Выпуклость и гладкость. Градиентный спуск

Методы оптимизации

Александр Безносиков

Московский физико-технический институт

14 сентября 2023

ФПМИ

Глобальный и локальный минимумы

Рассмотрим безусловную задачу:

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

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 $\mathbb{R}^d$, 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 и локальный минимумы

Рассмотрим безусловную задачу:

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

Локальный минимум

Точка x^* называется локальным минимумом функции f на $\mathbb{R}^d$, 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 $f(x^*) \le f(y)$.

Глобальный минимум

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

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 $\mathbb{R}^d$ $\mathcal{X}$, если существует $r > 0$ такое, что для любого $y \in B_2^d(r, x^*) \overset{\cap \mathcal{X}}{\triangleq} \{y \in \mathbb{R}^d \mid \|y - x^*\|_2 \leq r\}$ следует, что $f(x^*) \leq f(y)$.

Глобальный минимум

Точка x^* называется глобальным минимумом функции f на $\cancel{\mathbb{R}^d}$, если для любого $y \in \cancel{\mathbb{R}^d}$ 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 $y \in B_2^d(r, x^*) \cup \mathcal{X}$ 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 $\nabla f(x^*) = 0$.

Условие оптимальности: общий случай

Доказательство

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

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

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

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.

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.

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:

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

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:

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

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

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).
- На прошлой лекции стало понятно, что без дополнительных предположений искать глобальный минимум бессмысленно.

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

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

Выпуклость: определение

Определение выпуклой функции

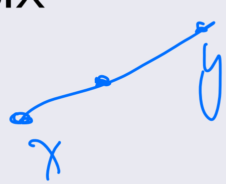
Пусть дана непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$. Будем говорить, что она является выпуклой, если для любых $x, y \in \mathbb{R}^d$ выполнено

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

На 4 семинаре будет еще одно определение (эквивалентное в случае дифференцируемых функций).

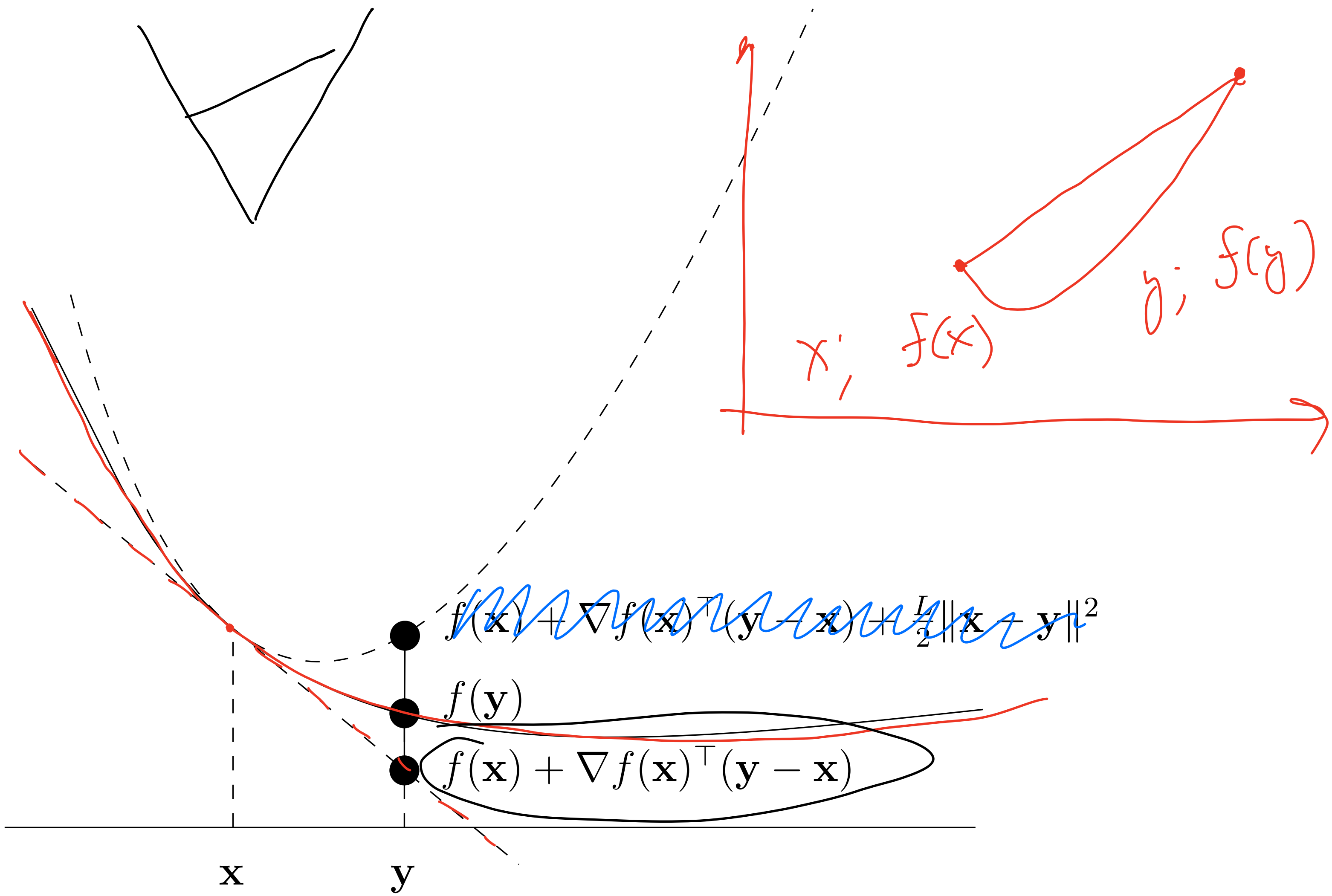
Определение выпуклой функции

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

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

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

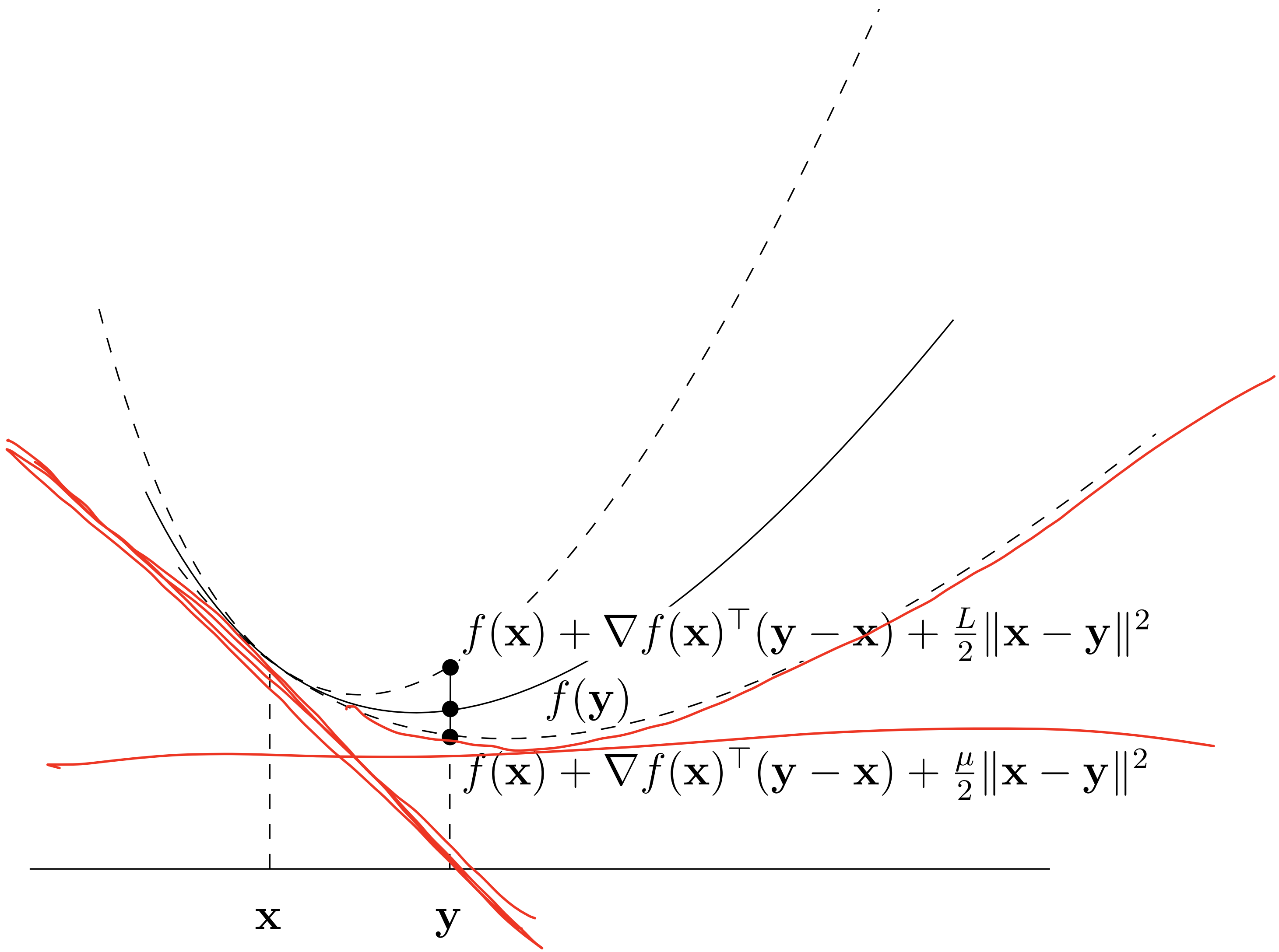
$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2.$$

Определение 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

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

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

Условие оптимальности: выпуклый случай

Теорема об условии оптимальности безусловной выпуклой задачи

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

Доказательство

Запишем определение выпуклости:

