}_k \parallel_2^2 \geq \sigma\underline{\alpha} \parallel \mathbf{g}_k \parallel_2^2.$$

Суммируя от $k = 0$ до N, получаем

$$\sigma\underline{\alpha} \sum_{k=0}^{N} \parallel \mathbf{g}_k \parallel_2^2 \leq f(\mathbf{x}^0) - f(\mathbf{x}^{N+1}) \leq f(\mathbf{x}^0) - f_{\text{low}},$$

где f_{low} — любая нижняя граница функции. Значит, ряд квадратов норм градиента сходится и $\parallel \mathbf{g}_k \parallel_2 \to 0$. Если существует предельная точка последовательности, непрерывность градиента делает её стационарной.

Утверждение не доказывает сходимость всей последовательности к единственной точке и не исключает седло в невыпуклой задаче. Для выпуклой функции стационарность уже даёт глобальный минимум. В основном квадратичном примере строгая выпуклость и оценки по собственным значениям позволяют заключить сходимость к $(1; -1)$.

У невыпуклой функции из раздела 6 градиент не является глобально липшицевым на всей плоскости, поскольку $12x^2 - 4$ неограниченно растёт. Поэтому сформулированную глобальную теорему нельзя применять к ней без уточнения области. Сами ручные вычисления и проверка Армихо в выбранных точках остаются корректными.

13 Критерии остановки и учебная программа

Основной критерий программы — $\parallel \nabla f(\mathbf{x}^k) \parallel_2 \leq \varepsilon_g$. Это означает приближённую стационарность. Дополнительно контролируются максимальное число итераций, конечность вычисленных значений и успешность поиска шага.

Малость $\parallel \mathbf{x}^{k+1} - \mathbf{x}^k \parallel$ может быть вызвана слишком малым шагом. Малость $|f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k)|$ может быть вызвана округлением. Поэтому такие проверки без градиента не подтверждают минимум.

Если известна сильная выпуклость с константой m и существует минимум, справедливы полезные оценки

$$\parallel \mathbf{x} - \mathbf{x}^* \parallel_2 \leq \frac{\parallel \nabla f(\mathbf{x}) \parallel_2}{m}, \quad f(\mathbf{x}) - f(\mathbf{x}^*) \leq \frac{\parallel \nabla f(\mathbf{x}) \parallel_2^2}{2m}.$$

Первая получается из сильной монотонности градиента и неравенства Коши — Буняковского. Для второй в неравенстве сильной выпуклости подставляют $\mathbf{y} = \mathbf{x}^*$ и оценивают снизу квадратичное выражение по $\mathbf{y} - \mathbf{x}$. Эти оценки позволяют выбрать точность по градиенту исходя из требуемой точности результата.

13.1 Порядок запуска

В архиве находится программа descent_foundations.py. Для запуска нужны Python 3.10 или новее, NumPy и Matplotlib:

```
python -m pip install numpy matplotlib
python descent_foundations.py
python descent_foundations.py --variant 7
python descent_foundations.py --nonconvex
python descent_foundations.py --nonconvex --start 0 0
```

По умолчанию решается квадратичный пример раздела 11. Для варианта используется начальная точка $(s + 2; t + 1)$. Невыпуклый пример стартует в $(0{,}2; 1)$. В Windows вместо python можно использовать py.

Программа сохраняет result.json с итоговым статусом и спектром Гессиана, iterations.csv с принятыми точками, trials.csv со всеми пробами шага и рисунок descent.png. Повторный запуск в ту же папку перезаписывает результаты. Другую папку задают параметром --output.

Ключевой фрагмент выбора шага выглядит так:

```
alpha = alpha0
slope = float(g @ p)
for j in range(max_trials):
    candidate = x + alpha * p
    value = float(f(candidate))
    bound = fx + sigma * alpha * slope
    if np.isfinite(value) and value <= bound:
        return alpha
    alpha *= beta
```

Полная версия дополнительно проверяет параметры, строгость направления спуска, размерности и изменение точки. Аналитический градиент задаётся отдельно; реализация не строит его автоматически.

13.2 Как читать результат

Статус gradient_tolerance означает выполнение критерия по градиенту, iteration_limit — достижение лимита. Ни один статус сам по себе не заменяет математическую классификацию точки. Подпрограмма curvature_report исследует только Гессиан; её результат имеет смысл для вывода об экстремуме после проверки стационарности.

Вычисленные собственные значения сравниваются с порогом 10^{-8}. Если значение близко к нулю, программа сообщает, что классификация по кривизне неубедительна. Для другой шкалы функции абсолютный порог надо изменить. Учебные параметры не являются универсальными настройками для любых задач.

Особенно полезно сравнить два запуска невыпуклого примера. Из $(0; 0)$ программа сразу останавливается: градиент нулевой, но Гессиан неопределён. Из $(0{,}2; 1)$ она движется к минимуму около $(1; 0)$. Начальная точка влияет на результат, а число итераций не доказывает глобальность найденного решения.

14 Самостоятельная работа

Используем 18 вариантов квадратичной функции из предыдущих лекций:

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

Начальная точка для этой лекции: $\mathbf{x}^0 = (s+2; t+1)$. Параметры дробления: $\alpha_0 = 1, \beta = 1/2$, $\sigma = 0{,}1$.

№	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
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. Найдите все стационарные точки своего варианта аналитически.
2. Выпишите Гессиан, угловые миноры и его собственные значения. Обоснуйте тип стационарной точки.
3. Докажите глобальность и единственность минимума, используя положительную определённость квадратичной формы.
4. Вычислите $f(\mathbf{x}^0)$, градиент и направление $\mathbf{p}_0 = -\nabla f(\mathbf{x}^0)$. Проверьте знак скалярного произведения.
5. Получите точный многочлен $\varphi(\alpha)$. Вручную выполните дробление до первого принятого шага. Запишите каждую пробу и границу Армихо.
6. Вычислите следующую точку и повторите расчёт для второй итерации. Сравните первый принятый шаг с точным минимумом на прямой.
7. Запустите программу для своего варианта. Сопоставьте ручные расчёты с CSV, укажите итоговую норму градиента и фактическую ошибку относительно $(s; t)$.

