Лекция 11

5. Численные методы многомерной оптимизации

5.1. Выпуклые множества и выпуклые функции

Пусть $\boldsymbol{x}$, $\boldsymbol{y} \in \mathbb{R}^n$. Множество

$$[\boldsymbol{x}, \boldsymbol{y}] = \{\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} : \ \alpha \in [0,1] \ \} \tag{1.1}$$

называется *отрезком*, соединяющим точки $\boldsymbol{x}$ и $\boldsymbol{y}$.

Множество $X \subset \mathbb{R}^n$ называется *выпуклым*, если вместе с любыми точками $\boldsymbol{x}$ и $\boldsymbol{y} \in X$ оно содержит и весь отрезок $[\boldsymbol{x}, \boldsymbol{y}]$.

Множество $\mathbb{R}^n$, очевидно, является выпуклым множеством. Пустое множество $\varnothing$ считается выпуклым множеством.

Теорема 1. *Пусть I — любое, конечное или бесконечное, множество индексов, X_i $(i \in I)$ — выпуклые множества. Тогда их пересечение $X = \bigcap\limits_{i \in I} X_i$ выпукло.*

Доказательство. Пусть $X \neq \varnothing$ и $\boldsymbol{x}$, $\boldsymbol{y} \in X$, $\alpha \in [0,1]$. По определению пересечения для любого $i \in I$ имеем $\boldsymbol{x}$, $\boldsymbol{y} \in X_i$ и значит $\boldsymbol{z} = \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in X_i$ так как X_i выпукло. Тогда $\boldsymbol{z} \in \bigcap\limits_{i \in I} X_i = X$, т. е. X выпукло. ∎

Функция $f(\boldsymbol{x})$, заданная на выпуклом множестве $X \subset \mathbb{R}^n$, называется *выпуклой*, если для любых точек $\boldsymbol{x}$, $\boldsymbol{y} \in X$ и любого $\alpha \in [0,1]$ выполняется неравенство

$$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \le \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}). \tag{1.2}$$

Функция $f(x)$ называется *строго выпуклой* если всех $\alpha \in (0,1)$ неравенство (1.2) выполняется как строгое.

Теорема 2. *Линейная комбинация выпуклых на выпуклом множестве X функций $f_i(\boldsymbol{x})$, $i = 1,\ldots, m$, с неотрицательными коэффициентами λ_i, т. е.*

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}), \quad \lambda_i \ge 0,$$

есть выпуклая на множестве X функция.

Доказательство. Для любых $\boldsymbol{x}$, $\boldsymbol{y} \in X$, $\alpha \in [0,1]$ имеем

$$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) = \sum_{i=1}^{m} \lambda_i f_i(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \le$$

$$\le \sum_{i=1}^{m} \lambda_i (\alpha f_i(\boldsymbol{x}) + (1-\alpha) f_i(\boldsymbol{y})) = \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}),$$

т. е. выполняется (1.2). Значит, функция $f(\boldsymbol{x})$ выпукла. ∎

Теорема 3. *Пусть $f(\boldsymbol{x})$ — выпуклая функция, заданная в пространстве $\mathbb{R}^n$. Тогда множество*

$$X = \{\boldsymbol{x} \in \mathbb{R}^n : \ f(\boldsymbol{x}) \le b\}$$

выпукло.

Доказательство. Пусть $\boldsymbol{x}, \boldsymbol{y} \in X$ и $\boldsymbol{z} = \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}$, $\alpha \in [0,1]$. Из выпуклости функции $f(\boldsymbol{x})$ следует, что $f(\boldsymbol{z}) \le \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) \le b$, т. е. $\boldsymbol{z} \in X$ и множество X — выпукло. ■

Следствие. *Пусть $f_i(\boldsymbol{x})$, $i = 1, \dots, m$, — выпуклые функции в $\mathbb{R}^n$. Тогда множество точек $\boldsymbol{x}$, удовлетворяющих системе неравенств*

$$f_i(\boldsymbol{x}) \le b_i, \quad i = 1, \dots, m,$$

выпукло.

Доказательство. Это следует из теорем 3.1 и 3.3. ■

Приведём свойства выпуклых функций, играющих важную роль в вопросах минимизации.

