# Методы оптимизации
# Лекция 7: Введение в методы оптимизации. Градиентный спуск

Александр Катруца

Факультет инноваций и высоких технологий
Физтех-школа прикладной математики и информатики



16 октября 2018 г.

# На прошлой лекции

- Использование выпуклости задачи при её решении
- Disciplined convex programming
- CVXPy
- ipopt

# Постановка задачи

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

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

- 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 $x^*$ точка локального минимума дифференцируемой функции $f(x)$, тогда

$$f'(x^*) = 0$$

# Некоторые аналитические результаты

## Необходимое условие первого порядка

Если $x^*$ точка локального минимума дифференцируемой функции $f(x)$, тогда

$$f'(x^*) = 0$$

## Необходимое условие второго порядка

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

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

# Некоторые аналитические результаты

## Необходимое условие первого порядка

Если $x^*$ точка локального минимума дифференцируемой функции $f(x)$, тогда

$$f'(x^*) = 0$$

## Необходимое условие второго порядка

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

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

## Достаточное условие

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

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

тогда $x^*$ является точкой строгого локального минимума функции $f(x)$

# 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:

$$\|x_k - x^*\|_2 < \varepsilon$$

# Критерии остановки

1. Сходимость по аргументу:

$$\|x_k - x^*\|_2 < \varepsilon$$

2. Сходимость по функции:

$$\|f_k - f^*\|_2 < \varepsilon$$

# Критерии остановки

1. Сходимость по аргументу:

$$\|x_k - x^*\|_2 < \varepsilon$$

2. Сходимость по функции:

$$\|f_k - f^*\|_2 < \varepsilon$$

3. Выполнение необходимого условия

$$\|f'(x_k)\|_2 < \varepsilon$$

# Критерии остановки

1. Сходимость по аргументу:

$$\|x_k - x^*\|_2 < \varepsilon$$

2. Сходимость по функции:

$$\|f_k - f^*\|_2 < \varepsilon$$

3. Выполнение необходимого условия

$$\|f'(x_k)\|_2 < \varepsilon$$

4. Зазор двойственности

$$p^* - d^* \leq \varepsilon$$

# 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

$$x_{k+1} = x_k + \alpha_k h_k$$

# Вычисление следующей точки

$$x_{k+1} = x_k + \alpha_k h_k$$

## Линейный поиск

1. Сначала выбирается направление $h_k$
2. Далее определяется «оптимальное» значение $\alpha_k$

# Вычисление следующей точки

$$x_{k+1} = x_k + \alpha_k h_k$$

## Линейный поиск

1. Сначала выбирается направление $h_k$
2. Далее определяется «оптимальное» значение $\alpha_k$

## Метод доверительных областей

1. Выбирается $\alpha$-окрестность $x_k$
2. В этой окрестности строится упрощённая **модель** целевой функции
3. Далее определяется направления $h_k$, минимизирующее модель целевой функции и не выводящее точку $x_k + h_k$ 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

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

где $\alpha < 0$ и $0 < C < \infty$

# Скорости сходимости

1. Сублинейная

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

где $\alpha < 0$ и $0 < C < \infty$

2. Линейная (геометрическая прогрессия)

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

где $q \in (0, 1)$ и $0 < C < \infty$

# Скорости сходимости

1. Сублинейная

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

где $\alpha < 0$ и $0 < C < \infty$

2. Линейная (геометрическая прогрессия)

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

где $q \in (0, 1)$ и $0 < C < \infty$

3. Сверхлинейная

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

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

# Скорости сходимости

1. Сублинейная

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

где $\alpha < 0$ и $0 < C < \infty$

2. Линейная (геометрическая прогрессия)

