

### Метод покоординатного спуска

В основе метода лежит последовательная оптимизация  $f(x)$  по каждому из параметров. Рассмотрим метод на примере функции двух переменных.

Пусть  $X_0 = (x_1^{(0)}; x_2^{(0)})$  — начальная точка поиска, которой задаются.

Каждая итерация разбивается на количество шагов, равное количеству переменных — в данном случае на два.

На каждом шаге изменяются значения одной конкретной переменной, все остальные переменные на это время становятся фиксированными параметрами. Таким образом, задача сводится к оптимизации функции одной переменной.

На примере:

Итерация первая, шаг первый:

$$x_1 = \text{var}, x_2^{(0)} = \text{const.}$$

Оптимизируем функцию  $f(x_1, x_2^{(0)})$ , находим ее минимум в точке  $x_1^*$ .

Итерация первая, шаг второй:

$$x_1^* = \text{const}, x_2 = \text{var.}$$

Оптимизируем функцию  $f(x_1^*, x_2)$ , находим ее минимум в точке  $x_2^*$ .

Итерация первая, проверка точности:

Проверяем поставленные условия по точности для  $f(x_1^*, x_2^*)$ , они могут быть следующими:

- $\Delta x_i = |x_i^* - x_i^{(0)}| \leq \varepsilon_{xi}$  — по каждой отдельной координате

- $\Delta X = \sqrt{\sum (x_i^* - x_i^{(0)})^2} \leq \varepsilon_X$  — по общему расстоянию

- $f'(x_i^*) \leq \varepsilon_{yi}$  — по значению частной производной целевой функции по каждой отдельной координате

Итерация вторая (и любая последующая) производится, если нужная точность не достигнута, и состоит из тех же действий, что и первая, только в качестве начальной точки поиска выступает найденная в предыдущей итерации.

Важно отметить, что метод покоординатного спуска неприменим, если в целевую функцию входит произведение каких-либо двух переменных.

Пример: определить экстремум функции  $f(x_1, x_2) = 4(x_1 - 5)^2 + (x_2 - 6)^2$  с точностью  $\varepsilon_{x1} = \varepsilon_{x2} = \varepsilon_{y1} = \varepsilon_{y2} = 0,1$ .

Зададимся начальным приближением:

$$x_1^{(0)} = 8$$

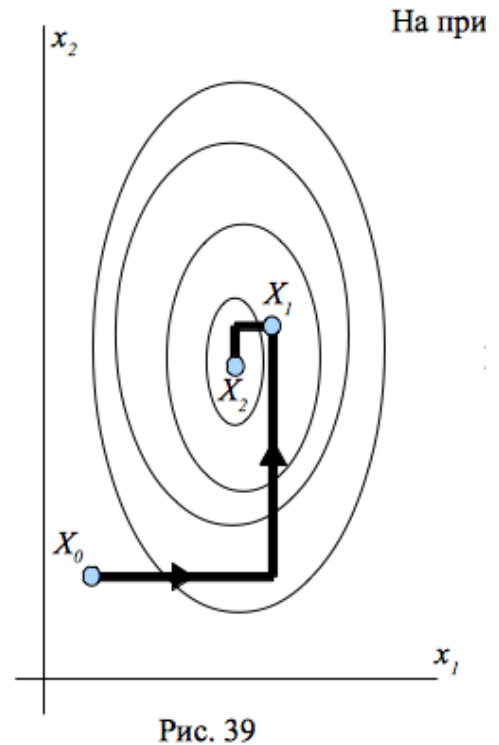
$$x_2^{(0)} = 8$$

$$f(X_0) = 4(8 - 5)^2 + (8 - 6)^2 = 40$$

Итерация первая, шаг первый:

$$x_1 = \text{var}, x_2^{(0)} = 8.$$

Найдем минимум функции  $f(x_1) = 4(x_1 - 5)^2 + (8 - 6)^2 = 4(x_1 - 5)^2 + 4$  с помощью метода Ньютона:



$$\begin{aligned}f'(x_1) &= 8(x_1 - 5) \\f''(x_1) &= 8\end{aligned}$$

В качестве начального приближения возьмем  $x_1^{(0)} = 8$ .

