Лекция № 4

Пример 1

Дано: $f(X) = x_1^2 + 3x_2^2$

Сделать 2 итерации методом градиентного спуска из начальной точки $X^0 = (0.5, 0.5)$

Решение:

Найдем градиент функции $\nabla f(X) = (2x_1; 6x_2)^T$

Итерация 0 алгоритма (соответствует начальной точке)

$X^0 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix};$

$f(X^0) = 0.5^2 + 3 \cdot 0.5^2 = 1$

$\nabla f(X^0) = \begin{pmatrix} 1 \\ 3 \end{pmatrix};$ $\quad \left\| \nabla f(X^0) \right\| = \sqrt{1^2 + 3^2} = 3.1623$

Итерация 1 алгоритма ($k = 0$)

$X^1 = X^0 - t_0 \nabla f(X^0)$

Зададим шаг $t_0 = 0.1$, тогда:

$X^1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - 0.1 \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 0.4 \\ 0.2 \end{pmatrix}$

Вычислим значение функции в точке X^1: $f(X^1) = 0.4^2 + 3 \cdot 0.2^2 = 0.28$

Т.к. $f(X^1) < f(X^0)$ шаг выбран удачно

$\nabla f(X^1) = \begin{pmatrix} 0.8 \\ 1.2 \end{pmatrix};$ $\quad \left\| \nabla f(X^1) \right\| = \sqrt{0.8^2 + 1.2^2} = 1.4422$

Итерация 2 алгоритма ($k = 1$)

$X^2 = X^1 - t_1 \nabla f(X^1)$

Зададим шаг $t_1 = 0.1$, тогда:

$X^2 = \begin{pmatrix} 0.4 \\ 0.2 \end{pmatrix} - 0.1 \begin{pmatrix} 0.8 \\ 1.2 \end{pmatrix} = \begin{pmatrix} 0.32 \\ 0.08 \end{pmatrix}$

Вычислим значение функции в точке X^2: $f(X^2) = 0.32^2 + 3 \cdot 0.08^2 = 0.1216$

Т.к. $f(X^2) < f(X^1)$ шаг выбран удачно

$\nabla f(X^2) = \begin{pmatrix} 0.64 \\ 0.48 \end{pmatrix};$ $\quad \left\| \nabla f(X^2) \right\| = \sqrt{0.64^2 + 0.48^2} = 0.8$

Пример 2

Дано: $f(X) = x_1^2 + 3x_2^2$

Сделать 1 итерацию методом наискорейшего градиентного спуска из начальной точки $X^0 = (0.5, 0.5)$

Решение:

Итерация 0 алгоритма (соответствует начальной точке) – см. пример 1.

Итерация 1 алгоритма ($k = 0$)

$$X^1 = X^0 - t_0 \nabla f(X^0)$$

$$X^1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - t_0 \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 0.5 - t_0 \\ 0.5 - 3t_0 \end{pmatrix}$$

Способ А вычисления шага t_0:

Вычислим значение функции в точке X^1:

$$f(X^1) = (0.5 - t_0)^2 + 3\cdot(0.5 - 3t_0)^2 = 0.25 - t_0 + t_0^2 + 0.75 - 9t_0 + 27t_0^2 = 28t_0^2 - 10t_0 + 1$$

Как видно функция в точке $f(X^1)$ зависит только от величины шага t_0, следовательно, можно записать:

$$f(X^1) = \varphi(t_0) = 28t_0^2 - 10t_0 + 1.$$

Найдем минимум функции $\varphi(t_0)$:

$$\frac{d\varphi(t_0)}{dt_0} = 56t_0 - 10 = 0 \quad \Rightarrow \quad t_0 = 0.17857$$ - значение шага

$$\frac{d^2\varphi(t_0)}{dt_0^2} = 56 > 0$$ - значит, функция $\varphi(t_0) = f(X^1)$ принимает минимальное значение

Окончательно:

$$X^1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - 0.17857 \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 0.3214 \\ -0.0357 \end{pmatrix}; \quad f(X^1) = 0.3214^2 + 3\cdot(-0.0357)^2 = 0.1071$$

$$\nabla f(X^1) = \begin{pmatrix} 0.6429 \\ -0.2143 \end{pmatrix}; \qquad \left\| \nabla f(X^1) \right\| = \sqrt{0.6429^2 + (-0.2143)^2} = 0.6776$$

