

5. ЛЕКЦИЯ. Методы первого порядка. Метод наискорейшего градиентного спуска.

Метод наискорейшего спуска отличается от метода градиентного спуска способом определения величины шага h_k . Величина шага задается не произвольно, а выбирается так, чтобы на каждой итерации достигалось максимально возможное уменьшение целевой функции $f(\bar{x})$ вдоль направления ее антиградиента $-\nabla f(\bar{x}^{(k)})$, вычисленного в точке $\bar{x}^{(k)}$. Величина шага h_k определяется из решения вспомогательной одномерной задачи минимизации

$$\varphi(h_k) = f\left(\bar{x}^{(k)} - h_k \nabla f(\bar{x}^{(k)})\right) \rightarrow \min_{h_k > 0},$$

которая может быть решена аналитически или численно. При квадратичной интерполяции целевой функции величину шага можно определить по формуле

$$h_k = \frac{\left(\nabla f(\bar{x}^{(k)}), \nabla f(\bar{x}^{(k)})\right)}{\left(H(\bar{x}^{(k)}) \nabla f(\bar{x}^{(k)}), \nabla f(\bar{x}^{(k)})\right)}. \quad (1)$$

где

$$H(\bar{x}^{(k)}) = \begin{bmatrix} \frac{\partial^2 f(\bar{x}^{(k)})}{\partial x_1^2} & \frac{\partial^2 f(\bar{x}^{(k)})}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(\bar{x}^{(k)})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\bar{x}^{(k)})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\bar{x}^{(k)})}{\partial x_2^2} & \dots & \frac{\partial^2 f(\bar{x}^{(k)})}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f(\bar{x}^{(k)})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\bar{x}^{(k)})}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(\bar{x}^{(k)})}{\partial x_n^2} \end{bmatrix} - \begin{matrix} \text{матрица} \\ \text{Гессе,} \\ \text{вычисленная} \\ \text{в точке } \bar{x}^{(k)}. \end{matrix}$$

На рисунке 1 представлена траектория приближения к точке минимума \bar{x}^* методом наискорейшего спуска для случая $n = 2$. Здесь каждая последующая точка находится как точка касания антиградиента целевой функции и линии уровня.

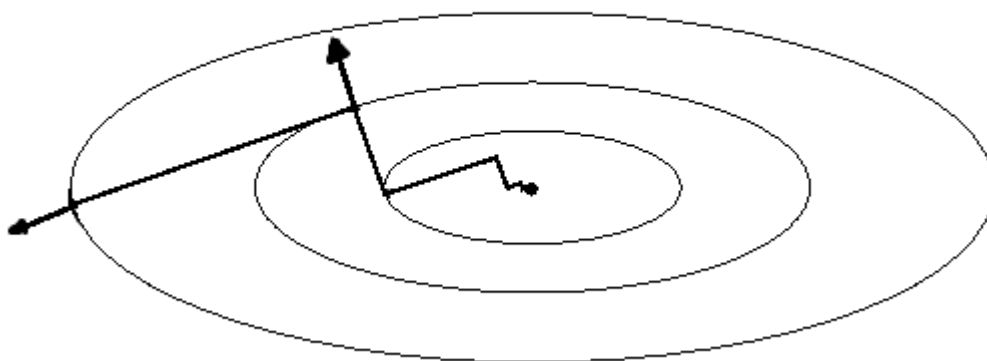


Рис. 1. Траектория движения к точке минимума в методе наискорейшего спуска

Алгоритм метода наискорейшего спуска

1. Задать размерность задачи оптимизации n , координаты начальной точки

$\overline{x^{(0)}} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$, точка поиска ε .

2. Положить счетчик числа $k = 0$.

3. Определить направление вектора градиента целевой функции $f(\bar{x})$ $\nabla f(\overline{x^{(k)}})$

$$= \left(\frac{\partial f(\overline{x^{(k)}})}{\partial x_1}, \frac{\partial f(\overline{x^{(k)}})}{\partial x_2}, \dots, \frac{\partial f(\overline{x^{(k)}})}{\partial x_n} \right) \text{ в точке}$$

$\overline{x^{(k)}} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$. Для вычисления координат вектора градиента

использовать разностную формулу (2).

4. Проверить условие окончания поиска

$$\left\| \nabla f(\overline{x^{(k)}}) \right\| = \sqrt{\sum_{i=1}^n \left(\frac{\partial f(\overline{x^{(k)}})}{\partial x_i} \right)^2} \leq \varepsilon.$$

Если условие выполнено, то расчет окончен $\bar{x}^* = \overline{x^{(k)}}$, иначе перейти к пункту 5.

