

Лекция 8

Луа Yaroshevskiy

13 мая 2023 г.

Содержание

1	Метод градиентного спуска	1
2	Метод наискорейшего спуска	3

Определение. Направление вектора p^k называется **направлением убывания** функции $f(x)$ в точке x^k , если при всех достаточно малых положительных α выполняется неравенство:

$$f(x^k + \alpha p^k) < f(x^k)$$

Теорема 0.1 (достаточное условие направления убывания). Пусть функция $f(x)$ дифференцируема в точке x^k . Если вектор p^k удовлетворяет условию:

$$(\nabla f(x^k), p^k) < 0$$

, то направление вектора p^k является направлением убывания

Доказательство. Из свойства дифференцируемости функции и условия данной теоремы следует, что

$$\begin{aligned} f(x^{k+1}) - f(x^k) &= f(x^k + \alpha p^k) - f(x^k) = (\nabla f(x^k), \alpha p^k) + o(\alpha) = \\ &= \alpha \left((\nabla f(x^k), p^k) + \frac{o(\alpha)}{\alpha} \right) < 0 \end{aligned}$$

, при всех достаточно малых $\alpha > 0$, т.е. p^k задает направление убывания функции $f(x)$ в точке x^k \square

Примечание. Геометрическая интерпретация $(\nabla f(x^k), p^k) < 0 \implies p^k$ составляет тупой угол с $\nabla f(x^k)$

$f(x)$ дифференцируема в E_n :

$$x^{k+1} = x^k + \alpha_k p^k \quad k = 0, 1, \dots \quad (1)$$

где p^k определяется с учетом информации о частных производных, а величина шага $\alpha_k > 0$, такова, что:

$$f(x^{k+1}) < f(x^k) \quad k = 0, 1, \dots \quad (2)$$

Останов итерационного процесса: $\|\nabla f(x^k)\| < \varepsilon$

1 Метод градиентного спуска

В 1: $p^k = -\nabla f(x^k)$ — предположение. Если $\nabla f(x^k) \neq 0$, то $(\nabla f(x^k), p^k) < 0$, следовательно p^k — направление убывания функции $f(x)$, в малой окрестности точки x^k направление p^k обеспечивает наискорейшее убывание этой функции. Таким образом можно найти такое $\alpha_k > 0$, что выполнится

2

Алгоритм 1 метод Градиентного спуска**Ввод** $\varepsilon > 0, \alpha > 0, x \in E_k, f(x)$

- 1: **повторять**
- 2: Вычисляем $\nabla f(x)$
- 3: **если** Выполнено условие достижения точности $\|\nabla f(x)\| < \varepsilon$ **тогда**
- 4: **Вернуть** $x^* := x, f^* := f(x^*)$
- 5: **конец если**
- 6: **повторять**
- 7: Найти $y := x - \alpha \nabla f(x)$
- 8: Вычислить $f(y)$
- 9: **если** $f(y) < f(x)$ **тогда**
- 10: $x := y$
- 11: $f(x) := f(y)$
- 12: Выйти из цикла
- 13: **иначе**
- 14: $\alpha := \frac{\alpha}{2}$
- 15: **конец если**
- 16: **конец повторять**
- 17: **конец повторять**

Примечание. В окрестности стационарной точки функции $f(x)$ величина $\|\nabla f(x)\|$ становится малой, это приводит к замедлению сходимости последовательности $\{x^k\}$. Поэтому в 1 иногда полагают

$$p^k = \frac{-\nabla f(x^k)}{\|\nabla f(x^k)\|}$$

Теорема 1.1. Пусть симметричная матрица A квадратичной функции $f(x)$ положительно определена, а l и L — наименьшее и наибольшее собственное значение A . Тогда при любых $\alpha \in (0, \frac{2}{L})$ и $x^0 \in E_n$

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

сходится к единственной точке глобального минимума x^* функции $f(x)$ линейно (со скоростью геометрической прогрессии)

$$\rho(x^k, x^*) \leq q^k \rho(x^0, x^*)$$

, где $q = \max\{|1 - \alpha l|, |1 - \alpha L|\}$

Доказательство. Т.к. A положительно определена, то $f(x)$ — сильно выпукла. Следовательно точка x^* — существует и единственна. $\nabla f(x^*) = 0$ в точке x^* , тогда

$$\nabla f(x^k) = Ax^k + b = Ax^k + b - \underbrace{Ax^* + b}_{\nabla f(x^*)} = A(x^k - x^*)$$

Оценим норму разности

$$\begin{aligned} \|x^k - x^*\| &= \|x^{k-1} - \alpha \nabla f(x^{k-1}) - x^*\| = \|x^{k-1} - x^* - \alpha A(x^{k-1} - x^*)\| = \\ &= \|(E - \alpha A)(x^{k-1} - x^*)\| \\ \|x^k - x^*\| &\leq \|E - \alpha A\| \cdot \|x^{k-1} - x^*\| \leq q \|x^{k-1} - x^*\| \leq q^k \|x^0 - x^*\| \end{aligned}$$

— из определения линейной сходимости, где q — оценка нормы матрицы через величину ее собственных значений

$$\|E - \alpha A\| \leq q = \max\{|1 - \alpha l|, |1 - \alpha L|\}$$

Если $\alpha \in (0; \frac{2}{L})$, то $q < 1$

q: $q^* = \frac{L-l}{L+l}$, при $\alpha = \alpha^* = \frac{2}{L+l}$. Т.к. $l < L$, то $1 - \alpha l = -(1 - \alpha L)$. От соотношения L и l существенно зависит число итераций градиентного метода при минимизации выпуклой квадратичной функции \square

Пример. $L = l > 0$, тогда точка минимума находится за один шаг

$$\begin{aligned} f(x) &= x_1^2 + x_2^2 \rightarrow \min \\ x^0 &= (1, 1)^T \quad \alpha = \alpha^* = \frac{2}{l + L} \end{aligned}$$

Решение.