Способ В вычисления шага t_0:

Вычислим градиент функции в точке X^1:

$$\nabla f(X^1) = \begin{pmatrix} 2(0.5 - t_0) \\ 6(0.5 - 3t_0) \end{pmatrix}$$

Воспользуемся условием $\nabla f(X^0) \perp \nabla f(X^1)$:

$$1\cdot 2(0.5 - t_0) + 3\cdot 6(0.5 - 3t_0) = 0$$

$$1 - 2t_0 + 9 - 54t_0 = 0$$

$$10 - 56t_0 = 0$$

$$t_0 = 0.17857$$

Окончательно:

$$X^1 = \begin{pmatrix}0.5\\0.5\end{pmatrix} - 0.17857\begin{pmatrix}1\\3\end{pmatrix} = \begin{pmatrix}0.3214\\-0.0357\end{pmatrix}; \quad f(X^1) = 0.3214^2 + 3\cdot(-0.0357)^2 = 0.1071$$

$$\nabla f(X^1) = \begin{pmatrix}0.6429\\-0.2143\end{pmatrix}; \qquad \left\|\nabla f(X^1)\right\| = \sqrt{0.6429^2 + (-0.2143)^2} = 0.6776$$

Пример 3

<u>Дано</u>: $f(X) = x_1^2 + 3x_2^2$

Сделать 2 итерации методом покоординатного спуска из начальной точки $X^0 = (0.5, 0.5)$

<u>Решение</u>:

<u>Итерация 0</u> алгоритма (соответствует начальной точке) – см. пример 1.

<u>Итерация 1</u> алгоритма (k=0)

$$X^1 = X^0 - t_0\left[\nabla f(X^0)\right]_{\text{пр на } x_2}$$

Для проекции градиента выберем направление оси x_2 .Зададим шаг $t_0 = 0.1$, тогда:

$$X^1 = \begin{pmatrix}0.5\\0.5\end{pmatrix} - 0.1\begin{pmatrix}0\\3\end{pmatrix} = \begin{pmatrix}0.5\\0.2\end{pmatrix}$$

Вычислим значение функции в точке X^1: $f(X^1) = 0.5^2 + 3\cdot 0.2^2 = 0.37$

Т.к. $f(X^1) < f(X^0)$ шаг выбран удачно

$$\nabla f(X^1) = \begin{pmatrix}1\\1.2\end{pmatrix}; \qquad \left\|\nabla f(X^1)\right\| = \sqrt{1^2 + 1.2^2} = 1.5621$$

<u>Итерация 2</u> алгоритма (k=1)

$$X^2 = X^1 - t_1\left[\nabla f(X^1)\right]_{\text{пр на } x_1}$$

Для проекции градиента выберем направление оси x_1. Зададим шаг $t_1 = 0.5$, тогда:

$$X^2 = \begin{pmatrix}0.5\\0.2\end{pmatrix} - 0.5\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}0\\0.2\end{pmatrix}$$

Вычислим значение функции в точке X^2: $f(X^2) = 0^2 + 3\cdot 0.2^2 = 0.12$

Т.к. $f(X^2) < f(X^1)$ шаг выбран удачно

$$\nabla f(X^2) = \begin{pmatrix}0\\1.2\end{pmatrix}; \qquad \left\|\nabla f(X^2)\right\| = \sqrt{0^2 + 1.2^2} = 1.2$$

Пример 4

Дано: $f(X) = x_1^2 + 3x_2^2$

Сделать 1 итерацию методом Гаусса-Зейделя из начальной точки $X^0 = (0.5, 0.5)$

Решение:

Итерация 0 алгоритма (соответствует начальной точке) – см. пример 1.