$$x_1^* = x_1^{(0)} - \frac{f'(x_1^{(0)})}{f''(x_1^{(0)})} = 8 - \frac{24}{8} = 5.$$

Проверим точность:

$$f'(x_1^*) = 8(5 - 5) = 0 < \varepsilon_{y1}.$$

Примечание: полезно знать, что если функция имеет квадратичную форму, то метод Ньютона находит ее экстремум с предельной точностью за одну итерацию.

Итерация первая, шаг второй:

$$x_1^* = 5, x_2 = var.$$

Найдем минимум функции  $f(x_2) = 4(5 - 5)^2 + (x_2 - 6)^2 = (x_2 - 6)^2$  с помощью метода Ньютона:

$$\begin{aligned}f'(x_2) &= 2(x_2 - 6) \\f''(x_2) &= 2\end{aligned}$$

В качестве начального приближения возьмем  $x_2^{(0)} = 8$ .

$$x_2^* = x_2^{(0)} - \frac{f'(x_2^{(0)})}{f''(x_2^{(0)})} = 8 - \frac{4}{2} = 6.$$

Проверим точность:

$$f'(x_2^*) = 2(6 - 6) = 0 < \varepsilon_{y2}.$$

Учитывая тот факт, что точности в каждом шаге были предельными, дальнейшие итерации производить не представляется возможным. Целевая функция имеет минимум, равный нулю, в точке (5; 6).

## 7.2.2. Методы первого порядка. Градиентные методы

### Понятие градиента

Если  $f(x_1, x_2, \dots, x_n)$  непрерывна и дифференцируема, то градиентом этой функции называется  $n$ -мерный вектор, элементами которого являются первые частные производные по каждой переменной:

$$\text{grad} f(x_1, x_2, \dots, x_n) = \left\{ \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right\}.$$

Если  $\text{grad} f(X_0) \neq 0$ , то он ортогонален изолинии  $f(x) = \text{const} = f(X_0)$  и направлен в

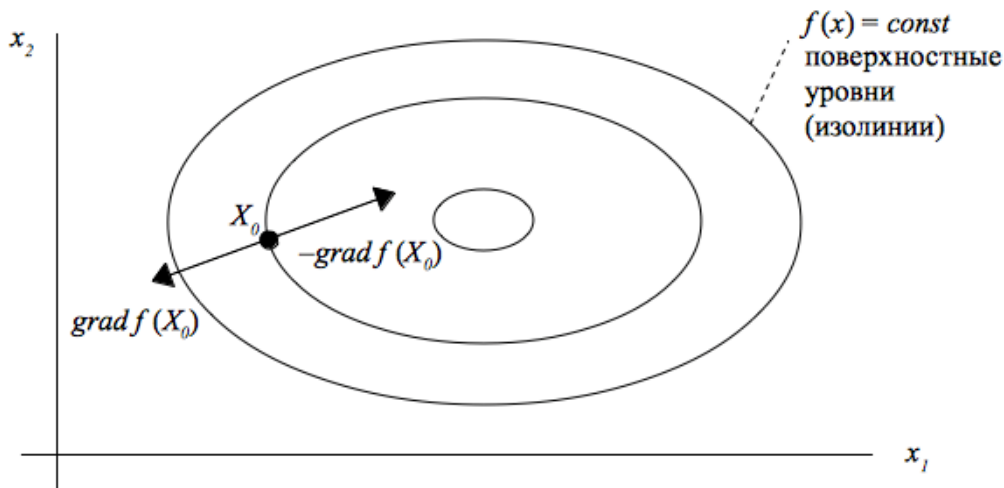


Рис. 40

сторону наискорейшего возрастания функции. Тогда антиградиент,  $-grad f(X_0)$ , направлен в сторону ее скорейшего убывания.