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

где $q \in (0, 1)$ и $0 < C < \infty$

3. Сверхлинейная

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

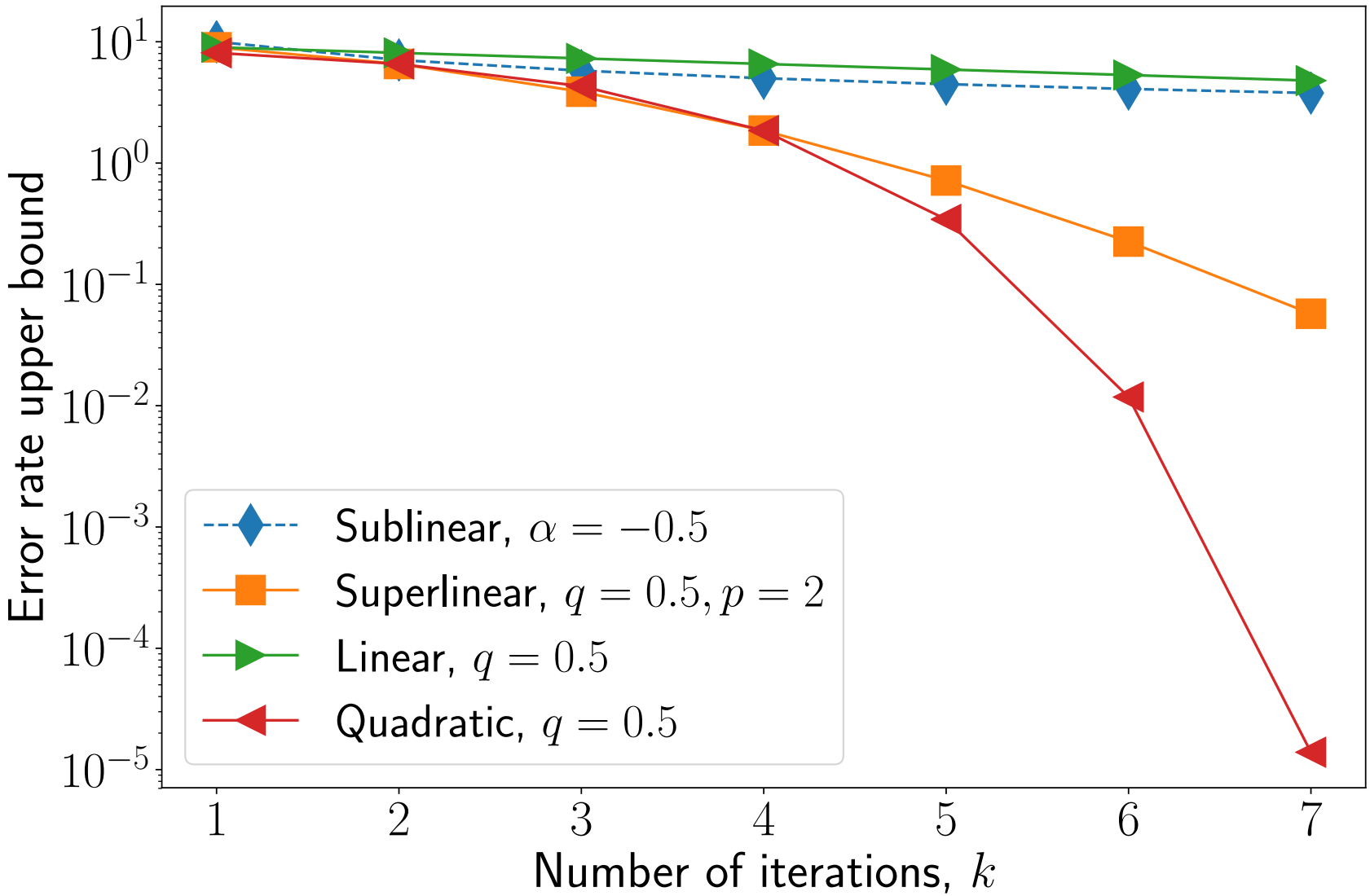
где $q \in (0, 1)$, $0 < C < \infty$ и $p > 1$

4. Квадратичная

$$\|x_{k+1}-x^*\|_2 \leq C\|x_k-x^*\|_2^2, \qquad \text{или} \qquad \|x_{k+1}-x^*\|_2 \leq Cq^{2^k}$$

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 $f(x)$

# Классификация методов

## Порядок метода

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

# Классификация методов

## Порядок метода

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

# Классификация методов

## Порядок метода

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

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

# Классификация методов

## Порядок метода

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

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

## Использование истории

# Классификация методов

## Порядок метода

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

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

## Использование истории

1. Одношаговые методы

$$x_{k+1} = \Phi(x_k)$$

# Классификация методов

## Порядок метода

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

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

## Использование истории

1. Одношаговые методы

$$x_{k+1} = \Phi(x_k)$$

2. Многошаговые методы

$$x_{k+1} = \Phi(x_k, x_{k-1}, ...)$$

# Главное

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

# Методы спуска

$$x_{k+1} = x_k + \alpha_k h_k$$

так что

$$f(x_{k+1}) < f(x_k)$$

# Методы спуска

$$x_{k+1} = x_k + \alpha_k h_k$$

так что

$$f(x_{k+1}) < f(x_k)$$

## Определение

Направление $h_k$ называется *направлением убывания*

# Методы спуска

$$x_{k+1} = x_k + \alpha_k h_k$$

так что

$$f(x_{k+1}) < f(x_k)$$

## Определение

Направление $h_k$ называется *направлением убывания*

## Замечание

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

# Градиентный спуск

Глобальная оценка сверху на функцию $f$ в точке $x_k$:

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

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

# Градиентный спуск

Глобальная оценка сверху на функцию $f$ в точке $x_k$:

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

где $\lambda_{\max}(f''(x)) \leq L$ для всех допустимых $x$.
Справа – квадратичная форма, точка минимума которой имеет аналитическое выражение:

$$\begin{aligned} g'(y^*) &= 0 \\ f'(x_k) + L(y^* - x_k) &= 0 \\ y^* &= x_k - \frac{1}{L} f'(x_k) \equiv x_{k+1} \end{aligned}$$

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

$$f(x_{k+1}) \leq f(x_k) + \langle f'(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_2^2 =$$

$$f(x_k) - \alpha_k \|f'(x_k)\|_2^2 + \frac{L\alpha_k^2}{2}\|f'(x_k)\|_2^2 =$$

$$f(x_k) - \left(\alpha_k - \frac{L\alpha_k^2}{2}\right)\|f'(x_k)\|_2^2$$

# Сходимость к стационарной точке

$$f(x_{k+1}) \leq f(x_k) + \langle f'(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_2^2 =$$

$$f(x_k) - \alpha_k \|f'(x_k)\|_2^2 + \frac{L\alpha_k^2}{2}\|f'(x_k)\|_2^2 =$$

$$f(x_k) - \left(\alpha_k - \frac{L\alpha_k^2}{2}\right)\|f'(x_k)\|_2^2$$

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

# Сходимость к стационарной точке

$$f(x_{k+1}) \leq f(x_k) + \langle f'(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_2^2 =$$

$$f(x_k) - \alpha_k \|f'(x_k)\|_2^2 + \frac{L\alpha_k^2}{2}\|f'(x_k)\|_2^2 =$$

$$f(x_k) - \left(\alpha_k - \frac{L\alpha_k^2}{2}\right)\|f'(x_k)\|_2^2$$

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

# Сходимость к стационарной точке

$$f(x_{k+1}) \leq f(x_k) + \langle f'(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_2^2 =$$

$$f(x_k) - \alpha_k \|f'(x_k)\|_2^2 + \frac{L\alpha_k^2}{2}\|f'(x_k)\|_2^2 =$$

$$f(x_k) - \left(\alpha_k - \frac{L\alpha_k^2}{2}\right)\|f'(x_k)\|_2^2$$

- Условие убывания: $\alpha_k - \frac{L\alpha_k^2}{2} > 0 \Rightarrow \alpha_k < \frac{2}{L}$
- $\alpha_k^* = \arg\max\limits_{\alpha_k} \left(\alpha_k - \frac{L\alpha_k^2}{2}\right) = \frac{1}{L}$
- $f(x_k) - f(x_{k+1}) \geq \frac{1}{2L}\|f'(x_k)\|_2^2$

# Сходимость к стационарной точке

$$f(x_{k+1}) \leq f(x_k) + \langle f'(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_2^2 =$$

$$f(x_k) - \alpha_k \|f'(x_k)\|_2^2 + \frac{L\alpha_k^2}{2}\|f'(x_k)\|_2^2 =$$

$$f(x_k) - \left(\alpha_k - \frac{L\alpha_k^2}{2}\right)\|f'(x_k)\|_2^2$$

- Условие убывания: $\alpha_k - \frac{L\alpha_k^2}{2} > 0 \Rightarrow \alpha_k < \frac{2}{L}$
- $\alpha_k^* = \arg\max\limits_{\alpha_k} \left(\alpha_k - \frac{L\alpha_k^2}{2}\right) = \frac{1}{L}$
- $f(x_k) - f(x_{k+1}) \geq \frac{1}{2L}\|f'(x_k)\|_2^2$
- $\frac{1}{2L}\sum\limits_{k=0}^{T}\|f'(x_k)\|_2^2 \leq f(x_0) - f(x_{T+1}) \leq f(x_0) - f^*$

