Выпуклость. Гладкость

Методы оптимизации

Александр Безносиков

Московский физико-технический институт

11 сентября 2025

ФПМИ

Глобальный и локальный минимумы

Рассмотрим безусловную задачу: $\min_{x \in \mathbb{R}^d} f(x)$.

Локальный минимум

Точка x^* называется локальным минимумом функции f на $\mathbb{R}^d$ (локальным решением задачи минимизации f на $\mathbb{R}^d$), если существует $r > 0$ такое, что для любого $y \in B_2^d(r, x^*) = \{y \in \mathbb{R}^d \mid \|y - x^*\|_2 \leq r\}$ следует, что $f(x^*) \leq f(y)$.

Глобальный минимум

Точка x^* называется глобальным минимумом функции f на $\mathbb{R}^d$ (глобальным решением задачи минимизации f на $\mathbb{R}^d$), если для любого $y \in \mathbb{R}^d$ следует, что $f(x^*) \leq f(y)$.

Определение можно обобщить и до локального/глобального минимума на множестве $\mathcal{X}$, т.е. для задачи вида $\min_{x \in \mathcal{X}} f(x)$. Для этого надо брать $y \in B_2^d(r, x^*) \cap \mathcal{X}$ и $y \in \mathcal{X}$ в соответствующих определениях.

Условие оптимальности: общий случай

Теорема об условии оптимальности локального минимума

Пусть x^* – локальный минимумом функции f на $\mathbb{R}^d$, тогда если f дифференцируема, то $\nabla f(x^*) = 0$.

Легко проверить, что обратное неверно.

Условие оптимальности: общий случай

Доказательство

Пойдем от противного и предположим $\nabla f(x^*) \neq 0$. Разложим в ряд в окрестности локального минимума:

$$f(x) = f(x^*) + \langle \nabla f(x^*), x - x^* \rangle + o(\|x - x^*\|_2),$$

где $\lim_{x \to x^*} \frac{o(\|x - x^*\|_2)}{\|x - x^*\|_2} = 0$.
Рассмотрим $\tilde{x} = x^* - \lambda \nabla f(x^*)$. Цель: выбрать λ, чтобы $\tilde{x}$ попал в нужную окрестность из определения локального минимума. Понятно, что такое λ можно найти. Тогда с одной стороны:

$$f(\tilde{x}) \geq f(x^*), \quad \text{и}$$

$$\begin{aligned} f(\tilde{x}) &= f(x^*) + \langle \nabla f(x^*), \tilde{x} - x^* \rangle + o(\|\tilde{x} - x^*\|_2) \\ &= f(x^*) - \lambda \|\nabla f(x^*)\|^2 + o(\lambda \|\nabla f(x^*)\|_2) \end{aligned}$$

Условие оптимальности: общий случай

Доказательство

Набросим еще одно ограничение на "малость" λ. Пусть теперь еще выполнено, что $|o(\lambda\|\nabla f(x^*)\|_2)| \leq \frac{\lambda}{2}\|\nabla f(x^*)\|_2^2$. Тогда для подобранного $\lambda > 0$

$$f(\tilde{x}) \leq f(x^*) - \frac{\lambda}{2}\|\nabla f(x^*)\|^2$$

Пришли к противоречию, что x^* – локальный минимум.

Локальный и глобальный минимум

- Наша цель – глобальный минимум (или точка близкая к нему в некотором смысле).
- Заветная мечта – придумать метод решающий все задачи оптимизации. Выглядит нереалистично, но чем черт не шутит.
- Но уже на прошлой лекции мы поняли, что ее в полной мере не осуществить.
- Нам нужны дополнительные предположения на целевую задачу, чтобы построить оптимистичную теорию.

