Вводная лекция Методы оптимизации

Александр Безносиков

Московский физико-технический институт

7 сентября 2023



Команда курса: лектор

- Безносиков Александр Николаевич
- кандидат физико-математических наук
- 10+ статей по оптимизации на А*/А конференции или в Q1 журналах в области ML; более 30 статей всего
- заведующий лабораторией МФТИ-Яндекс, исследователь в Иннополисе, Сколтехе, ВШЭ, Яндексе, ИППИ РАН, МВZUAI (ОАЭ)
- сайт: anbeznosikov.github.io
- почта: beznosikov.an@phystech.edu, anbeznosikov@gmail.com
- tg: @abeznosikov

Команда курса: семинаристы

- Бородич Екатерина Дмитриевна 6 статей (2 в А*/А и Q1)
 почта: borodich.ed@phystech.edu tg: @e borodich
- Кормаков Георгий Владимирович 2 статьи, большой опыт в DS почта: Georgiy.Kormakov@skoltech.ru tg: @gkormakov
- Корнилов Никита Максимович 2 статьи почта: kornilov.nm@phystech.edu tg: @Tugnir
- Курузов Илья Алексеевич 2 статьи, несколько промышленных проектов
 - почта: kuruzov.ia@phystech.edu tg: @EliasKur
- Лобанов Александр Владимирович 10 статей почта: lobbsasha@mail.ru

Команда курса: лектор и семинаристы

- Метелев Дмитрий Сергеевич 3 статьи (одна из них на А*) почта: metelev.ds@phystech.edu tg: @FlyBot239
- Моложавенко Александр Александрович
 почта: molozhavenko.aa@phystech.edu tg: @MetaMelon
- Чежегов Савелий Андреевич 4 статьи почта: chezhegov.sa@phystech.edu tg: @Savochak
- Юдин Никита Евгеньевич 2 статьи почта: iudin.ne@phystech.edu tg: @nikitayudin
- Ярмошик Демьян Валерьевич 8 статей почта: yarmoshik.dv@phystech.edu tg: @niquepolice

Команда курса: ассистенты

- Андреев Артем 1 статья
 почта: andreev.artem@phystech.edu tg: @artyomandreyev
- Былинкин Дмитрий 1 статья почта: bylinkin.da@phystech.edu tg: @lxstsvund
- Богданов Александр 1 статья
 почта: bogdanov.ai@phystech.edu tg: @d0dos
- Рябинин Евгений почта: riabinin.ev@phystech.edu tg: @REV757
- Ткаченко Светлана 1 статья почта: tkachenko.s@phystech.edu tg: @Aikiseito
- Шестаков Александр 1 статья почта: shestakov.aa@phystech.edu tg: @brigada al

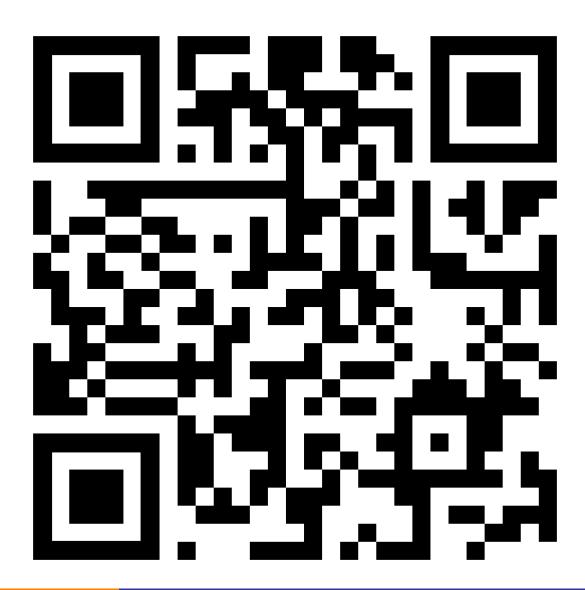
Telegram беседа



Диск с материалами



Форма обратной связи



Правила игры: система оценивания

Вид активности	Баллы
тесты на 10 минут в начале	3
каждого семинара по теме	
прошлой лекции и прошлого	
семинара	
домашнее задание (выдается	3
каждые 2 недели)	
контрольная работа в	3
середине семестра по темам	
лекций и семинаров	
коллоквиум в конце семестра	3
по темам лекций и семинаров	
разбор статьи	2
Итого:	14

Правила игры: система оценивания

- Для получения оценки удовлетворительно и выше необходимо, чтобы было выполнено хотя бы одно из следующих условий: оценка за коллоквиум ≥ 1 , оценка за $\mathsf{KP} \geq 1$.
- Для получения оценки хорошо и выше необходимо, чтобы было выполнено оба следующих условия: оценка за коллоквиум ≥ 1 , оценка за $\mathsf{KP} \geq 1$.