5. Вычислить шаг h_k по формуле (1), используя результаты вычислений пункта 3 и разностные формулы (3), (4).

6. Определить координаты точки

$$\overline{x^{(k+1)}} = \overline{x^{(k)}} - h_k \nabla f(\overline{x^{(k)}})$$

положить $k = k + 1$ и перейти к пункту 3.

Пример. Найти минимум целевой функции

$$f(\bar{x}) = x_1^2 - x_1 x_2 + 3x_2^2 - x_1$$

методом наискорейшего градиентного спуска с точностью $\varepsilon = 0,1$.

Решение. За начальную точку примем

$$\overline{x^{(0)}} = \left(x_1^{(0)}, x_2^{(0)}\right)^T = (0, 0)^T.$$

Найдем градиент функции в произвольной точке $\overline{x^{(k)}} = \left(x_1^{(k)}, x_2^{(k)}\right)^T$.

$$\nabla f\left(\overline{x^{(k)}}\right) = \left(\frac{\partial f(\overline{x^{(k)}})}{\partial x_1}; \frac{\partial f(\overline{x^{(k)}})}{\partial x_2}\right)^T = \left(2x_1^{(k)} - x_2^{(k)} - 1; -x_1^{(k)} + 6x_2^{(k)}\right)^T.$$

Итерация $k = 0$. Вычислим градиент функции $\nabla f\left(\overline{x^{(0)}}\right)$ в начальной точке

$\overline{x^{(0)}} : \nabla f\left(\overline{x^{(0)}}\right) = (-1; 0)^T$. Определим координаты точки $\overline{x^{(1)}}$ по формуле

$$\overline{x^{(1)}} = \overline{x^{(0)}} - h_0 \nabla f\left(\overline{x^{(0)}}\right) = (0; 0)^T - h_0(-1; 0)^T = (h_0; 0)^T$$

Найдем значение h_0 из условия

$$\varphi(h_0) = f\left(\overline{x^{(0)}} - h_0 \nabla f\left(\overline{x^{(0)}}\right)\right) \rightarrow \min_{h_0 > 0},$$

Подставляя координаты точки $\overline{x^{(1)}}$ в целевую функцию $f(\bar{x})$, получаем

$$\varphi(h_0) = f\left(\overline{x^{(1)}}\right) = f(h_0, 0) = h_0^2 - h_0 - \text{функция одной переменной}.$$

Запишем необходимое условие безусловного экстремума функции $\varphi(h_0)$:

$$\frac{\partial \varphi(h_0)}{\partial h_0} = 2h_0 - 1 = 0$$

В результате решения уравнения определим $h_0 = \frac{1}{2}$. Проверим выполнение

достаточных условий экстремума: $\frac{\partial^2 \varphi(h_0)}{\partial h_0^2} = 2 > 0$. Найденное значение величины

шага h_0 обеспечивает минимум функции $\varphi(h_0)$. По данной величине шага

находим координаты точки $\overline{x^{(1)}}$:

$$\overline{x^{(1)}} = \overline{x^{(0)}} - h_0 \nabla f(\overline{x^{(0)}}) = (0; 0)^T - \frac{1}{2} (-1; 0)^T = \left(\frac{1}{2}; 0\right)^T. \text{ Проверим}$$

условие окончания процесса поиска. Для этого вычислим градиент целевой

$$\text{функции } \nabla f(\overline{x^{(1)}}) \text{ в точке } \overline{x^{(1)}} : \nabla f(\overline{x^{(1)}}) = \left(0; -\frac{1}{2}\right)^T.$$

Так как норма вектора градиента $\|\nabla f(\overline{x^{(1)}})\| = \frac{1}{2} > \varepsilon$, то переходим к следующей итерации.

