

# Методы оптимизации

## Лекция 7: Введение в методы оптимизации.

### Градиентный спуск

Александр Катруца

Физтех-школа прикладной математики и информатики  
Московский физико-технический институт



13 ноября 2021 г.

## На прошлой лекции

- ▶ Использование выпуклости задачи при её решении
- ▶ Disciplined convex programming
- ▶ CVXPY
- ▶ ipopt

## Постановка задачи

$$\begin{aligned} & \min_{\mathbf{x} \in S} f_0(\mathbf{x}) \\ \text{s.t. } & f_j(\mathbf{x}) = 0, \quad j = 1, \dots, m \\ & g_k(\mathbf{x}) \leq 0, \quad k = 1, \dots, p \end{aligned}$$

где  $S \subseteq \mathbb{R}^n$ ,  $f_j : S \rightarrow \mathbb{R}$ ,  $j = 0, \dots, m$ ,  $g_k : S \rightarrow \mathbb{R}$ ,  $k = 1, \dots, p$

- ▶ Все функции как минимум непрерывны
- ▶ Задачи нелинейной оптимизации в общем случае являются **численно неразрешимыми!**

# Некоторые аналитические результаты

# Некоторые аналитические результаты

## Необходимое условие первого порядка

Если  $\mathbf{x}^*$  точка локального минимума дифференцируемой функции  $f(\mathbf{x})$ , тогда

$$f'(\mathbf{x}^*) = 0$$

# Некоторые аналитические результаты

## Необходимое условие первого порядка

Если  $\mathbf{x}^*$  точка локального минимума дифференцируемой функции  $f(\mathbf{x})$ , тогда

$$f'(\mathbf{x}^*) = 0$$

## Необходимое условие второго порядка

Если  $\mathbf{x}^*$  точка локального минимума дважды дифференцируемой функции  $f(\mathbf{x})$ , тогда

$$f'(\mathbf{x}^*) = 0 \quad \text{и} \quad f''(\mathbf{x}^*) \succeq 0$$

# Некоторые аналитические результаты

## Необходимое условие первого порядка

Если  $\mathbf{x}^*$  точка локального минимума дифференцируемой функции  $f(\mathbf{x})$ , тогда

$$f'(\mathbf{x}^*) = 0$$

## Необходимое условие второго порядка

Если  $\mathbf{x}^*$  точка локального минимума дважды дифференцируемой функции  $f(\mathbf{x})$ , тогда

$$f'(\mathbf{x}^*) = 0 \quad \text{и} \quad f''(\mathbf{x}^*) \succeq 0$$

## Достаточное условие

Пусть  $f(\mathbf{x})$  дважды дифференцируемая функция, и пусть точка  $\mathbf{x}^*$  удовлетворяет условиям

$$f'(\mathbf{x}^*) = 0 \quad f''(\mathbf{x}^*) \succ 0,$$

тогда  $\mathbf{x}^*$  является точкой строгого локального минимума функции  $f(\mathbf{x})$

# Особенности численного решения

- ▶ Точно решить задачу принципиально невозможно из-за погрешности машинной арифметики



# Особенности численного решения

- ▶ Точно решить задачу принципиально невозможно из-за погрешности машинной арифметики
- ▶ Необходимо задать критерий обнаружения решения

# Особенности численного решения

- ▶ Точно решить задачу принципиально невозможно из-за погрешности машинной арифметики
- ▶ Необходимо задать критерий обнаружения решения
- ▶ Необходимо определить, какую информацию о задаче использовать

# Общая схема

- ▶ Начальная точка  $x_0$
- ▶ Желаемая точность  $\varepsilon$

```
def GeneralScheme(x, epsilon):  
    while StopCriterion(x) > epsilon:  
        OracleResponse = RequestOracle(x)  
        UpdateInformation(I, x, OracleResponse)  
        x = NextPoint(I, x)  
    return x
```

# Вопросы

1. Какие критерии остановки могут быть?

# Вопросы

1. Какие критерии остановки могут быть?
2. Что такое оракул и зачем он нужен?

# Вопросы

1. Какие критерии остановки могут быть?
2. Что такое оракул и зачем он нужен?
3. Что такое информационная модель?

# Вопросы

1. Какие критерии остановки могут быть?
2. Что такое оракул и зачем он нужен?
3. Что такое информационная модель?
4. Как вычисляется новая точка?

# Критерии остановки

1. Сходимость по аргументу:

$$\|\mathbf{x}_k - \mathbf{x}^*\|_2 < \varepsilon$$



# Критерии остановки

1. Сходимость по аргументу:

$$\|\mathbf{x}_k - \mathbf{x}^*\|_2 < \varepsilon$$

2. Сходимость по функции:

$$|f_k - f^*| < \varepsilon$$

# Критерии остановки

1. Сходимость по аргументу:

$$\|\mathbf{x}_k - \mathbf{x}^*\|_2 < \varepsilon$$

2. Сходимость по функции:

$$|f_k - f^*| < \varepsilon$$

3. Выполнение необходимого условия

$$\|f'(\mathbf{x}_k)\|_2 < \varepsilon$$

# Критерии остановки

1. Сходимость по аргументу:

$$\|\mathbf{x}_k - \mathbf{x}^*\|_2 < \varepsilon$$

2. Сходимость по функции:

$$|f_k - f^*| < \varepsilon$$

3. Выполнение необходимого условия

$$\|f'(\mathbf{x}_k)\|_2 < \varepsilon$$

4. Зазор двойственности

$$f_k - g(\boldsymbol{\lambda}_k, \boldsymbol{\mu}_k) \leq \varepsilon$$

Что такое оракул?