Правила игры: комментарии

- Ни один из видов активности не является 100% обязательным, но смотри дополнительные правила выше.
- Тесты будут проводиться в начале каждого семинара.
- Обновленные конспекты и ДЗ будут появляться на диске в выходные (ДЗ раз в 2 недели).
- При подозрении в списывании ДЗ баллы за конкретное домашнее задание обнуляются у всех авторов, подозреваемых в списывании (в том числе и у тех, кто дал списать).
- КР проводится в середине семестра на лекции. Программа контрольной соответствует программе курса, изученной в рамках лекций и семинаров за 8 недель.

Правила игры: комментарии

- Коллоквиум проходит в конце семестра во время последнего семинара и на зачетной неделе (на выбор). Программа коллоквиума соответствует всей программе курса, изученной в рамках лекций и семинаров. Принимают коллоквиум семинарист и несколько приглашенных преподавателей. Процедура коллоквиума соответствует процедуре проведения обычного устного экзамена на Физтехе с билетами, дополнительными вопросами/задачами и беседой в рамках курса.
- Разбор статьи предполагает полный разбор популярной статьи в области оптимизации. Необходимо разобраться в идее, в доказательствах, а также воспроизвести численные эксперименты, добавив к ним свои. Отчетность состоит из двух пунктов: 15 минутное выступление с презентацией на семинарском занятии, а также часовой доклад на семинаре лабораторий МФТИ-Яндекс и ММО. Заявки на разбор статьи принимаются до 31 юктября. ≥

• **Bonpoc**: какие приложения задач оптимизации знаете/уже встречали?

• **Boпpoc:** какие приложения задач оптимизации знаете/уже встречали? Приложений у оптимизации масса: машинное обучение, анализ данных, статистика, финансы, логистика, планирование, управление и многие другие.

- **Bonpoc**: какие приложения задач оптимизации знаете/уже встречали? Приложений у оптимизации масса: машинное обучение, анализ данных, статистика, финансы, логистика, планирование, управление и многие другие.
- Оптимизация часто выступает инструментом во многих прикладных задачах, люди пользуются готовыми решениями/пакетами/"черными ящиками", который способны решать задачи оптимизации. Цель курса как познакомиться с такими "черными ящиками", так и заглянуть внутрь и понять, что лежит внутри (в том числе с токи зрения теории).

Немного истории

$$Ax-b = 0 \quad \text{min} ||Ax-b||_2^2$$

- 1847: Коши и градиентный спуск для линейных систем
- 1950ые: линейное программирование (быстро перешло в нелинейное программирование), появление стохастических методов
- 1980ые: появление теории для общих задач.
- 2010ые: задачи оптимизации большого размера, теория стохастических методов



- ullet $Q\subseteq \mathbb{R}^d$ подмножество d-мерного пространства
- ullet $f:Q o \mathbb{R}$ некоторая функция, заданная на множестве Q
- В качестве & берётся (<) либо (=)
- ullet $g_i(x):Q o \mathbb{R},\ i=1,\ldots,m$ функции, задающие ограничения

$$\min_{\substack{g_i(x) \& 0, \\ i=1,\dots,m, \\ x \in Q}} f(x) \tag{1}$$

- ullet $Q\subseteq \mathbb{R}^d$ подмножество d-мерного пространства
- ullet $f:Q o \mathbb{R}$ некоторая функция, заданная на множестве Q
- В качестве & берётся ≤ либо =
- ullet $g_i(x):Q o \mathbb{R}$, $i=1,\ldots,m$ функции, задающие ограничения

Вопрос: что можно сказать про эту задач? сложная ли эта задача?

Задачи оптимизации. Первые наблюдения.

- ① В общем случае задачи оптимизации могут не иметь решения. Например, задача $\min_{x \in \mathbb{R}} x$ не имеет решения.
- 2 Задачи оптимизации часто нельзя решить аналитически.
- ③ Их сложность зависит от вида целевой функции f, множества Q и может зависеть от размерности x.

Задачи оптимизации. Первые наблюдения.

- ① В общем случае задачи оптимизации могут не иметь решения. Например, задача $\min_{x \in \mathbb{R}} x$ не имеет решения.
- 2 Задачи оптимизации часто нельзя решить аналитически.
- ③ Их сложность зависит от вида целевой функции f, множества Q и может зависеть от размерности x.

Если же задача оптимизации имеет решение, то на практике её обычно решают, вообще говоря, приближённо. Для этого применяются специальные алгоритмы, которые и называют методами оптимизации.

