Методы оптимизации. Семинар 6. Субдифференциальное исчисление.

Корнилов Никита Максимович

Московский физико-технический институт

8 октября 2024г

Субдифференциал

Обощение свойств градиента на точки недифференцируемости.

Definition (Субградиент)

Пусть $f : E \to \mathbb{R}$ функция, определенная на подмножестве E евклидового пространства V и $x_0 \in E$. Вектор $s \in V$ называется *субградиентом функции f в точке x_0*, если

$$f(x) \geq f(x_0) + \langle s, x - x_0 \rangle, \qquad x \in E.$$

- Множество всех субград. в точке — *субдифференциал f в точке x_0*, обозначение $\partial f(x_0)$.
- Если $\partial f(x_0) \neq \emptyset$, то *функция f называется субдифференцируемой в точке x_0*.

Посчитаем по определению

Example

Пусть $f : \mathbb{R} \to \mathbb{R}$ функция $f(x) = |x|$. Посчитайте $\partial f(x)$ для $x \in \mathbb{R}$.

Свойства субдифференциала для **выпуклых функций**

1. Субдифференциал — это выпуклое замкнутое множество (возможно, пустое)
$\partial f(x) = \cap_{x \in E}\{s \in V : f(x) \geq f(x_0) + \langle s, x - x_0 \rangle\}$

Свойства субдифференциала для **выпуклых функций**

1. Субдифференциал — это выпуклое замкнутое множество (возможно, пустое)
$\partial f(x) = \cap_{x \in E}\{s \in V : f(x) \geq f(x_0) + \langle s, x - x_0 \rangle\}$
2. Пусть $x_0 \in int(E)$ и функция f выпуклая. Тогда функция дифференцируемая в точке x_0 когда и только когда субдифференциал состоит из одного элемента $\partial f(x_0) = \{\nabla f(x_0)\}$.
3. Пусть $x_0 \in int(E)$ и функция f выпуклая. Тогда $\partial f(x_0)$ не пуст и является выпуклым компактом.

Свойства субдифференциала для **выпуклых функций**

1. Субдифференциал — это выпуклое замкнутое множество (возможно, пустое)
$\partial f(x) = \cap_{x \in E}\{s \in V : f(x) \geq f(x_0) + \langle s, x - x_0 \rangle\}$
2. Пусть $x_0 \in int(E)$ и функция f выпуклая. Тогда функция дифференцируемая в точке x_0 когда и только когда субдифференциал состоит из одного элемента $\partial f(x_0) = \{\nabla f(x_0)\}$.
3. Пусть $x_0 \in int(E)$ и функция f выпуклая. Тогда $\partial f(x_0)$ не пуст и является выпуклым компактом.
4. **Критерий глобального минимума**. Точка x_0 — глобальный минимум функции f на E когда и только когда $0 \in \partial f(x_0)$.

Свойства субдифференциала для **произвольных функций**

1. Субдифференциал — это всё ещё выпуклое замкнутое множество (возможно, пустое)
2. Пусть $x_0 \in int(E)$ и f дифференцируема в x_0. Тогда либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$.

Свойства субдифференциала для **произвольных функций**

1. Субдифференциал — это всё ещё выпуклое замкнутое множество (возможно, пустое)
2. Пусть $x_0 \in int(E)$ и f дифференцируема в x_0. Тогда либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$.
3. **Критерий глобального минимума**. Точка x_0 — глобальный минимум функции f на E когда и только когда $0 \in \partial f(x_0)$.

Примеры с дифф функциями

Example

Пусть $f : \mathbb{R} \to \mathbb{R}$ функция $f(x) = \sin(x)$ на $E = \left[\frac{\pi}{2}, 2\pi\right]$. Найдите $\partial f(x)$.

Многомерный пример

Example (Норма)

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ функция $f(x) = \|x\|$. Найдите $\partial f(x)$.
Hint. По определению, сопряженная норма $\|g\|_* := \sup_{\|x\| \le 1} \langle g, x \rangle$.

(Суб)Дифференцируемость и выпуклость

- Если функция f субдифференцируемая на E, то она выпукла на E
- В общем случае из выпуклости не следует субдиф $f(x) = -\sqrt{x}$.
- Если функция f выпуклая, то она может не дифференцируема только в счетном числе точек.

Субдифференциальное исчисление I

- **Умножение на константу.** Пусть $x_0 \in E, c \geq 0$, тогда

$$\partial \left[c \cdot f \right] (x_0) = c \cdot \partial f (x_0)$$

Субдифференциальное исчисление I

- **Умножение на константу.** Пусть $x_0 \in E, c \geq 0$, тогда
$$\partial\left[c \cdot f\right](x_0) = c \cdot \partial f(x_0)$$

- **Сумма.** Пусть f определена на E, g определена на G, а $x_0 \in E \cap G$, тогда
$$\partial(f+g)(x_0) \supseteq \partial f(x_0) + \partial g(x_0)$$
Если f, g — выпуклы, а $E \cap int(G) \neq \emptyset$, то
$$\partial(f+g)(x_0) = \partial f(x_0) + \partial g(x_0)$$
Обобщение на m выпуклых функций и $c_i \geq 0, i \in \overline{1, m}$
$$\partial\left(\sum_{i=1}^{m} f_i\right)(x_0) = \sum_{i=1}^{m} c_i \partial f_i(x_0).$$

- **Конечный максимум.** Пусть $f_i : E_i \to \mathbb{R}$ функции для $i = \overline{1, m}$ и $f(x) := \max_{i=\overline{1,m}} f_i(x)$. Тогда для $x_0 \in \cap_{i=1}^m E_i$

$$\partial f(x_0) \supseteq \overline{\mathrm{Conv}} \left(\bigcup_{i \in I(x_0)} \partial f_i(x_0) \right),$$