Абсолютная величина градиента равна геометрической сумме векторов, равных значениям частных производных и направленных по соответствующим координатным осям:

$$|grad f| = \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 + \dots + \left(\frac{\partial f}{\partial x_n}\right)^2}.$$

### Градиентный метод

Рассматривается задача оптимизации непрерывной выпуклой функции:

$$f(X) = f(x_1, x_2, \dots, x_n).$$

Поиск оптимума начинается с некоторого исходного допустимого решения  $X_0$ . Если  $grad f(X_0) \neq 0$ , то это означает, что его можно улучшить, то есть, найти такую точку  $X^*$ , где  $|grad f(X^*)| < |grad f(X_0)|$ .

Алгоритм метода сводится к нахождению последовательности допустимых решений  $X_k$  таких, чтобы  $f(X_{k+1}) < f(X_k)$ .

Очередное решение задачи получается из уравнения:  $X_{k+1} = X_k \pm \Delta h \cdot grad f(X_k)$  (знак зависит от того, максимум ищем, +, или минимум, -). Очевидно, встает задача выбора шага  $\Delta h$ . Существуют различные способы этого выбора, это и определяет вариант градиентного метода.

#### 1. Градиентный метод с конечным шагом

Величина шага  $\Delta h$  задается постоянной в начале расчета. В таком случае длина смещения расчетной точки зависит только от значения градиента в этой точке (направление смещения в любом случае задается градиентом, как вектором):

$$\Delta X = X_{k+1} - X_k = \Delta h \cdot grad f(X_k).$$

Такой метод выбора шага отличается тем, что по мере приближения к экстремуму целевой функции ее частные производные уменьшаются, следовательно, уменьшается градиент, а с ним уменьшается и смещение расчетной точки.

Суть метода сводится к постоянному перенахождению следующей расчетной точки по формуле:

$$X_{k+1} = X_k \pm \Delta h \cdot grad f(X_k).$$

В качестве критерия окончания расчетов выбирается условие того, что длина градиента становится много меньше заданной величины точности:

$$\sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}\right)^2} \ll \varepsilon_y.$$

Особенность траектории движения расчетной точки в том, что каждый отрезок ее траектории перпендикулярен к той поверхности уровня (изолинии), от которой он начался.

Важный момент в реализации данного метода — выбор величины шага  $\Delta h$ . Если взять шаг слишком малым, то нахождение оптимума потребует долгих вычислений градиента в каждой из многочисленных точек. Если же взять его очень большим — возможен незатухающий процесс блуждания расчетной точки вокруг истинного оптимума (она будет постоянно его «перескакивать»), и это в особенности касается таких функций, у которых близкие к оптимуму поверхности уровней сильно вытянуты по одной из осей.

## 2. Метод наискорейшего спуска

Заключается в том, что на каждой итерации  $i$  вычисляется такое значение  $\Delta h_i$ , чтобы минимизировалась функция:

$$f(x_i - \Delta h_i \text{grad} f(x_i)).$$

Это — вложенная в каждую итерацию исходной задачи дополнительная задача одномерной оптимизации. Решение ее осуществляется любым подходящим методом на усмотрение инженера.

Чаще всего используют самый простой — метод дробления шага. В нем шаг  $\Delta h_i$  разбивается на  $m$  частей, вычисляются значения  $f\left(x_i - \frac{j}{m} \cdot \Delta h_i \text{grad} f(x_i)\right)$  для всех  $j$  от единицы до  $m$ , из них выбирается минимальное, и соответствующее ему  $j$  пускается в дальнейший расчет.

Метод наискорейшего спуска требует меньшего количества расчетов и имеет более быструю сходимость, нежели градиентный метод с конечным шагом. Однако, они разделяют общий недостаток: из-за того, что градиент всегда перпендикулярен поверхности уровня, градиентные методы работают медленно в случаях, когда поверхности уровня функции вытянуты по какой-то оси.

