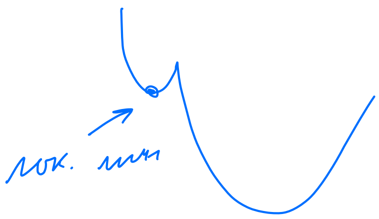
Минимум

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

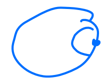
Локальный минимум

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 $y \in B_2^d(r, x^*) = \{y \in \mathbb{R}^d \mid \|y - x^*\|_2 \le r\}$ следует, что $f(x^*) \le f(y)$.

$\forall y \in B_2^d \cap X$



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



$\forall y \in X$

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 $f(x^*) \le f(y)$.

Теорема об условии оптимальности локального минимума

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

Обратное неверно

Док-во: пусть $\nabla f(x^*) \neq 0$

Рассм. разложение в ряд в окр. x^*

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

$$\lim_{x \to x^*} \frac{o(\|x - x^*\|_2)}{\|x - x^*\|_2} = 0$$

Рассм. $\hat{x} = x^* - \lambda \nabla f(x^*)$ $\quad \lambda > 0$

можно подобр. λ: $\hat{x} \in$ окр., где x^* явл. лок. мин

$f(\hat{x}) \ge f(x^*)$ из опр. лок. мин.

с другой стороны $\hat{x}$ в разложение

$$f(\hat{x}) = f(x^*) + \underbrace{\langle \nabla f(x^*), -\lambda \nabla f(x^*) \rangle}_{-\lambda \|\nabla f(x^*)\|_2^2} + o(\lambda \|\nabla f(x^*)\|_2)$$

подберем λ : $o(\lambda \|\nabla f(x^*)\|_2) \le \frac{\lambda}{2} \|\nabla f(x^*)\|_2^2$

$$\begin{aligned} f(\hat{x}) &= f(x^*) - \lambda \|\nabla f(x^*)\|_2^2 + o(\lambda \|\nabla f(x^*)\|_2) \\ &\le f(x^*) - \underbrace{\frac{\lambda}{2}}_{>0} \underbrace{\|\nabla f(x^*)\|_2^2}_{>0 \text{ в силу пред}} \end{aligned}$$

$\Downarrow$

$f(\hat{x}) < f(x^*)$ противоречие! ■

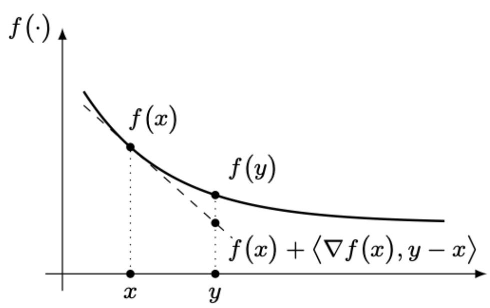
Выпуклость

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

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

Физический смысл:

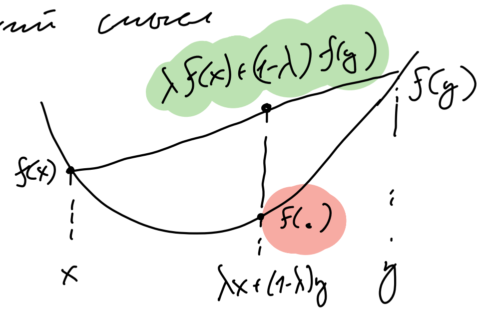


Определение выпуклой функции

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

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

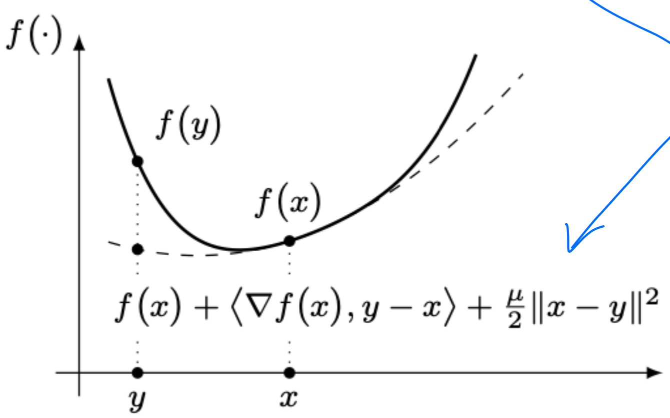
Физический смысл



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

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



Более сильное предпол. функция подпир. параболой

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

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

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 $\mathbb{R}^d$.

Док-во:

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

подст. x, x^*, x^*, x, x^*

$$f(x) \geq f(x^*) + \underbrace{\langle \nabla f(x^*); x - x^* \rangle}_{0}$$

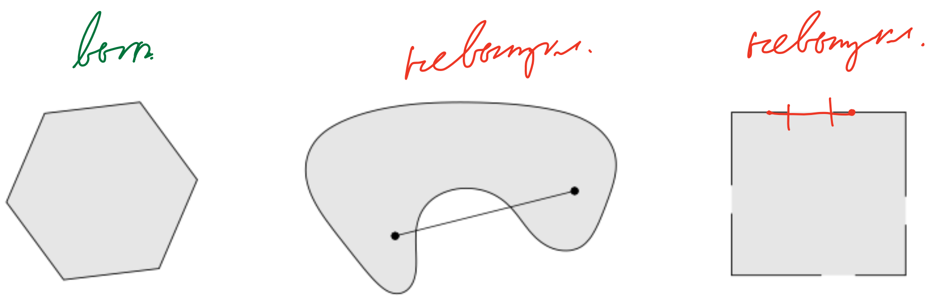
$f(x) \geq f(x^*)$ — по опред. глоб. минимум

от $\min\limits_{x \in \mathbb{R}^d} f(x) \Rightarrow \min\limits_{x \in \mathcal{X}} f(x)$

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

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



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.