Выпуклость: определение

Определение выпуклой функции

Пусть дана непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Будем говорить, что она является выпуклой, если для любых $x, y \in \mathbb{R}^d$ выполнено

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$$

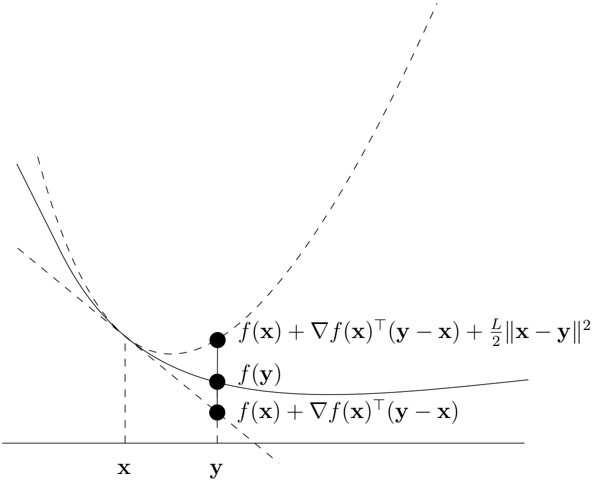
На 5 семинаре будет еще одно определение (эквивалентное в случае дифференцируемых функций).

Определение выпуклой функции

Будем говорить, что она является выпуклой, если для любых $x, y \in \mathbb{R}^d$ и для любого $\lambda \in [0; 1]$ выполнено

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$$

Выпуклость



Ограничение снизу на поведение.

Сильная выпуклость: определение

Определение μ-сильно выпуклой функции

Пусть дана непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Будем говорить, что она является μ-сильно выпуклой ($\mu > 0$), если для любых $x, y \in \mathbb{R}^d$ выполнено

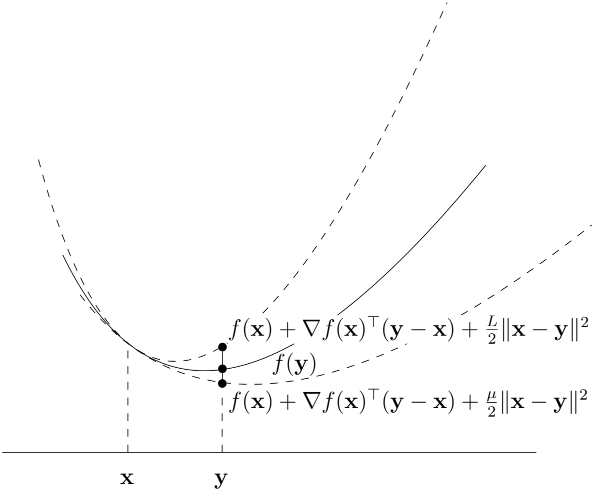
$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|x - y\|_2^2.$$

Определение μ-сильно выпуклой функции

Будем говорить, что она является μ-сильно выпуклой, если для любых $x, y \in \mathbb{R}^d$ и для любого $\lambda \in [0; 1]$ выполнено

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y) - \lambda(1 - \lambda)\frac{\mu}{2} \|x - y\|_2^2$$

Сильная выпуклость



Более сильное ограничение снизу на поведение.

Условие оптимальности: выпуклый случай

Теорема об условии оптимальности безусловной выпуклой задачи

Пусть дана выпуклая непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Если для некоторой точки $x^* \in \mathbb{R}^d$ верно, что $\nabla f(x^*) = 0$, то x^* – глобальный минимум f на всем $\mathbb{R}^d$.

Доказательство

Запишем определение выпуклости:

$$f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle = f(x^*).$$

В обратную сторону уже доказывали выше для произвольных функций.

Выпуклое множество: определение

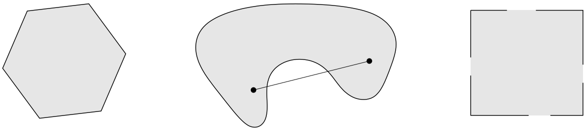
Определение выпуклого множества

Множество $\mathcal{X}$ называется выпуклым, если для любых $x, y \in \mathcal{X}$ и для любого $\lambda \in [0; 1]$ следует, что

$$\lambda x + (1 - \lambda) y \in \mathcal{X}.$$

Смысл: вместе с любыми двумя точками множества в множество входит и отрезок с концами в этих точках.
Подробнее на 4 семинаре.

Выпуклое множество: пример



Вопрос: какие множества здесь выпуклые? 1 (смотрите на границы 3)
Вопрос: понятие выпуклости функции можно обобщить на множество $\mathcal{X}$ (необязательно $\mathbb{R}^d$), но важно, чтобы множество $\mathcal{X}$ было выпуклым. Зачем?