# Что такое оракул?

## Почти определение

Оракулом называют некоторое абстрактное устройство, которое отвечает на последовательные вопросы метода

# Что такое оракул?

## Почти определение

Оракулом называют некоторое абстрактное устройство, которое отвечает на последовательные вопросы метода

## Аналогия из ООП

- ▶ оракул – это виртуальный метод базового класса
- ▶ каждая задача – производный класс
- ▶ оракул определяется для каждой задачи отдельно согласно общему определению в базовом классе

# Что такое оракул?

## Почти определение

Оракулом называют некоторое абстрактное устройство, которое отвечает на последовательные вопросы метода

## Аналогия из ООП

- ▶ оракул – это виртуальный метод базового класса
- ▶ каждая задача – производный класс
- ▶ оракул определяется для каждой задачи отдельно согласно общему определению в базовом классе

## Концепция чёрного ящика

1. Единственной информацией, получаемой в ходе работы итерационного метода, являются ответы оракула
2. Ответы оракула являются **локальными**

# Информация о задаче

1. Каждый ответ оракула даёт **локальную** информацию о поведении функции в точке
2. Агрегируя все полученные ответы оракула, обновляем информацию о **глобальном** виде целевой функции:
  - ▶ кривизна
  - ▶ направление убывания
  - ▶ etc



Вычисление следующей точки

## Вычисление следующей точки

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{h}_k$$

## Вычисление следующей точки

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{h}_k$$

### Линейный поиск

1. Сначала выбирается направление  $\mathbf{h}_k$
2. Далее определяется «оптимальное» значение  $\alpha_k$

# Вычисление следующей точки

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{h}_k$$

## Линейный поиск

1. Сначала выбирается направление  $\mathbf{h}_k$
2. Далее определяется «оптимальное» значение  $\alpha_k$

## Метод доверительных областей

1. Выбирается  $\alpha$ -окрестность  $\mathbf{x}_k$
2. В этой окрестности строится упрощённая **модель** целевой функции
3. Далее определяется направления  $\mathbf{h}_k$ , минимизирующее модель целевой функции и не выводящее точку  $\mathbf{x}_k + \mathbf{h}_k$  за пределы области

# Как сравнивать методы оптимизации?

Для заданного класса задач сравнивают следующие величины:

## 1. Сложность

- ▶ аналитическая: число обращений к оракулу для решения задачи с точностью  $\varepsilon$
- ▶ арифметическая: общее число всех вычислений, необходимых для решения задачи с точностью  $\varepsilon$

## 2. Скорость сходимости

## 3. Эксперименты

# Скорости сходимости

## 1. Сублинейная

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq Ck^\alpha,$$

где  $\alpha < 0$  и  $0 < C < \infty$

# Скорости сходимости

## 1. Сублинейная

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq Ck^\alpha,$$

где  $\alpha < 0$  и  $0 < C < \infty$

## 2. Линейная (геометрическая прогрессия)

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq Cq^k,$$

где  $q \in (0, 1)$  и  $0 < C < \infty$

# Скорости сходимости

## 1. Сублинейная

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq Ck^\alpha,$$

где  $\alpha < 0$  и  $0 < C < \infty$

## 2. Линейная (геометрическая прогрессия)

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq Cq^k,$$

где  $q \in (0, 1)$  и  $0 < C < \infty$

## 3. Сверхлинейная

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq Cq^{k^p},$$

где  $q \in (0, 1)$ ,  $0 < C < \infty$  и  $p > 1$



# Скорости сходимости

## 1. Сублинейная

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq Ck^\alpha,$$

где  $\alpha < 0$  и  $0 < C < \infty$

## 2. Линейная (геометрическая прогрессия)

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq Cq^k,$$

где  $q \in (0, 1)$  и  $0 < C < \infty$

## 3. Сверхлинейная

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq Cq^{k^p},$$

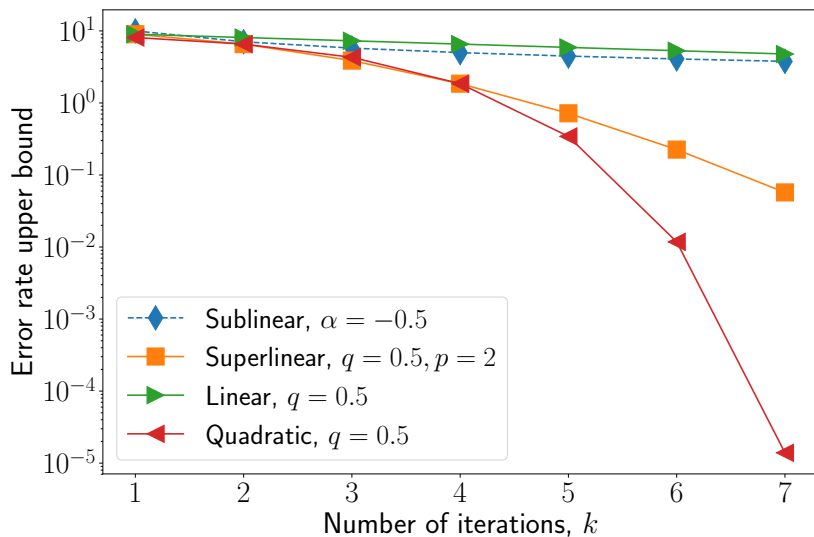
где  $q \in (0, 1)$ ,  $0 < C < \infty$  и  $p > 1$

## 4. Квадратичная

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq C\|\mathbf{x}_k - \mathbf{x}^*\|_2^2, \quad \text{или} \quad \|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq Cq^{2^k}$$