Итерация $k = 1$. Определим координаты точки $\overline{x^{(2)}}$ по формуле:

$$\overline{x^{(2)}} = \overline{x^{(1)}} - h_1 \nabla f(\overline{x^{(1)}}) = \left(\frac{1}{2}; 0\right)^T - h_1 \left(0; -\frac{1}{2}\right)^T = \left(\frac{1}{2}; \frac{h_1}{2}\right)^T$$

Найдем значение h_1 из условия

$$\varphi(h_1) = f\left(\overline{x^{(1)}} - h_1 \nabla f(\overline{x^{(1)}})\right) \rightarrow \min_{h_1 > 0},$$

Подставляя координаты точки $\overline{x^{(2)}}$ в целевую функцию $f(\bar{x})$, получаем

$$\varphi(h_1) = f(\overline{x^{(2)}}) = f(h_1, 0) = \frac{3}{4} h_1^2 - \frac{1}{4} h_1 - \frac{1}{4} \quad - \text{ функция одной переменной.}$$

Запишем необходимое условие безусловного экстремума функции $\varphi(h_1)$:

$$\frac{\partial \varphi(h_1)}{\partial h_1} = \frac{3}{4} h_1 - \frac{1}{4} = 0$$

В результате решения уравнения определим $h_1 = \frac{1}{6}$. Проверим выполнение

$$\text{достаточных условий экстремума : } \frac{\partial^2 \varphi(h_0)}{\partial h_0^2} = \frac{3}{2} > 0.$$

Найденное значение величины шага h_1 обеспечивает минимум функции

$\varphi(h_1)$. По данной величине шага находим координаты точки $\overline{x^{(2)}}$:

$$\overline{x^{(2)}} = \overline{x^{(1)}} - h_1 \nabla f(\overline{x^{(1)}}) = \left(\frac{1}{2}; 0\right)^T - \frac{1}{6} \left(0; -\frac{1}{2}\right)^T = \left(\frac{1}{2}; \frac{1}{12}\right)^T.$$

Проверим условие окончания процесса поиска. Для этого вычислим

$$\text{градиент целевой функции } \nabla f(\overline{x^{(2)}}) \text{ в точке } \overline{x^{(2)}} : \nabla f(\overline{x^{(2)}}) = \left(-\frac{1}{12}; 0\right).$$

$$\text{Так как норма вектора градиента } \|\nabla f(\overline{x^{(2)}})\| = \frac{1}{12} < \varepsilon, \text{ то требуемая точность}$$

достигнута и точка $\overline{x^{(2)}}$ есть найденное приближение точки минимума $\overline{x^*} = (\frac{1}{2}; \frac{1}{12})^T, f(\overline{x^*}) = -0,27$.

Метод покоординатного спуска

Суть данного метода заключается в следующем. Выбирается произвольная начальная точка

$$\overline{x^{(0)}} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$$

Из области определения целевой функции $f(\bar{x})$.

В первой итерации фиксируются все координаты целевой функции, кроме первой x_1 . Находится минимум функции одной переменной $f(x_1, x_2^{(0)}, \dots, x_n^{(0)})$ в направлении проекций вектора антиградиента $-f(\overline{x^{(0)}}) \cdot \bar{e}_1$ на ось x_1 , где \bar{e}_1 — единичный вектор этой оси. В результате решения одномерной задачи оптимизации находится точка $\overline{x^{(1)}} = (x_1^{(1)}, x_2^{(0)}, \dots, x_n^{(0)})$, в которой целевая функция $f(x_1, x_2^{(0)}, \dots, x_n^{(0)})$ принимает минимальное значение по координате x_1 .

В следующей итерации фиксируют все координаты целевой функции, кроме координаты

$$x_2, f(x_1^{(1)}, x_2, \dots, x_n^{(0)}).$$

Снова решается одномерная задача оптимизации и находится точка

$\overline{x^{(2)}} = (x_1^{(1)}, x_2^{(1)}, x_3^{(0)}, \dots, x_n^{(0)})$, в которой целевая функция $f(x_1^{(1)}, x_2, \dots, x_n^{(0)})$ принимает минимум вдоль направления проекции вектора антиградиента $-f(\overline{x^{(1)}}) \cdot \bar{e}_2$ на ось x_2 , здесь \bar{e}_2 — единичный вектор этой оси.

Аналогично находятся точки минимума по координатам x_3, x_4, \dots, x_n .

Одномерный поиск на каждой итерации проводится в соответствии с формулой

$$\overline{x^{(k+1)}} = \overline{x^{(k)}} - h_k \nabla f(\overline{x^{(k)}}) \cdot \bar{e}_{k+1},$$

где h_k — величина шага, обеспечивающая максимально возможное уменьшение целевой функции $f(\bar{x})$,

k — номер итерации $k = \overline{0, (n-1)}$. После завершения n итераций точка $\overline{x^{(n)}}$ берется за начальную точку $\overline{x^{(0)}} = \overline{x^{(n)}}$ и итерации повторяются снова.