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

Условие оптимальности: выпуклый случай

Теорема об условии оптимальности безусловной выпуклой задачи

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

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

Выпуклое множество: определение

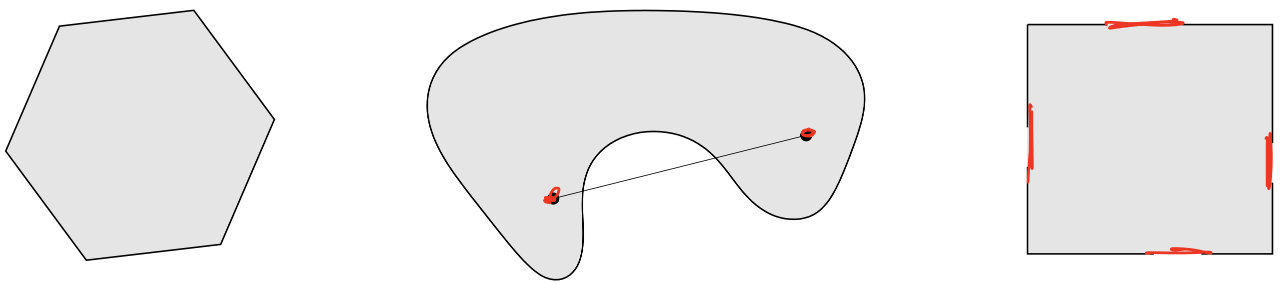
Определение выпуклого множества

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

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

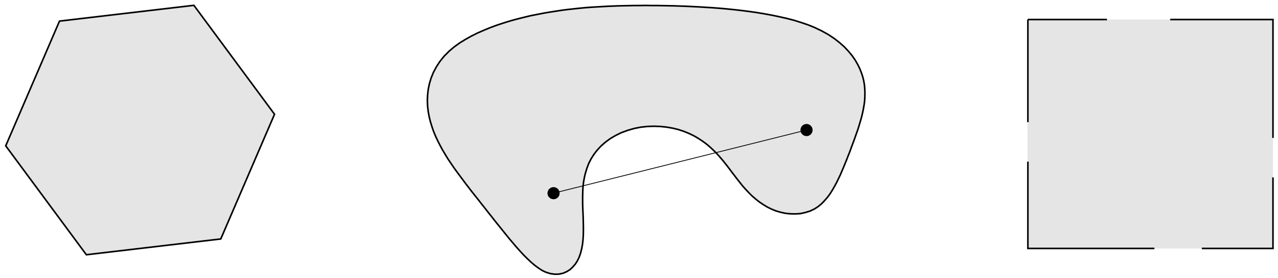
Смысл: вместе с любыми двумя точками множества в множество входит и отрезок с концами в этих точках.
Подробнее на 3 семинаре.

Выпуклое множество: пример



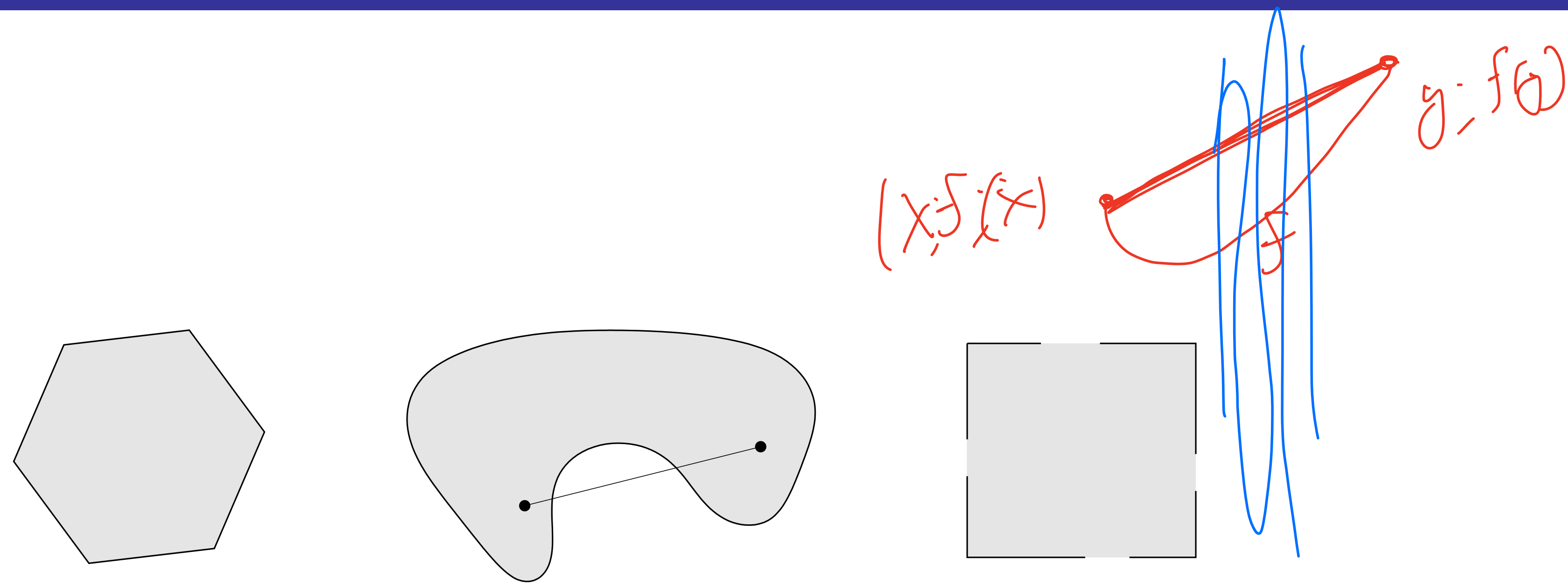
Вопрос: какие множества здесь выпуклые?

Выпуклое множество: пример



Вопрос: какие множества здесь выпуклые? 1 и 3

Выпуклое множество: пример



Вопрос: какие множества здесь выпуклые? 1 и 3

Вопрос: понятие выпуклости функции можно обобщить на множество $\mathcal{X}$ (необязательно $\mathbb{R}^d$), но важно, чтобы множество $\mathcal{X}$ было выпуклым. Зачем?

Условие оптимальности: выпуклый случай

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

Теорема об условии оптимальности условной выпуклой задачи

Пусть дана выпуклая непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to R$ и выпуклое множество $\mathcal{X}$. Тогда $x^* \in \mathcal{X}$ – минимум f на $\mathcal{X}$ тогда и только тогда, когда для всех $x \in \mathcal{X}$ выполнено

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0.$$

Условие оптимальности: выпуклый случай

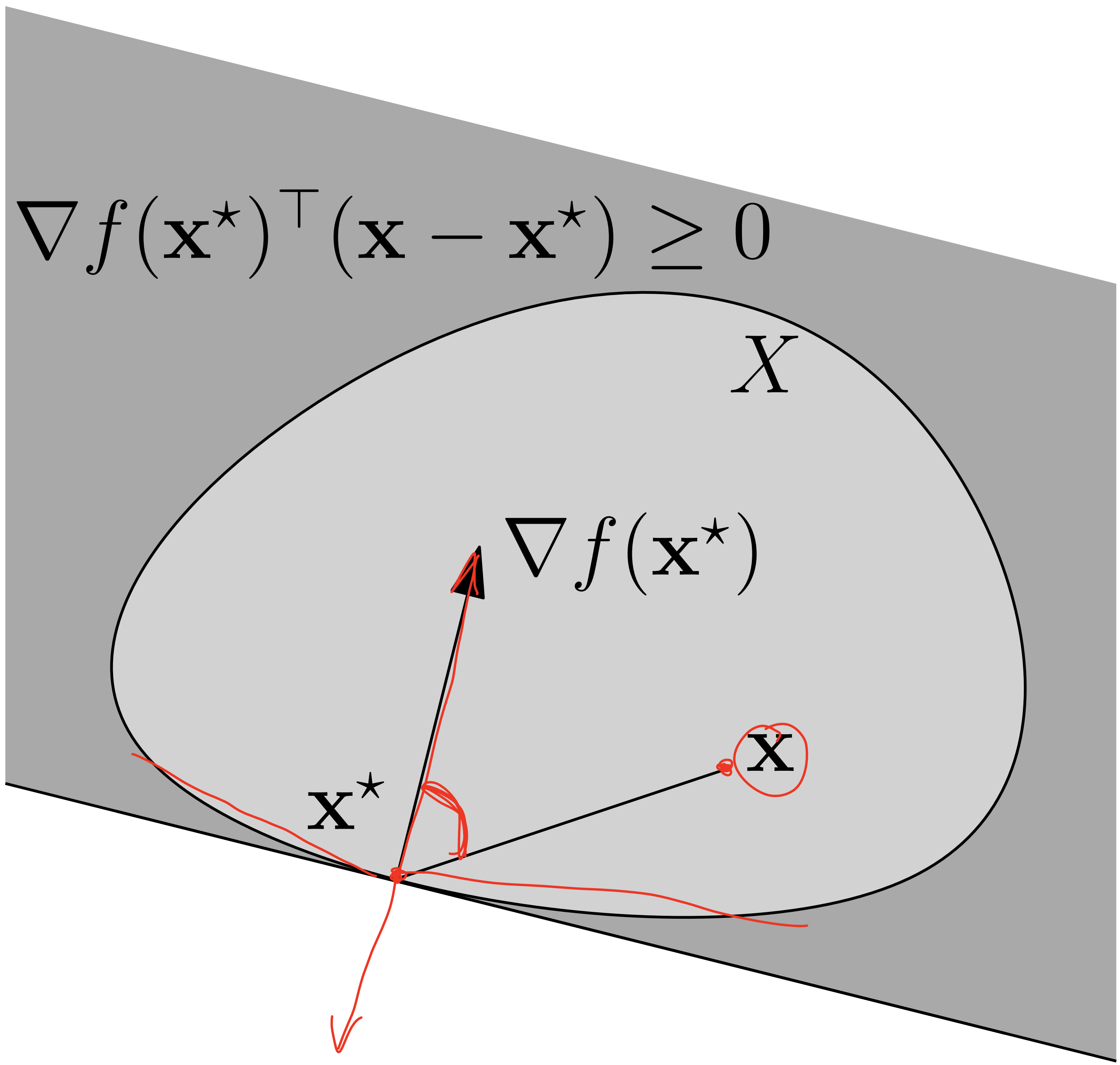
Теорема об условии оптимальности условной выпуклой задачи

Пусть дана выпуклая непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to R$ и выпуклое множество $\mathcal{X}$. Тогда $x^* \in \mathcal{X}$ – минимум f на $\mathcal{X}$ тогда и только тогда, когда для всех $x \in \mathcal{X}$ выполнено

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0.$$

Доказательство в пособии и через 2 лекции. Сегодня пока не пригодится.