где  $q \in (0, 1)$  и  $0 < C < \infty$

# Сравнение скоростей сходимости



# Значение теорем сходимости

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

## Что дают теоремы сходимости

- ▶ класс задач, для которых применим метод

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

## Что дают теоремы сходимости

- ▶ класс задач, для которых применим метод
  - ▶ выпуклость

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

## Что дают теоремы сходимости

- ▶ класс задач, для которых применим метод
  - ▶ выпуклость
  - ▶ гладкость

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

## Что дают теоремы сходимости

- ▶ класс задач, для которых применим метод
  - ▶ выпуклость
  - ▶ гладкость
- ▶ качественное поведение метода

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

## Что дают теоремы сходимости

- ▶ класс задач, для которых применим метод
  - ▶ выпуклость
  - ▶ гладкость
- ▶ качественное поведение метода
  - ▶ существенно ли начальное приближение



# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

## Что дают теоремы сходимости

- ▶ класс задач, для которых применим метод
  - ▶ выпуклость
  - ▶ гладкость
- ▶ качественное поведение метода
  - ▶ существенно ли начальное приближение
  - ▶ по какому функционалу есть сходимость

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

## Что дают теоремы сходимости

- ▶ класс задач, для которых применим метод
  - ▶ выпуклость
  - ▶ гладкость
- ▶ качественное поведение метода
  - ▶ существенно ли начальное приближение
  - ▶ по какому функционалу есть сходимость
- ▶ оценку скорости сходимости

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

## Что дают теоремы сходимости

- ▶ класс задач, для которых применим метод
  - ▶ выпуклость
  - ▶ гладкость
- ▶ качественное поведение метода
  - ▶ существенно ли начальное приближение
  - ▶ по какому функционалу есть сходимость
- ▶ оценку скорости сходимости
  - ▶ теоретическая оценка без проведения экспериментов

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

## Что дают теоремы сходимости

- ▶ класс задач, для которых применим метод
  - ▶ выпуклость
  - ▶ гладкость
- ▶ качественное поведение метода
  - ▶ существенно ли начальное приближение
  - ▶ по какому функционалу есть сходимость
- ▶ оценку скорости сходимости
  - ▶ теоретическая оценка без проведения экспериментов
  - ▶ определение факторов, которые влияют на сходимость

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

## Что дают теоремы сходимости

- ▶ класс задач, для которых применим метод
  - ▶ выпуклость
  - ▶ гладкость
- ▶ качественное поведение метода
  - ▶ существенно ли начальное приближение
  - ▶ по какому функционалу есть сходимость
- ▶ оценку скорости сходимости
  - ▶ теоретическая оценка без проведения экспериментов
  - ▶ определение факторов, которые влияют на сходимость
  - ▶ иногда заранее можно выбрать число итераций для достижения заданной точности

# Значение теорем сходимости

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

Что НЕ дают теоремы сходимости

- ▶ сходимость метода *ничего не говорит* о целесообразности его применения

# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

Что НЕ дают теоремы сходимости

- ▶ сходимость метода *ничего не говорит* о целесообразности его применения
- ▶ оценки сходимости зависят от неизвестных констант



# Значение теорем сходимости

(Б.Т. Поляк Введение в оптимизацию, гл. 1, § 6)

Что НЕ дают теоремы сходимости

- ▶ сходимость метода *ничего не говорит* о целесообразности его применения
- ▶ оценки сходимости зависят от неизвестных констант
- ▶ учёт ошибок округления и точности решения вспомогательных задач

# Классификация методов

# Классификация методов

## Порядок метода

- ▶ Методы нулевого порядка: оракул возвращает только значение функции  $f(\mathbf{x})$

# Классификация методов

## Порядок метода

- ▶ Методы нулевого порядка: оракул возвращает только значение функции  $f(\mathbf{x})$
- ▶ Методы первого порядка: оракул возвращает значение функции  $f(\mathbf{x})$  и её градиент  $f'(\mathbf{x})$

# Классификация методов

## Порядок метода

- ▶ Методы нулевого порядка: оракул возвращает только значение функции  $f(\mathbf{x})$
- ▶ Методы первого порядка: оракул возвращает значение функции  $f(\mathbf{x})$  и её градиент  $f'(\mathbf{x})$
- ▶ Методы второго порядка: оракул возвращает значение функции  $f(\mathbf{x})$ , её градиент  $f'(\mathbf{x})$  и гессиан  $f''(\mathbf{x})$ .

# Классификация методов

## Порядок метода

- ▶ Методы нулевого порядка: оракул возвращает только значение функции  $f(\mathbf{x})$
- ▶ Методы первого порядка: оракул возвращает значение функции  $f(\mathbf{x})$  и её градиент  $f'(\mathbf{x})$
- ▶ Методы второго порядка: оракул возвращает значение функции  $f(\mathbf{x})$ , её градиент  $f'(\mathbf{x})$  и гессиан  $f''(\mathbf{x})$ .

**Q:** существуют ли методы более высокого порядка?

# Классификация методов

## Порядок метода

- ▶ Методы нулевого порядка: оракул возвращает только значение функции  $f(\mathbf{x})$
- ▶ Методы первого порядка: оракул возвращает значение функции  $f(\mathbf{x})$  и её градиент  $f'(\mathbf{x})$
- ▶ Методы второго порядка: оракул возвращает значение функции  $f(\mathbf{x})$ , её градиент  $f'(\mathbf{x})$  и гессиан  $f''(\mathbf{x})$ .

**Q:** существуют ли методы более высокого порядка?

**A:** да, но их использование пока не столь широко распространено. Основная работа<sup>1</sup>

---

<sup>1</sup>Nesterov Y. Implementable tensor methods in unconstrained convex optimization // Mathematical Programming. – 2019. – С. 1-27.