Пример: методом наискорейшего спуска оптимизировать функцию:

$$f(X) = 4(x_1 - 5)^2 + (x_2 - 6)^2$$

с точностью  $\varepsilon_y = 1$ .

$$\frac{\partial f}{\partial x_1} = 8x_1 - 40, \quad \frac{\partial f}{\partial x_2} = 2x_2 - 12.$$

Начальное приближение:  $X_0 = (8, 10)$ .

$$f(X_0) = 52.$$

$$X_{k+1} = X_k \pm \Delta h_i \cdot \text{grad} f(X_k).$$

Первая итерация:

$$x_1^{(1)} = x_1^{(0)} - \Delta h_1 \cdot \frac{\partial f}{\partial x_1^{(0)}} = 8 - 24\Delta h_1$$

$$x_2^{(1)} = x_2^{(0)} - \Delta h_1 \cdot \frac{\partial f}{\partial x_2^{(0)}} = 10 - 8\Delta h_1$$

Нужно подобрать такое  $\Delta h_1$ , чтобы  $f(x_1^{(1)}, x_2^{(1)})$  было минимальным. Подставим их выражения в целевую функцию, и минимизируем ее по переменной  $\Delta h_1$ :

$$f(\Delta h_1) = 4(8 - 24\Delta h_1 - 5)^2 + (10 - 8\Delta h_1 - 6)^2 = 2304(\Delta h_1 - 0,125)^2 + 64(\Delta h_1 - 0,5)^2$$

$$f'(\Delta h_1) = 4608(\Delta h_1 - 0,125) + 128(\Delta h_1 - 0,5) = 4736\Delta h_1 - 640$$

$$f'(\Delta h_1) = 0: \Delta h_1 = 0,135$$

$$x_1^{(1)} = 8 - 24 \cdot 0,135 = 4,76$$

$$x_2^{(1)} = 10 - 8 \cdot 0,135 = 8,92$$

Проверка точности:

$$\sqrt{\sum_1^2 \left( \frac{\partial f}{\partial x_i^{(1)}} \right)^2} = \sqrt{(8 \cdot 4,76 - 40)^2 + (2 \cdot 8,92 - 12)^2} = 6,148 > \varepsilon_y, \text{ продолжаем}$$

расчет.

Вторая итерация:

$$x_1^{(2)} = x_1^{(1)} - \Delta h_2 \cdot \frac{\partial f}{\partial x_1^{(1)}} = 4,76 + 1,92\Delta h_2$$

$$x_2^{(2)} = x_2^{(1)} - \Delta h_2 \cdot \frac{\partial f}{\partial x_2^{(1)}} = 8,92 - 5,84\Delta h_2$$

$$\begin{aligned} f(\Delta h_2) &= 4(4,76 + 1,92\Delta h_2 - 5)^2 + (8,92 - 5,84\Delta h_2 - 6)^2 = \\ &= 14,746(\Delta h_2 - 0,125)^2 + 34,106(\Delta h_2 - 0,5)^2 \\ f'(\Delta h_2) &= 29,492(\Delta h_2 - 0,125) + 68,212(\Delta h_2 - 0,5) = 97,704\Delta h_2 - 37,793 \end{aligned}$$

$$f'(\Delta h_2) = 0: \Delta h_2 = 0,387$$

$$x_1^{(2)} = 5,503$$

$$x_2^{(2)} = 6,660$$

$$\sqrt{(8 \cdot 5,503 - 40)^2 + (2 \cdot 6,66 - 12)^2} = 4,235 > \varepsilon_y, \text{ продолжаем расчет.}$$