Теорема 4. *Пусть $f(\boldsymbol{x})$ — выпуклая функция, заданная на выпуклом множестве X. Тогда любой её локальный минимум на множестве X является одновременно и глобальным.*

Доказательство. Предположим противное, т. е. пусть $\boldsymbol{x}_0$ — точка локального, а $\boldsymbol{x}_*$ — точка глобального минимума $f(\boldsymbol{x})$ на множестве X, $\boldsymbol{x}_* \ne \boldsymbol{x}_0$ и $f(\boldsymbol{x}_0) > f(\boldsymbol{x}_*)$. Отсюда с учётом выпуклости функции имеем:

$$f(\alpha\boldsymbol{x}_* + (1-\alpha)\boldsymbol{x}_0) \le \alpha f(\boldsymbol{x}_*) + (1-\alpha) f(\boldsymbol{x}_0) < f(\boldsymbol{x}_0).$$

При $\alpha \to +0$ точка $\boldsymbol{x} = \alpha\boldsymbol{x}_* + (1-\alpha)\boldsymbol{x}_0$ попадёт в сколь угодно малую окрестность точки $\boldsymbol{x}_0$. Поэтому полученное неравенство $f(\boldsymbol{x}) < f(\boldsymbol{x}_0)$ противоречит предположению о том, что $\boldsymbol{x}_0$ — точка локального минимума. ■

Теорема 5. *Глобальный минимум строго выпуклой функции $f(\boldsymbol{x})$, заданной на на выпуклом множестве X, может достигаться лишь в единственной точке.*

Доказательство. Предположим, что $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$ — две различные точки глобального минимума. Из строгой выпуклости $f(\boldsymbol{x})$ следует, что для всех $\alpha \in (0,1)$ выполняется строгое неравенство

$$f(\alpha\boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) < \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2) = f_* = \min_{\boldsymbol{x}\in X} f(\boldsymbol{x}),$$

что противоречит предположению о том, что точки $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$ — точки глобального минимума. ■

Функция $f(\boldsymbol{x})$, заданная в пространстве $\mathbb{R}^n$, называется *сильно выпуклой*, если существует такое число $l > 0$ (*константа сильной выпуклости*), что для всех $\boldsymbol{x}$ и $\boldsymbol{y} \in \mathbb{R}^n$ и любого $\alpha \in [0,1]$ выполняется неравенство:

$$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \le \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y}) - \alpha(1-\alpha)\, l\, |\boldsymbol{x} - \boldsymbol{y}|^2. \tag{1.3}$$

5.2. Выпуклые квадратичные функции

Функция вида

$$f(\boldsymbol{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_{ij} x_i x_j + \sum_{j=1}^{n} b_j x_j + c \tag{2.1}$$

называется *квадратичной функцией n переменных*. Положив $a_{ij} = \alpha_{ij} + \alpha_{ji}$, получим симметрическую матрицу $A = (a_{ij})$, с помощью которой выражение (2.1) можно записать в другой форме

$$f(\boldsymbol{x}) = \frac{1}{2} < A\boldsymbol{x}, \boldsymbol{x} > + < \boldsymbol{b}, \boldsymbol{x} > + c \tag{2.2}$$

где $\boldsymbol{b} = (b_1, \ldots, b_n)^t \in \mathbb{R}^n$ — вектор коэффициентов b_j.

Пример 1. Функция

$$f(\boldsymbol{x}) = 2x_1^2 - 2x_1x_2 + 3x_1x_3 + x_2^2 - 2x_2x_3 + 4x_3^2 + x_1 - x_2 + 3x_3 + 5$$

является квадратичной. Запишем её матрицу A, вектор $\boldsymbol{b}$ и коэффициент c из (2.2):

$$A = \begin{pmatrix} 4 & -2 & 3 \\ -2 & 2 & -2 \\ 3 & -2 & 8 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}, \quad c = 5.$$

Перечислим основные свойства квадратичных функций.

1. Для градиента квадратичной функции справедлива формула:

$$\nabla f(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}. \tag{2.3}$$

Доказательство. Имеем

$$\frac{\partial}{\partial x_k} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} x_i x_j = \frac{\partial}{\partial x_k} \left\{ \sum_{i=1}^{n} \alpha_{ii} x_i^2 + \sum_{i \neq j} \alpha_{ij} x_i x_j \right\} =$$

$$= 2\alpha_{kk} x_k + \sum_{i \neq k} (\alpha_{ik} + \alpha_{ki}) x_i = 2 \sum_{i=1}^{n} a_{ki} x_i = 2(A\boldsymbol{x})_k$$