Минимумы выпуклых функций

Теорема о минимумах выпуклых функций

Рассмотрим задачу

$$\min_{x \in \mathcal{X}} f(x),$$

где f – выпуклая, $\mathcal{X}$ - выпуклое. Тогда всякий локальный минимум f на $\mathcal{X}$ является и глобальным.

Минимумы выпуклых функций

Доказательство

Пусть x^* – локальный минимум. Рассмотрим точку вида

$$x_\lambda = \lambda x + (1-\lambda)x^*,$$

где x – произвольная точка из $\mathcal{X}$. **Вопрос:** что можно сказать про x_λ? $x_\lambda \in \mathcal{X}$ в силу выпуклости $\mathcal{X}$. Подберем $\lambda > 0$ достаточно малым, что x_λ попадает в окрестность, где x^* локальный минимум.Тогда уже по выпуклости f

$$f(x^*) \le f(x_\lambda) \le \lambda f(x) + (1-\lambda)f(x^*).$$

Вопрос: что получили? $f(x) \ge f(x^*)$. В силу произвольности $x \in \mathcal{X}$ минимум из локального превратился в глобальный.

Минимумы выпуклых функций

Теорема о минимумах выпуклых функций

Рассмотрим задачу

$$\min_{x \in \mathcal{X}} f(x),$$

где f – выпуклая, $\mathcal{X}$ - выпуклое. Тогда множество точек минимума $\mathcal{X}^*$ выпукло.

Минимумы выпуклых функций

Доказательство

Пустое множество и множество из 1 точки выпуклы. Пусть теперь $x_1^*, x_2^* \in \mathcal{X}^*$. Рассмотрим $x_\lambda^* = \lambda x_1^* + (1-\lambda)x_2^*$, где $\lambda \in [0;1]$. $x_\lambda^* \in \mathcal{X}$ в силу выпуклости $\mathcal{X}$.
В силу выпуклости функции f:

$$f^* \leq f(x_\lambda^*) \leq \lambda f(x_1^*) + (1-\lambda) f(x_2^*) = f^*.$$

Откуда $f(x_\lambda^*) = f^*$, а значит $x_\lambda^* \in \mathcal{X}^*$.

Минимумы выпуклых функций

Теорема о минимумах выпуклых функций

Рассмотрим задачу

$$\min_{x \in \mathcal{X}} f(x),$$

где f – *сильно* выпуклая, $\mathcal{X}$ - выпуклое. Тогда множество точек минимума $\mathcal{X}^*$ может состоять только из одного элемента.

Минимумы выпуклых функций

Доказательство

От противного: пусть есть $x_1^* \neq x_2^* \in \mathcal{X}^*$. Рассмотрим $x_\lambda^* = \lambda x_1^* + (1-\lambda)x_2^*$, где $\lambda \in (0;1)$. Опять же $x_\lambda^* \in \mathcal{X}$ в силу выпуклости $\mathcal{X}$.

Но теперь в силу сильной выпуклости функции f:

$$f^* \leq f(x_\lambda^*) \leq \lambda f(x_1^*) + (1-\lambda)f(x_2^*) - \lambda(1-\lambda)\frac{\mu}{2}\|x_1^* - x_2^*\|_2^2$$
$$= f^* - \lambda(1-\lambda)\frac{\mu}{2}\|x_1^* - x_2^*\|_2^2.$$

Последнее слагаемое < 0 в силу выбора $x_1^* \neq x_2^*$ и $\lambda \in (0;1)$. Противоречие.

Минимумы выпуклых функций

Теорема о минимумах выпуклых функций

Рассмотрим задачу

$$\min_{x \in \mathcal{X}} f(x),$$

где f – *сильно* выпуклая, $\mathcal{X}$ - выпуклое. Тогда множество точек минимума $\mathcal{X}^*$ может состоять только из одного элемента.

- На самом деле для сильно выпуклой функции можно доказать, что решение строго единственное (т.е. добавить к предыдущей теореме существование). Это следует из того, что мы снизу всегда подперты параболой. Смотри док-во в конспекте.