Условие оптимальности: суть



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 $\mathcal{X}$.

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^*?

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 $\mathcal{X}$.

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.

Минимумы выпуклых функций

Доказательство

Пусть x^* – локальный минимум. Рассмотрим точку вида

$1-\lambda$ λ

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

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~~ $\to 1$, что x_λ попадает в окрестность, где x^* локальный минимум. Тогда уже по выпуклости f

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

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

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

Вопрос: что получили?

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

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

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

Минимумы выпуклых функций

$(x_1 - x_2)^2$

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.

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 $\mathcal{X}$.

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:

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

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 $x^* \in \mathcal{X}^*$.

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 $\mathcal{X}^*$.

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:

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

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

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:

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

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 эквивалентном 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

$$\langle \nabla f(x) - \nabla f(x), x - y \rangle \geq 0.$$

Сильная выпуклость: больше фактов

Теорема об еще одном эквивалентном 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

$$\langle \nabla f(x) - \nabla f(\cancel{x}\,y), x - y \rangle \geq \cancel{0}\,\mu \|x-y\|_2^2.$$

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

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

Сильная выпуклость: больше фактов

Теорема об еще одном эквивалентном 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

$$\langle \nabla f(x) - \nabla f(x), x - y \rangle \geq 0.$$

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

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

Оба факта доказаны в пособии. Второй пригодится для ДЗ.

Гладкость: определение

Определение L-гладкой функции

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

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

Гладкость: определение

Определение L-гладкой функции

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

Теорема (свойство L - гладкой функции)

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

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

Гладкость: свойства

Доказательство

Начнем с формулы Ньютона-Лейбница

$\|x-y\|$

$r(b) = y$ $\qquad$ $r(a) = x$

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

$$
\int f'(x+\tau(y-x)); (y-x)\, d\tau
$$

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

$$
f(y) \quad - f(x)
$$

$$
\int_a^b \langle \nabla f(r(t)); dr(t) \rangle = f(r(b)) - f(r(a))
$$

$$
dr(t) = (y-x)\, d\tau
$$

$r(t)$

y

$x+\tau(y-x)$

x

$\|x-y\|$

Гладкость: свойства

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: свойства

Доказательство

Применим КБШ:

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

$$\langle a; b \rangle \leq \|a\|_2 \|b\|_2$$

$$\leq \int\limits_0^1 \|\nabla f(x + \tau(y - x)) - \nabla f(x)\|_2 \|y - x\|_2$$

Гладкость: свойства

Доказательство

Применим КБШ:

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

$$\leq \int\limits_0^1 \|\nabla f(x + \tau(y - x)) - \nabla f(x)\|_2 \|y - x\|_2$$

Далее определение L-гладкости:

$$|f(y) - f(x) - \langle \nabla f(x), y - x\rangle| \leq L\|y - x\|_2^2 \int\limits_0^1 \tau d\tau$$

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

Гладкость: свойства

Теорема (свойства L - гладкой выпуклой функции)

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

$$0 \leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{L}{2}\|x - y\|_2^2$$

и

$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y).$$

Гладкость: свойства

Теорема (свойства L - гладкой выпуклой функции)

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

$$0 \leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{L}{2} \|x - y\|_2^2$$

и

$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y).$$

Доказательство

Доказательство первого факта следует из выпуклости и предыдущего свойства гладкости: подмодульное выражение справедливо из-за выпуклости.

Гладкость: свойства

Доказательство

Рассмотрим функцию $\phi(y) = f(y) - \langle \nabla f(x), y \rangle$. **Вопрос:** является ли она L_ϕ-гладкой? выпуклой?

$\nabla f(x)$ ax

$$\|\nabla \phi(y_1) - \nabla \phi(y_2)\|_2 \le$$

$$= \|\nabla f(y_1) - \nabla f(y_2) - \nabla f(x) + \nabla f(x)\|_2$$

$$= \|\nabla f(y_1) - \nabla f(y_2)\|_2 \le L_f \|y_1 - y_2\|_2$$

$$\nabla \phi(y) = \nabla f(y) - \nabla f(x)$$

Гладкость: свойства

Доказательство

Рассмотрим функцию $\phi(y) = f(y) - \langle \nabla f(x), y \rangle$. **Вопрос:** является ли она L_ϕ-гладкой? выпуклой? Да на оба вопроса и $L_\phi = L$ (проверка по определению).

Гладкость: свойства

Доказательство

Рассмотрим функцию $\phi(y) = f(y) - \langle \nabla f(x), y \rangle$. **Вопрос:** является ли она L_ϕ-гладкой? выпуклой? Да на оба вопроса и $L_\phi = L$ (проверка по определению). Также можно заметить, что $y^* = x$ – минимум.

Вопрос: почему? $\nabla \phi(y^*) = \nabla f(\underbrace{y^*}_{=x}) - \nabla f(x) = 0$

Доказательство

Рассмотрим функцию $\phi(y) = f(y) - \langle \nabla f(x), y \rangle$. **Вопрос:** является ли она L_ϕ-гладкой? выпуклой? Да на оба вопроса и $L_\phi = L$ (проверка по определению). Также можно заметить, что $y^* = x$ – минимум.

Вопрос: почему? $\nabla \phi(y^*) = \nabla \phi(x) = 0$. Воспользуемся первым пунктом теоремы: $f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{L}{2}\|x - y\|_2^2$ с $\left(y = y - \frac{1}{L}\nabla\phi(y), x = y, f = \phi\right)$. Тогда

ny ny

$$\phi\left(y - \frac{1}{L}\nabla\phi(y)\right) - \phi(y) - \left\langle \nabla\phi(y), -\frac{1}{L}\nabla\phi(y)\right\rangle \leq \frac{1}{2L}\|\nabla\phi(y)\|_2^2$$

После небольшой перестановки:

$$\phi\left(y - \frac{1}{L}\nabla\phi(y)\right) \leq \phi(y) - \frac{1}{2L}\|\nabla\phi(y)\|_2^2$$

Гладкость: свойства

Доказательство

Тогда получаем, зная, что $y^* = x$ – минимум:

$$\phi(x) = \phi(y^*) \leq \phi\left(y - \frac{1}{L}\nabla\phi(y)\right) \leq \phi(y) - \frac{1}{2L}\|\nabla\phi(y)\|_2^2$$

Гладкость: свойства

Доказательство

Тогда получаем, зная, что $y^* = x$ – минимум:

$$\phi(x) = \phi(y^*) \leq \phi\left(y - \frac{1}{L}\nabla\phi(y)\right) \leq \phi(y) - \frac{1}{2L}\|\nabla\phi(y)\|_2^2$$

Подставляя ϕ:

$$f(x) - \langle \nabla f(x), x\rangle \leq f(y) - \langle \nabla f(x), y\rangle - \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2$$

Гладкость: свойства

Доказательство

Тогда получаем, зная, что $y^* = x$ – минимум:

$$\phi(x) = \phi(y^*) \leq \phi\left(y - \frac{1}{L}\nabla\phi(y)\right) \leq \phi(y) - \frac{1}{2L}\|\nabla\phi(y)\|_2^2$$

Подставляя ϕ:

$$f(x) - \langle\nabla f(x), x\rangle \leq f(y) - \langle\nabla f(x), y\rangle - \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2$$

Осталось переставить:

$$f(x) + \langle\nabla f(x), y - x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y)$$

Гладкость: свойства

Доказательство

Тогда получаем, зная, что $y^* = x$ – минимум:

$$\phi(x) = \phi(y^*) \leq \phi\left(y - \frac{1}{L}\nabla\phi(y)\right) \leq \phi(y) - \frac{1}{2L}\|\nabla\phi(y)\|_2^2$$

Подставляя ϕ:

$$f(x) - \langle \nabla f(x), x\rangle \leq f(y) - \langle \nabla f(x), y\rangle - \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2$$

Осталось переставить:

$$f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y)$$

Вопрос: а пользовались вообще здесь выпуклость?