Далее

$$\frac{\partial}{\partial x_k} \sum_{j=1}^{n} b_j x_j = b_k.$$

Следовательно, равенство (2.3) доказано. ■

2. Матрица Гессе квадратичной функции (2.2) совпадает с матрицей A:

$$f''(\boldsymbol{x}) = A. \tag{2.4}$$

Доказательство. Вычислим элемент матрицы Гессе:

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial x_l \, \partial x_k} = \frac{\partial}{\partial x_l} \left(\frac{\partial f(\boldsymbol{x})}{\partial x_k} \right) = \frac{\partial}{\partial x_l} \left(\sum_{i=1}^{n} a_{ki} x_i + b_k \right) = a_{kl}.$$

Равенство (2.4) доказано. ■

2. Квадратичной функция (2.2) с положительно определённой матрицей A сильно выпукла.

Доказательство. ■

Пример 2. Рассмотрим квадратичную функцию из примера 1. Матрица Гессе $f''(\boldsymbol{x}) = A$ — положительно определена, так как

$$\Delta_1 = 4 > 0; \ \Delta_2 = \begin{vmatrix} 4 & -2 \\ -2 & 2 \end{vmatrix} = 4 > 0; \ \Delta_3 = \det A = 22 > 0.$$

Следовательно, $f(\boldsymbol{x})$ сильно выпукла по свойству **3** квадратичных функций.

5.3. Общие принципы n-мерной минимизации

Для численного решения задач безусловной минимизации:

$$f(\boldsymbol{x}) \to \min, \quad \boldsymbol{x} \in \mathbb{R}^n$$

разработано много алгоритмов, использующих итерационные процедуры вида

$$\boldsymbol{x}^{k+1} = \Phi(\boldsymbol{x}^k, \boldsymbol{x}^{k-1}, \ldots, \boldsymbol{x}^0), \quad \boldsymbol{x}^0 \in \mathbb{R}^n, \tag{3.1}$$

позволяющие при определённых условиях построить последовательность $\{\boldsymbol{x}^k\}$ такую, что

$$\lim_{k\to\infty} f(\boldsymbol{x}^k) = \begin{cases} f_* = \min\limits_{\mathbb{R}^n} f(\boldsymbol{x}), & X_* \neq \varnothing, \\ f_* = \inf\limits_{\mathbb{R}^n} f(\boldsymbol{x}), & X_* = \varnothing, \end{cases} \tag{3.2}$$

где X_* — множество точек глобального минимума функции $f(\boldsymbol{x})$. Последовательность $\{\boldsymbol{x}^k\}$, удовлетворяющая требованию (3.2), называется *минимизирующей* для функции $f(\boldsymbol{x})$.

Минимизирующая последовательность может и не сходиться к точке минимума (см. пример 4.1.3). Важной характеристикой сходящихся минимизирующих последовательностей является *скорость сходимости.*

Последовательность $\{\boldsymbol{x}^k\}$ сходится к точке $\boldsymbol{x}_*$ *линейно* (*со скоростью геометрической прогрессии*), если существует такое число $q \in (0, 1)$, что выполняется неравенство

$$\left|\boldsymbol{x}^k - \boldsymbol{x}_*\right| \leq q \left|\boldsymbol{x}^{k-1} - \boldsymbol{x}_*\right| \tag{3.3}$$

Последовательность $\{\boldsymbol{x}^k\}$ сходится к точке $\boldsymbol{x}_*$ *сверхлинейно*, если

$$\left|\boldsymbol{x}^k - \boldsymbol{x}_*\right| \leq q_k \left|\boldsymbol{x}^{k-1} - \boldsymbol{x}_*\right|, \quad q_k \to +0, \quad k \to \infty. \tag{3.4}$$

Последовательность $\{\boldsymbol{x}^k\}$ сходится к точке $\boldsymbol{x}_*$ с *квадратичной скоростью*, если существует константа C такая, что

$$\left|\boldsymbol{x}^k - \boldsymbol{x}_*\right| \leq C \left|\boldsymbol{x}^{k-1} - \boldsymbol{x}_*\right|^2. \tag{3.5}$$

Конкретный вычислительный процесс на основе (3.1), в котором может получаться, вообще говоря, бесконечная последовательность $\{\boldsymbol{x}^k\}$, необходимо дополнять условием остановки (критерием окончания счёта). На практике часто пользуются следующими условиями:

$$\left|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\right| < \varepsilon_1, \tag{3.6}$$

$$\left| f(\boldsymbol{x}^{k+1}) - f(\boldsymbol{x}^k) \right| < \varepsilon_2, \tag{3.7}$$

$$\left| \nabla f(\boldsymbol{x}^k) \right| < \varepsilon_3, \tag{3.8}$$

где ε_1, ε_2, ε_3 — заранее заданные параметры точности.