# Использование истории

## 1. Одношаговые методы

$$\mathbf{x}_{k+1} = \Phi(\mathbf{x}_k)$$



# Использование истории

## 1. Одношаговые методы

$$\mathbf{x}_{k+1} = \Phi(\mathbf{x}_k)$$

## 2. Многошаговые методы

$$\mathbf{x}_{k+1} = \Phi(\mathbf{x}_k, \mathbf{x}_{k-1}, \dots)$$

- ▶ Введение в численные методы оптимизации
- ▶ Общая схема работы метода
- ▶ Способы сравнения методов оптимизации
- ▶ Зоопарк задач и методов

# Методы спуска

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{h}_k$$

так что

$$f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$$

## Методы спуска

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{h}_k$$

так что

$$f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$$

### Определение

Направление  $\mathbf{h}_k$  называется *направлением убывания*

# Методы спуска

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{h}_k$$

так что

$$f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$$

## Определение

Направление  $\mathbf{h}_k$  называется *направлением убывания*

## Замечание

Существуют методы, которые не требуют монотонного убывания функции от итерации к итерации

# $L$ -гладкая функция

## Определение

Пусть  $L > 0$ . Функция  $f$  называется  $L$ -гладкой, если она дифференцируема и выполнено следующее неравенство

$$\|f'(\mathbf{x}) - f'(\mathbf{y})\|_* \leq L\|\mathbf{x} - \mathbf{y}\|,$$

для любых  $\mathbf{x}, \mathbf{y}$ . Норма  $\|\cdot\|_*$  является сопряжённой нормой для  $\|\cdot\|$

# $L$ -гладкая функция

## Определение

Пусть  $L > 0$ . Функция  $f$  называется  $L$ -гладкой, если она дифференцируема и выполнено следующее неравенство

$$\|f'(\mathbf{x}) - f'(\mathbf{y})\|_* \leq L\|\mathbf{x} - \mathbf{y}\|,$$

для любых  $\mathbf{x}, \mathbf{y}$ . Норма  $\|\cdot\|_*$  является сопряжённой нормой для  $\|\cdot\|$

- ▶  $L$  является константой Липшица для градиента

# $L$ -гладкая функция

## Определение

Пусть  $L > 0$ . Функция  $f$  называется  $L$ -гладкой, если она дифференцируема и выполнено следующее неравенство

$$\|f'(\mathbf{x}) - f'(\mathbf{y})\|_* \leq L\|\mathbf{x} - \mathbf{y}\|,$$

для любых  $\mathbf{x}, \mathbf{y}$ . Норма  $\|\cdot\|_*$  является сопряжённой нормой для  $\|\cdot\|$

- ▶  $L$  является константой Липшица для градиента
- ▶ Интерес представляет минимально возможное  $L$ , при котором выполняется неравенство в определении



# Лемма (descent lemma)

## Descent lemma

Пусть  $f$   $L$ -гладкая функция. Тогда для любых пар  $\mathbf{x}, \mathbf{y}$

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

## Доказательство

# Лемма (descent lemma)

## Descent lemma

Пусть  $f$   $L$ -гладкая функция. Тогда для любых пар  $\mathbf{x}, \mathbf{y}$

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

## Доказательство

- По формуле Ньютона-Лейбница

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \langle f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle dt$$

# Лемма (descent lemma)

## Descent lemma

Пусть  $f$   $L$ -гладкая функция. Тогда для любых пар  $\mathbf{x}, \mathbf{y}$

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

## Доказательство

- По формуле Ньютона-Лейбница

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \langle f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle dt$$

- $f(\mathbf{y}) - f(\mathbf{x}) =$   
 $\langle f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \int_0^1 \langle f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle dt$

# Лемма (descent lemma)

## Descent lemma

Пусть  $f$   $L$ -гладкая функция. Тогда для любых пар  $\mathbf{x}, \mathbf{y}$

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

## Доказательство

- ▶ По формуле Ньютона-Лейбница

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \langle f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle dt$$

- ▶  $f(\mathbf{y}) - f(\mathbf{x}) = \langle f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \int_0^1 \langle f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle dt$
- ▶  $|f(\mathbf{y}) - f(\mathbf{x}) - \langle f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| \leq \int_0^1 |\langle f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| dt \leq \int_0^1 \|f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f'(\mathbf{x})\| \|\mathbf{y} - \mathbf{x}\| dt$

# Лемма (descent lemma)

## Descent lemma

Пусть  $f$   $L$ -гладкая функция. Тогда для любых пар  $\mathbf{x}, \mathbf{y}$

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

## Доказательство

- ▶ По формуле Ньютона-Лейбница

$$f(\mathbf{y}) - f(\mathbf{x}) = \int_0^1 \langle f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle dt$$

- ▶  $f(\mathbf{y}) - f(\mathbf{x}) = \langle f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \int_0^1 \langle f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle dt$
- ▶  $|f(\mathbf{y}) - f(\mathbf{x}) - \langle f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| \leq \int_0^1 |\langle f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| dt \leq \int_0^1 \|f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f'(\mathbf{x})\| \|\mathbf{y} - \mathbf{x}\| dt$
- ▶  $\int_0^1 \|f'(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - f'(\mathbf{x})\| \|\mathbf{y} - \mathbf{x}\| dt \leq \int_0^1 tL \|\mathbf{y} - \mathbf{x}\|^2 dt = \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2$

## $L$ -гладкость и свойства гессиана

### Критерий второго порядка

Пусть  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  дважды непрерывно дифференцируема, тогда для заданного  $L > 0$  следующие условия эквивалентны