Сильная выпуклость: больше фактов

Теорема об еще одном определении сильной выпуклости

Пусть функция $f : \mathbb{R}^d \to \mathbb{R}$ непрерывно дифференцируема на $\mathbb{R}^d$. Тогда функция f является μ-сильно выпуклой тогда и только тогда, когда для любых $x, y \in \mathbb{R}^d$ выполнено

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|_2^2.$$

Теорема о критерии сильной выпуклости

Пусть функция $f : \mathbb{R}^d \to \mathbb{R}$ дважды непрерывно дифференцируема на $\mathbb{R}^d$. Тогда функция f является μ-сильно выпуклой тогда и только тогда, когда для любого $x \in \mathbb{R}^d$ выполнено

$$\nabla^2 f(x) \succeq \mu I.$$

Оба факта доказаны в пособии. Второй пригодится для ДЗ.

Гладкость: определение

Определение L-гладкой функции

Пусть дана непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Будем говорить, что данная функция имеет L-Липшицев градиент (говорить, что она является L-гладкой), если для любых $x, y \in \mathbb{R}^d$ выполнено

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2.$$

Определение L-гладкости можно писать и в не евклидовой норме. Поэтому формально в предыдущем определении можно указывать, что имеется в виду L-гладкость в терминах $\| \cdot \|_2$.

Гладкость: свойство

Теорема (свойство L - гладкой функции)

Пусть дана L - гладкая функция $f : \mathbb{R}^d \to \mathbb{R}$. Тогда для любых $x, y \in \mathbb{R}^d$ выполнено

$$|f(y) - f(x) - \langle \nabla f(x), y - x \rangle| \leq \frac{L}{2} \|x - y\|_2^2.$$

Гладкость: свойство

Доказательство

Начнем с формулы Ньютона-Лейбница

$$f(y) - f(x) = \int\limits_0^1 \langle \nabla f(x + \tau(y - x)), y - x \rangle d\tau$$

$$= \langle \nabla f(x), y - x \rangle + \int\limits_0^1 \langle \nabla f(x + \tau(y - x)) - \nabla f(x), y - x \rangle d\tau$$

Тогда

$$|f(y) - f(x) - \langle \nabla f(x), y - x \rangle| = \left| \int\limits_0^1 \langle \nabla f(x + \tau(y - x)) - \nabla f(x), y - x \rangle d\tau \right|$$

$$\leq \int\limits_0^1 |\langle \nabla f(x + \tau(y - x)) - \nabla f(x), y - x \rangle| d\tau$$

Гладкость: свойство

Доказательство

Применим КБШ:

$$|f(y) - f(x) - \langle \nabla f(x), y - x\rangle| \leq \int_0^1 |\langle \nabla f(x + \tau(y - x)) - \nabla f(x), y - x\rangle| d\tau$$

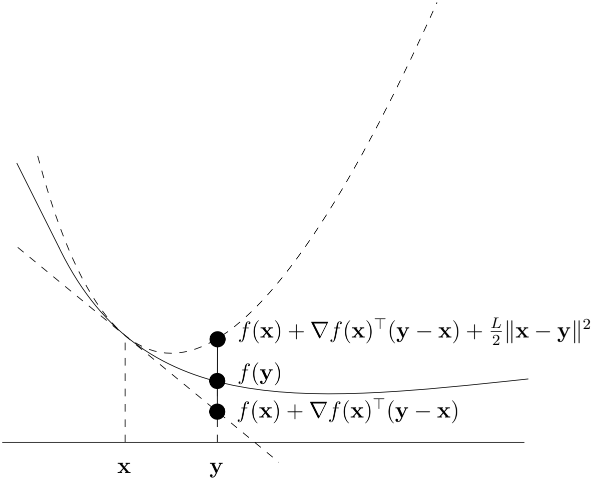
$$\leq \int_0^1 \|\nabla f(x + \tau(y - x)) - \nabla f(x)\|_2 \|y - x\|_2$$

Далее определение L-гладкости:

$$|f(y) - f(x) - \langle \nabla f(x), y - x\rangle| \leq L\|y - x\|_2^2 \int_0^1 \tau d\tau$$

$$= \frac{L}{2}\|x - y\|_2^2$$

Гладкость: физический смысл



Ограничение сверху на поведение (рост) – растет не слишком быстро.

Гладкость: физический смысл