Ниже будут рассматриваться вычислительные алгоритмы простейших процедур (3.1), основанные на рекуррентных формулах вида

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha_k \boldsymbol{p}^k, \quad k = 0, 1, \ldots, \tag{3.9}$$

где $\boldsymbol{p}^k$ — направление поиска точки $\boldsymbol{x}^{k+1}$ из точки $\boldsymbol{x}^k$, а число α_k — величина шага, которая выбирается так, чтобы выполнялось условие

$$f\big(\boldsymbol{x}^{k+1}\big) < f\big(\boldsymbol{x}^k\big). \tag{3.10}$$

Эти алгоритмы различаются способом построения вектора $\boldsymbol{p}^k$ и выбора шага α_k.

Будем говорить, что в итерационном процессе (3.9) производится *исчерпывающий спуск*, если величина шага α_k находится из решения одномерной задачи минимизации

$$\Phi_k(\alpha) \to \min, \quad \Phi_k(\alpha) = f\big(\boldsymbol{x}^k + \alpha \boldsymbol{p}^k\big). \tag{3.11}$$

Таким образом, при исчерпывающем спуске на каждом шаге полностью реализуется возможность уменьшить значение целевой функции $f(\boldsymbol{x})$ при перемещении из точки $\boldsymbol{x}^k$ в направлении $\boldsymbol{p}^k$. Величина шага α_k может быть найдена методами одномерной оптимизации.

Теорема 1. *Для дифференцируемой в $\mathbb{R}^n$ функции $f(\boldsymbol{x})$ в итерационном процессе* (3.9) *с выбором шага α_k в соответствии с* (3.11) *для всех $k \geq 1$ выполняется условие*

$$< \nabla f(\boldsymbol{x}^{k+1}), \boldsymbol{p}^k >= 0. \tag{3.12}$$

Доказательство. Запишем необходимое условие минимума функции одной переменной $\Phi_k(\alpha)$ из (3.11):

$$\frac{d\Phi_k(\alpha)}{d\alpha} = \sum_{j=1}^{n} \frac{\partial f(\boldsymbol{x}^{k+1})}{\partial x_j} \frac{dx_j^{k+1}}{d\alpha} = 0.$$

Учитывая, что $x_j^{k+1} = x_j^k + \alpha p_j^k$, получаем условие (3.12). ■

Свойство (3.12) позволяет в явном виде найти величину α_k для квадратичной функции.

Теорема 2. *Для квадратичной функции*

$$f(\boldsymbol{x}) = \frac{1}{2} < A\boldsymbol{x}, \boldsymbol{x} > + < \boldsymbol{b}, \boldsymbol{x} > + c$$

величина α_k исчерпывающего спуска в итерационном процессе (3.9) *будет*

$$\alpha_k = -\frac{< \nabla f(\boldsymbol{x}^k), \boldsymbol{p}^k >}{< A\boldsymbol{p}^k, \boldsymbol{p}^k >} = -\frac{< A\boldsymbol{x}^k + \boldsymbol{b}, \boldsymbol{p}^k >}{< A\boldsymbol{p}^k, \boldsymbol{p}^k >} \tag{3.13}$$

Доказательство. Умножив равенство (3.9) слева на матрицу A квадратичной функции $f(\boldsymbol{x})$ и прибавив к обеим частям вектор $\boldsymbol{b}$, получим:

$$A\boldsymbol{x}^{k+1} + \boldsymbol{b} = A\boldsymbol{x}^k + \boldsymbol{b} + \alpha_k A\boldsymbol{p}^k.$$

Учитывая, что $\nabla f(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}$, имеем $\nabla f(\boldsymbol{x}^{k+1}) = \nabla f(\boldsymbol{x}^k) + \alpha_k A\boldsymbol{p}^k$. Подставляя выражение для $\nabla f(\boldsymbol{x}^{k+1})$ в равенство (3.12), получаем формулу (3.13). ■

Направление вектора $\boldsymbol{p}$ называется *направлением убывания* функции $f(\boldsymbol{x})$ в точке $\boldsymbol{x}$, если при всех достаточно малых $\alpha > 0$ выполняется неравенство $f(\boldsymbol{x}+\alpha\boldsymbol{p}) < f(\boldsymbol{x})$

Теорема 3. *Пусть функция $f(\boldsymbol{x})$ дифференцируема в точке $\boldsymbol{x}^k$. Если вектор $\boldsymbol{p}^k$ удовлетворяет условию*

$$< \nabla f(\boldsymbol{x}^k), \boldsymbol{p}^k >< 0, \tag{3.14}$$

то направление вектора $\boldsymbol{p}^k$ является направлением убывания.