Итерация 1 алгоритма (k=0)

$$X^1 = X^0 - t_0 \left[\nabla f(X^0)\right]_{\text{пр на } x_2}$$

Для проекции градиента выбирается направление оси x_2.

$$X^1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - t_0 \begin{pmatrix} 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 - 3t_0 \end{pmatrix}$$

Используем способ А вычисления шага t_0:

Вычислим значение функции в точке X^1:

$$f(X^1) = 0.5^2 + 3 \cdot (0.5 - 3t_0)^2 = 0.25 + 0.75 - 9t_0 + 27t_0^2 = 27t_0^2 - 9t_0 + 1$$

Как видно функция в точке $f(X^1)$ зависит только от величины шага t_0, следовательно, можно записать:

$$f(X^1) = \varphi(t_0) = 27t_0^2 - 9t_0 + 1.$$

Найдем минимум функции $\varphi(t_0)$:

$$\frac{d\varphi(t_0)}{dt_0} = 54t_0 - 9 = 0 \quad \Rightarrow \quad t_0 = \frac{1}{6}$$ - значение шага

$$\frac{d^2\varphi(t_0)}{dt_0^2} = 54 > 0$$ - значит, функция $\varphi(t_0) = f(X^1)$ принимает минимальное значение

Окончательно:

$$X^1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - \frac{1}{6} \cdot \begin{pmatrix} 0 \\ 3 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}; \qquad f(X^1) = 0.5^2 + 3 \cdot 0^2 = 0.25$$

$$\nabla f(X^1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \qquad \left\| \nabla f(X^1) \right\| = \sqrt{1^2 + 0^2} = 1$$

Пример 5

Дано: $f(X) = x_1^2 + 3x_2^2$

Сделать 2 итерации методом сопряженных градиентов из начальной точки $X^0 = (0.5, 0.5)$

Решение:

Итерация 0 алгоритма (соответствует начальной точке) – см. пример 1.

Итерация 1 алгоритма (k=0)

Результаты итерации совпадают с 1-й итерацией метода наискорейшего градиентного спуска – см. пример 2.

$$X^1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - 0.17857 \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 0.3214 \\ -0.0357 \end{pmatrix}; \quad f(X^1) = 0.3214^2 + 3 \cdot (-0.0357)^2 = 0.1071$$

$$\nabla f(X^1) = \begin{pmatrix} 0.6429 \\ -0.2143 \end{pmatrix}; \qquad \left\| \nabla f(X^1) \right\| = \sqrt{0.6429^2 + (-0.2143)^2} = 0.6776$$

Итерация 2 алгоритма (k=1)

$$X^2 = X^1 + t_1 d^1 \qquad d^1 = -\nabla f(X^1) + \beta_0 d^0$$

$$\beta_0 = \frac{\left\| \nabla f(X^1) \right\|^2}{\left\| \nabla f(X^0) \right\|^2} = \frac{0.6776^2}{3.1623^2} = 0.0459$$

$$d^1 = -\begin{pmatrix} 0.6429 \\ -0.2143 \end{pmatrix} + 0.0459 \begin{pmatrix} -1 \\ -3 \end{pmatrix} = \begin{pmatrix} -0.6888 \\ 0.0766 \end{pmatrix}$$

$$X^2 = \begin{pmatrix} 0.3214 \\ -0.0357 \end{pmatrix} + t_1 \begin{pmatrix} -0.6888 \\ 0.0766 \end{pmatrix} = \begin{pmatrix} 0.3214 - 0.6888 t_1 \\ -0.0357 + 0.0766 t_1 \end{pmatrix}$$

Используем способ А вычисления шага t_1:

Вычислим значение функции в точке X^2:

$$f(X^2) = (0.3214 - 0.6888 t_1)^2 + 3(-0.0357 + 0.0766 t_1)^2 = 0.492 t_1^2 - 0.4592 t_1 + 0.10711$$

$$f(X^2) = \varphi(t_1) = 0.492 t_1^2 - 0.4592 t_1 + 0.10711.$$

Найдем минимум функции $\varphi(t_1)$:

$$\frac{d\varphi(t_1)}{dt_1} = 0.984 t_1 - 0.4592 = 0 \quad \Rightarrow \quad t_1 = 0.46666 = \frac{7}{15}$$ - значение шага

$\dfrac{d^2\varphi(t_1)}{dt_1^2} = 0.984 > 0$ - значит, функция $\varphi(t_1) = f(X^2)$ принимает минимальное значение

Окончательно:

$$X^2 = \begin{pmatrix} 0.3214 \\ -0.0357 \end{pmatrix} + \frac{7}{15} \cdot \begin{pmatrix} -0.6888 \\ 0.0766 \end{pmatrix} = \begin{pmatrix} -0.00004 \\ 0.00005 \end{pmatrix}; \qquad f(X^2) = 0$$

$$\nabla f(X^2) = \begin{pmatrix} -0.00008 \\ 0.0003 \end{pmatrix}; \qquad \left\| \nabla f(X^2) \right\| = 0.0003$$

Очевидно, что на 2-й итерации найдены координаты стационарной точки с точностью $\varepsilon = 0.0005$.

