Методы оптимизации. Задание 2

Сергей Володин, 374 гр.

задано 2016.03.29

Задача 1

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

1. Имеем $||\nabla f(x) - \nabla f(y)||_* \leqslant \beta ||x-y||$. Тогда
2. Обозначим $\mu(t) = f(y+t(x-y))\colon [0,1] \to \mathbb{R}$. Поскольку f дифференцируема, μ также дифференцируема как композиция дифференцируемых функций. Тогда $\mu(1) = \mu(0) + \int\limits_0^1 \mu'(t)dt$. Подставим определение μ, получим формулу Ньютона-Лейбница для f на отрезке $[y, x] \in \mathbb{R}^n$: $f(x) = f(y) + \int\limits_0^1 dt \nabla^T f(y + t(x-y))(x-y)$.
3. Рассмотрим величину $\alpha \overset{\text{def}}{=} f(x) - f(y) - \nabla^T f(y)(x-y)$ и докажем, что $\alpha \leqslant 0$:
4. $\alpha = \int\limits_0^1 dt \nabla^T f(y+t(x-y))(x-y) - \nabla^T f(y)$. Внесем второе слагаемое под интеграл, получим
$$\alpha = \int\limits_0^1 dt \left(\nabla^T f(y+t(x-y)) - \nabla^T f(y)\right)(x-y)$$
5. $|\alpha| \leqslant \int\limits_0^1 dt |\underbrace{\left(\nabla^T f(y+t(x-y)) - \nabla^T f(y)\right)}_{A}(x-y)|$.
6. A— оператор, действующий на $x-y$. Поскольку f — β-гладкая, т.е.
$$\forall x, y \hookrightarrow ||\nabla f(x) - \nabla f(y)||_* \leqslant \beta ||x-y||,$$
Получаем $||A||_* = \sup\limits_{x \in \mathbb{R}} \frac{|Ax|}{||x||} \leqslant \beta ||y - t(x-y) - y|| = \beta t ||x-y||$, откуда $|A(x-y)| \leqslant ||A||_* ||x-y|| \leqslant \beta t ||x-y||^2$
7. Получаем $|\alpha| \leqslant \int\limits_0^1 \beta t ||x-y||^2 dt = \frac{\beta}{2}||x-y||^2$ ■

Задача 2

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

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

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

Переобозначим $k \to t-k$ во втором слагаемом: $\boxed{=} \left((1-\gamma_0)\cdot\ldots\cdot(1-\gamma_t) + \sum\limits_{k=1}^t \gamma_k^2 \prod\limits_{j=k+1}^t (1-\gamma_j)\right)\cdot \mu \overset{\text{def}}{=} \hat{\delta}_{t+1}$, где $\mu = \frac{\beta R^2}{2}$.

Обозначим

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

Тогда

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

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

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

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
$$\frac{\partial \Psi_{t+1}}{\partial \gamma_l} = (1-\gamma_t)\frac{\partial \Psi_t}{\partial \gamma_l} = ... = (1-\gamma_t)\cdot ... \cdot(1-\gamma_{l+1})\frac{\partial \Psi_{l+1}}{\partial \gamma_l} = 0$$
Считаем, что $\gamma_t \neq 1, ..., \gamma_2 \neq 1$, 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
$$\gamma_{l+1} = \frac{1}{2}\Psi_{l+1} = \frac{1}{2}(\Psi_l - \frac{\Psi_l^2}{4}) = \frac{1}{2}(2\gamma_l - \gamma_l^2) = \gamma_l - \frac{\gamma_l^2}{2}$$
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 $f^*(p) \overset{\text{def}}{=} \sup\limits_{x\in\mathbb{R}^n}(p^Tx - f(x))$, $p \in E^*$. 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 $p \in \partial f(x) \Leftrightarrow \forall\, x' \in \mathbb{R}^n \hookrightarrow f(x') \geqslant f(x) + p^T(x'-x) \Leftrightarrow \forall\, x' \in \mathbb{R}^n \hookrightarrow f(x') - p^Tx' \geqslant f(x) - p^Tx \Leftrightarrow \forall x' \in \mathbb{R}^n \hookrightarrow p^Tx' - f(x') \leqslant p^Tx - f(x')$.

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

Тогда получаем, что $f^*(p) = p^Tx - f(x) \Leftrightarrow p \in \partial f(x)$. То есть,
$$p \in \partial f(x) \Leftrightarrow p^Tx = f(x) + f^*(p)$$
Аналогично,
$$x \in \partial f^*(p) \Leftrightarrow x^Tp = f^*(p) + f^{**}(x)$$
Поскольку $f^{**}(x) = f(x)$, получаем
$$x \in \partial f^*(p) \Leftrightarrow p^Tx = f(x) + f^*(p) \Leftrightarrow p \in \partial f(x) \Leftrightarrow x \in (\partial f)^{-1}(p)$$
То есть,
$$\partial f^* = (\partial f)^{-1}$$