Доказательство. Из дифференцируемости функции $f(\boldsymbol{x})$ и условия (3.14) следует, что

$$f(\boldsymbol{x}^k + \alpha\boldsymbol{p}^k) - f(\boldsymbol{x}^k) =< \nabla f(\boldsymbol{x}^k), \alpha\boldsymbol{p}^k > + o(\alpha) = \alpha\left(< \nabla f(\boldsymbol{x}^k), \alpha\boldsymbol{p}^k > + \frac{o(\alpha)}{\alpha}\right) < 0$$

при достаточно малых $\alpha > 0$. Таким образом, мы видим, что вектор $\boldsymbol{p}^k$ задаёт направлением убывания функции $f(\boldsymbol{x})$. ■

Лекция 12

5.4. Метод (циклического) покоординатного спуска

Этот метод заключается в последовательной минимизации целевой функции $f(\boldsymbol{x})$ сначала по направлению первого базисного вектора $\boldsymbol{e}^1$, затем второго — $\boldsymbol{e}^2$ и т. д. После окончания минимизации по направлению последнего базисного вектора $\boldsymbol{e}^n$ цикл повторяется.

Опишем этот алгоритм.

0. Выбрать $\boldsymbol{x} \in \mathbb{R}^n$, критерий достижения точности (например, (3.6) или (3.7)), величину ε. Найти $f(\boldsymbol{x})$, положить $j = 1$.

1. Решить задачу одномерной оптимизации

$$\Phi(\alpha) = f(\boldsymbol{x} + \alpha \boldsymbol{e}^j) \to \min, \quad \alpha \in \mathbb{R},$$

т. е. найти α_*. Положить $\hat{\boldsymbol{x}} = \boldsymbol{x} + \alpha_* \boldsymbol{e}^j$, вычислить $f(\hat{\boldsymbol{x}})$.

2. Если $j < n$, то положить $\boldsymbol{x} = \hat{\boldsymbol{x}}$, $j = j + 1$ и перейти к **1**, иначе — перейти к **3**.

3. Проверить условие достижения точности $|\boldsymbol{x} - \hat{\boldsymbol{x}}| < \varepsilon$ или $|f(\boldsymbol{x}) - f(\hat{\boldsymbol{x}})| < \varepsilon$. Если оно выполняется, то положить $\boldsymbol{x}_* = \hat{\boldsymbol{x}}$, $f_* = f(\hat{\boldsymbol{x}})$ и закончить поиск. Иначе — положить $\boldsymbol{x} = \hat{\boldsymbol{x}}$, $f(\boldsymbol{x}) = f(\hat{\boldsymbol{x}})$, $j = 1$ и перейти к **1**.

Пример 1. Рассмотрим задачу

$$f(\boldsymbol{x}) = x_1^2 + x_2^2 \ \to \ \min, \quad \boldsymbol{x} \in \mathbb{R}^2.$$

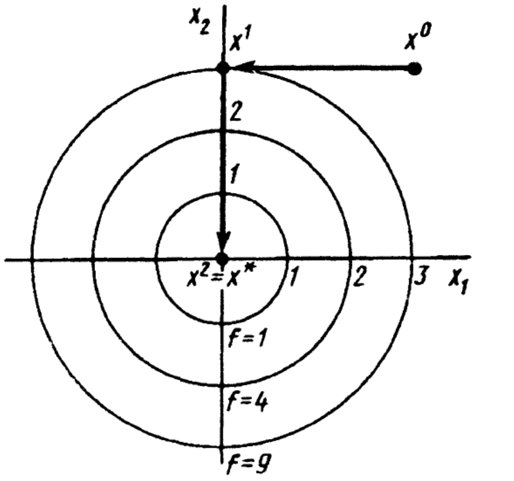
Линии уровня этой целевой функции — окружности с центром в начале координат. Выберем произвольную начальную точку $\boldsymbol{x}$, например $\boldsymbol{x} = (3,3)^t$. Очевидно, два шага исчерпывающего спуска сначала по направлению $\boldsymbol{e}^1$, затем — $\boldsymbol{e}^2$ приведут в точку минимума $((3,3) \to (0,3) \to (0,0))$.