Методы оптимизации

- Нет смысла искать лучший метод для решения конкретной задачи. Например, лучший метод для решений задачи $\min_{x \in \mathbb{R}^d} \|x\|^2$ сходится за 1 итерацию: этот метод просто всегда выдаёт ответ $x^* = 0$. Очевидно, что для других задач такой метод не пригоден.
- Эффективность метода определяется для класса задач, т.к. обычно численные методы разрабатываются для *приближённого* решения множества однотипных задач.
- Метод разрабатывается для класса задач ⇒ метод не может иметь с самого начала полной информации о задаче. Вместо этого метод использует модель задачи, например, формулировку задачи, описание функциональных компонент, множества, на котором происходит оптимизация и т.д.

Вопрос: Какого рода вопросы хочется задавать оракулу?

Вопрос: Какого рода вопросы хочется задавать оракулу?

Вопрос: Какого рода вопросы хочется задавать оракулу?

Примеры оракулов

- Оракул нулевого порядка в запрашиваемой точке x возвращает значение целевой функции f(x).
- Оракул первого порядка в запрашиваемой точке возвращает значение функции f(x) и её градиент в данной точке $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n}\right)$.
- Оракул второго порядка в запрашиваемой точке возвращает значение и градиент функции $f(x), \nabla f(x)$, а также её гессиан в данной точке $\left(\nabla f^2(x)\right)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_i}$.

Входные данные: начальная точка x^0 (0 — верхний индекс), требуемая точность решения задачи $\varepsilon > 0$.

Входные данные: начальная точка x^0 (0 – верхний индекс), требуемая точность решения задачи $\varepsilon > 0$.

Настройка. Задать k=0 (счётчик итераций) и $l_{-1}=\varnothing$ (накапливаемая информационная модель решаемой задачи).

Входные данные: начальная точка x^0 (0 — верхний индекс), требуемая точность решения задачи $\varepsilon > 0$. Настройка. Задать k = 0 (счётчик итераций) и $I_{-1} = \varnothing$ (накапливаемая информационная модель решаемой задачи). Основной цикл

 $oldsymbol{0}$ Задать вопрос к оракулу \mathcal{O} в точке x^k .

Входные данные: начальная точка x^0 (0 — верхний индекс), требуемая точность решения задачи $\varepsilon > 0$.

Настройка. Задать k=0 (счётчик итераций) и $I_{-1}=\varnothing$ (накапливаемая информационная модель решаемой задачи).

Основной цикл

- $oldsymbol{0}$ Задать вопрос к оракулу \mathcal{O} в точке x^k .
- **2** Пересчитать информационную модель: $(I_k) = I_{k-1} \cup (x_k, \mathcal{O}(x^k))$.

Входные данные: начальная точка x^0 (0 — верхний индекс), требуемая точность решения задачи $\varepsilon > 0$.

Настройка. Задать k=0 (счётчик итераций) и $I_{-1}=\varnothing$ (накапливаемая информационная модель решаемой задачи).

Основной цикл

- $oldsymbol{0}$ Задать вопрос к оракулу \mathcal{O} в точке x^k .
- **2** Пересчитать информационную модель: $I_k = I_{k-1} \cup (x_k, \mathcal{O}(x^k))$.
- ③ Применить правило метода \mathcal{M} для получения новой точки x^{k+1} по модели I_k .

Входные данные: начальная точка x^0 (0 — верхний индекс), требуемая точность решения задачи $\varepsilon > 0$.

Настройка. Задать k=0 (счётчик итераций) и $I_{-1}=\varnothing$ (накапливаемая информационная модель решаемой задачи).

Основной цикл

- $oldsymbol{0}$ Задать вопрос к оракулу \mathcal{O} в точке x^k .
- **2** Пересчитать информационную модель: $I_k = I_{k-1} \cup (x_k, \mathcal{O}(x^k))$.
- (3) Применить правило метода ${\mathcal M}$ для получения новой точки x^{k+1} по модели I_k .
- 4 Проверить критерий остановки $\mathcal{T}_{arepsilon}$. Если критерий выполнен, то выдать ответ $ar{x}$, иначе положить k:=k+1 и вернуться на шаг 1.

Рассмотрим задачу оптимизации

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}),\tag{2}$$

где функция f(x) дифференцируема. Предположим, что в любой точке мы можем посчитать её градиент.

Рассмотрим задачу оптимизации

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}),\tag{2}$$

где функция f(x) дифференцируема. Предположим, что в любой точке мы можем посчитать её градиент.

