

Методы оптимизации. Задание 2

Сергей Володин, 374 гр.

задано 2016.03.29

Задача 1

(2016.03.29) Доказать: Пусть f — β -гладкая. Тогда $\forall x, y \hookrightarrow f(x) \leq f(y) + \nabla^T f(y)(x - y) + \frac{\beta}{2} \|x - y\|^2$.

1. Имеем $\|\nabla f(x) - \nabla f(y)\|_* \leq \beta \|x - y\|$. Тогда
2. Обозначим $\mu(t) = f(y + t(x - y))$: $[0, 1] \rightarrow \mathbb{R}$. Поскольку f дифференцируема, μ также дифференцируема как композиция дифференцируемых функций. Тогда $\mu(1) = \mu(0) + \int_0^1 \mu'(t) dt$. Подставим определение μ , получим формулу Ньютона-Лейбница для f на отрезке $[y, x] \in \mathbb{R}^n$: $f(x) = f(y) + \int_0^1 dt \nabla^T f(y + t(x - y))(x - y)$.
3. Рассмотрим величину $\alpha \stackrel{\text{def}}{=} f(x) - f(y) - \nabla^T f(y)(x - y)$ и докажем, что $\alpha \leq 0$:
4. $\alpha = \int_0^1 dt \nabla^T f(y + t(x - y))(x - y) - \nabla^T f(y)(x - y)$. Внесем второе слагаемое под интеграл, получим

$$\alpha = \int_0^1 dt (\nabla^T f(y + t(x - y)) - \nabla^T f(y))(x - y)$$

5. $|\alpha| \leq \int_0^1 dt \underbrace{|\nabla^T f(y + t(x - y)) - \nabla^T f(y)|}_A (x - y)$.

6. A — оператор, действующий на $x - y$. Поскольку f — β -гладкая, т.е.

$$\forall x, y \hookrightarrow \|\nabla f(x) - \nabla f(y)\|_* \leq \beta \|x - y\|,$$

Получаем $\|A\|_* = \sup_{x \in \mathbb{R}} \frac{|Ax|}{\|x\|} \leq \beta \|y - t(x - y) - y\| = \beta t \|x - y\|$, откуда $|A(x - y)| \leq \|A\|_* \|x - y\| \leq \beta t \|x - y\|$

7. Получаем $|\alpha| \leq \int_0^1 \beta t \|x - y\|^2 dt = \frac{\beta}{2} \|x - y\|^2 \blacksquare$

Задача 2

Определим $\delta_t \stackrel{\text{def}}{=} f(x_t) - f(x^*)$, где x_t — t -я точка в алгоритме Frank-Wolfe. Получена оценка

$$\delta_{t+1} \leq \frac{\beta R^2}{2} \left(\prod_{k=1}^t (1 - \gamma_k) + \sum_{k=1}^t \gamma_{t-k}^2 \prod_{j=t-k+1}^t (1 - \gamma_j) \right) \equiv$$

Оценить выражение как функцию γ_t и выбрать γ_t как минимум этой функции.

Переобозначим $k \rightarrow t - k$ во втором слагаемом: $\equiv \left((1 - \gamma_0) \cdot \dots \cdot (1 - \gamma_t) + \sum_{k=1}^t \gamma_k^2 \prod_{j=k+1}^t (1 - \gamma_j) \right) \cdot \mu \stackrel{\text{def}}{=} \hat{\delta}_{t+1}$, где $\mu = \frac{\beta R^2}{2}$.

Обозначим

$$\Psi_{t+1} \stackrel{\text{def}}{=} \frac{\hat{\delta}_{t+1}}{\mu} = (1 - \gamma_1) \dots (1 - \gamma_t) + \sum_{k=1}^t \gamma_k^2 \prod_{j=k+1}^t (1 - \gamma_j)$$

Тогда

$$\Psi_{t+1} = (1 - \gamma_t) \Psi_t + \gamma_t^2$$

причем $\Psi_2 = 1$ (так как $\delta_2 \leq \frac{\beta R^2}{2}$, поскольку f — β -гладкая) и $\Psi_{t+1} = \Psi_{t+1}(\gamma_1, \dots, \gamma_t)$

Найдем минимум функции $\Psi_{t+1}(\gamma_1, \dots, \gamma_t)$.