Третья итерация:

$$x_1^{(3)} = x_1^{(2)} - \Delta h_3 \cdot \frac{\partial f}{\partial x_1^{(2)}} = 5,503 - 4,024\Delta h_3$$

$$x_2^{(3)} = x_2^{(2)} - \Delta h_3 \cdot \frac{\partial f}{\partial x_2^{(2)}} = 6,66 - 1,32\Delta h_3$$

$$\begin{aligned} f(\Delta h_3) &= 4(5,503 - 4,024\Delta h_3 - 5)^2 + (6,66 - 1,32\Delta h_3 - 6)^2 = \\ &= 64,77(\Delta h_3 - 0,125)^2 + 1,74(\Delta h_3 - 0,5)^2 \\ f'(\Delta h_3) &= 129,54(\Delta h_3 - 0,125) + 3,48(\Delta h_3 - 0,5) = 133,02\Delta h_3 - 17,933 \end{aligned}$$

$$f'(\Delta h_3) = 0: \Delta h_3 = 0,135$$

$$x_1^{(3)} = 4,960$$

$$x_2^{(3)} = 6,482$$

$$\sqrt{(8 \cdot 4,96 - 40)^2 + (2 \cdot 6,482 - 12)^2} = 1,016 > \varepsilon_y, \text{ продолжаем расчет.}$$

Четвертая итерация:

$$x_1^{(4)} = x_1^{(3)} - \Delta h_4 \cdot \frac{\partial f}{\partial x_1^{(3)}} = 4,96 + 0,32\Delta h_4$$

$$x_2^{(4)} = x_2^{(3)} - \Delta h_4 \cdot \frac{\partial f}{\partial x_2^{(3)}} = 6,482 - 0,964\Delta h_4$$

$$\begin{aligned} f(\Delta h_4) &= 4(4,96 + 0,32\Delta h_4 - 5)^2 + (6,482 - 0,964\Delta h_4 - 6)^2 \\ &= 0,410(\Delta h_4 - 0,125)^2 + 0,929(\Delta h_4 - 0,5)^2 \\ f'(\Delta h_4) &= 0,82(\Delta h_4 - 0,125) + 1,858(\Delta h_4 - 0,5) = 2,678\Delta h_4 - 1,032 \end{aligned}$$

$$f'(\Delta h_4) = 0: \Delta h_4 = 0,385$$

$$x_1^{(4)} = 5,083$$

$$x_2^{(4)} = 6,110$$

$$\sqrt{(8 \cdot 5,083 - 40)^2 + (2 \cdot 6,11 - 12)^2} = 0,699 < \varepsilon_y, \text{ расчет закончен.}$$

Оптимум целевой функции:

$$f(5,083, 6,11) = 4(5,083 - 5)^2 + (6,11 - 6)^2 = 0,040.$$

### 7.2.3. Методы второго порядка

#### Метод Ньютона

Оптимизация непрерывной выпуклой функции нескольких переменных  $f(X) = f(x_1, x_2, \dots, x_n)$  методом Ньютона сводится к вычислению последовательности допустимых решений  $X_k$  таких, чтобы  $f(X_{k+1}) < f(X_k)$ , причем каждое  $X_{k+1}$  получается из выражения:

$$X_{k+1} = X_k - \Delta h \cdot \frac{f'(X_k)}{f''(X_k)}.$$

В качестве критерия окончания расчетов применяем уже использованное ранее условие:

$$\sqrt{\sum_{i=1}^n \left( \frac{\partial f}{\partial x_i} \right)^2} \ll \varepsilon_y.$$

Пример: определить экстремум функции  $f = 4(x_1 - 5)^2 + (x_2 - 6)^2$  с точностью  $\varepsilon_y = 0,1$

Определим производные функции:

$$\frac{\partial f}{\partial x_1} = 8x_1 - 40, \frac{\partial f}{\partial x_2} = 2x_2 - 12$$

$$\frac{\partial^2 f}{\partial x_1^2} = 8, \frac{\partial^2 f}{\partial x_2^2} = 2$$

Зададимся начальным приближением:

$$X_0 = (8, 10)$$