Алгоритм 1 Градиентный спуск с постоянным размером шага

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

1: **for**
$$k = 0, 1, ..., K - 1$$
 do

2. Bычислить
$$\nabla f(x^k)$$

3:
$$x^{k+1} = x^k - \gamma \nabla f(x^k)$$
 \leftarrow

$$x^{k+1} = x^k - \gamma \nabla f(x^k) \quad \longleftarrow$$

4: end for

Выход: x^K

Рассмотрим задачу оптимизации

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}),\tag{2}$$

где функция f(x) дифференцируема. Предположим, что в любой точке мы можем посчитать её градиент.

Алгоритм 1 Градиентный спуск с постоянным размером шага

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

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

Выход: x^K

Вопрос: в чем Алгоритм 1 отличается от определения общей итеративной схемы?

Рассмотрим задачу оптимизации

$$\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}),\tag{2}$$

где функция f(x) дифференцируема. Предположим, что в любой точке мы можем посчитать её градиент.

Алгоритм 1 Градиентный спуск с постоянным размером шага

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

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

Вопрос: в чем Алгоритм 1 отличается от определения общей итеративной схемы? В итеративной схеме использовался $\mathcal{T}_{\varepsilon}$.

20 / 39

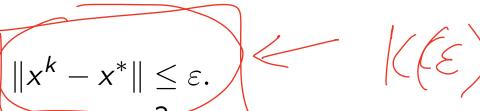
Критерии останова

• По аргументу:

$$\|x^k - x^*\| \le \varepsilon.$$

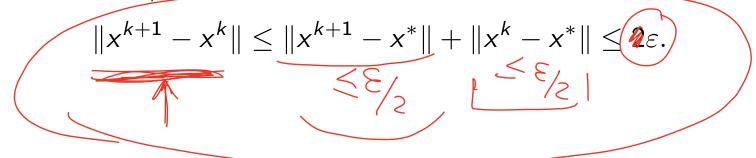
Вопрос: какие проблемы тут видим?

• По аргументу:



Вопрос: какие проблемы тут видим?

• x^* — неизвестно, но можно так



• По аргументу:

$$||x^k - (x^*)|| \le \varepsilon.$$

Вопрос: какие проблемы тут видим?

• x^* — неизвестно, но можно так

$$||x^{k+1} - x^k|| \le ||x^{k+1} - x^*|| + ||x^k - x^*|| \le 2\varepsilon.$$

• x^* – не уникально. Тогда можно поменять критерий



$$\chi_1 = \chi_2$$



• По аргументу:

$$||x^k - x^*|| \le \varepsilon.$$

Вопрос: какие проблемы тут видим?

• x^* — неизвестно, но можно так

$$||x^{k+1} - x^k|| \le ||x^{k+1} - x^*|| + ||x^k - x^*|| \le 2\varepsilon.$$

- x^* не уникально. Тогда можно поменять критерий
- По функции:

$$f(x^k) - f^* \le \varepsilon$$
.

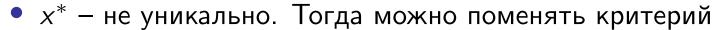
• По аргументу:

$$||x^k - x^*|| \le \varepsilon.$$

Вопрос: какие проблемы тут видим?

• x^* — неизвестно, но можно так

естно, но можно так
$$\|x^{k+1}-x^k\| \leq \|x^{k+1}-x^*\| + \|x^k-x^*\| \leq 2arepsilon.$$



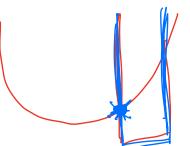
• По функции:

$$f(x^k) - f^* \le \varepsilon$$
.

• По норме градиента:

$$\|\nabla f(x^k)\| \leq \varepsilon.$$

Вопрос: когда такой критерий можно использовать?



• По аргументу:

$$||x^k - x^*|| \le \varepsilon.$$

Вопрос: какие проблемы тут видим?

• x^* — неизвестно, но можно так

$$||x^{k+1} - x^k|| \le ||x^{k+1} - x^*|| + ||x^k - x^*|| \le 2\varepsilon.$$

- x^* не уникально. Тогда можно поменять критерий
- По функции:

$$f(x^k) - f^* \le \varepsilon$$
.

• По норме градиента:

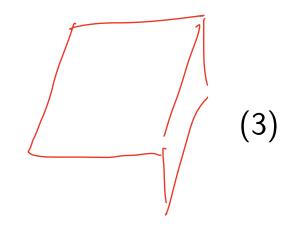
$$\|\nabla f(x^k)\| \leq \varepsilon.$$

Вопрос: когда такой критерий можно использовать? В безусловной оптимизации

Сложность методов оптимизации

- **Аналитическая/Оракульная сложность** число обращений к оракулу, необходимое для решения задачи с точностью ε .
- Арифметическая/Временна'я сложность общее число вычислений (включая работу оракула), необходимых для решения задачи с точностью ε .