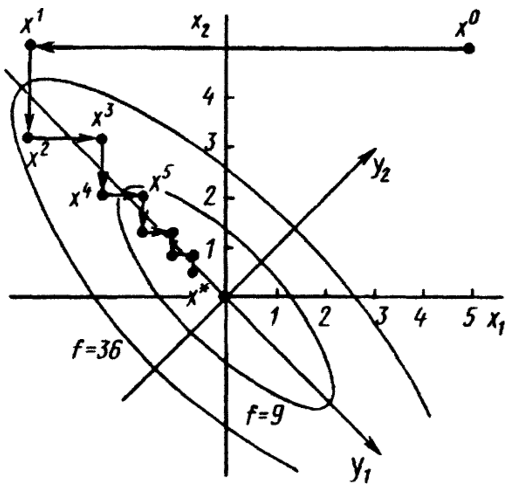
Пример 2. Рассмотрим задачу

$$f(\boldsymbol{x}) = 5x_1^2 + 8x_1x_2 + 5x_2^2 \ \to \ \min, \quad \boldsymbol{x} \in \mathbb{R}^2..$$

Ниже приведены результаты 5 итераций алгоритма покоординатного спуска.

	x_1	x_2	$f(\boldsymbol{x})$
0	5	5	450
1_1	−4	5	45
1_2	−4	3.2	28.8
2_1	−2.56	3.2	18.432
2_2	−2.56	2.048	11.79648
3_1	−1.6384	2.048	7.54975
3_2	−1.6384	1.31072	4.83183
4_1	−1.04858	1.31072	3.09238
4_2	−1.04858	0.83886	1.97912
5_1	−0.67109	0.83886	1.26664
5_2	−0.67109	0.53687	0.81065





5.5. Метод градиентного спуска

Все итерационные процессы вида

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha_k \nabla f(\boldsymbol{x}^k), \quad \alpha_k > 0, \quad k = 0, 1, \ldots, \tag{5.1}$$

в которых направление движения совпадает с антиградиентом функции, называются *градиентными методами* и отличаются друг от друга способами выбора шага α_k.

Существуют много различных способов выбора α_k, но наиболее распространены два: первый называется *методом с дроблением шага* и связан с проверкой на каждой итерации некоторого неравенства; во втором производится исчерпывающий спуск — *метод наискорейшего спуска.*

Рассмотрим процесс (5.1). Первая проблема, которая возникает при его реализации, — это выбор шага α_k. Достаточно малый шаг α_k обеспечит убывание функции, т. е. выполнение неравенства

$$f\big(\boldsymbol{x}^k - \alpha_k \nabla f(\boldsymbol{x}^k)\big) < f(\boldsymbol{x}^k), \tag{5.2}$$

но может привести к неприемлемо большому количеству итераций, необходимых для достижения точки минимума. С другой стороны, слишком большой шаг может вызвать неожиданный рост функции (невыполнение условия (5.2)) либо привести к колебаниям около точки минимума.

Пример 1. Рассмотрим задачу минимизации функции $f(x) = ax^2$, где a — некоторое положительное число. Тогда формула (5.1) принимает вид

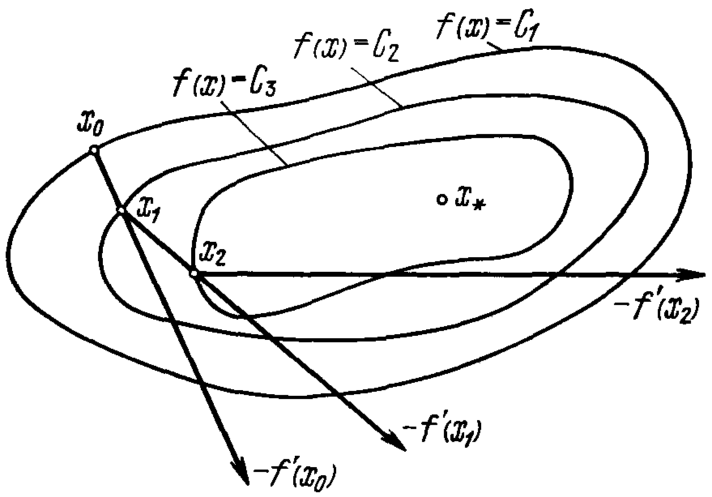
$$x^{k+1} = x^k - \alpha_k \cdot 2ax^k = (1 - 2\alpha_k a)x^k.$$

Очевидно, что при постоянном шаге α_k соответствующий процесс будет сходиться, если $0 < \alpha_k < 1/a$, и расходиться для $\alpha_k > 1/a$. Если положить $\alpha_k = 1/a$, то $x_1 = -x_0$, $x_2 = x_0$, $x_3 = -x_0$ и т. д. Процесс будет расходящимся.

