Методы оптимизации. Семинар 7
Субдифференциальное исчисление.

Корнилов Никита Максимович

Московский физико-технический институт

16 октября 2025г

Субдифференциал

Обощение свойств градиента на точки недифференцируемости.

Definition (Субградиент)

Пусть дана функция $f : U \to \overline{\mathbb{R}}$ и точка $x_0 \in \operatorname{dom} f$. Элемент $s \in U$ называется *субградиентом функции f в точке x_0*, если

$$f(x) \geq f(x_0) + \langle s, x - x_0 \rangle, \quad \forall x \in \operatorname{dom} f.$$

- Множество всех субград. в точке $x_0 \in \operatorname{dom} f$ называется *субдифференциалом f в точке x_0* и обозначается как $\partial f(x_0)$. Для $x_0 \notin \operatorname{dom} f$ считаем $\partial f(x_0) = \emptyset$.
- 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 модуля $f(x) = |x|$. Посчитайте субдифференциал $\partial f(0)$ в точке 0.

Example

Пусть $f : \mathbb{S}^n \to \mathbb{R}$ задаётся как $f(X) = \lambda_{\max}(X)$. Покажите, что $uu^\top \in \partial f(X)$, где u — нормированный собственный вектор, соответствующий наибольшему собственному числу $\lambda_{\max}(X)$.

Example

Пусть даны функции $f : \mathbb{R}^n \to \mathbb{R}$ и $h : \mathbb{R}^n \to \mathbb{R}^d$. По определению функция $g : \mathbb{R}^d \to \mathbb{R}$ задана как

$$g(\lambda) = -\min{}_{x \in \mathbb{R}^n} \left\{ f(x) + \langle \lambda, h(x) \rangle \right\}.$$

Покажите, что $-h(x_0) \in \partial g(\lambda_0)$ для $\lambda_0 \in \mathrm{dom}\, g$, где x_0 — точка, в которой достигается min в $g(\lambda_0)$.

Свойства субдифференциала для **выпуклых функций**

1. Субдифференциал — это выпуклое замкнутое множество (возможно, пустое) $\partial f(x_0) = \cap_{x \in \text{dom } f}\{s \in U : f(x) \geq f(x_0) + \langle s, x - x_0 \rangle\}$
2. Пусть $x_0 \in int(\text{dom } f)$ и функция f выпуклая. Тогда $\partial f(x_0)$ не пуст и является выпуклым компактом (f - субдиф на $int(\text{dom } f)$).
3. Пусть $x_0 \in int(\text{dom } f)$ и функция f выпуклая. Тогда функция дифференцируемая в точке x_0 тогда и только тогда, когда субдифференциал состоит из одного элемента $\partial f(x_0) = \{\nabla f(x_0)\}$.
4. **Критерий глобального минимума**. Точка x_0 — глобальный минимум функции $f : U \to \mathbb{R}$ тогда и только тогда, когда $0 \in \partial f(x_0)$.
5. **Критерий условного минимума**. Точка $x_0 \in Q$ — условный минимум функции $f : U \to \mathbb{R}$ на $Q \subset U$ тогда и только тогда, когда $\exists g \in \partial f(x_0) : \langle g, x - x_0 \rangle \geq 0, \forall x \in Q$.

Свойства субдифференциала для **произвольных функций**

1. Субдифференциал — это всё ещё выпуклое замкнутое множество (возможно, пустое).
2. Пусть $x_0 \in int(\text{dom } f)$ и f дифференцируема в x_0. Тогда либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$.
   Только в случае выпуклых функций множество дифференцируемых точек является подмножеством субдифференцируемых!
3. **Критерий глобального минимума**. Точка x_0 — глобальный минимум функции $f : U \to \mathbb{R}$ тогда и только тогда, когда $0 \in \partial f(x_0)$.

Примеры

Example

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

Example

Пусть $f : \left[0, \frac{3\pi}{2}\right] \to \mathbb{R}$ функция $f(x) = \cos(x)$. Найдите $\partial f(x)$.

Example (Норма)

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

(Суб)Дифференцируемость и выпуклость

- Если функция f имеет открытый $\mathrm{dom}\, f$, то субдифференцируемость эквивалентна выпуклости.
- В общем случае для неоткрытых $\mathrm{dom}\, f$ из выпуклости не следует субдифференцируемость на граничных точках ($f(x) = -\sqrt{x}$).
- Но из субдифференцируемости всегда следует выпуклость.
- Если функция f выпуклая, то она может быть не дифференцируема только в счетном числе точек из int $\mathrm{dom}\, f$.