где $I(x_0)$ – множество индексов, на которых достигается max.
Если f_i выпуклые и $x_0 \in \cap_{i=1}^m int(E_i)$, тогда

$$\partial f(x_0) = \mathrm{Conv} \left(\bigcup_{i \in I(x_0)} \partial f_i(x_0) \right).$$

- **Бесконечный максимум.** Пусть $f_i : E_i \to \mathbb{R}$ функции для $i \in I$, где I — произвольное множество. При этом $f(x) := \max_{i \in I} f_i(x)$, определенная на $E := \{x \in \cap_{i=1}^m E_i, \max_{i \in I} f_i(x) < +\infty\}$. Тогда для $x_0 \in E$

$$\partial f(x_0) \supseteq \overline{\mathrm{Conv}} \left(\bigcup_{i \in I(x_0)} \partial f_i(x_0) \right).$$

Если f_i выпуклые, I компактно и $i \to f_i(x)$ полунепрерывна сверху на I для любого x, тогда для $x_0 \in int(E)$

$$\partial f(x_0) = \mathrm{Conv} \left(\bigcup_{i \in I(x_0)} \partial f_i(x_0) \right).$$

Субдифференциальное исчисление IX

- **Композиция.** Пусть $g_i(x) : \mathbb{R}^n \to \mathbb{R}$ выпуклые на E для $i \in \overline{1, m}$, $g(x) : \mathbb{R}^n \to \mathbb{R}^m = (g_1(x), \ldots, g_m(x))^\top$, $\phi(u) : \mathbb{R}^m \to \mathbb{R}$ неубывающая выпуклая функция на $U \supseteq g(E)$. Тогда для композиции $f(x) : \mathbb{R}^n \to \mathbb{R} := \phi(g(x))$ и $x_0 \in int(E)$

$$\partial f(x_0) = \bigcup_{p \in \phi(g(x_0))} \sum_{i=1}^{m} p_i \partial g_i(x_0).$$

Если ϕ дифф, то

$$\partial f(x_0) = \sum_{i=1}^{m} \frac{\partial \phi}{\partial u_i} \partial g_i(x_0).$$

В частности, аффинное преобразование $g(x) = Ax + b$ и выпуклая f

$$\partial \left(f(Ax + b) \right) (x_0) = A^\top \partial f(Ax_0 + b)$$

Примеры

Example (Модуль через max)

Посчитайте $\partial f(x)$ для $f(x) : \mathbb{R} \to \mathbb{R} = |x|$ через субдифф максимума.

Example

Посчитайте $\partial f(x)$ для $f(x) : \mathbb{R} \to \mathbb{R} = |x + 1| + |x - 1|$.

Example

Посчитайте $\partial f(x)$ для $f(x) : \mathbb{R}^n \to \mathbb{R} = \exp(\|Ax + b\|_1)$

Example (Норма через max)

Посчитайте $\partial f(x)$ для $f(x) : \mathbb{R}^n \to \mathbb{R} = \|x\|$ через субдифф max.

Moreau-Yosida Envelope

- $f(x)$ выпуклая, но негладкая
- Moreau-Yosida envelope($\lambda > 0$)

$$M_{\lambda f}(x) = \inf_u \left(f(u) + \frac{1}{2\lambda} \|u - x\|_2^2 \right)$$

- Посчитаем $M_{\lambda f}$ для $|x|$

Moreau-Yosida Envelope

- $f(x)$ выпуклая, но негладкая
- Moreau-Yosida envelope($\lambda > 0$)

$$M_{\lambda f}(x) = \inf_u \left(f(u) + \frac{1}{2\lambda}\|u - x\|_2^2 \right)$$

- Посчитаем $M_{\lambda f}$ для $|x|$

$$M_{\lambda|\cdot|}(x) = \begin{cases} \frac{x^2}{2\lambda}, & |x| \le \lambda \\ |x| - \lambda/2, & |x| \ge \lambda. \end{cases}$$

Свойства Moreau-Yosida Envelope

- $M_{\lambda f}(x)$ — выпуклая (infimal convolution)
- $M^*_{\lambda f}(y) = f^*(y) + \frac{\lambda}{2}\|y\|_2^2$ λ-сильно выпуклая
- $M_{\lambda f}(x) = M^{**}_{\lambda f}(x)$
- Сопряженная к λ-сильно выпуклой — $\frac{1}{\lambda}$-гладкая
- Можно построить градиентные методы над $M_{\lambda f}$

$$\nabla_x M_{\lambda f}(x) = \frac{1}{\lambda}(x - u^*), \quad u^* = \arg\min\left(f(u) + \frac{1}{2\lambda}\|u - x\|_2^2\right)$$

- **Множество минимумов f и $M_{\lambda f}$ совпадают**

Субградиентный спуск

Задача

$$\min_{x \in \mathbb{R}^n} f(x)$$

Субградиентный спуск

$$x^{k+1} = x^k - \gamma^k g^k, \quad \gamma^k = \frac{1}{(1+k)}$$

где $g^k \in \partial f(x^k)$. Ответ даём как $\min_{k=\overline{1,K}} f(x^k)$.
Оптимальная сходимость для M-липшецевой функции f по значению

$$\mu\text{-сильно выпуклая:} \quad O\left(\frac{M^2}{\mu K}\right)$$

$$\text{Выпуклая:} \quad O\left(\frac{MR}{\sqrt{K}}\right)$$