Док-во: x^* – лок. минимум

$x_\lambda = \lambda x + (1-\lambda) x^*$ $\quad x \in \mathcal{X}$

$\in \mathcal{X}$ $\quad \lambda \in [0; 1]$

пусть λ подбирается так, что $x_\lambda \in$ окр. лок. мин. x^*

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

по опред. лок. минимума

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

опред. выпуклости

$$\lambda f(x^*) \le \lambda f(x) \qquad \lambda > 0$$

$$f(x^*) \le f(x) \quad \text{для любых } x \in \mathcal{X} \qquad \blacksquare$$

- для $\mathbb{R}^d$ f выпуклой: $\nabla f(x^*) = 0 \iff x^*$ – глоб. минимум

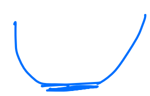
Теорема о минимумах выпуклых функций

Рассмотрим задачу

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

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

Док-во:



- пустое мн-во выпуклое
- один – выпукло
- пусть $x_1^*, x_2^* \in \mathcal{X}^*$ решения задачи

$$x_\lambda^* = \lambda x_1^* + (1-\lambda) x_2^* \qquad \lambda \in [0;1]$$

$\in \mathcal{X}$

$$\min_{x \in \mathcal{X}} f(x) \le f(x_\lambda^*) \le \lambda f(x_1^*) + (1-\lambda) f(x_2^*)$$

опред. вып.

$$= \lambda \min_{\mathcal{X}} f(x) + (1-\lambda) \min_{\mathcal{X}} f(x)$$

$$= \min_{\mathcal{X}} f(x)$$

$$f(x_\lambda^*) = \min_{x \in \mathcal{X}} f(x) \qquad \blacksquare$$

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.

обязательно состоит из 1 элемента

Док-во: от противного $X_1^* \neq X_2^* \in X^*$

$$\exists \quad \underbrace{x_\lambda^*}_{\in X} = \lambda x_1^* + (1-\lambda) x_2^* \qquad \lambda \in [0;1]$$

$$\min_{x\in X} f(x) \le f(x_\lambda^*) \underset{\text{опред. сил. вып.}}{\le} \lambda \underbrace{f(x_1^*)}_{\min f} + (1-\lambda)\underbrace{f(x_2^*)}_{\min f} - \lambda(1-\lambda)\frac{\mu}{2}\|x_1^* - x_2^*\|_2^2$$

$$\le \min_{x\in X} f(x) - \underbrace{\frac{\lambda(1-\lambda)\mu}{2}}_{>0}\underbrace{\|x_1^* - x_2^*\|_2^2}_{>0}$$

$\lambda \in (0;1)$

$\min f(x) \le \min f(x) - \underset{>0}{\delta}$ противоречие ∎

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

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

знак полож. опред.

ед. матрица

$$\left(\nabla^2 f(x) - \mu I\right) \succcurlyeq 0$$

Гладкость

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

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

Липшицевость градиента

Теорема (свойство L - гладкой функции)

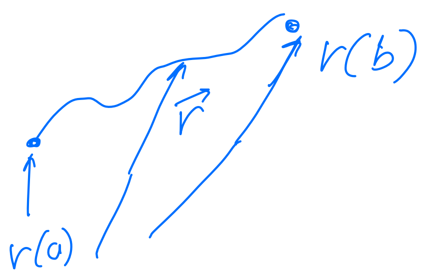
Пусть дана L - гладкая функция $f : \mathbb{R}^d \to \mathbb{R}$. Тогда для любых $x, y \in \mathbb{R}^d$ выполнено

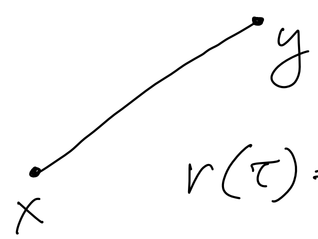
$$|f(y) - f(x) - \langle \nabla f(x), y - x\rangle| \le \frac{L}{2}\|x - y\|_2^2.$$

Док-во:

$$f(y) - f(x) = \int_0^1 \langle \nabla f(x + \tau(y-x)); y - x\rangle d\tau$$

$$f(\vec{r}(b)) - f(\vec{r}(a)) = \int_a^b \langle \nabla f(r(t)) \, d\vec{r}(t)\rangle$$





$$r(\tau) = x + \tau(y-x)$$
$$\tau \in [0;1]$$

$$= \int_0^1 \langle \nabla f(x + \tau(y-x)) - \nabla f(x); y - x\rangle d\tau$$

$$+ \underbrace{\int_0^1 \langle \nabla f(x); y - x\rangle d\tau}_{\langle \nabla f(x); y - x\rangle}$$

$$|f(y) - f(x) - \langle \nabla f(x); y - x \rangle|$$

$$= \left| \int_0^1 \langle \nabla f(x + \tau(y-x)) - \nabla f(x); y - x \rangle d\tau \right|$$

$|\Sigma| \leq \Sigma|\,|$

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

КБШ: $|\langle a; b \rangle| \leq \|a\|_2 \cdot \|b\|_2$

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

L – липшицевый градиент

$$\leq \int_0^1 L \, \|\tau(y - x)\|_2 \cdot \|y - x\|_2 \, d\tau$$

$$= \int_0^1 L \|y - x\|_2^2 \cdot \tau \, d\tau$$

$\int_0^1 \tau \, d\tau = \frac{1}{2}$

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

∎

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

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

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

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

L-гладкость

сильн. вып.

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

$L \geq \mu$

$\mu = 0$ вып.

Физический смысл: подпорка сверху параболой