Гладкость: свойства

Доказательство

Тогда получаем, зная, что $y^* = x$ – минимум:

$$\phi(x) = \phi(y^*) \leq \phi\left(y - \frac{1}{L}\nabla\phi(y)\right) \leq \phi(y) - \frac{1}{2L}\|\nabla\phi(y)\|_2^2$$

Подставляя ϕ:

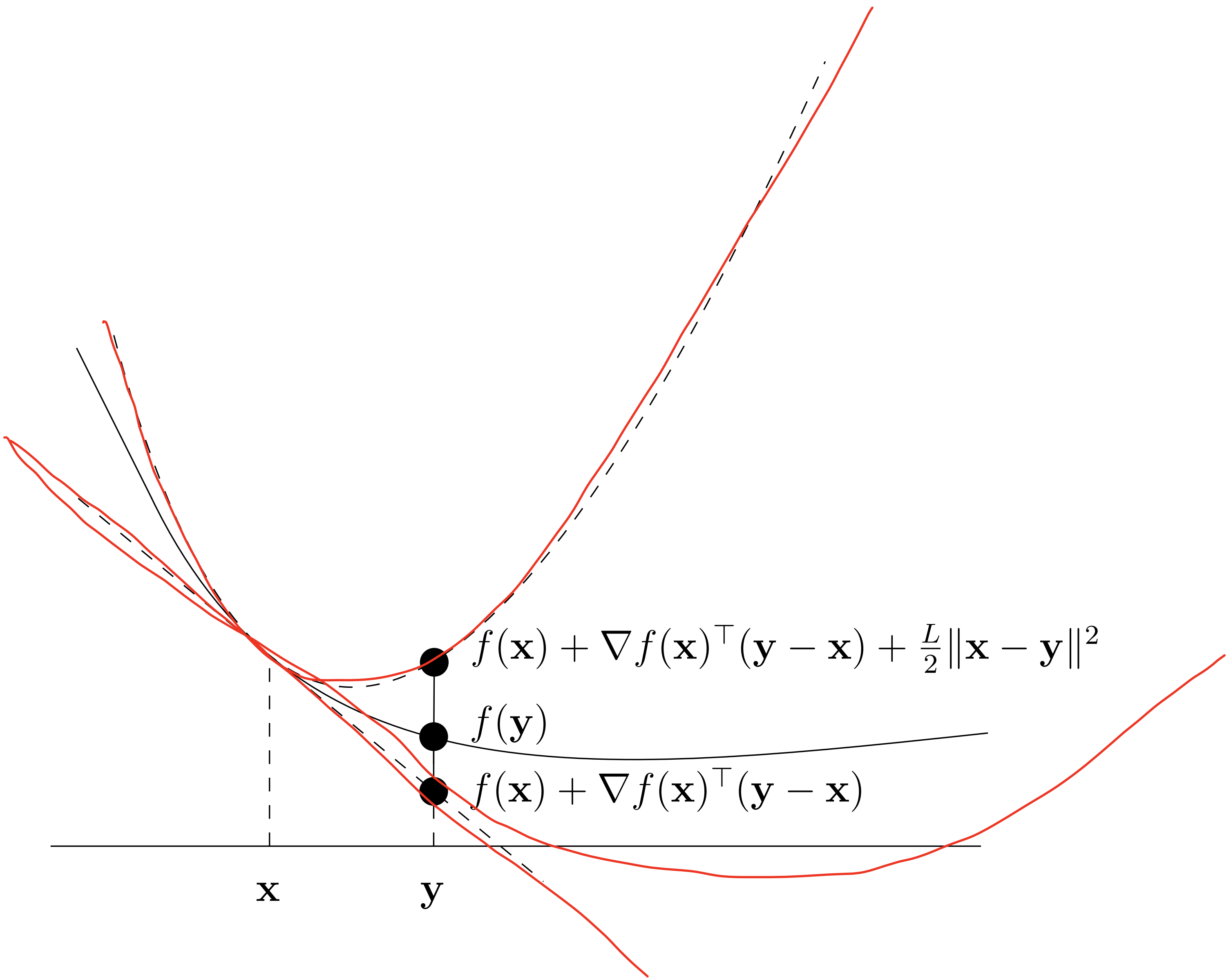
$$f(x) - \langle \nabla f(x), x \rangle \leq f(y) - \langle \nabla f(x), y \rangle - \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2$$

Осталось переставить:

$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y)$$

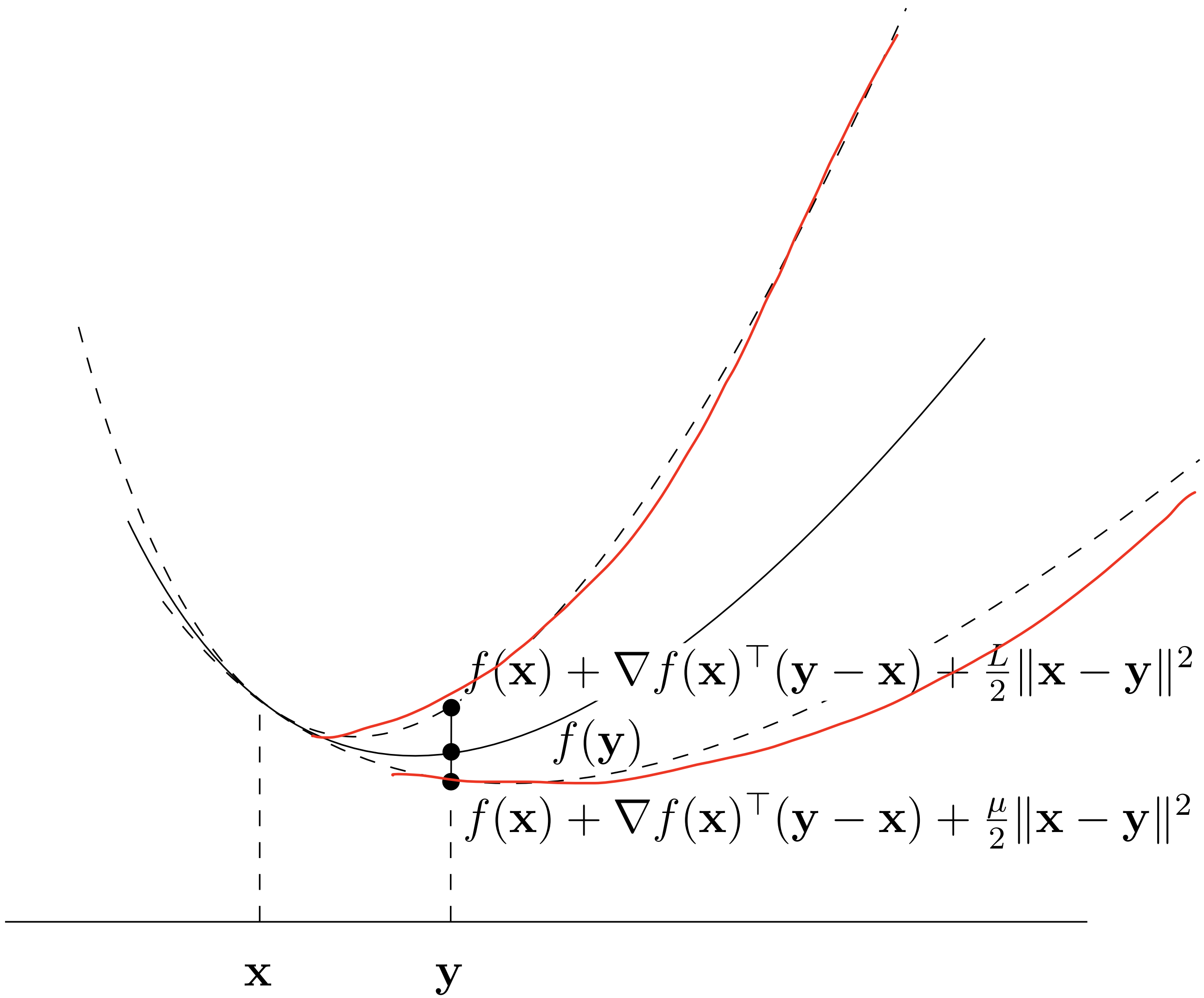
Вопрос: а пользовались вообще здесь выпуклость? да, $\nabla(y^*) = 0 \;\Rightarrow\; y^*$ – минимум.

Гладкость: физический смысл



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

Гладкость: физический смысл



Градиентный спуск

- **Задача:** найти решение безусловной оптимизации:

$$\min_{x \in \mathbb{R}^d} f(x). \tag{1}$$

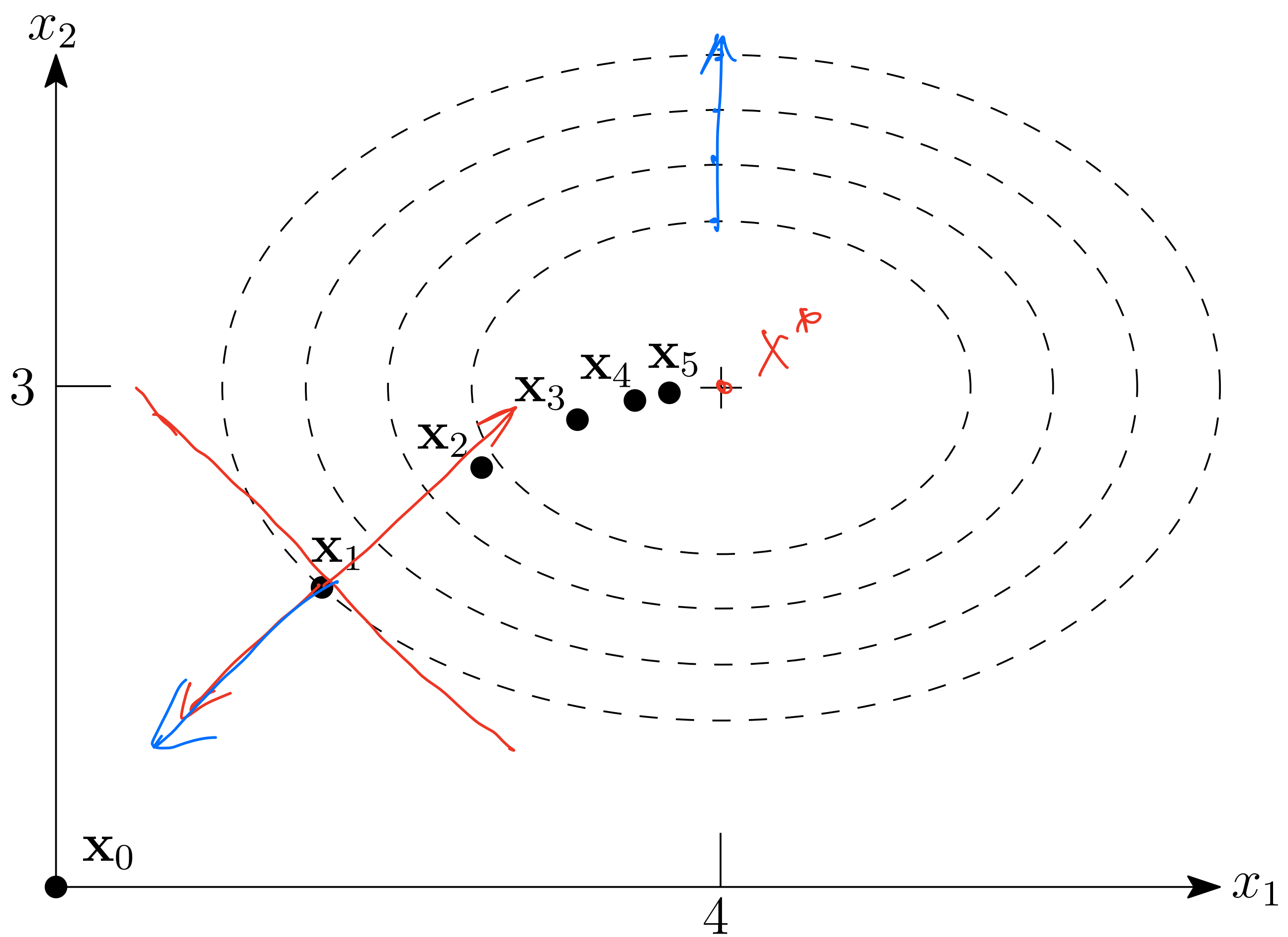
Алгоритм 1 Градиентный спуск

Вход: размеры шага $\{\gamma_k\} > 0$, стартовая точка $x^0 \in \mathbb{R}^d$, количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**
2: Вычислить $\nabla f(x^k)$
3: $x^{k+1} = x^k - \boxed{\gamma_k \nabla f(x^k)}$
4: **end for**

