ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

Субградиенты. Негладкие задачи

МЕТОДЫ ВЫПУКЛОЙ ОПТИМИЗАЦИИ

НЕДЕЛЯ 13

Даня Меркулов
Пётр Остроухов

Субградиент. Негладкие задачи.

Семинар

Оптимизация для всех! ЦУ

Повторение основных понятий

Повторение основных понятий

Основные понятия

Для множества $E \in \mathbb{R}^n$ и функции $f : E \to \mathbb{R}$:

- Вектор $g \in \mathbb{R}^n$ называется **субградиентом** функции f в точке $x \in E$, если $\forall y \in E$

$$f(y) \geq f(x) + g^T(y - x)$$

Повторение основных понятий

Основные понятия

Для множества $E \in \mathbb{R}^n$ и функции $f : E \to \mathbb{R}$:

- Вектор $g \in \mathbb{R}^n$ называется **субградиентом** функции f в точке $x \in E$, если $\forall y \in E$

$$f(y) \geq f(x) + g^T(y - x)$$

- Множество $\partial f(x)$ называется **субдифференциалом** функции f в точке $x \in E$, если:

$$\partial f(x) = \{g \in \mathbb{R}^n \mid f(y) \geq f(x) + g^T(y - x)\} \forall y \in E$$

Повторение основных понятий

Основные понятия

Для множества $E \in \mathbb{R}^n$ и функции $f : E \to \mathbb{R}$:

- Вектор $g \in \mathbb{R}^n$ называется **субградиентом** функции f в точке $x \in E$, если $\forall y \in E$

$$f(y) \geq f(x) + g^T(y - x)$$

- Множество $\partial f(x)$ называется **субдифференциалом** функции f в точке $x \in E$, если:

$$\partial f(x) = \{g \in \mathbb{R}^n \mid f(y) \geq f(x) + g^T(y - x)\} \forall y \in E$$

- $f(\cdot)$ называется **субдифференцируемой** в точке $x \in E$, если $\partial f(x) \neq \emptyset$

Связь между субдифференцируемостью и выпуклостью

Связь между субдифференцируемостью и выпуклостью

Если $f : E \to \mathbb{R}$ субдифференцируема на **выпуклом** подмножестве $S \in E$, то f выпукла на S.

- Обратное, вообще говоря, неверно

Связь между субдифференцируемостью и выпуклостью

Связь между субдифференцируемостью и выпуклостью

Если $f : E \to \mathbb{R}$ субдифференцируема на **выпуклом** подмножестве $S \in E$, то f выпукла на S.

- Обратное, вообще говоря, неверно
- Нет смысла искать субградиент невыпуклой функции.

Связь между субдифференцируемостью и дифференцируемостью

Связь между субдифференцируемостью и дифференцируемостью

1) Если $f : E \to \mathbb{R}$ выпукла и дифференцируема в точке $x \in \text{int } E$, то $\partial f(x) = \{\nabla f(x)\}$

Связь между субдифференцируемостью и дифференцируемостью

Связь между субдифференцируемостью и дифференцируемостью

1) Если $f : E \to \mathbb{R}$ выпукла и дифференцируема в точке $x \in \operatorname{int} E$, то $\partial f(x) = \{\nabla f(x)\}$
2) Если $f : E \to \mathbb{R}$ выпукла и для $x \in \operatorname{int} E$ $\partial f(x) = \{s\}$, то f дифференцируема в точке x и $\nabla f(x) = s$

Связь между субдифференцируемостью и дифференцируемостью

Связь между субдифференцируемостью и дифференцируемостью

1) Если $f : E \to \mathbb{R}$ выпукла и дифференцируема в точке $x \in \text{int } E$, то $\partial f(x) = \{\nabla f(x)\}$
2) Если $f : E \to \mathbb{R}$ выпукла и для $x \in \text{int } E$ $\partial f(x) = \{s\}$, то f дифференцируема в точке x и $\nabla f(x) = s$

- Искать субдифференциал дифференцируемой функции — это излишество.

Субградиентный спуск

Субградиентный метод — это простой алгоритм для минимизации выпуклой функции, которая не является дифференцируемой:

$$x_{k+1} = x_k - \alpha_k g_k$$

Здесь $g_k \in \partial f(x_k)$ и $\alpha_k > 0$. Существует несколько известных стратегий выбора α_k для этого метода:

- $\alpha_k = \alpha$ – фиксированный (для G-Липшицевых функций это может дать постоянную невязку $\frac{G^2\alpha}{2}$ между f^* и f_k)

Субградиентный спуск

Субградиентный метод — это простой алгоритм для минимизации выпуклой функции, которая не является дифференцируемой:

$$x_{k+1} = x_k - \alpha_k g_k$$

Здесь $g_k \in \partial f(x_k)$ и $\alpha_k > 0$. Существует несколько известных стратегий выбора α_k для этого метода:

- $\alpha_k = \alpha$ – фиксированный (для G-Липшицевых функций это может дать постоянную невязку $\frac{G^2\alpha}{2}$ между f^* и f_k)
- $\alpha_k = \dfrac{\gamma}{\|g_k\|_2}$ – постоянная длина шага ($\|x_{k+1} - x_k\| = \gamma$)

Субградиентный спуск

Субградиентный метод — это простой алгоритм для минимизации выпуклой функции, которая не является дифференцируемой:

$$x_{k+1} = x_k - \alpha_k g_k$$

Здесь $g_k \in \partial f(x_k)$ и $\alpha_k > 0$. Существует несколько известных стратегий выбора α_k для этого метода:

- $\alpha_k = \alpha$ – фиксированный (для G-Липшицевых функций это может дать постоянную невязку $\frac{G^2\alpha}{2}$ между f^* и f_k)
- $\alpha_k = \dfrac{\gamma}{\|g_k\|_2}$ – постоянная длина шага ($\|x_{k+1} - x_k\| = \gamma$)
- α_k : $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$ и $\sum_{k=1}^{\infty} \alpha_k = \infty$ – квадратично суммируемая, но не суммируемая

Субградиентный спуск

Субградиентный метод — это простой алгоритм для минимизации выпуклой функции, которая не является дифференцируемой:

$$x_{k+1} = x_k - \alpha_k g_k$$

Здесь $g_k \in \partial f(x_k)$ и $\alpha_k > 0$. Существует несколько известных стратегий выбора α_k для этого метода:

- $\alpha_k = \alpha$ – фиксированный (для G-Липшицевых функций это может дать постоянную невязку $\frac{G^2\alpha}{2}$ между f^* и f_k)
- $\alpha_k = \dfrac{\gamma}{\|g_k\|_2}$ – постоянная длина шага ($\|x_{k+1} - x_k\| = \gamma$)
- α_k : $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$ и $\sum_{k=1}^{\infty} \alpha_k = \infty$ – квадратично суммируемая, но не суммируемая
- α_k : $\lim_{k\to\infty} \alpha_k = 0$ и $\sum_{k=1}^{\infty} \alpha_k = \infty$ – несуммируемая убывающая

Субградиентный спуск

Субградиентный метод — это простой алгоритм для минимизации выпуклой функции, которая не является дифференцируемой:

$$x_{k+1} = x_k - \alpha_k g_k$$

Здесь $g_k \in \partial f(x_k)$ и $\alpha_k > 0$. Существует несколько известных стратегий выбора α_k для этого метода:

- $\alpha_k = \alpha$ – фиксированный (для G-Липшицевых функций это может дать постоянную невязку $\frac{G^2\alpha}{2}$ между f^* и f_k)
- $\alpha_k = \dfrac{\gamma}{\|g_k\|_2}$ – постоянная длина шага ($\|x_{k+1} - x_k\| = \gamma$)
- $\alpha_k : \sum_{k=1}^{\infty} \alpha_k^2 < \infty$ и $\sum_{k=1}^{\infty} \alpha_k = \infty$ – квадратично суммируемая, но не суммируемая
- $\alpha_k : \lim_{k\to\infty} \alpha_k = 0$ и $\sum_{k=1}^{\infty} \alpha_k = \infty$ – несуммируемая убывающая
- $\alpha_k = \dfrac{f(x_k) - f^*}{\|g_k\|_2^2}$ – шаг Поляка