# Сходимость к стационарной точке

$$f(x_{k+1}) \leq f(x_k) + \langle f'(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_2^2 =$$

$$f(x_k) - \alpha_k \|f'(x_k)\|_2^2 + \frac{L\alpha_k^2}{2}\|f'(x_k)\|_2^2 =$$

$$f(x_k) - \left(\alpha_k - \frac{L\alpha_k^2}{2}\right)\|f'(x_k)\|_2^2$$

- Условие убывания: $\alpha_k - \frac{L\alpha_k^2}{2} > 0 \Rightarrow \alpha_k < \frac{2}{L}$
- $\alpha_k^* = \arg\max\limits_{\alpha_k} \left(\alpha_k - \frac{L\alpha_k^2}{2}\right) = \frac{1}{L}$
- $f(x_k) - f(x_{k+1}) \geq \frac{1}{2L}\|f'(x_k)\|_2^2$
- $\frac{1}{2L}\sum\limits_{k=0}^{T}\|f'(x_k)\|_2^2 \leq f(x_0) - f(x_{T+1}) \leq f(x_0) - f^*$
- $f$ ограничена снизу, $\|f'(x_k)\|_2 \to 0,\ k \to \infty$

# Сходимость для выпуклой функции

## Теорема

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

$$f(x_{k+1}) - f^* \leq \frac{2L\|x - x_0\|_2^2}{k+4} = \mathcal{O}(1/k)$$

# Сходимость для сильно выпуклой функции

- Следствие сильной выпуклости

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

- Минимизируя обе части по $z$

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

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

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|f'(x_k)\|_2^2$$

- И наконец получим линейную сходимость

$$f(x_{k+1}) - f^* \leq \left(1 - \frac{1}{\kappa}\right)(f(x_k) - f^*)$$

# 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

$$f(x_k) - f^* \leq \frac{L}{2}\left(\frac{L-\mu}{L+\mu}\right)^{2k} \|x_0 - x^*\|_2^2$$

# 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 $f''(x)$.

**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 $f''(x)$.

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

- При $\kappa \gg 1$, $q^* \to 1 \Rightarrow$ ооооочень *медленная* сходимости.
  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 $f''(x)$.
**Q**: что такое число обусловленности матрицы?

- При $\kappa \gg 1$, $q^* \to 1 \Rightarrow$ оооочень *медленная* сходимости. Например при $\kappa = 100$: $q^* \approx 0.98$
- При $\kappa \simeq 1$, $q^* \to 0 \Rightarrow$ *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 $f''(x)$.

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

- При $\kappa \gg 1$, $q^* \to 1 \Rightarrow$ ооооочень *медленная* сходимости. Например при $\kappa = 100$: $q^* \approx 0.98$
- При $\kappa \simeq 1$, $q^* \to 0 \Rightarrow$ *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 существют такие «плохие» функции, для которых выполнены следующие оценки **снизу**

# Нижние оценки сходимости

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

- для выпуклых функций с Липшицевым градиентом

$$f(x_{k+1}) - f^* \geq \frac{3L\|x_0 - x^*\|_2^2}{32(k+1)^2}$$

# Нижние оценки сходимости

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

- для выпуклых функций с Липшицевым градиентом

$$f(x_{k+1}) - f^* \geq \frac{3L\|x_0 - x^*\|_2^2}{32(k+1)^2}$$

- для сильно выпуклых функций с Липшицевым градиентом

$$f(x_{k+1}) - f^* \geq \frac{\mu}{2}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2k}\|x_0 - x^*\|_2^2$$

# Нижние оценки сходимости

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

- для выпуклых функций с Липшицевым градиентом

$$f(x_{k+1}) - f^* \geq \frac{3L\|x_0 - x^*\|_2^2}{32(k+1)^2}$$

- для сильно выпуклых функций с Липшицевым градиентом

$$f(x_{k+1}) - f^* \geq \frac{\mu}{2}\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{2k}\|x_0 - x^*\|_2^2$$

Эти оценки справедливы для таких методов, что

$$x_{k+1} = x_0 + \mathbf{span}(f'(x_0), \ldots, f'(x_k))$$

# Оптимальные методы

Про методы, которые в той или иной степени достигают нижних оценок, будет в следующий раз

- метод сопряжённых градиентов
- метод тяжёлого щарика
- градиентный метод Нестерова

# Резюме

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