- ▶  $f$  является  $L$ -гладкой
- ▶  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

# $L$ -гладкость и свойства гессиана

## Критерий второго порядка

Пусть  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  дважды непрерывно дифференцируема, тогда для заданного  $L > 0$  следующие условия эквивалентны

- ▶  $f$  является  $L$ -гладкой
- ▶  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

## Доказательство

# $L$ -гладкость и свойства гессиана

## Критерий второго порядка

Пусть  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  дважды непрерывно дифференцируема, тогда для заданного  $L > 0$  следующие условия эквивалентны

- ▶  $f$  является  $L$ -гладкой
- ▶  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

## Доказательство

1. Пусть  $f$   $L$ -гладкая



# $L$ -гладкость и свойства гессиана

## Критерий второго порядка

Пусть  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  дважды непрерывно дифференцируема, тогда для заданного  $L > 0$  следующие условия эквивалентны

- ▶  $f$  является  $L$ -гладкой
- ▶  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

## Доказательство

1. Пусть  $f$   $L$ -гладкая

- ▶ Для любого направления  $\mathbf{d}$  и  $\alpha > 0$ :  
$$\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2 \leq \alpha L \|\mathbf{d}\|_2$$

# $L$ -гладкость и свойства гессиана

## Критерий второго порядка

Пусть  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  дважды непрерывно дифференцируема, тогда для заданного  $L > 0$  следующие условия эквивалентны

- ▶  $f$  является  $L$ -гладкой
- ▶  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

## Доказательство

1. Пусть  $f$   $L$ -гладкая

- ▶ Для любого направления  $\mathbf{d}$  и  $\alpha > 0$ :  
$$\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2 \leq \alpha L \|\mathbf{d}\|_2$$
- ▶ 
$$\lim_{\alpha \rightarrow +0} \frac{\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2}{\alpha} = \|f''(\mathbf{x})\mathbf{d}\|_2 \leq L \|\mathbf{d}\|_2$$

# $L$ -гладкость и свойства гессиана

## Критерий второго порядка

Пусть  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  дважды непрерывно дифференцируема, тогда для заданного  $L > 0$  следующие условия эквивалентны

- ▶  $f$  является  $L$ -гладкой
- ▶  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

## Доказательство

1. Пусть  $f$   $L$ -гладкая

- ▶ Для любого направления  $\mathbf{d}$  и  $\alpha > 0$ :  
$$\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2 \leq \alpha L \|\mathbf{d}\|_2$$
- ▶ 
$$\lim_{\alpha \rightarrow +0} \frac{\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2}{\alpha} = \|f''(\mathbf{x})\mathbf{d}\|_2 \leq L \|\mathbf{d}\|_2$$
- ▶ Так как это выполнено для любого  $\mathbf{d}$ , то  $\|f''(\mathbf{x})\|_2 \leq L$

# $L$ -гладкость и свойства гессиана

## Критерий второго порядка

Пусть  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  дважды непрерывно дифференцируема, тогда для заданного  $L > 0$  следующие условия эквивалентны

- ▶  $f$  является  $L$ -гладкой
- ▶  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

## Доказательство

1. Пусть  $f$   $L$ -гладкая

- ▶ Для любого направления  $\mathbf{d}$  и  $\alpha > 0$ :  
$$\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2 \leq \alpha L \|\mathbf{d}\|_2$$
- ▶ 
$$\lim_{\alpha \rightarrow +0} \frac{\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2}{\alpha} = \|f''(\mathbf{x})\mathbf{d}\|_2 \leq L \|\mathbf{d}\|_2$$
- ▶ Так как это выполнено для любого  $\mathbf{d}$ , то  $\|f''(\mathbf{x})\|_2 \leq L$

2. Пусть  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

# $L$ -гладкость и свойства гессиана

## Критерий второго порядка

Пусть  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  дважды непрерывно дифференцируема, тогда для заданного  $L > 0$  следующие условия эквивалентны

- ▶  $f$  является  $L$ -гладкой
- ▶  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

## Доказательство

1. Пусть  $f$   $L$ -гладкая

- ▶ Для любого направления  $\mathbf{d}$  и  $\alpha > 0$ :  
$$\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2 \leq \alpha L \|\mathbf{d}\|_2$$
- ▶ 
$$\lim_{\alpha \rightarrow +0} \frac{\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2}{\alpha} = \|f''(\mathbf{x})\mathbf{d}\|_2 \leq L \|\mathbf{d}\|_2$$
- ▶ Так как это выполнено для любого  $\mathbf{d}$ , то  $\|f''(\mathbf{x})\|_2 \leq L$

2. Пусть  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

- ▶ По формуле Ньютона-Лейбница  
$$f'(\mathbf{y}) - f'(\mathbf{x}) = \int_0^1 f''(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) dt$$

# $L$ -гладкость и свойства гессиана

## Критерий второго порядка

Пусть  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  дважды непрерывно дифференцируема, тогда для заданного  $L > 0$  следующие условия эквивалентны

- ▶  $f$  является  $L$ -гладкой
- ▶  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

## Доказательство

1. Пусть  $f$   $L$ -гладкая

- ▶ Для любого направления  $\mathbf{d}$  и  $\alpha > 0$ :  
$$\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2 \leq \alpha L \|\mathbf{d}\|_2$$
- ▶  $\lim_{\alpha \rightarrow +0} \frac{\|f'(\mathbf{x} + \alpha\mathbf{d}) - f'(\mathbf{x})\|_2}{\alpha} = \|f''(\mathbf{x})\mathbf{d}\|_2 \leq L \|\mathbf{d}\|_2$
- ▶ Так как это выполнено для любого  $\mathbf{d}$ , то  $\|f''(\mathbf{x})\|_2 \leq L$