В методе градиентного спуска с дроблением шага величина α_k выбирается так, чтобы было выполнено следующее неравенство:

$$f\big(\boldsymbol{x}^k - \alpha_k \nabla f(\boldsymbol{x}^k)\big) - f(\boldsymbol{x}^k) \leq -\varepsilon\alpha_k \big|\nabla f(\boldsymbol{x}^k)\big|^2, \tag{5.3}$$

где $0 < \varepsilon < 1$ — произвольно выбранная постоянная (одна и та же для всех итераций). Очевидно, что требование (5.3) на выбор шага более жёсткое, чем условие (5.3), но имеет тот же смысл: функция должна убывать от итерации к итерации.



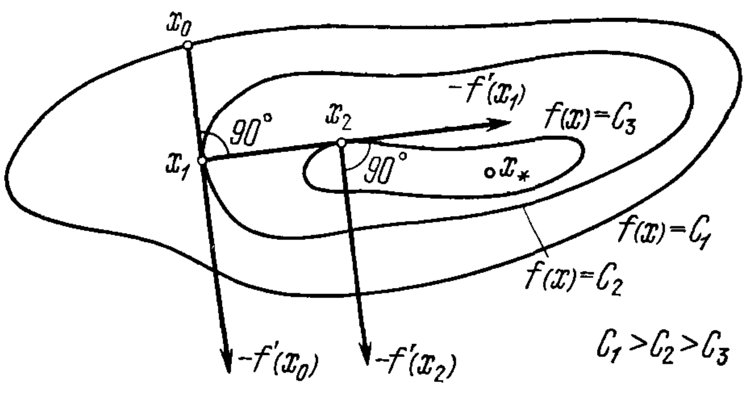
Процесс (5.1) с выбором шага, удовлетворяющего неравенству (5.3), протекает следующим образом. Выбираем число $\alpha > 0$, одно и то же для всех итераций. На k-й итерации проверяем выполнение неравенства (5.3) при $\alpha_k = \alpha$. Если оно выполнено, полагаем $\alpha_{k+1} = \alpha$ и переходим к следующей итерации. Если нет, то шаг α_k дробим до тех пор, пока оно не выполнится. Геометрически градиентный спуск с дроблением шага изображен на рисунке. Здесь изображены линии уровня функции $f(\boldsymbol{x})$, имеющей минимум в точке x_*, причём $C_1 > C_2 > C_3 > \ldots$, и некоторая зигзагообразная траектория $x_0x_1 \ldots x_k$, ортогональная в каждой точке $x_0, x_1, \ldots, x_k$ соответствующим линиям уровня и приводящая из начальной точки x_0 в точку минимума x_*.

5.6. Метод наискорейшего спуска

Процесс, на каждой итерации которого шаг выбирается из условия минимума функции $f(\boldsymbol{x})$ в направлении движения, т. е.

$$f\big(\boldsymbol{x}^k - \alpha_k \nabla f(\boldsymbol{x}^k)\big) = \min_{\alpha \geq 0} f\big(\boldsymbol{x}^k - \alpha \nabla f(\boldsymbol{x}^k)\big), \tag{6.1}$$

называется *методом наискорейшего спуска.* В этом варианте градиентного спуска на каждой итерации требуется решать задачу одномерной минимизации. Разумеется, этот способ выбора α_k сложнее, чем рассмотренные в предыдущем пункте.



Геометрическая интерпретация метода наискорейшего спуска представлена на рисунке. В этом методе, в отличие от обычного градиентного спуска, направление движения из точки x_k касается линии уровня в точке x_{k+1}. Последовательность точек $x_0, x_1, \ldots, x_k, \ldots$ зигзагообразно приближается к точке минимума x_*, причём звенья этого зигзага ортогональны между собой. В самом деле, шаг α выбирается из условия минимизации по α функции

$$\varphi(\alpha) = f\big(\boldsymbol{x}^k - \alpha\nabla f(\boldsymbol{x}^k)\big),$$

и поэтому

$$\frac{d\varphi(\alpha_k)}{d\alpha} = -\nabla f(\boldsymbol{x}^{k+1})\nabla f(\boldsymbol{x}^k) = 0.$$

Таким образом, направления спуска на двух последовательных итерациях взаимно ортогональны.