Общий алгоритм

Для выпуклых функций:

1. Для внутренних дифференцируемых точек считаем градиент, он и есть субдифференциал $\partial f(x_0) = \{\nabla f(x_0)\}$.
2. Для недифф, сложно дифференцируемых или граничных точек — смотрим по субдифф исчислению или по определению (или на график).

Для невыпуклых функций:
Внимательно изучаем каждую внутреннюю и граничную точку по определению или графику.

Субдифференциальное исчисление I

Для простоты изложения считаем, что dom всех функций это U.

- **Умножение на константу.** Пусть дана произвольная функция f, точка $x_0 \in U$ и положительный коеф $c \geq 0$, тогда

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

- **Сумма.** Пусть даны *выпуклые* функции f и g на U, тогда

$$\partial(f + g)(x_0) = \partial f(x_0) + \partial g(x_0).$$

Обобщение на m выпуклых функций и $c_i \geq 0, i \in \overline{1, m}$

$$\partial \left(\sum_{i=1}^{m} f_i \right)(x_0) = \sum_{i=1}^{m} c_i \partial f_i(x_0).$$

Если f, g — произвольные, то верно только

$$\partial(f + g)(x_0) \supseteq \partial f(x_0) + \partial g(x_0).$$

Примеры

Example

Посчитайте $\partial f(x)$ для $f(x) : \mathbb{R} \to \mathbb{R} = 2|x+1| + |x-1|$.

- **Аффинное преобразование.** Аффинное преобразование $g(x) = Ax + b$ и *выпуклая* f в точке x_0 дают
$$\partial\left(f(Ax+b)\right)(x_0) = A^\top \partial f(Ax_0 + b).$$
- **Неубывающая композиция.** Пусть $g_i(x) : \mathbb{R}^n \to \mathbb{R}$ *выпуклые* функции для всех $i \in \overline{1, m}$, а $\phi(u) : \mathbb{R}^m \to \mathbb{R}$ *неубывающая выпуклая* функция. Тогда для композиции $f(x) : \mathbb{R}^n \to \mathbb{R} := \phi(g_1(x), \dots, g_m(x))$ и точки x_0 верно
$$\partial f(x_0) = \bigcup_{p \in \partial \phi(g(x_0))} \sum_{i=1}^{m} p_i \partial g_i(x_0).$$
Если ϕ дифф, то
$$\partial f(x_0) = \sum_{i=1}^{m} \frac{\partial \phi}{\partial u_i} \partial g_i(x_0).$$

Примеры

Example

Посчитайте $\partial f(x)$ для $f(x) : \mathbb{R}^n \to \mathbb{R} = \exp(\|Ax + b\|_2)$.

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

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

где $I(x_0) := \{i \in \overline{1, m} : f_i(x_0) = f(x_0)\}$ – множество индексов, на которых достигается max.
Если f_i — произвольные, то верно только

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

- **Бесконечный максимум.** Пусть даны *выпуклые* функции $f_i : U \to \mathbb{R}$ функции для $i \in I$, где I — произвольное *компактное* множество такое, что $i \to f_i(x)$ полунепрерывна сверху на I. Тогда для $f(x) := \max_{i \in I} f_i(x)$ и точки $x_0 \in U$ верно

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

где $I(x_0) := \{i \in I : f_i(x_0) = f(x_0)\}$ – множество индексов, на которых достигается max.

Если f_i и I произвольные, то верно только

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

Примеры

Example (Модуль через max)

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

Example (Норма через max)

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

Example (ℓ_1-норма)

Посчитайте $\partial f(x)$ для $f(x) : \mathbb{R}^n \to \mathbb{R} = \|x\|_1$ через сумму.

Moreau-Yosida Envelope

- $f(x) : \mathbb{R}^n \to \mathbb{R}$ выпуклая, но негладкая.
- Moreau-Yosida envelope($\lambda > 0$) позволяет сделать новую выпуклую гладкую функцию

$$M_{\lambda f}(x) = \inf_{u \in \mathbb{R}^n} \left(f(u) + \frac{1}{2\lambda} \|u - x\|_2^2 \right).$$

- 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) и $\frac{1}{\lambda}$-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.**