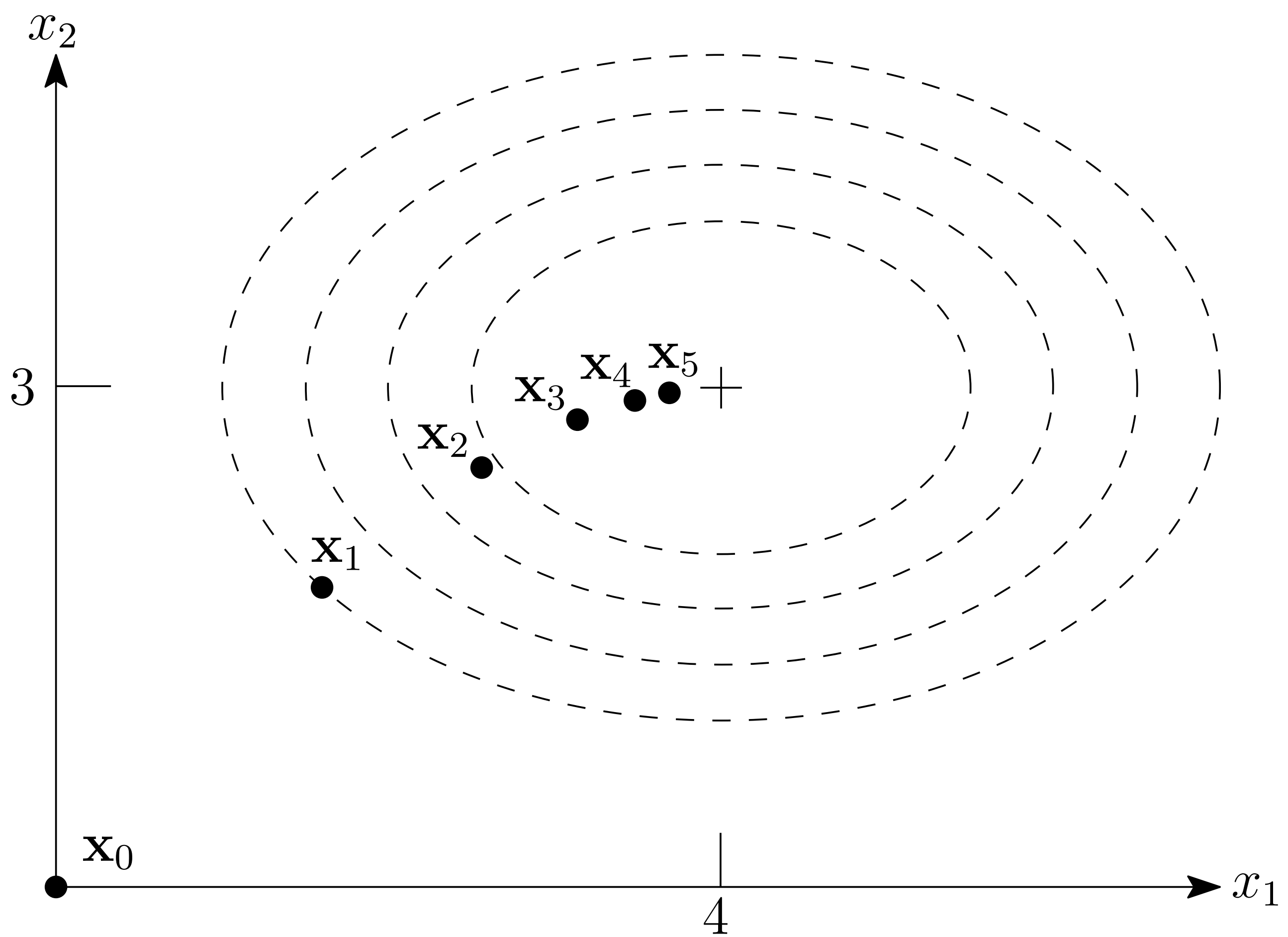
Выход: x^K

Пример



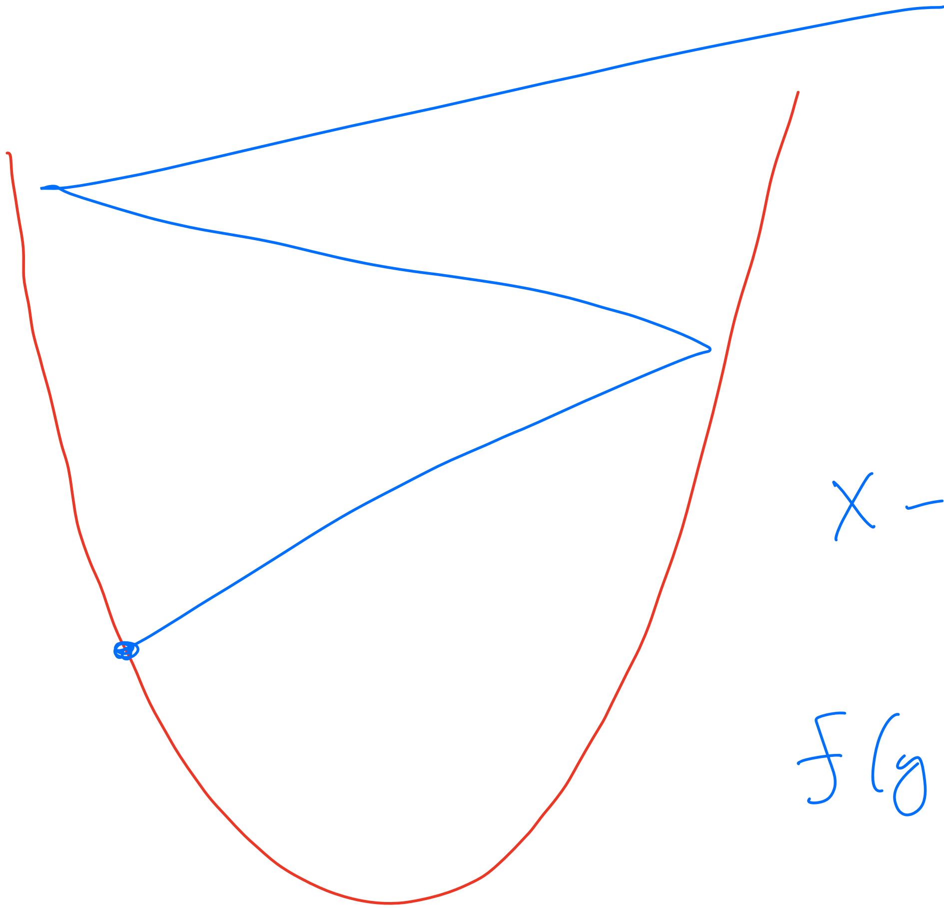
Вопрос: куда направлен градиент в точке x_1?

Пример



Вопрос: куда направлен градиент в точке x_1? направление роста

Зачем нужен шаг?



$$\frac{1}{2}x^2$$

$$x - \gamma_k x$$

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

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Знаем, что для сильно выпуклых функций решение уникально, попытаемся оценить, как меняется расстояние до него. Подставим итерацию:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 &= \|x^k - \gamma_k \nabla f(x^k) - x^*\|_2^2 \\
&= \|x^k - x^*\|_2^2 \underbrace{- 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle}_{} + \gamma_k^2 \|\nabla f(x^k)\|_2^2
\end{aligned}
$$

$-2\gamma_k \langle \nabla f(x^k) - \nabla f(x^*), x^k - x^* \rangle$ | $+\gamma_k^2 \|\nabla f(x^k) - \nabla f(x^*)\|_2^2$

$\le -\mu\gamma_k \|x^k - x^*\|_2^2$ | $\le \gamma_k^2 L^2 \|x^k - x^*\|_2^2$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Знаем, что для сильно выпуклых функций решение уникально, попытаемся оценить, как меняется расстояние до него. Подставим итерацию:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 &= \|x^k - \gamma_k \nabla f(x^k) - x^*\|_2^2 \\
&= \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle + \gamma_k^2 \|\nabla f(x^k)\|_2^2
\end{aligned}
$$

Вопрос: что дальше?

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Знаем, что для сильно выпуклых функций решение уникально, попытаемся оценить, как меняется расстояние до него. Подставим итерацию:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 &= \|x^k - \gamma_k \nabla f(x^k) - x^*\|_2^2 \\
&= \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle + \gamma_k^2 \|\nabla f(x^k)\|_2^2
\end{aligned}
$$

Вопрос: что дальше? Вспоминаем, что у нас есть гладкость $\|\nabla f(x) + \nabla f(y)\|_2^2 \leq L^2 \|x - y\|_2^2$ и сильная выпуклость в виде $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|_2^2$.