Задача 1

Question

Найти субдифференциал функции

$$f(x) = -\sqrt{x}$$

Правила субдифференцирования

1) $f : E \to \mathbb{R}$, $x \in E$, $c > 0$

$$\Rightarrow \partial(cf)(x) = c\partial f(x)$$

Правила субдифференцирования

1) $f : E \to \mathbb{R}$, $x \in E$, $c > 0$

$$\Rightarrow \partial(cf)(x) = c\partial f(x)$$

2) $f : F \to \mathbb{R}$, $g : G \to \mathbb{R}$, $x \in F \bigcap G$

$$\Rightarrow \partial(f+g)(x) \supseteq \partial f(x) + \partial g(x)$$

Правила субдифференцирования

1) $f : E \to \mathbb{R}$, $x \in E$, $c > 0$

$$\Rightarrow \partial(cf)(x) = c\partial f(x)$$

2) $f : F \to \mathbb{R}$, $g : G \to \mathbb{R}$, $x \in F \bigcap G$

$$\Rightarrow \partial(f+g)(x) \supseteq \partial f(x) + \partial g(x)$$

3) $T : V \to W = Ax + b$, $g : W \to \mathbb{R}$, $x_0 \in V$

$$\Rightarrow \partial(g \circ T)(x_0) \supseteq A^* \, \partial(g)(T(x_0))$$

Правила субдифференцирования

1) $f : E \to \mathbb{R}$, $x \in E$, $c > 0$

$$\Rightarrow \partial(cf)(x) = c\partial f(x)$$

2) $f : F \to \mathbb{R}$, $g : G \to \mathbb{R}$, $x \in F \bigcap G$

$$\Rightarrow \partial(f + g)(x) \supseteq \partial f(x) + \partial g(x)$$

3) $T : V \to W = Ax + b$, $g : W \to \mathbb{R}$, $x_0 \in V$

$$\Rightarrow \partial(g \circ T)(x_0) \supseteq A^* \, \partial(g)(T(x_0))$$

4) $f(x) = \text{max}(f_1(x), \dots, f_m(x))$, $I(x) = \{i \in 1 \dots m | f_i(x) = f(x)\}$

$$\Rightarrow \partial f(x) \supseteq \text{Conv}(\bigcup_{i \in I(x)} \partial f_i(x))$$

Правила субдифференцирования

1) $f : E \to \mathbb{R}$, $x \in E$, $c > 0$

$$\Rightarrow \partial(cf)(x) = c\partial f(x)$$

2) $f : F \to \mathbb{R}$, $g : G \to \mathbb{R}$, $x \in F \bigcap G$

$$\Rightarrow \partial(f+g)(x) \supseteq \partial f(x) + \partial g(x)$$

3) $T : V \to W = Ax + b$, $g : W \to \mathbb{R}$, $x_0 \in V$

$$\Rightarrow \partial(g \circ T)(x_0) \supseteq A^* \, \partial(g)(T(x_0))$$

4) $f(x) = \text{max}(f_1(x), \dots, f_m(x))$, $I(x) = \{i \in 1 \dots m | f_i(x) = f(x)\}$

$$\Rightarrow \partial f(x) \supseteq \text{Conv}(\bigcup_{i \in I(x)} \partial f_i(x))$$

Правила субдифференцирования

1) $f: E \to \mathbb{R}$, $x \in E$, $c > 0$

$$\Rightarrow \partial(cf)(x) = c\partial f(x)$$

2) $f: F \to \mathbb{R}$, $g: G \to \mathbb{R}$, $x \in F \bigcap G$

$$\Rightarrow \partial(f+g)(x) \supseteq \partial f(x) + \partial g(x)$$

3) $T: V \to W = Ax + b$, $g: W \to \mathbb{R}$, $x_0 \in V$

$$\Rightarrow \partial(g \circ T)(x_0) \supseteq A^* \partial(g)(T(x_0))$$

4) $f(x) = \text{max}(f_1(x), \dots, f_m(x))$, $I(x) = \{i \in 1 \dots m | f_i(x) = f(x)\}$

$$\Rightarrow \partial f(x) \supseteq \text{Conv}(\bigcup_{i \in I(x)} \partial f_i(x))$$

Когда достигается равенство?

Если вышеупомянутые функции выпуклы и x является внутренней точкой, то все неравенства превращаются в равенства.

Задача 2

Question

Найти субдифференциал функции $h(x)$, если

$$h(x) = \begin{cases} f(x) \equiv -\sqrt{x} & \text{при } x \geq 0 \\ -\sqrt{-x} & \text{при } x \leq 0 \end{cases}$$

Задача 3

Question

1) Найти субдифференциал функции $f(x) = ||Ax - b||_1$;

Задача 3

Question

1) Найти субдифференциал функции $f(x) = ||Ax - b||_1$;
2) Для задачи $f(x) = \frac{1}{2}||Ax - b||_2^2 + \lambda||x||_1 \to \min_x$ сказать, какие значения λ приводят к $x_{opt} = 0$

Задача 4

Exercise

Мы хотим найти вектор весов $w \in \mathbb{R}^d$, определяющий линейный классификатор $sign(w^T x)$. Используем SVM с l_2-регуляризацией:

$$L(w) = \frac{\lambda}{2}\|w\|^2 + \frac{1}{N}\sum_{i=1}^{N} \max\{0,\ 1 - y_i w^T x_i\}$$

Здесь $\ell_i(w) = \max\{0,\ 1 - y_i w^T x_i\}$ — это hinge loss. Покажите, что субградиент в любой точке задается выражением:

$$g(w) = \lambda w + \frac{1}{N}\sum_{i=1}^{N} \begin{cases} -y_i x_i, & \text{если } y_i w^T x_i < 1, \\ 0, & \text{иначе} \end{cases}$$

и посмотрите пример кода здесь.

Задача 5

Question

Найти субдифференциал $\partial f(x)$ функции $f(x) = \exp(|x-1| + |x+1|)$ для всех $x \in \mathbb{R}$