Поскольку Ψ_t не зависит от γ_t , получаем:

$$\frac{\partial \Psi_{t+1}}{\partial \gamma_t} = -\Psi_t + 2\gamma_t = 0$$

Откуда

$$\gamma_t = \frac{\Psi_t}{2}$$

Пусть $l < t$. Найдем

$$\frac{\partial \Psi_{t+1}}{\partial \gamma_l} = (1 - \gamma_t) \frac{\partial \Psi_t}{\partial \gamma_l} = \dots = (1 - \gamma_t) \cdot \dots \cdot (1 - \gamma_{l+1}) \frac{\partial \Psi_{l+1}}{\partial \gamma_l} = 0$$

Считаем, что $\gamma_t \neq 1, \dots, \gamma_2 \neq 1$, откуда

$$\frac{\partial \Psi_{l+1}}{\partial \gamma_l} = 0 \Rightarrow \gamma_l = \frac{\Psi_l}{2}$$

Пусть $l \in \overline{2, t}$. Подставим γ_l в Ψ_{l+1} :

$$\Psi_{l+1} = (1 - \gamma_l) \Psi_l + \gamma_l^2 = \Psi_l - \frac{\Psi_l^2}{4}$$

Найдем

$$\gamma_{l+1} = \frac{1}{2} \Psi_{l+1} = \frac{1}{2} \left(\Psi_l - \frac{\Psi_l^2}{4} \right) = \frac{1}{2} (2\gamma_l - \gamma_l^2) = \gamma_l - \frac{\gamma_l}{2}$$

Получаем рекуррентную последовательность:

$$\begin{cases} \gamma_1 = 1 \\ \gamma_{l+1} = \gamma_l - \frac{\gamma_l^2}{2} \end{cases}$$

??? Сюда не подходит $\gamma_t = \frac{2}{t+1}$

Задача 3

$E = (\mathbb{R}^n, \|\cdot\|)$. Определим $f^*(p) \stackrel{\text{def}}{=} \sup_{x \in \mathbb{R}^n} (p^T x - f(x))$, $p \in E^*$. Найти субдифференциал $\partial f^*(p)$.

Пусть $f \in C(\mathbb{R}^n)$, f — выпуклая. Тогда $f^{**}(x) = f(x)$.

Пусть $x \in \mathbb{R}^n$. Рассмотрим $p \in \partial f(x) \Leftrightarrow \forall x' \in \mathbb{R}^n \hookrightarrow f(x') \geq f(x) + p^T(x' - x) \Leftrightarrow \forall x' \in \mathbb{R}^n \hookrightarrow f(x') - p^T x' \geq f(x) - p^T x \Leftrightarrow \forall x' \in \mathbb{R}^n \hookrightarrow p^T x' - f(x') \leq p^T x - f(x)$.

По определению, $f^*(p) = \sup_{x \in \mathbb{R}^n} (p^T x - f(x))$

Тогда получаем, что $f^*(p) = p^T x - f(x) \Leftrightarrow p \in \partial f(x)$. То есть,

$$p \in \partial f(x) \Leftrightarrow p^T x = f(x) + f^*(p)$$

Аналогично,

$$x \in \partial f^*(p) \Leftrightarrow x^T p = f^*(p) + f^{**}(x)$$

Поскольку $f^{**}(x) = f(x)$, получаем

$$x \in \partial f^*(p) \Leftrightarrow p^T x = f(x) + f^*(p) \Leftrightarrow p \in \partial f(x) \Leftrightarrow x \in (\partial f)^{-1}(p)$$

То есть,

$$\partial f^* = (\partial f)^{-1}$$