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Знаем, что для сильно выпуклых функций решение уникально, попытаемся оценить, как меняется расстояние до него. Подставим итерацию:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 &= \|x^k - \gamma_k \nabla f(x^k) - x^*\|_2^2 \\
&= \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle + \gamma_k^2 \|\nabla f(x^k)\|_2^2
\end{aligned}
$$

Вопрос: что дальше? Вспоминаем, что у нас есть гладкость $\|\nabla f(x) + \nabla f(y)\|_2^2 \leq L^2 \|x - y\|_2^2$ и сильная выпуклость в виде $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|_2^2$. Достаточно только вспомнить условие оптимальности $\nabla f(x^*) = 0$.

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Знаем, что для сильно выпуклых функций решение уникально, попытаемся оценить, как меняется расстояние до него. Подставим итерацию:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 &= \|x^k - \gamma_k \nabla f(x^k) - x^*\|_2^2 \\
&= \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle + \gamma_k^2 \|\nabla f(x^k)\|_2^2
\end{aligned}
$$

Вопрос: что дальше? Вспоминаем, что у нас есть гладкость $\|\nabla f(x) + \nabla f(y)\|_2^2 \leq L^2 \|x - y\|_2^2$ и сильная выпуклость в виде $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|_2^2$. Достаточно только вспомнить условие оптимальности $\nabla f(x^*) = 0$.

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 =& \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k) - \nabla f(x^*), x^k - x^* \rangle \\
&+ \gamma_k^2 \|\nabla f(x^k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Гладкость $\|\nabla f(x) + \nabla f(y)\|_2^2 \leq L^2\|x - y\|_2^2$ и сильная выпуклость в виде $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu\|x - y\|_2^2$:

$$\begin{aligned}
\|x^{k+1} - x^*\|_2^2 =& \|x^k - x^*\|_2^2 - 2\gamma_k\langle \nabla f(x^k) - \nabla f(x^*), x^k - x^*\rangle \\
& + \gamma_k^2\|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \\
\leq& \|x^k - x^*\|_2^2 - 2\gamma_k\mu\|x^k - x^*\|_2^2 + \gamma_k^2 L^2\|x^k - x^*\|_2^2 \\
=& \underbrace{(1 - 2\gamma_k\mu + \gamma_k^2 L^2)}_{< 1}\|x^k - x^*\|_2^2
\end{aligned}$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Гладкость $\|\nabla f(x) + \nabla f(y)\|_2^2 \leq L^2\|x - y\|_2^2$ и сильная выпуклость в виде $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu\|x - y\|_2^2$:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 =& \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k) - \nabla f(x^*), x^k - x^* \rangle \\
& + \gamma_k^2 \|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \\
\leq& \|x^k - x^*\|_2^2 - 2\gamma_k \mu \|x^k - x^*\|_2^2 + \gamma_k^2 L^2 \|x^k - x^*\|_2^2 \\
=& (1 - 2\gamma_k \mu + \gamma_k^2 L^2) \|x^k - x^*\|_2^2
\end{aligned}
$$

Вопрос: а что мы хотим теперь?

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Гладкость $\|\nabla f(x) + \nabla f(y)\|_2^2 \le L^2\|x-y\|_2^2$ и сильная выпуклость в виде $\langle \nabla f(x) - \nabla f(y), x-y\rangle \ge \mu\|x-y\|_2^2$:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 =& \|x^k - x^*\|_2^2 - 2\gamma_k\langle \nabla f(x^k) - \nabla f(x^*), x^k - x^*\rangle \\
&+ \gamma_k^2\|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \\
\le& \|x^k - x^*\|_2^2 - 2\gamma_k\mu\|x^k - x^*\|_2^2 + \gamma_k^2 L^2\|x^k - x^*\|_2^2 \\
=& (1 - 2\gamma_k\mu + \gamma_k^2 L^2)\|x^k - x^*\|_2^2
\end{aligned}
$$

Вопрос: а что мы хотим теперь? $(1 - 2\gamma_k\mu + \gamma_k^2 L^2) < 1$. Как подобрать?

$\gamma_k = \frac{\mu}{L^2}$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Гладкость $\|\nabla f(x) + \nabla f(y)\|_2^2 \leq L^2\|x - y\|_2^2$ и сильная выпуклость в виде $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu\|x - y\|_2^2$:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 =& \|x^k - x^*\|_2^2 - 2\gamma_k\langle \nabla f(x^k) - \nabla f(x^*), x^k - x^*\rangle \\
& + \gamma_k^2\|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \\
\leq& \|x^k - x^*\|_2^2 - 2\gamma_k\mu\|x^k - x^*\|_2^2 + \gamma_k^2 L^2\|x^k - x^*\|_2^2 \\
=& (1 - 2\gamma_k\mu + \gamma_k^2 L^2)\|x^k - x^*\|_2^2
\end{aligned}
$$

Вопрос: а что мы хотим теперь? $(1 - 2\gamma_k\mu + \gamma_k^2 L^2) < 1$. Как подобрать? $\arg\min_{\gamma_k}(1 - 2\gamma_k\mu + \gamma_k^2 L^2)$?

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Гладкость $\|\nabla f(x) + \nabla f(y)\|_2^2 \leq L^2\|x - y\|_2^2$ и сильная выпуклость в виде $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu\|x - y\|_2^2$:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 =& \|x^k - x^*\|_2^2 - 2\gamma_k\langle \nabla f(x^k) - \nabla f(x^*), x^k - x^*\rangle \\
& + \gamma_k^2\|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \\
\leq& \|x^k - x^*\|_2^2 - 2\gamma_k\mu\|x^k - x^*\|_2^2 + \gamma_k^2 L^2\|x^k - x^*\|_2^2 \\
=& (1 - 2\gamma_k\mu + \gamma_k^2 L^2)\|x^k - x^*\|_2^2 \quad = \left(1 - \frac{\mu^2}{L^2}\right)
\end{aligned}
$$

Вопрос: а что мы хотим теперь? $(1 - 2\gamma_k\mu + \gamma_k^2 L^2) < 1$. Как подобрать? $\arg\min_{\gamma_k}(1 - 2\gamma_k\mu + \gamma_k^2 L^2)$? $\gamma_k = \frac{\mu}{L^2}$ и $(1 - 2\gamma_k\mu + \gamma_k^2 L^2) = 1 - \frac{\mu^2}{L^2}$.

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Итого:

$$\|x^{k+1} - x^*\|_2^2 \leq \left(1 - \frac{\mu^2}{L^2}\right) \|x^k - x^*\|_2^2$$

$$\leq \left(1 - \frac{\mu^2}{L^2}\right)^2 \|x^{k-1} - x^*\|_2^2$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Итого:

$$\|x^{k+1} - x^*\|_2^2 \leq \left(1 - \frac{\mu^2}{L^2}\right) \|x^k - x^*\|_2^2$$

Запустим рекурсию:

$$\|x^k - x^*\|_2^2 \leq \left(1 - \frac{\mu^2}{L^2}\right)^k \|x^0 - x^*\|_2^2$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Итого:

$$\|x^{k+1} - x^*\|_2^2 \leq \left(1 - \frac{\mu^2}{L^2}\right) \|x^k - x^*\|_2^2$$

Запустим рекурсию:

$$\|x^k - x^*\|_2^2 \leq \left(1 - \frac{\mu^2}{L^2}\right)^k \|x^0 - x^*\|_2^2$$

Вопрос: какая это скорость сходимости?

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Итого:

$$\|x^{k+1} - x^*\|_2^2 \leq \left(1 - \frac{\mu^2}{L^2}\right) \|x^k - x^*\|_2^2$$

Запустим рекурсию:

$$\|x^k - x^*\|_2^2 \leq \left(1 - \frac{\mu^2}{L^2}\right)^k \|x^0 - x^*\|_2^2$$

Вопрос: какая это скорость сходимости? Линейная.

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Итого:

$$\|x^{k+1} - x^*\|_2^2 \leq \left(1 - \frac{\mu^2}{L^2}\right) \|x^k - x^*\|_2^2$$

Запустим рекурсию:

$$\underbrace{\|x^k - x^*\|_2^2}_{} \leq \underbrace{\left(1 - \frac{\mu^2}{L^2}\right)^k \|x^0 - x^*\|_2^2}_{} \leq \varepsilon$$

Вопрос: какая это скорость сходимости? Линейная. А как получить оценку на число итераций?

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Итого:

$$\|x^{k+1} - x^*\|_2^2 \leq \left(1 - \frac{\mu^2}{L^2}\right) \|x^k - x^*\|_2^2$$

Запустим рекурсию:

$$\|x^k - x^*\|_2^2 \leq \underbrace{\left(1 - \frac{\mu^2}{L^2}\right)}^k \|x^0 - x^*\|_2^2$$

Вопрос: какая это скорость сходимости? Линейная. А как получить оценку на число итераций? (Здесь просто нужно вспомнить разложение экспоненты в ряд)

$1 - x + \frac{x^2}{2}$

$1 - x \leq \exp(-x)$ $x \in (0;1)$

$$\|x^k - x^*\|_2^2 \leq \left(1 - \frac{\mu^2}{L^2}\right)^k \|x^0 - x^*\|_2^2 \leq \exp\left(-\frac{\mu^2}{L^2} \cdot k\right) \|x^0 - x^*\|_2^2$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

С предыдущего слайда:

$$\|x^k - x^*\|_2^2 \leq \exp\left(-\frac{\mu^2}{L^2} \cdot k\right) \|x^0 - x^*\|_2^2$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

С предыдущего слайда:

$$\|x^k - x^*\|_2^2 \le \exp\left(-\frac{\mu^2}{L^2} \cdot k\right) \|x^0 - x^*\|_2^2$$

Мы хотим, чтобы гарантированно

$$\|x^k - x^*\|_2^2 \le \exp\left(-\frac{\mu^2}{L^2} \cdot k\right) \|x^0 - x^*\|_2^2 \le \varepsilon^2$$

Тогда логарифмируем и получаем

$$k \ge \frac{L^2}{\mu^2} \log\left(\frac{\|x^0 - x^*\|_2^2}{\varepsilon^2}\right)$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

С предыдущего слайда:

$$\|x^k - x^*\|_2^2 \leq \exp\left(-\frac{\mu^2}{L^2} \cdot k\right) \|x^0 - x^*\|_2^2$$

Мы хотим, чтобы гарантированно

$$\|x^k - x^*\|_2^2 \leq \exp\left(-\frac{\mu^2}{L^2} \cdot k\right) \|x^0 - x^*\|_2^2 \leq \varepsilon^2$$

Тогда логарифмируем и получаем

$$k \geq \frac{L^2}{\mu^2} \log\left(\frac{\|x^0 - x^*\|_2^2}{\varepsilon^2}\right)$$

Итого: Not great, not terrible – можно лучше. Пример того, как в получении верхних оценок можно «загрубить».