2. Пусть  $\|f''(\mathbf{x})\|_2 \leq L$  для любого  $\mathbf{x}$

- ▶ По формуле Ньютона-Лейбница  
$$f'(\mathbf{y}) - f'(\mathbf{x}) = \int_0^1 f''(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) dt$$
- ▶  $\|f'(\mathbf{y}) - f'(\mathbf{x})\|_2 \leq \left( \int_0^1 \|f''(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))\|_2 dt \right) \|\mathbf{y} - \mathbf{x}\|_2 \leq L \|\mathbf{y} - \mathbf{x}\|_2$

## Градиентный спуск

Глобальная оценка сверху на функцию  $f$  в точке  $\mathbf{x}_k$ :

$$f(\mathbf{y}) \leq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{y} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{y} - \mathbf{x}_k\|_2^2 \equiv g(\mathbf{y}),$$

где  $\lambda_{\max}(f''(\mathbf{x})) \leq L$  для всех допустимых  $\mathbf{x}$ .

## Градиентный спуск

Глобальная оценка сверху на функцию  $f$  в точке  $\mathbf{x}_k$ :

$$f(\mathbf{y}) \leq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{y} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{y} - \mathbf{x}_k\|_2^2 \equiv g(\mathbf{y}),$$

где  $\lambda_{\max}(f''(\mathbf{x})) \leq L$  для всех допустимых  $\mathbf{x}$ .

Справа – квадратичная форма, точка минимума которой имеет аналитическое выражение:

$$g'(\mathbf{y}^*) = 0$$

$$f'(\mathbf{x}_k) + L(\mathbf{y}^* - \mathbf{x}_k) = 0$$

$$\mathbf{y}^* = \mathbf{x}_k - \frac{1}{L} f'(\mathbf{x}_k) \equiv \mathbf{x}_{k+1}$$

Этот способ позволяет оценить значение шага как  $\frac{1}{L}$ .



# Выбор шага

- ▶ Постоянный  $\alpha_k \equiv \text{const} < \frac{2}{L}$
- ▶ Убывающая последовательность, такая что  $\sum_{k=1}^{\infty} \alpha_k = \infty$ ,  
например  $\frac{1}{k}$ ,  $\frac{1}{\sqrt{k}}$ , etc
- ▶ Адаптивный поиск: правила Армихо, Вольфа, Гольдштейна и другие
- ▶ Наискорейший спуск: поиск лучшего  $\alpha_k$

## Важно

Лучший размер шага даёт не столь существенное теоретическое ускорение сходимости

## Сходимость к стационарной точке

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 = \\ &f(\mathbf{x}_k) - \alpha_k \|f'(\mathbf{x}_k)\|_2^2 + \frac{L\alpha_k^2}{2} \|f'(\mathbf{x}_k)\|_2^2 = \\ &f(\mathbf{x}_k) - \left( \alpha_k - \frac{L\alpha_k^2}{2} \right) \|f'(\mathbf{x}_k)\|_2^2 \end{aligned}$$

## Сходимость к стационарной точке

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 = \\ &f(\mathbf{x}_k) - \alpha_k \|f'(\mathbf{x}_k)\|_2^2 + \frac{L\alpha_k^2}{2} \|f'(\mathbf{x}_k)\|_2^2 = \\ &f(\mathbf{x}_k) - \left( \alpha_k - \frac{L\alpha_k^2}{2} \right) \|f'(\mathbf{x}_k)\|_2^2 \end{aligned}$$

► Условие убывания:  $\alpha_k - \frac{L\alpha_k^2}{2} > 0 \Rightarrow \alpha_k < \frac{2}{L}$

## Сходимость к стационарной точке

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 = \\ &f(\mathbf{x}_k) - \alpha_k \|f'(\mathbf{x}_k)\|_2^2 + \frac{L\alpha_k^2}{2} \|f'(\mathbf{x}_k)\|_2^2 = \\ &f(\mathbf{x}_k) - \left( \alpha_k - \frac{L\alpha_k^2}{2} \right) \|f'(\mathbf{x}_k)\|_2^2 \end{aligned}$$

- ▶ Условие убывания:  $\alpha_k - \frac{L\alpha_k^2}{2} > 0 \Rightarrow \alpha_k < \frac{2}{L}$
- ▶  $\alpha_k^* = \arg \max_{\alpha_k} \left( \alpha_k - \frac{L\alpha_k^2}{2} \right) = \frac{1}{L}$

## Сходимость к стационарной точке

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 = \\ &f(\mathbf{x}_k) - \alpha_k \|f'(\mathbf{x}_k)\|_2^2 + \frac{L\alpha_k^2}{2} \|f'(\mathbf{x}_k)\|_2^2 = \\ &f(\mathbf{x}_k) - \left( \alpha_k - \frac{L\alpha_k^2}{2} \right) \|f'(\mathbf{x}_k)\|_2^2 \end{aligned}$$

- ▶ Условие убывания:  $\alpha_k - \frac{L\alpha_k^2}{2} > 0 \Rightarrow \alpha_k < \frac{2}{L}$
- ▶  $\alpha_k^* = \arg \max_{\alpha_k} \left( \alpha_k - \frac{L\alpha_k^2}{2} \right) = \frac{1}{L}$
- ▶  $f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \frac{1}{2L} \|f'(\mathbf{x}_k)\|_2^2$