8. Для общего невыпуклого примера сравните старты $(0;0)$, $(0{,}2;1)$ и $(-0{,}2;1)$. Объясните различие результатов через стационарные точки и симметрию функции.

Содержание отчёта: исходная функция; аналитическое решение; проверка Гессиана; доказательство глобальности; две ручные итерации с журналом проб; программные результаты; график траектории; вывод о точности и критерии остановки.

14.1 Формулы для самопроверки

Во всех вариантах $a > 0$ и $ac - b^2 > 0$. Поэтому единственная стационарная точка $(s;t)$ является глобальным минимумом со значением d. Гессиан и его собственные значения равны

$$H = \begin{pmatrix} 2a & 2b \\ 2b & 2c \end{pmatrix},$$

$$\lambda_{\min} = a + c - \sqrt{(a-c)^2 + 4b^2}, \quad \lambda_{\max} = a + c + \sqrt{(a-c)^2 + 4b^2}.$$

Можно взять $m = \lambda_{\min}$, $L = \lambda_{\max}$. В начальной точке

$$f_0 = 4a + 4b + c + d, \quad \mathbf{g}_0 = (4a + 2b;\, 4b + 2c)^{\mathsf{T}}.$$

Для $\mathbf{p}_0 = -\mathbf{g}_0$ справедливо точное равенство

$$\varphi(\alpha) = f_0 - \alpha \parallel \mathbf{g}_0 \parallel_2^2 + \frac{\alpha^2}{2} \mathbf{g}_0^{\mathsf{T}} H \mathbf{g}_0.$$

Точный минимум на прямой достигается при

$$\alpha_{\text{exact}} = \frac{\parallel \mathbf{g}_0 \parallel_2^2}{\mathbf{g}_0^{\mathsf{T}} H \mathbf{g}_0}.$$

При $\alpha > 0$ условие Армихо для квадратичной функции эквивалентно

$$\alpha \le 2(1 - \sigma)\alpha_{\text{exact}}.$$

Эта формула позволяет независимо проверить правильность дробления, но не заменяет журнал вычислений в отчёте.

15 Контрольные вопросы и короткие задачи

1. Почему нулевой градиент является необходимым, но не достаточным условием минимума?
2. Какие предпосылки нужны для применения условия первого порядка?
3. Чем положительная определённость отличается от полуопределённости?
4. Почему при нулевом определителе Гессиана нужен дополнительный анализ?
5. Когда локальный минимум можно признать глобальным?
6. Как проверяется строгое направление спуска?
7. Почему слишком большой шаг вдоль антиградиента опасен?
8. Что регулируют параметры σ и β в правиле Армихо?
9. Почему дробление заканчивается за конечное число проб в точной арифметике?
10. Почему монотонного убывания функции недостаточно для сходимости к стационарной точке?
11. Какие условия использованы при доказательстве стремления нормы градиента к нулю?

12. Что означает остановка программы в седловой точке?

Задача 1. Исследуйте функцию $f(x,y) = x^2 + xy + y^2 - 4x - 5y$.

Решение. Система $2x + y = 4, x + 2y = 5$ даёт $(x,y) = (1,2)$. Гессиан имеет вид $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, миноры 2 и 3 положительны. Квадратичная функция строго выпукла и коэрцитивна. Минимум единственный и глобальный, $f(1,2) = -7$.

Задача 2. Классифицируйте начало координат для $x^4 + y^2$.

Решение. Градиент равен нулю, Гессиан diag(0,2) вырожден. Тест положительной определённости не работает. Но функция неотрицательна и равна нулю только в начале координат: это строгий глобальный минимум.

Задача 3. Для $f(x,y) = (x^2 + 4y^2)/2$ из точки $(1;1)$ выполните первую итерацию Армихо с параметрами лекции.

Решение. $f_0 = 2{,}5$, $\mathbf{g} = (1;4)$, $\mathbf{p} = (-1;-4)$, $\varphi(\alpha) = 2{,}5 - 17\alpha + 32{,}5\alpha^2$. Пробы 1 и 0,5 отклоняются: значения равны 18 и 2,125, границы — 0,8 и 1,65. Проба 0,25 принимается: $0{,}28125 \le 2{,}075$. Новая точка $(0{,}75;\,0)$.

Задача 4. Для сильно выпуклой функции известны $m = 2$ и $\| \nabla f(\mathbf{x}) \|_2 \le 10^{-3}$. Оцените ошибки.

Ответ: расстояние до минимума не превосходит $5 \cdot 10^{-4}$, а ошибка значения функции не превосходит $2{,}5 \cdot 10^{-7}$.

16 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, § 1–2: выпуклость, необходимые и достаточные условия безусловного экстремума.

[2] Kochenderfer M. J., Wheeler T. A. Algorithms for Optimization. Cambridge, MA: MIT Press, 2019. Chapter 4, § 4.2–4.3, pp. 54–57: поиск шага и условие достаточного уменьшения. Обозначения параметров в лекции унифицированы: σ — доля уменьшения, β — коэффициент дробления. Утверждения о сходимости сформулированы с явными предпосылками.

[3] Рабочая программа дисциплины «Методы оптимизации», 2024. Образовательная программа 6B06103 «Математическое и компьютерное моделирование». Тема 12 «Методы многомерной оптимизации. Классический метод минимизации функции многих переменных. Теоретические основы методов спуска».