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



- $B_d = \{x \in \mathbb{R}^d | 0 \le x_i \le 1, \quad i = 1, ..., d\}$
- ullet Функция f(x) является M-липшицевой на B_d относительно ℓ_∞ -нормы:

$$\forall x, y \mid f(x) - f(y) \mid \leq M \|x - y\|_{\infty} = M \max_{i=1,\dots,d} |x_i - y_i|. \tag{4}$$

Наблюдение

Множество B_d является ограниченным и замкнутым, т.е. компактом, а из липшицевости функции f следует и её непрерывность, поэтому задача (3) имеет решение, ибо непрерывная на компакте функция достигает своих минимального и максимального значений. Пусть $f_*^{\kappa} = \min_{x \in B_d} f(x)$.

- Класс методов. Для данной задачи рассмотрим методы нулевого порядка.

Рассмотрим один из самых простых способов решения этой задачи — метод равномерного перебора.

Алгоритм 2 Метод равномерного перебора

Вход: целочисленный параметр перебора $p \geq 1$

- 1: Сформировать $(p+1)^d$ точек вида $x_{(i_1,\ldots,i_d)}=\left(\frac{i_1}{p},\frac{i_2}{p},\ldots,\frac{i_d}{p}\right)^{\perp}$, где $(i_1,\ldots,i_d)\in\{0,1,\ldots,p\}^n$
- 2: Среди точек $x_{(i_1,...,i_d)}$ найти точку \bar{x} с наименьшим значением целевой функции f .

Выход: $\bar{x}, f(\bar{x})$

Теорема 1

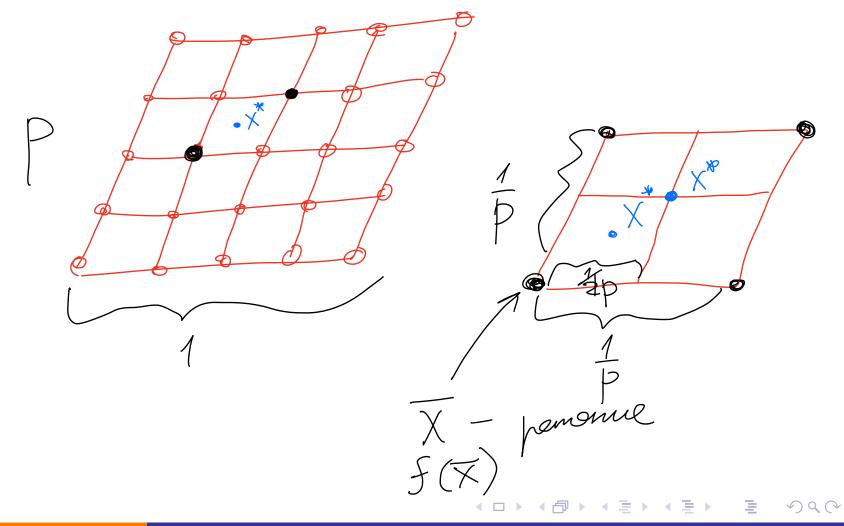
Алгоритм 2 с параметром p возвращает такую точку \bar{x} , что

$$f(\bar{x}) - f_* \le \frac{M}{2p},\tag{5}$$

откуда следует, что методу равномерного перебора нужно в худшем случае

$$\left(\left|\frac{M}{2\varepsilon}\right|+1\right)^d\tag{6}$$

обращений к оракулу, чтобы гарантировать $f(\bar{x}) - f_* \leq \varepsilon$.



Доказательство Теоремы 1

Пусть x_*^* — решение задачи (точка минимума функции f). Тогда в построенной «сетке» из точек найдётся такая точка $x_{(i_1,...,i_d)}$, что $x := x_{(i_1,...,i_d)} \le x^* \le x_{(i_1+1,...,i_d+1)} =: y$, где знак « \le » применяется покомпонентно.

Доказательство Теоремы 1

Пусть x_* — решение задачи (точка минимума функции f). Тогда в построенной «сетке» из точек найдётся такая точка $x_{(i_1,\ldots,i_d)}$, что $x:=x_{(i_1,\ldots,i_d)}\leq x^*\leq x_{(i_1+1,\ldots,i_d+1)}=:y$, где знак « \leq » применяется покомпонентно. Во-первых, $y_i-x_i=\frac{1}{p}$ и $x_i^*\in[x_i,y_i]$ для всех $i=1,\ldots,d$.