## Сходимость к стационарной точке

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 = \\ &= f(\mathbf{x}_k) - \alpha_k \|f'(\mathbf{x}_k)\|_2^2 + \frac{L\alpha_k^2}{2} \|f'(\mathbf{x}_k)\|_2^2 = \\ &= f(\mathbf{x}_k) - \left( \alpha_k - \frac{L\alpha_k^2}{2} \right) \|f'(\mathbf{x}_k)\|_2^2 \end{aligned}$$

- ▶ Условие убывания:  $\alpha_k - \frac{L\alpha_k^2}{2} > 0 \Rightarrow \alpha_k < \frac{2}{L}$
- ▶  $\alpha_k^* = \arg \max_{\alpha_k} \left( \alpha_k - \frac{L\alpha_k^2}{2} \right) = \frac{1}{L}$
- ▶  $f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \frac{1}{2L} \|f'(\mathbf{x}_k)\|_2^2$
- ▶  $\frac{1}{2L} \sum_{k=0}^T \|f'(\mathbf{x}_k)\|_2^2 \leq f(\mathbf{x}_0) - f(\mathbf{x}_{T+1}) \leq f(\mathbf{x}_0) - f^*$

## Сходимость к стационарной точке

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \frac{L}{2} \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 = \\ &= f(\mathbf{x}_k) - \alpha_k \|f'(\mathbf{x}_k)\|_2^2 + \frac{L\alpha_k^2}{2} \|f'(\mathbf{x}_k)\|_2^2 = \\ &= f(\mathbf{x}_k) - \left( \alpha_k - \frac{L\alpha_k^2}{2} \right) \|f'(\mathbf{x}_k)\|_2^2 \end{aligned}$$

- ▶ Условие убывания:  $\alpha_k - \frac{L\alpha_k^2}{2} > 0 \Rightarrow \alpha_k < \frac{2}{L}$
- ▶  $\alpha_k^* = \arg \max_{\alpha_k} \left( \alpha_k - \frac{L\alpha_k^2}{2} \right) = \frac{1}{L}$
- ▶  $f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}) \geq \frac{1}{2L} \|f'(\mathbf{x}_k)\|_2^2$
- ▶  $\frac{1}{2L} \sum_{k=0}^T \|f'(\mathbf{x}_k)\|_2^2 \leq f(\mathbf{x}_0) - f(\mathbf{x}_{T+1}) \leq f(\mathbf{x}_0) - f^*$
- ▶  $f$  ограничена снизу,  $\|f'(\mathbf{x}_k)\|_2 \rightarrow 0, k \rightarrow \infty$

# Сходимость для выпуклой функции

## Теорема

Пусть  $f$  выпуклая функция с Липшицевым градиентом и  $\alpha = \frac{1}{L}$ , тогда градиентный спуск сходится как

$$f(\mathbf{x}_{k+1}) - f^* \leq \frac{2L\|\mathbf{x}^* - \mathbf{x}_0\|_2^2}{k+4} = \mathcal{O}(1/k)$$



# Сходимость для сильно выпуклой функции

- ▶ Следствие сильной выпуклости

$$f(\mathbf{z}) \geq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{z} - \mathbf{x}_k \rangle + \frac{\mu}{2} \|\mathbf{z} - \mathbf{x}_k\|_2^2$$

# Сходимость для сильно выпуклой функции

- ▶ Следствие сильной выпуклости

$$f(\mathbf{z}) \geq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{z} - \mathbf{x}_k \rangle + \frac{\mu}{2} \|\mathbf{z} - \mathbf{x}_k\|_2^2$$

- ▶ Минимизируя обе части по  $\mathbf{z}$

$$f(\mathbf{x}^*) \geq f(\mathbf{x}_k) - \frac{1}{2\mu} \|f'(\mathbf{x}_k)\|_2^2, \quad \|f'(\mathbf{x}_k)\|_2^2 \geq 2\mu(f(\mathbf{x}_k) - f^*)$$

## Сходимость для сильно выпуклой функции

- ▶ Следствие сильной выпуклости

$$f(\mathbf{z}) \geq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{z} - \mathbf{x}_k \rangle + \frac{\mu}{2} \|\mathbf{z} - \mathbf{x}_k\|_2^2$$

- ▶ Минимизируя обе части по  $\mathbf{z}$

$$f(\mathbf{x}^*) \geq f(\mathbf{x}_k) - \frac{1}{2\mu} \|f'(\mathbf{x}_k)\|_2^2, \quad \|f'(\mathbf{x}_k)\|_2^2 \geq 2\mu(f(\mathbf{x}_k) - f^*)$$

- ▶ Вспомним, что для  $\alpha_k \equiv \frac{1}{L}$

$$f^* \leq f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{1}{2L} \|f'(\mathbf{x}_k)\|_2^2$$

## Сходимость для сильно выпуклой функции

- ▶ Следствие сильной выпуклости

$$f(\mathbf{z}) \geq f(\mathbf{x}_k) + \langle f'(\mathbf{x}_k), \mathbf{z} - \mathbf{x}_k \rangle + \frac{\mu}{2} \|\mathbf{z} - \mathbf{x}_k\|_2^2$$

- ▶ Минимизируя обе части по  $\mathbf{z}$

$$f(\mathbf{x}^*) \geq f(\mathbf{x}_k) - \frac{1}{2\mu} \|f'(\mathbf{x}_k)\|_2^2, \quad \|f'(\mathbf{x}_k)\|_2^2 \geq 2\mu(f(\mathbf{x}_k) - f^*)$$

- ▶ Вспомним, что для  $\alpha_k \equiv \frac{1}{L}$

$$f^* \leq f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{1}{2L} \|f'(\mathbf{x}_k)\|_2^2$$

- ▶ И наконец получим линейную сходимость

$$f(\mathbf{x}_{k+1}) - f^* \leq \left(1 - \frac{1}{\kappa}\right) (f(\mathbf{x}_k) - f^*)$$

# Теорема для сильно выпуклой функции

## Теорема

Пусть  $f$  с Липшицевым градиентом и  $\mu$  сильно выпукла,  $\alpha_k = \frac{2}{\mu+L}$ , тогда градиентный спуск сходится как

$$f(\mathbf{x}_k) - f^* \leq \frac{L}{2} \left( \frac{L - \mu}{L + \mu} \right)^{2k} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2$$

Что влияет на линейную скорость сходимости?

$$q^* = \frac{L - \mu}{L + \mu} = \frac{L/\mu - 1}{L/\mu + 1} = \frac{\kappa - 1}{\kappa + 1},$$

где  $\kappa$  - оценка числа обусловленности  $f''(\mathbf{x})$ .