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Стартуем аналогично:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 &= \|x^k - \gamma_k \nabla f(x^k) - x^*\|_2^2 \\
&= \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle + \gamma_k^2 \|\nabla f(x^k)\|_2^2 \\
&= \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle \\
&\quad + \gamma_k^2 \|\nabla f(x^k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

$$\le \gamma_k^2 \cdot 2L \left(f(x^k) - f(x^*) + \cancel{\langle \nabla f(x^*), x^k - x^* \rangle} \right)$$

$$\le -2\gamma_k \left(\frac{\mu}{2} \|x^k - x^*\|^2 + f(x^k) - f(x^*) \right)$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Стартуем аналогично:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 =& \|x^k - \gamma_k \nabla f(x^k) - x^*\|_2^2 \\
=& \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle + \gamma_k^2 \|\nabla f(x^k)\|_2^2 \\
=& \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle \\
& + \gamma_k^2 \|\nabla f(x^k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

Но сделаем тоньше. Сильная выпуклость в виде:
$-\langle \nabla f(x), x - y \rangle \leq \frac{\mu}{2}\|x - y\|_2^2 + f(x) - f(y)$:

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Стартуем аналогично:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 =& \|x^k - \gamma_k \nabla f(x^k) - x^*\|_2^2 \\
=& \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle + \gamma_k^2 \|\nabla f(x^k)\|_2^2 \\
=& \|x^k - x^*\|_2^2 - 2\gamma_k \langle \nabla f(x^k), x^k - x^* \rangle \\
& + \gamma_k^2 \|\nabla f(x^k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

Но сделаем тоньше. Сильная выпуклость в виде:
$-\langle \nabla f(x), x - y \rangle \leq \frac{\mu}{2}\|x - y\|_2^2 + f(x) - f(y)$:

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 \leq& \|x^k - x^*\|_2^2 - 2\gamma_k \left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*) \right) \\
& + \gamma_k^2 \|\nabla f(x^k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Дальше гладкость, но в виде:
$\|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \leq 2L\left(f(x^k) - f(x^*)\right)$. **Вопрос:** все ли верно в этом свойстве?

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Дальше гладкость, но в виде: $\|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \le 2L\left(f(x^k) - f(x^*)\right)$. **Вопрос:** все ли верно в этом свойстве? Да, использовано, что $\nabla f(x^*) = 0$. Получаем

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 \le & \|x^k - x^*\|_2^2 - 2\gamma_k\left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right) \\
& + 2\gamma_k^2 L(f(x^k) - f(x^*)) \\
= & (1 - \gamma_k\mu)\|x^k - x^*\|_2^2 + 2\gamma_k(\gamma_k L - 1)(f(x^k) - f(x^*))
\end{aligned}
$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Дальше гладкость, но в виде:
$\|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \leq 2L\left(f(x^k) - f(x^*)\right)$. **Вопрос:** все ли верно в этом свойстве? Да, использовано, что $\nabla f(x^*) = 0$. Получаем

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 \leq & \|x^k - x^*\|_2^2 - 2\gamma_k\left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right) \\
& + 2\gamma_k^2 L(f(x^k) - f(x^*)) \\
= & (1 - \gamma_k\mu)\|x^k - x^*\|_2^2 + 2\gamma_k(\gamma_k L - 1)(f(x^k) - f(x^*))
\end{aligned}
$$

Вопрос: что осталось?

$\gamma_k \leq \frac{1}{L}$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

Дальше гладкость, но в виде:
$\|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \leq 2L\left(f(x^k) - f(x^*)\right)$. **Вопрос:** все ли верно в этом свойстве? Да, использовано, что $\nabla f(x^*) = 0$. Получаем

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 \leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right) \\
& + 2\gamma_k^2 L(f(x^k) - f(x^*)) \\
= & (1 - \gamma_k \mu)\|x^k - x^*\|_2^2 + 2\gamma_k(\gamma_k L - 1)\underbrace{(f(x^k) - f(x^*))}_{\geq 0}
\end{aligned}
$$

Вопрос: что осталось? $(\gamma_k L - 1) \leq 1$. А значит $\gamma_k \leq \frac{1}{L}$.

$$
\|x^{k+1} - x^*\|_2^2 \leq (1 - \gamma_k \mu)\|x^k - x^*\|_2^2
$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Доказательство

С предыдущего слайда:

$$\|x^{k+1} - x^*\|_2^2 \leq (1 - \gamma_k \mu)\|x^k - x^*\|_2^2$$

Запускаем рекурсию:

$$\|x^k - x^*\|_2^2 \leq \prod_{i=0}^{k-1}(1 - \gamma_i \mu)\|x^0 - x^*\|_2^2$$

С постоянным шагом $\gamma_k = \gamma = \frac{1}{L}$:

$$\|x^k - x^*\|_2^2 \leq \left(1 - \frac{\mu}{L}\right)^k \|x^0 - x^*\|_2^2$$

Сходимость: L-гладкие и μ-сильно выпуклые функции

Теорема сходимость градиентного спуска для L-гладких и μ-сильно выпуклых функций

Пусть задача задача безусловной оптимизации (1) с L-гладкой, μ-сильно выпуклой целевой функцией f решается с помощью градиентного спуска. Тогда справедлива следующая оценка сходимости

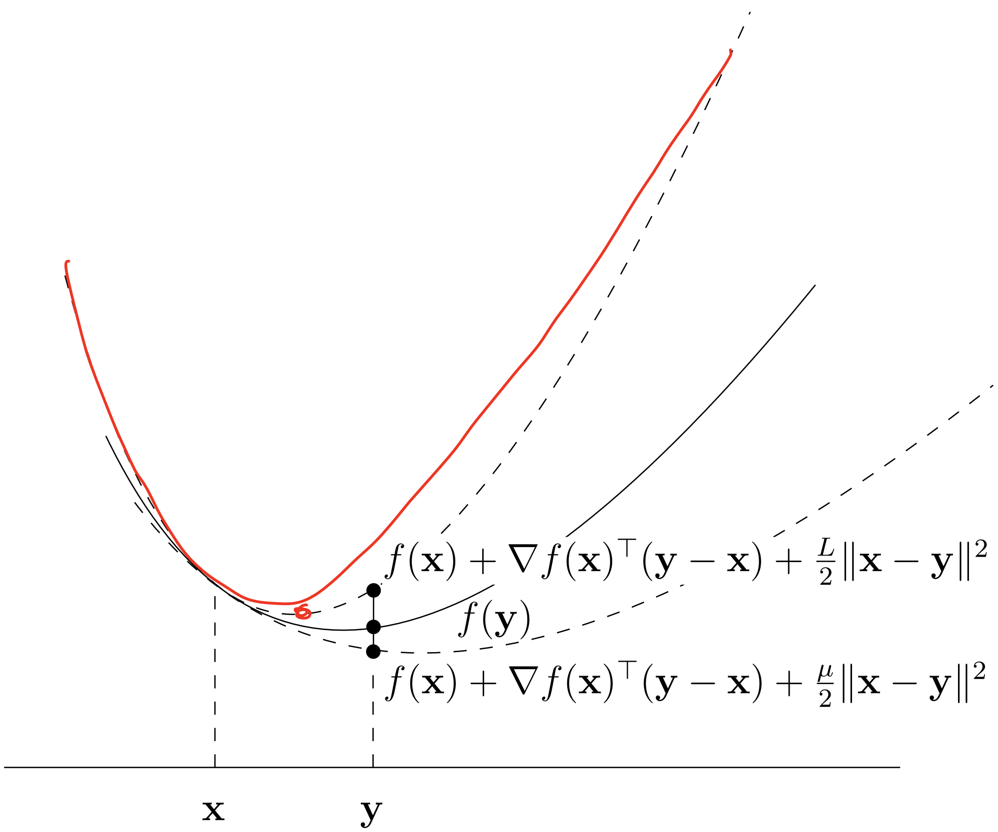
$$\|x^k - x^*\|_2^2 \leq \left(1 - \frac{\mu}{L}\right)^k \|x^0 - x^*\|_2^2.$$

Более того, чтобы добиться точности ε по аргументу, необходимо

$$k = O\left(\frac{L}{\mu} \log \frac{\|x^0 - x^*\|_2}{\varepsilon}\right) = \tilde{O}\left(\frac{L}{\mu}\right)$$ итераций.