Доказательство Теоремы 1

Пусть x_* — решение задачи (точка минимума функции f). Тогда в построенной «сетке» из точек найдётся такая точка $x_{(i_1,\ldots,i_d)}$, что $x:=x_{(i_1,\ldots,i_d)}\leq x^*\leq x_{(i_1+1,\ldots,i_d+1)}=:y$, где знак « \leq » применяется покомпонентно. Во-первых, $y_i-x_i=\frac{1}{p}$ и $x_i^*\in[x_i,y_i]$ для всех $i=1,\ldots,d$. Кроме того, рассмотрим точки \hat{x} и \tilde{x} такие, что $\hat{x}=\frac{x+y}{2}$ и

$$ilde{x}_i = egin{cases} y_i, & ext{если } x_i^* \geq \hat{x}_i, \ x_i, & ext{иначе.} \end{cases}$$

Доказательство Теоремы 1 (продолжение)

Заметим, что $ilde{x}$ принадлежит «сетке» и $| ilde{x}_i - x_i^*| \leq rac{1}{2p}$, а значит,

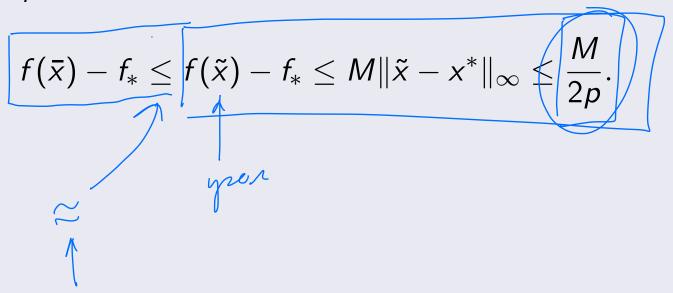
$$\|\tilde{x}-x^*\|_{\infty}\leq \frac{1}{2p}.$$

Доказательство Теоремы 1 (продолжение)

Заметим, что \tilde{x} принадлежит «сетке» и $|\tilde{x}_i - x_i^*| \leq \frac{1}{2p}$, а значит, $\|\tilde{x} - x^*\|_{\infty} \leq \frac{1}{2p}$. Поскольку $f(\bar{x}) \leq f(\tilde{x})$ (по определению), получаем

Доказательство Теоремы 1 (продолжение)

Заметим, что \tilde{x} принадлежит «сетке» и $|\tilde{x}_i - x_i^*| \leq \frac{1}{2p}$, а значит, $\|\tilde{x} - x^*\|_{\infty} \leq \frac{1}{2p}$. Поскольку $f(\bar{x}) \leq f(\tilde{x})$ (по определению), получаем



Доказательство Теоремы 1 (продолжение)

Заметим, что \tilde{x} принадлежит «сетке» и $|\tilde{x}_i - x_i^*| \leq \frac{1}{2p}$, а значит, $\|\tilde{x} - x^*\|_{\infty} \leq \frac{1}{2p}$. Поскольку $f(\bar{x}) \leq f(\tilde{x})$ (по определению), получаем

$$f(\overline{x}) - f_* \leq f(\overline{x}) - f_* \leq M \|\overline{x} - x^*\|_{\infty} \leq \frac{M}{2p}.$$

Выписанная выше оценка достигается методом равномерного перебора за $(p+1)^d$ обращений к оракулу. Следовательно, чтобы гарантировать $f(\bar{x})-f_*\leq \varepsilon$, необходимо взять $p=\left\lfloor\frac{M}{2\varepsilon}\right\rfloor+1$, т.е. метод сделает $\left(\left\lfloor\frac{M}{2\varepsilon}\right\rfloor+1\right)^d$ обращений к оракулу.

Вопрос: хороший результат получили или нет?

• Предположим M=2, d=13 И $\varepsilon=0.01$, то есть размерность задачи сравнительно небольшая и точность решения задачи не слишком высокая.

- Предположим M=2, d=13 И $\varepsilon=0.01$, то есть размерность задачи сравнительно небольшая и точность решения задачи не слишком высокая.
- Необходимое число обращений к оракулу:

$$\left(\left\lfloor\frac{M}{2\varepsilon}\right\rfloor+1\right)^d=102^{13}>10^{26}$$

- Предположим M=2, d=13 И $\varepsilon=0.01$, то есть размерность задачи сравнительно небольшая и точность решения задачи не слишком высокая.
- Необходимое число обращений к оракулу: $\left(\left\lfloor \frac{M}{2\varepsilon} \right\rfloor + 1\right)^d = 102^{13} > 10^{26}$.
- Производительность компьютера: 10^{11} арифметических операций в секунду.