Пример 1. Рассмотрим функцию $f(\boldsymbol{x}) = x_1^2 + 100x_2^2$ и используем метод наискорейшего спуска для решения задачи её минимизации из начальной точки $\boldsymbol{x}^0 = (1,1)^t$. Имеем

	x_1	x_2	$f(\boldsymbol{x})$	$\nabla f(\boldsymbol{x})$
0	1	1	101	200.01
1	0.9899990	−0.0000990	0.9800990	1.9800970
2	0.00970397	0.00970396	0.00951085	1.9408890
3	0.00960692	$-9.6069\cdot 10^{-7}$	0.00009229	0.0192148

5.7. Метод Ньютона

Пусть функция $f(\boldsymbol{x})$ дважды дифференцируема в $\mathbb{R}^n$. Тогда для неё можно записать разложение по формуле Тейлора в окрестности точки $\boldsymbol{x}^k$:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^k) + < f'(\boldsymbol{x}^k), \boldsymbol{x}-\boldsymbol{x}^k > + \frac{1}{2} < f''(\boldsymbol{x}^k)(\boldsymbol{x}-\boldsymbol{x}^k), \boldsymbol{x}-\boldsymbol{x}^k > + o(|\boldsymbol{x}-\boldsymbol{x}^k|^2). \quad (7.1)$$

Отсюда видно, что повеление функции $f(\boldsymbol{x})$ с точностью до величины порядка $o(|\boldsymbol{x}-\boldsymbol{x}^k|^2)$ может быть описано квадратичной функцией

$$\Phi_k(\boldsymbol{x}) = \frac{1}{2} < f''(\boldsymbol{x}^k)(\boldsymbol{x}-\boldsymbol{x}^k), (\boldsymbol{x}-\boldsymbol{x}^k) > + < f'(\boldsymbol{x}^k), \boldsymbol{x}-\boldsymbol{x}^k > + f(\boldsymbol{x}^k). \quad (7.2)$$

Минимизируем функцию $\Phi_k(\boldsymbol{x})$ вместо $f(\boldsymbol{x})$. Найдём её точку минимума $\boldsymbol{x}^{k+1}$ из условия $\Phi_k'(\boldsymbol{x}) = 0$:

$$\Phi_k'(\boldsymbol{x}) = f''(\boldsymbol{x}^k)(\boldsymbol{x}-\boldsymbol{x}^k) + f'(\boldsymbol{x}^k) = 0. \quad (7.3)$$

Пусть матрица Гессе $f''(\boldsymbol{x})$ положительно определена при всех $\boldsymbol{x} \in \mathbb{R}^n$ и, следовательно, невырождена ($\det f''(\boldsymbol{x}) > 0$). Тогда существует обратная матрица $[f''(\boldsymbol{x})]^{-1}$. Отметим, что квадратичная функция (7.2) с положительно определённой матрицей $f''(\boldsymbol{x}^k)$ сильно выпукла и уравнение (7.3) определяет единственную точку глобального минимума функции $\Phi_k(\boldsymbol{x})$. Умножим слева обе части равенства (7.3) на матрицу $[f''(\boldsymbol{x}^k)]^{-1}$ и найдём точку минимума $\boldsymbol{x}^{k+1}$ квадратичной функции (7.2), аппроксимирующей $f(\boldsymbol{x})$ в окрестности точки $\boldsymbol{x} = \boldsymbol{x}^k$:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - [f''(\boldsymbol{x}^k)]^{-1} \cdot f'(\boldsymbol{x}^k), \quad k = 0, 1 \ldots \quad (7.5)$$

Итерационный процесс (7.5), начатый из произвольной точки $\boldsymbol{x}^0 \in \mathbb{R}^n$, называется *методом Ньютона минимизации функции многих переменных.*

Очевидно, что для квадратичной функции с положительно определённой матрицей A применение метода Ньютона обеспечивает получение точки глобального минимума ровно за один шаг из любой точки $\boldsymbol{x}^0 \in \mathbb{R}^n$.

Недостатком метода Ньютона является необходимость вычисления и обращения матрицы Гессе на каждой итерации.