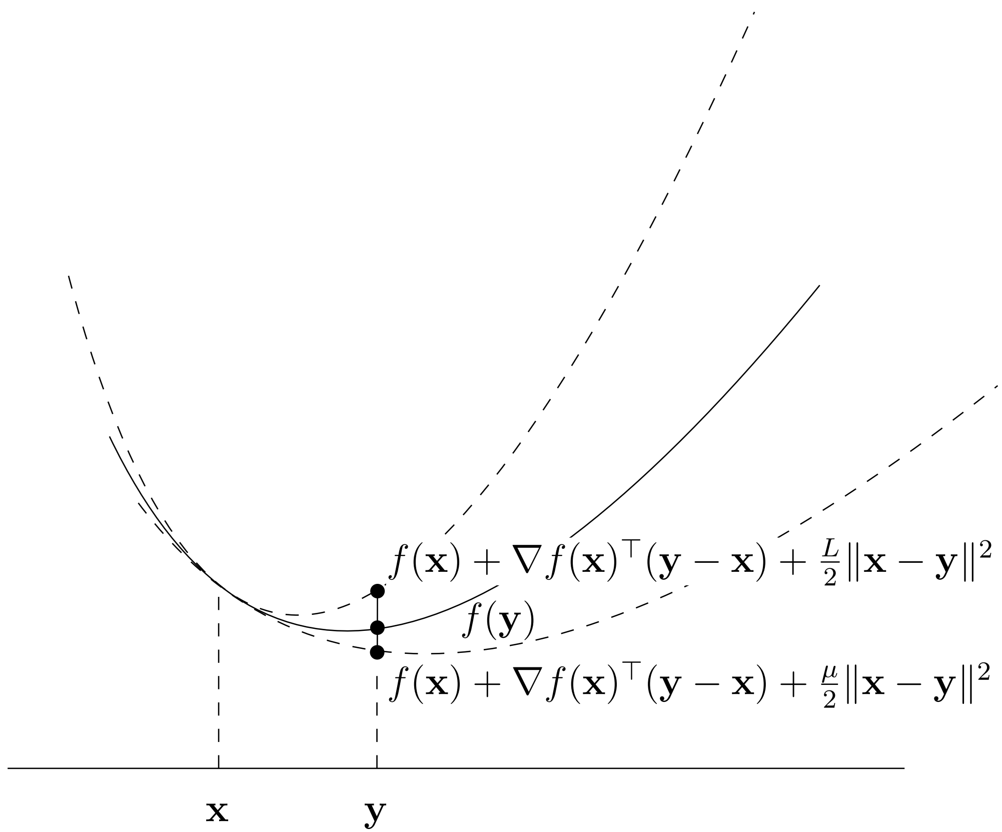
Мы будем использовать O-нотацию, чтобы "убирать" численные фактор и $\tilde{O}$-нотацию, чтобы убирать еще и log-факторы.

Немного интуиции доказательства



$\frac{\mu}{L}$

Немного интуиции доказательства



Шагаем, исходя из свойств верхней границы (L) – чтобы гарантированно не "улететь", и перемещаемся в худшем случае, исходя из свойств нижней границы (μ).

Сходимость

	μ-сильно выпуклая	выпуклая	невыпуклая
L-гладкая	$O\left(\frac{L}{\mu}\log\frac{\|x^0-x^*\|_2}{\varepsilon}\right)$	$O\left(\frac{L\|x^0-x^*\|_2^2}{\varepsilon}\right)$	$O\left(\frac{L(f(x^0)-f^*)}{\varepsilon^2}\right)$
M-липшецева	$O\left(\frac{M^2}{\mu^2\varepsilon}\right)$	$O\left(\frac{M^2\|x^0-x^*\|_2^2}{\varepsilon^2}\right)$	1 лекция

- В сильно выпуклом случае по аргументу: $\|x - x^*\|_2 \le \varepsilon$,
- В выпуклом случае по функции (решение x^* может быть не единственно): $f(x) - f^* \le \varepsilon$,
- В невыпуклом случае (сходимость к какой-то стационарной точке): $\|\nabla f(x)\|_2 \le \varepsilon$.

$\varepsilon \sim \frac{1}{k}$

$k \sim \frac{1}{\varepsilon}$

Сходимость

	μ-сильно выпуклая	выпуклая	невыпуклая
L-гладкая	$O\left(\frac{L}{\mu}\log\frac{\\|x^0-x^*\\|_2}{\varepsilon}\right)$	$O\left(\frac{L\\|x^0-x^*\\|_2^2}{\varepsilon}\right)$	$O\left(\frac{L(f(x^0)-f^*)}{\varepsilon^2}\right)$
M-липшецева	$O\left(\frac{M^2}{\mu^2\varepsilon}\right)$	$O\left(\frac{M^2\\|x^0-x^*\\|_2^2}{\varepsilon^2}\right)$	1 лекция

- В сильно выпуклом случае по аргументу: $\|x - x^*\|_2 \leq \varepsilon$,
- В выпуклом случае по функции (решение x^* может быть не единственно): $f(x) - f^* \leq \varepsilon$,
- В невыпуклом случае (сходимость к какой-то стационарной точке): $\|\nabla f(x)\|_2 \leq \varepsilon$.
- Градиентный спуск оптимален (**вопрос:** что это значит?) в негладком случае, а также в гладком невыпуклом.
- Наш анализ градиентного спуска в сильно выпуклом случае неулучшаем с точностью до численных множителей.
- В гладком выпуклом и сильно выпуклом случаях возможны улучшения, но для этого нужен другой метод (3 лекция).

Подбор шага: Поляк-Шор

Уже получали

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 \leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right) \\
& + \gamma_k^2 \|\nabla f(x^k)\|_2^2 \\
\leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(f(x^k) - f(x^*)\right) + \gamma_k^2 \|\nabla f(x^k)\|_2^2
\end{aligned}
$$

Подбор шага: Поляк-Шор

Уже получали

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 \leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right) \\
& + \gamma_k^2 \|\nabla f(x^k)\|_2^2 \\
\leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(f(x^k) - f(x^*)\right) + \gamma_k^2 \|\nabla f(x^k)\|_2^2
\end{aligned}
$$

Вопрос: как можно подобрать γ_k оптимально в этой ситуации?

Подбор шага: Поляк-Шор

Уже получали

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 \leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right) \\
& + \gamma_k^2 \|\nabla f(x^k)\|_2^2 \\
\leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(f(x^k) - f(x^*)\right) + \gamma_k^2 \|\nabla f(x^k)\|_2^2
\end{aligned}
$$

Вопрос: как можно подобрать γ_k оптимально в этой ситуации?
$\arg\min_{\gamma_k} \left(-2\gamma_k \left(f(x^k) - f(x^*)\right) + \gamma_k^2 \|\nabla f(x^k)\|_2^2\right)$?

Подбор шага: Поляк-Шор

Уже получали

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 \leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*) \right) \\
& + \gamma_k^2 \|\nabla f(x^k)\|_2^2 \\
\leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(f(x^k) - f(x^*) \right) + \gamma_k^2 \|\nabla f(x^k)\|_2^2
\end{aligned}
$$

Вопрос: как можно подобрать γ_k оптимально в этой ситуации?
$\arg\min_{\gamma_k} \left(-2\gamma_k \left(f(x^k) - f(x^*) \right) + \gamma_k^2 \|\nabla f(x^k)\|_2^2 \right)$?

$$
\gamma_k = \frac{f(x^k) - f(x^*)}{\|\nabla f(x^k)\|_2^2}
$$

Вопрос: какие видите проблемы?

Подбор шага: Поляк-Шор

Уже получали

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 \leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right) \\
& + \gamma_k^2 \|\nabla f(x^k)\|_2^2 \\
\leq & \|x^k - x^*\|_2^2 - 2\gamma_k \left(f(x^k) - f(x^*)\right) + \gamma_k^2 \|\nabla f(x^k)\|_2^2
\end{aligned}
$$

Вопрос: как можно подобрать γ_k оптимально в этой ситуации?
$\arg\min_{\gamma_k} \left(-2\gamma_k \left(f(x^k) - f(x^*)\right) + \gamma_k^2 \|\nabla f(x^k)\|_2^2\right)$?

$$
\gamma_k = \frac{f(x^k) - f(x^*)}{\|\nabla f(x^k)\|_2^2}
$$

Вопрос: какие видите проблемы? $f(x^*)$ – иногда известно, а иногда можно оценить.

Подбор шага

- Шаг Поляка-Шора:

$$\gamma_k = \frac{f(x^k) - f(x^*)}{\alpha\|\nabla f(x^k)\|_2^2}, \quad \alpha \geq 1 \quad \text{(надо подбирать)}$$

Подбор шага

- Шаг Поляка-Шора:

$$\gamma_k = \frac{f(x^k) - f(x^*)}{\alpha \|\nabla f(x^k)\|_2^2}, \quad \alpha \geq 1 \quad \text{(надо подбирать)}$$

- Наискорейший спуск:

$$\gamma_k = \arg\min_{\gamma} f(x^k - \gamma \nabla f(x^k))$$

Подбор шага

- Шаг Поляка-Шора:

$$\gamma_k = \frac{f(x^k) - f(x^*)}{\alpha \|\nabla f(x^k)\|_2^2}, \quad \alpha \geq 1 \quad \text{(надо подбирать)}$$

- Наискорейший спуск:

$$\gamma_k = \arg\min_{\gamma} f(x^k - \gamma \nabla f(x^k))$$

Вопрос: как решать?

Подбор шага

- Шаг Поляка-Шора:

$$\gamma_k = \frac{f(x^k) - f(x^*)}{\alpha \|\nabla f(x^k)\|_2^2}, \quad \alpha \geq 1 \quad \text{(надо подбирать)}$$

- Наискорейший спуск:

$$\gamma_k = \arg\min_{\gamma} f(x^k - \gamma \nabla f(x^k))$$

Вопрос: как решать? Иногда есть явная формула, а так нужно решать одномерную задачу.

Подбор шага

- Шаг Поляка-Шора:

$$\gamma_k = \frac{f(x^k) - f(x^*)}{\alpha \|\nabla f(x^k)\|_2^2}, \quad \alpha \geq 1 \quad \text{(надо подбирать)}$$

- Наискорейший спуск:

$$\gamma_k = \arg\min_{\gamma} f(x^k - \gamma \nabla f(x^k))$$

  Вопрос: как решать? Иногда есть явная формула, а так нужно решать одномерную задачу.

- Правила Армихо, Вульфа и Гольдстейна (на 8 семинаре).
- Адаптивный подбор, например, онлайн оценка локальной константы L (на 8 семинаре).