- Предположим M=2, d=13 И $\varepsilon=0.01$, то есть размерность задачи сравнительно небольшая и точность решения задачи не слишком высокая.
- Необходимое число обращений к оракулу: $\left(\left\lfloor \frac{M}{2\varepsilon} \right\rfloor + 1\right)^d = 102^{13} > 10^{26}$.
- Производительность компьютера: 10^{11} арифметических операций в секунду.
- Общее время: хотя бы 10^{15} секунд, что больше 30 миллионов лет.

• **Bonpoc**: что мы сейчас получили? верхнюю или нижнюю оценку? что такое верхняя оценка?

- **Bonpoc:** что мы сейчас получили? верхнюю или нижнюю оценку? что такое верхняя оценка?
- Верхняя оценка гарантии нахождения решения определенным методом из рассматриваемого класса методов (например, методы с оракулом нулевого порядка) для любой задачи из класса (липшецева целевая функция).
- Нижняя оценка гарантия, что <u>существует</u> «плохая» задача из класса, что <u>любой</u> метод из класса методов будет сходиться не лучше утверждает нижняя оценка.
- Возникает вопрос: может мы плохо вывели верхнюю оценку (неидеальный анализ), может ли предложить другой метод из рассматриваемого класса, который будет находить приближённое решение существенно быстрее? На этот вопрос и даст ответ

Теорема 2

Пусть $\varepsilon < \frac{M}{2}$. Тогда аналитическая сложность описанного класса задач, т.е. аналитическая сложность метода на «худшей» для него задаче из данного класса, составляет по крайней мере

$$\mathcal{N} = \left(\left| \frac{M}{2\varepsilon} \right| \right)^d$$
 вызовов оракула. (7)

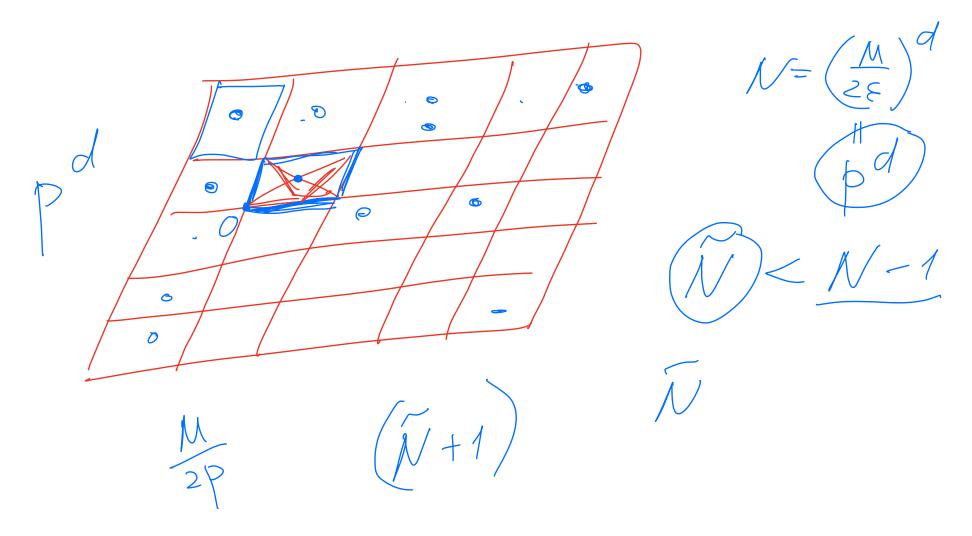


Схема доказательства Теоремы 2

Пусть $p = \lfloor \frac{M}{2\varepsilon} \rfloor$. Доказываем от противного: предположим, что существует такой метод, который решает задачу за $N < p^d$ обращений к оракулу, чтобы решить задачу с точностью ε (по функции).

Схема доказательства Теоремы 2

Пусть $p = \left \lfloor \frac{M}{2\varepsilon} \right \rfloor$. Доказываем от противного: предположим, что существует такой метод, который решает задачу за $N < p^d$ обращений к оракулу, чтобы решить задачу с точностью ε (по функции). Построим такую функцию, на которой метод не сможет найти ε -решение, при помощи сопротивляющегося оракула: на каждый запрос метода о значении функции в запрашиваемой точке, оракул будет возвращать 0.