В качестве условия окончания поиска используется близость к нулю нормы градиента $||\nabla f(\overline{x^{(k+1)}})|| \leq \varepsilon$.

Величина шага h_k для каждого значения k определяется из решения вспомогательной одномерной задачи минимизации

$$f(h_k) = f(\overline{x^{(k)}} - h_k \nabla f(\overline{x^{(k)}})) \rightarrow \min_{h_k > 0}, \text{ которая может быть решена}$$

аналитически или численно. При квадратичной интерполяции целевой функции величину шага h_k можно определить по формуле (1). При ее использовании необходимо учитывать текущие значения переменных на каждой итерации.

На рисунке 2. представлена иллюстрация последовательных приближений к точке минимума $\overline{x^*}$ методом покоординатного спуска для случая двух переменных. Траектория спуска производится по ломаной линии, состоящей из отрезков прямых, параллельных осям координат.

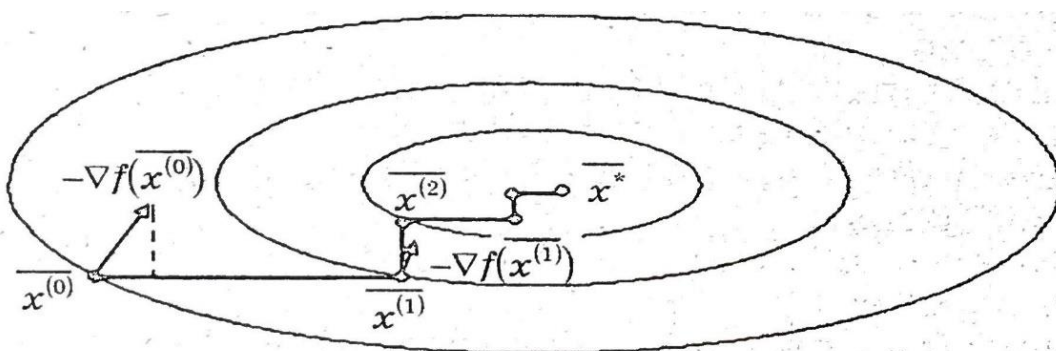


Рис. 2.19. Графическая иллюстрация поиска точки минимума методом покоординатного спуска

Пример. Найти минимум целевой функции

$$f(\bar{x}) = x_1^2 - x_1 x_2 + 3x_2^2 - x_1$$

методом покоординатного спуска с точностью $\varepsilon = 0,1$.

Решение. За начальную точку примем

$$\overline{x^0} = (x_1^{(0)}, x_2^{(0)})^T = (0, 0)^T.$$

Найдем градиент функции в произвольной точке $\overline{x^{(k)}} = (x_1^{(k)}, x_2^{(k)})^T$

$$\nabla f(\bar{x}^{(k)}) = \left(\frac{\partial f(\bar{x}^{(k)})}{\partial x_1}; \frac{\partial f(\bar{x}^{(k)})}{\partial x_2} \right)^T = (2x_1^{(k)} - x_2^{(k)} - 1; -x_1^{(k)} + 6x_2^{(k)})^T.$$

Итерация $k = 0$. Спуск по координате x_1 . Фиксируем координату $x_2 = x_2^{(0)} = 0$ в целевой функции $f(\bar{x})$. Тогда

$$f(x_1, x_2^{(0)}) = x_1^2 - x_1 x_2^{(0)} + 3(x_2^{(0)})^2 - x_1 = \varphi(x_1) - \text{функция одной}$$

переменной. Вычислим градиент функции $\nabla f(\bar{x}^{(0)})$ в начальной точке

$$\bar{x}^{(0)}: \nabla f(\bar{x}^{(0)}) = (-1; 0)^T.$$

Определим координату x_1 по формуле:

$$x_1 = x_1^{(0)} - h_0 \nabla f(\bar{x}^{(0)}) \cdot \bar{e}_1 = 0 - h_0 (-1; 0)^T \begin{pmatrix} 1 \\ 0 \end{pmatrix} = h_0$$