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \implies l = L = 2$$

$$\alpha^* = \frac{2}{2+2} = \frac{1}{2}$$

$$x^1 = x^0 - \frac{1}{2} \operatorname{grad} f(x^0) = (0, 0)^T$$

$$x^1 = x^*$$

Примечание. При $l = L$ — линии уровня $f(x)$ — концентрические окружности

Примечание. $L \gg l > 0$ — линии уровня $f(x)$ — эллипсы

Пример.

$$f(x) = x_1^2 + 100x_2^2 \rightarrow \min$$

$$x^0 = (1, 1)^T$$

$$\alpha = \alpha^*$$

Решение.

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 200 \end{pmatrix} \implies l = 2, L = 200$$

Линии уровня — эллипсы сильно вытянутые вдоль оси Ox_1

$$\alpha = \alpha^* = \frac{2}{202} = \frac{1}{101}$$

$$-\nabla f(x^0) = (-2, -200)^T$$

— сильно отличается от $x^* - x^0$

$$x^* - x^0 = (-1, -1)^T$$

— направление точки глобального минимума

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

$$\nabla f(x^k) = (2x_1, 200x_2)^T$$

$$\begin{cases} x_1^{k+1} = \frac{99}{101} x_1^k \\ x_2^{k+1} = -\frac{99}{101} x_2^k \end{cases}$$

— закон изменения координат точек, минимизирующей последовательности. $\{x^k\}$ — сходится медленно

Определение. Число обусловленности для симметричной положительно определенной матрицы $\mu = \frac{L}{l}$. Оно характеризует вытянутость линий уровня $f(x) = C$

- Если μ велико, то линии уровня сильно вытянуты и говорят что функция имеет **овражный** характер (резко меняется по одним направлением и слабо по другим) \implies Полохо обусловленная задача
- Если $\mu \sim 1$, то линии уровня близки к окружностям и задача является хорошо обусловленной

2 Метод наискорейшего спуска

После вычисления в начальной точке градиента функции делают в направлении антиградиента не маленький шаг, а передвигаются до тех пор, пока функция убывает. Достигнув точки минимума на выбранном направлении снова вычисляют градиент функции и повторяют описанную процедуру

$$p^k = -\nabla f(x^k)$$

α_k — находится из решения задачи одномерной оптимизации

$$\Phi_k(\alpha) \rightarrow \min$$

$$\Phi_k(\alpha) = f(x^k - \alpha \nabla f(x^k)) \quad \alpha > 0 \quad (3)$$

Алгоритм 2 метод наискорейшего спуска**Ввод** $\varepsilon > 0, x \in E_k, f(x)$

- 1: **повторять**
- 2: Вычислить $\nabla f(x)$
- 3: **если** $\|\nabla f(x)\| < \varepsilon$ **тогда**
- 4: **Вернуть** $x^* := x, f^* := f(x)$
- 5: **конец если**
- 6: Решить задачу одномерной оптимизации 3 для $x^k := x$, т.е. найти α^*
- 7: $x := x - \alpha^* \nabla f(x)$
- 8: **конец повторять**

Определение. Ненулевые векторы p^1, \dots, p^k называются сопряженными относительно матрицы A размера $n \times n$ или A -ортогональными, если

$$(Ap^i, p^j) = 0 \quad i \neq j \quad i, j = 1, \dots, k$$

Примечание. Система из n векторов p^1, \dots, p^n сопряженных относительно положительно определенной матрицы A линейно независима

Примечание. n ненулевых A -ортгональных векторов образуют базис в E_n . Рассмотрим минимизацию квадратичной функции в E_n

$$f(x) = \frac{1}{2}(Ax, x) + (b, x) + c$$

A — положительно определенная. Итерационный процесс

$$x^k = x^{k-1} + \alpha_k p^k \quad k = 1, 2, \dots \quad (4)$$

, где p^k — A -ортогональные

Примечание. Если в итерационном процессе 4 на каждом шаге используется исчерпывающий спуск, то

$$\alpha_k = -\frac{(\nabla f(x^0), p^k)}{(Ap^k, p^k)} \quad k = 1, 2, \dots \quad (5)$$

Доказательство.

$$x^k = x^{k-1} + \alpha_k p^k = x^0 + \sum_{i=1}^k \alpha_i p^i \quad (6)$$

$$\nabla f(x) = Ax + b$$

$$\nabla f(x^k) = \nabla f(x^0) + \sum_{i=0}^k \alpha_i Ap^i$$

умножим на p^k и учитываем $(\nabla f(x^k), p^k) = 0$, A -ортогональность p^k

$$(\nabla f(x^0), p^k) + \alpha_k (Ap^k, p^k) = 0$$

, т.к. A — положительно определена, то $(Ap^k, p^k) > 0$, и для α_k :

$$\alpha_k = -\frac{(\nabla f(x^0), p^k)}{(Ap^k, p^k)}$$

□