**Q:** что такое число обусловленности матрицы?

## Что влияет на линейную скорость сходимости?

$$q^* = \frac{L - \mu}{L + \mu} = \frac{L/\mu - 1}{L/\mu + 1} = \frac{\kappa - 1}{\kappa + 1},$$

где  $\kappa$  - оценка числа обусловленности  $f''(\mathbf{x})$ .

**Q:** что такое число обусловленности матрицы?

- ▶ При  $\kappa \gg 1$ ,  $q^* \rightarrow 1 \Rightarrow$  оооочень медленная сходимость.  
Например при  $\kappa = 100$ :  $q^* \approx 0.98$

## Что влияет на линейную скорость сходимости?

$$q^* = \frac{L - \mu}{L + \mu} = \frac{L/\mu - 1}{L/\mu + 1} = \frac{\kappa - 1}{\kappa + 1},$$

где  $\kappa$  - оценка числа обусловленности  $f''(\mathbf{x})$ .

**Q:** что такое число обусловленности матрицы?

- ▶ При  $\kappa \gg 1$ ,  $q^* \rightarrow 1 \Rightarrow$  *оооочень медленная сходимости*.  
Например при  $\kappa = 100$ :  $q^* \approx 0.98$
- ▶ При  $\kappa \simeq 1$ ,  $q^* \rightarrow 0 \Rightarrow$  *ускорение сходимости*. Например при  $\kappa = 4$ :  $q^* = 0.6$



## Что влияет на линейную скорость сходимости?

$$q^* = \frac{L - \mu}{L + \mu} = \frac{L/\mu - 1}{L/\mu + 1} = \frac{\kappa - 1}{\kappa + 1},$$

где  $\kappa$  - оценка числа обусловленности  $f''(\mathbf{x})$ .

**Q:** что такое число обусловленности матрицы?

- ▶ При  $\kappa \gg 1$ ,  $q^* \rightarrow 1 \Rightarrow$  *очень медленная* сходимости. Например при  $\kappa = 100$ :  $q^* \approx 0.98$
- ▶ При  $\kappa \simeq 1$ ,  $q^* \rightarrow 0 \Rightarrow$  *ускорение* сходимости. Например при  $\kappa = 4$ :  $q^* = 0.6$

**Q:** какая геометрия у этого требования?

# Can we do better?

## Что нам известно

- ▶ Для выпуклых функций с Липшицевым градиентом градиентный спуск сходится как  $\mathcal{O}(1/k)$
- ▶ Для сильно выпуклых функций с Липшицевым градиентом градиентный спуск сходится с линейной скоростью  $q = \frac{\kappa-1}{\kappa+1}$

**Q:** есть ли методы, которые сходятся быстрее, и как это выяснить?

## Нижние оценки сходимости

Для обоих классов функций существуют такие «плохие» функции, для которых выполнены следующие оценки **снизу**

## Нижние оценки сходимости

Для обоих классов функций существуют такие «плохие» функции, для которых выполнены следующие оценки **снизу**

- ▶ для выпуклых функций с Липшицевым градиентом

$$f(\mathbf{x}_{k+1}) - f^* \geq \frac{3L\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2}{32(k+1)^2}$$

## Нижние оценки сходимости

Для обоих классов функций существуют такие «плохие» функции, для которых выполнены следующие оценки **снизу**

- ▶ для выпуклых функций с Липшицевым градиентом

$$f(\mathbf{x}_{k+1}) - f^* \geq \frac{3L\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2}{32(k+1)^2}$$

- ▶ для сильно выпуклых функций с Липшицевым градиентом

$$f(\mathbf{x}_{k+1}) - f^* \geq \frac{\mu}{2} \left( \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2k} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2$$

## Нижние оценки сходимости

Для обоих классов функций существуют такие «плохие» функции, для которых выполнены следующие оценки **снизу**

- ▶ для выпуклых функций с Липшицевым градиентом

$$f(\mathbf{x}_{k+1}) - f^* \geq \frac{3L\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2}{32(k+1)^2}$$

- ▶ для сильно выпуклых функций с Липшицевым градиентом

$$f(\mathbf{x}_{k+1}) - f^* \geq \frac{\mu}{2} \left( \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2k} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2$$

Эти оценки справедливы для таких методов, что

$$\mathbf{x}_{k+1} = \mathbf{x}_0 + \text{span}(f'(\mathbf{x}_0), \dots, f'(\mathbf{x}_k))$$

# Оптимальные методы

Про методы, которые в той или иной степени достигают нижних оценок, будет рассказано на следующей лекции:

- ▶ метод сопряжённых градиентов
- ▶ метод тяжёлого шарика
- ▶ градиентный метод Нестерова

- ▶ Общая схема работы методов оптимизации
- ▶ Скорости сходимости
- ▶ Градиентный спуск
- ▶ Свойства и сходимость
- ▶ Нижние оценки