Подставляя координату точки x_1 , в целевую функцию $f(\bar{x})$, получаем

$$\varphi(h_0) = f(h_0, 0) = h_0^2 - h_0 - \text{функция одной переменной.}$$

Запишем необходимое условие безусловного экстремума функции

$$\varphi(h_0): \frac{\partial \varphi(h_0)}{\partial h_0} = 2h_0 - 1 = 0. \text{ В результате решения уравнения определим } h_0 = \frac{1}{2}.$$

$$\text{Проверим выполнение достаточных условий экстремума: } \frac{\partial^2 \varphi(h_0)}{\partial h_0^2} = 2 > 0.$$

Найденное значение величины шага h_0 обеспечивает минимум функции $\varphi(h_0)$. По данной величине шага находим координату точки $x_1^{(1)}$:

$$x_1^{(1)} = x_1^{(0)} - h_0 \nabla f(\bar{x}^{(0)}) \cdot \bar{e}_1 = 0 - \frac{1}{2} (-1; 0)^T \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}.$$

Таким образом, переходим от точки $\bar{x}^{(0)} = (x_1^{(0)}, x_2^{(0)})$ к точке

$\bar{x}^{(1)} = (x_1^{(1)}, x_2^{(0)})^T = (\frac{1}{2}, 0)$, в которой целевая функция $f(\bar{x})$ принимает наименьшее значение по координате x_1 .

Проверим условие окончания процесса поиска для этого вычислим градиент целевой функции $\nabla f(\bar{x}^{(1)})$ в точке $\bar{x}^{(1)}: \nabla f(\bar{x}^{(1)}) = (0; -\frac{1}{2})^T$. Так как норма вектора градиента $\|\nabla f(\bar{x}^{(1)})\| = \frac{1}{2} > \varepsilon$, то переходим к следующей итерации.

Итерация $k = 1$. Спуск по координате x_2 . Рассмотрим целевую функцию этой переменной

$$f(x_1^{(1)}, x_2) = f\left(\frac{1}{2}, x_2\right) = \frac{1}{4} - \frac{1}{2}x_2 + 3x_2^2 - \frac{1}{2} = 3x_2^2 - \frac{1}{2}x_2 - \frac{1}{4} = \varphi(x_2).$$

Определим координату x_2 по формуле:

$$x_2 = x_2^{(0)} - h_1 \nabla f(\bar{x}^{(1)}) \cdot \bar{e}_2 = 0 - h_1 \left(0; -\frac{1}{2}\right)^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{h_1}{2}.$$

Подставляя координату x_2 в целевую функцию $f(\bar{x})$, получаем $\varphi(h_1) = f(x_1^{(1)}, h_1) = \frac{3}{4}h_1^2 - \frac{1}{4}h_1 - \frac{1}{4}$ — функция одной переменной.

Запишем необходимое условие безусловного экстремума функции

$$\varphi(h_1): \frac{\partial \varphi(h_1)}{\partial h_1} = \frac{3}{2}h_1 - \frac{1}{4} = 0. \text{ В результате решения уравнения определим } h_1 =$$

$$\frac{1}{6}. \text{ Проверим выполнение достаточных условий экстремума: } \frac{\partial^2 \varphi(h_1)}{\partial h_1^2} = \frac{3}{2} > 0.$$

Найденное значение величины шага h_1 обеспечивает минимум функции $\varphi(h_1)$. По данной величине шага находим координаты точки x_2 :

$$x_2 = x_2^{(0)} - h_1 \nabla f(\bar{x}^{(1)}) \cdot \bar{e}_2 = 0 - \frac{1}{6} \left(0; -\frac{1}{2}\right)^T \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{12}$$

Таким образом, переходим от точки $\bar{x}^{(1)} = (x_1^{(1)}, x_2^{(0)})^T$ к точке

$\bar{x}^{(2)} = (x_1^{(1)}, x_2^{(1)})^T = \left(\frac{1}{2}, \frac{1}{12}\right)^T$, в которой целевая функция $f(\bar{x})$ принимает наименьшее значение по координате x_2 .

Проверим условие окончания процесса поиска. Для этого вычислим градиент целевой функции $\nabla f(\bar{x}^{(2)})$ в точке $\bar{x}^{(2)}$: $\nabla f(\bar{x}^{(2)}) = \left(-\frac{1}{12}, 0\right)$. Так как норма вектора градиента $\|\nabla f(\bar{x}^{(2)})\| = \frac{1}{12} < \varepsilon$, то требуемая точность достигнута и точка $\bar{x}^{(2)}$ есть найденное приближение точки минимума $\bar{x}^* = \left(\frac{1}{2}, \frac{1}{12}\right)^T, f(\bar{x}^*) = -0,27$.