Методы оптимизации. Задание 1: Субградиентный спуск

Сергей Володин, 374 гр.

задано 2016.02.09

Задача 1

Делаем проекцию при $k \in K$. $Q \ni x^* = \underset{x \in Q}{\arg\min} f(x)$. Рассмотрим $k+1$-ю итерацию:

1. Пусть $k+1 \in K$. Тогда $x_{k+1} \overset{\text{def}}{=} \pi_Q(x_k - \alpha_k g^k)$
2. Иначе $x_{k+1} = x_k - \alpha_k g^k$.

Здесь $g^k \in \partial f(x_k)$. В первом случае $||x_{k+1} - x^*|| = ||\pi_Q(x_k - \alpha_k g^k) - x^*|| \leqslant ||x_k - \alpha_k g^k - x^*||$ по свойству проекции (расстояние до $x^* \in Q$ не увеличивается при проектировании на Q).

Тогда в обоих случаях $||x_{k+1} - x^*||^2 \leqslant ||x_k - \alpha_k g^k - x^*||^2 = ||x_k^2 - x^*||^2 + \alpha_k^2 ||g^k||^2 - 2\alpha_k (g^k, x_k - x^*) \overset{g \in \partial f(x_k)}{\leqslant} ||x_k - x^*||^2 + \alpha_k^2 ||g^k||^2 - 2\alpha_k (f(x_k) - f(x^*))$.

Последовательно подставим (индукция) оценку разности для x_k, ..., x_1 в x_{k+1}, получим такую же оценку, как и для обычного метода градиентного спуска, т.е. $f_{\text{best}} - f_* \leqslant \frac{RL}{\sqrt{k}}$

Оценка в худшем случае (равенство) не изменится.

Задача 2

Ответ: да, верно, да, может. Приведем пример $f\colon \mathbb{R}^n \to \mathbb{R}$ — выпуклая, $x_0 \in \mathbb{R}^n, a \in \partial f(x_0)$, x_0 — не точка минимума f, $-a$ — не направление убывания f.

$f\left(\left\|\begin{matrix} x_1 \\ x_2 \end{matrix}\right\|\right) \overset{\text{def}}{=} |x_1| + |x_2| \colon \mathbb{R}^2 \to \mathbb{R}$. Точка $x_0 \overset{\text{def}}{=} \left\|\begin{matrix} 1 & 0 \end{matrix}\right\|^T$ Тогда

1. f — выпуклая: пусть $t_1, t_2 \in \mathbb{R}_+$, $x, y \in \mathbb{R}^2$. $f(t_1 x + t_2 y) = |t_1 x_1 + t_2 y_1| + |t_1 x_2 + t_2 y_2| \leqslant t_1 |x_1| + t_2 |y_1| + t_1 |x_2| + t_2 |y_2| = t_1(|x_1| + |x_2|) + t_2(|y_1| + |y_2|) = t_1 f(x) + t_2 f(y)$. Возьмем $t_1 \in [0, 1]$, $t_2 = 1 - t_1$, получим определение выпуклой функции.
2. Пусть $a = \left\|\begin{matrix} 1 & 1 \end{matrix}\right\|^T$. Докажем, что $a \in \partial f(x_0)$. Фиксируем $x \in \mathbb{R}^2$. $f(x) - f(x_0) = |x_1| + |x_2| - 1 = 1 \cdot (|x_1| - 1) + 1 \cdot (|x_2|) \equiv (a, x - x_0)$. То есть, верно:
$$\forall x \in \mathbb{R}^2 \hookrightarrow f(x) - f(x_0) \geqslant (a, x - x_0)$$
То есть, a — субградиент.
3. $-a$ — не направление убывания в x_0. Пусть $t \in (0, 1)$. Рассмотрим $f(x_0 - ta) = |1 - t| + |-t| = 1 - t + t = 1$. Получаем $\forall t \in (0, 1) \hookrightarrow f(x_0 - ta) = f(x_0)$. Получаем,
$$\forall t_0 > 0 \, \exists t \overset{\text{def}}{=} \min\{1/2, t_0/2\} < t_0 \colon f(x_0 - ta) \geqslant f(x_0)$$
Это отрицания определения направления убывания.
4. x_0 — не точка минимума f: $f(x_0) = |1| + |0| = 1$, $f(0) = 0 < 1 = f(x_0)$.

Задача 3

Пусть $f(x_1, x_2) \colon \mathbb{R}^{n_1 + n_2} \to \mathbb{R}$, $g \colon \mathbb{R}^{n_1} \to \mathbb{R}$, $h \colon \mathbb{R}^{n_2} \to \mathbb{R}$, $\forall x_1, x_2 \hookrightarrow f(x_1, x_2) = g(x_1) + h(x_2)$, f, g, h — выпуклые и непрерывно дифференцируемые.

Пусть g достигает минимума в x_1^*, h — в x_2^*. Тогда $f(x_1, x_2) = g(x_1) + h(x_2) \geqslant g(x_1^*) + h(x_1^*) = f(x_1^*, x_2^*)$, т.е. (x_1^*, x_2^*) — также минимум f.

Пусть (x_1^0, x_2^0) — точка старта алгоритма.

Пусть $||g'_{x_1}|| \leqslant R_1$, $||h'_{x_2}|| \leqslant R_2$, $||x_1^0 - x_1^*|| \leqslant L_1$, $||x_2^0 - x_2^*|| \leqslant L_2$.

Тогда $||f'_{x_1, x_2}|| = ||(g'(x_1), h'(x_2))|| = \sqrt{R_1^2 + R_2^2}$, $||x^0 - x^*|| = ||(x_1^0 - x_1^*, x_2^0 - x_2^*))|| = \sqrt{L_1^2 + L_2^2}$.

Алгоритм, минимизирующий f целиком, сделает (в худшем случае) $k = \frac{R^2 L^2}{\varepsilon^2}$ шагов. На каждом шаге он вычисляет g, h, g', h', т.е. он сделает $a = 4k = 4\frac{R^2 L^2}{\varepsilon^2}$ операций

Алгоритм, минимизирующий только g, сделает (в худшем случае) $k_1 = \frac{R_1^2 L_1^2}{\varepsilon^2/4}$ шагов (берем $\varepsilon/2$, чтобы в сумме с такой же погрешностью h дало ε). На каждой итерации вычисляется g, g', т.е. каждый шаг требует двух операций: $a_1 = 2\frac{R_1^2 L_1^2}{\varepsilon^2/4}$.

Аналогично $a_2 = 2\frac{R_2^2 L_2^2}{\varepsilon^2/4}$

Вычитаем, получаем выигрыш: $a - a_1 - a_2 = 4\frac{(R_2^2 - R_1^2)(L_1^2 - L_2^2)}{\varepsilon^2} > 0$ при $\begin{cases} L_1 > L_2 \\ R_2 > R_1 \end{cases}$ или $\begin{cases} L_1 < L_2 \\ R_2 < R_1 \end{cases}$