Методы второго порядка

(6) Метод Ньютона

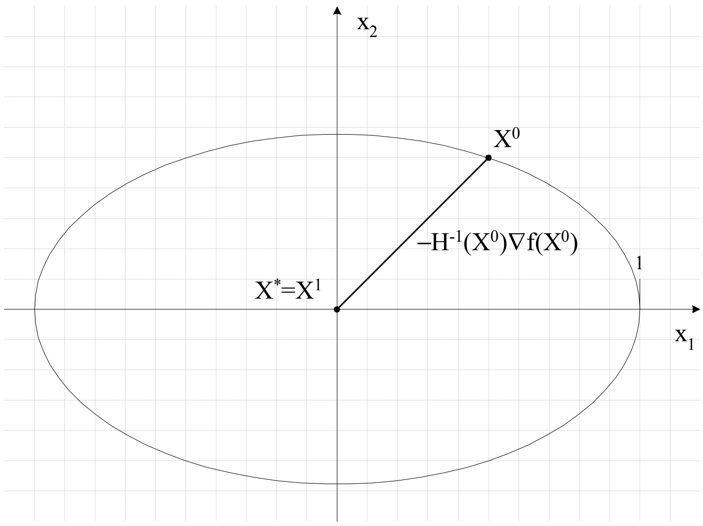
Алгоритм метода: $\boxed{X^{k+1} = X^k - H^{-1}(X^k)\nabla f(X^k)}$ (6.1)

здесь:

- $d^k = -H^{-1}(X^k)\nabla f(X^k)$ - направление спуска (6.2)
- $t_k = 1$ (6.3)

Особенностью метода Ньютона является то, что при $H(X^0) > 0$ метод позволяет отыскать минимум квадратичной функции **за одну итерацию**.

Геометрическая интерпретация метода для квадратичной функции:



Критерии окончания метода такие же, как и в методе градиентного спуска.

Сходимость метода Ньютона. Сходимость метода Ньютона доказана только для сильно выпуклых функций и существенно зависит от выбора начальной точки X^0. Можно утверждать, что метод Ньютона обеспечивает сходимость к точке минимума функции, только при условии выполнения условия $H(X^k) > 0$ в каждой точке последовательности $\{X^k\}$.

Пример 6

Дано: $f(X) = x_1^2 + 3x_2^2$

Сделать 1 итерацию методом Ньютона из начальной точки $X^0 = (0.5, 0.5)$

Решение:

Итерация 0 алгоритма (соответствует начальной точке)

$$X^0 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix};$$

$$f(X^0) = 0.5^2 + 3 \cdot 0.5^2 = 1$$

$$\nabla f(X^0)=\begin{pmatrix}1\\3\end{pmatrix};\qquad \left\|\nabla f(X^0)\right\|=\sqrt{1^2+3^2}=3.1623$$

$$H(X^0)=\begin{pmatrix}2&0\\0&6\end{pmatrix}\qquad H^{-1}(X^0)=\frac{1}{12}\begin{pmatrix}6&0\\0&2\end{pmatrix}=\begin{pmatrix}1/2&0\\0&1/6\end{pmatrix}$$

<u>**Итерация 1**</u> алгоритма (k=0)

$$X^1=X^0-H^{-1}(X^0)\nabla f(X^0)$$

$$X^1=\begin{pmatrix}0.5\\0.5\end{pmatrix}-\begin{pmatrix}1/2&0\\0&1/6\end{pmatrix}\cdot\begin{pmatrix}1\\3\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$$

$$\nabla f(X^1)=\begin{pmatrix}0\\0\end{pmatrix};\qquad \left\|\nabla f(X^1)\right\|=0$$

Очевидно, что на 1-й итерации найдены координаты стационарной точки.