Схема доказательства Теоремы 2

Пусть $p = \left \lfloor \frac{M}{2\varepsilon} \right \rfloor$. Доказываем от противного: предположим, что существует такой метод, который решает задачу за $N < p^d$ обращений к оракулу, чтобы решить задачу с точностью ε (по функции). Построим такую функцию, на которой метод не сможет найти ε -решение, при помощи сопротивляющегося оракула: на каждый запрос метода о значении функции в запрашиваемой точке, оракул будет возвращать 0. Тогда по принципу Дирихле найдётся такой «кубик» $B = \{x \mid \hat{x} \leq x \leq \hat{x} + \frac{1}{p}e\}$, где \hat{x} и $\hat{x} + \frac{1}{p}e$ — точки из «сетки» с шагом p, e — вектор из единиц.

Схема доказательства Теоремы 2 (продолжение)

Пусть
$$x^*$$
 — это центр «кубика» B , т.е. $x^* = \hat{x} + \frac{1}{2p}e$.

Схема доказательства Теоремы 2 (продолжение)

Пусть x^* — это центр «кубика» B, т.е. $x^* = \hat{x} + \frac{1}{2p}e$. Рассмотрим функцию $\bar{f}(x) = \min\{0, M\|x - x^*\|_{\infty} - \varepsilon\}$.

Схема доказательства Теоремы 2 (продолжение)

Пусть x^* — это центр «кубика» B, т.е. $x^* = \hat{x} + \frac{1}{2p}e$. Рассмотрим функцию $\bar{f}(x) = \min\{0, M\|x - x^*\|_{\infty} - \varepsilon\}$. Функция $\bar{f}(x)$ Липшицева с константой M относительно ℓ_{∞} -нормы и принимает своё минимальное значение $-\varepsilon$ в точке x^* .

Схема доказательства Теоремы 2 (продолжение)

Пусть x^* — это центр «кубика» B, т.е. $x^* = \hat{x} + \frac{1}{2p}e$. Рассмотрим функцию $\bar{f}(x) = \min\{0, M\|x - x^*\|_{\infty} - \varepsilon\}$. Функция $\bar{f}(x)$ Липшицева с константой M относительно ℓ_{∞} -нормы и принимает своё минимальное значение $-\varepsilon$ в точке x^* . Более того, функция $\bar{f}(x)$ отлична от нуля только внутри куба $B' = \{x \mid \|x - x^*\| \leq \frac{\varepsilon}{M}\}$, который лежит внутри куба B, т.к. $2p \leq \frac{M}{\varepsilon}$.

Схема доказательства Теоремы 2 (продолжение)

Пусть x^* — это центр «кубика» B, т.е. $x^* = \hat{x} + \frac{1}{2p}e$. Рассмотрим функцию $\bar{f}(x) = \min\{0, M\|x - x^*\|_{\infty} - \varepsilon\}$. Функция $\bar{f}(x)$ Липшицева с константой M относительно ℓ_{∞} -нормы и принимает своё минимальное значение $-\varepsilon$ в точке x^* . Более того, функция $\bar{f}(x)$ отлична от нуля только внутри куба $B' = \{x \mid \|x - x^*\| \leq \frac{\varepsilon}{M}\}$, который лежит внутри куба B, т.к. $2p \leq \frac{M}{\varepsilon}$. Следовательно, рассмотренный метод на данной функции не может найти ε -решение. Противоречие.

Схема доказательства Теоремы 2 (продолжение)

Пусть x^* — это центр «кубика» B, т.е. $x^* = \hat{x} + \frac{1}{2p}e$. Рассмотрим функцию $\bar{f}(x) = \min\{0, M\|x - x^*\|_{\infty} - \varepsilon\}$. Функция $\bar{f}(x)$ Липшицева с константой M относительно ℓ_{∞} -нормы и принимает своё минимальное значение $-\varepsilon$ в точке x^* . Более того, функция $\bar{f}(x)$ отлична от нуля только внутри куба $B' = \{x \mid \|x - x^*\| \leq \frac{\varepsilon}{M}\}$, который лежит внутри куба B, т.к. $2p \leq \frac{M}{\varepsilon}$. Следовательно, рассмотренный метод на данной функции не может найти ε -решение. Противоречие.

Итак, в указанном классе у любого метода оценки на скорость сходимости весьма пессимистичные. Возникает вопрос: какие свойства нужно потребовать от класса оптимизируемых функций, чтобы оценки стали более оптимистичными?

Схема доказательства Теоремы 2 (продолжение)

Пусть x^* — это центр «кубика» B, т.е. $x^* = \hat{x} + \frac{1}{2p}e$. Рассмотрим функцию $\bar{f}(x) = \min\{0, M\|x - x^*\|_{\infty} - \varepsilon\}$. Функция $\bar{f}(x)$ Липшицева с константой M относительно ℓ_{∞} -нормы и принимает своё минимальное значение $-\varepsilon$ в точке x^* . Более того, функция $\bar{f}(x)$ отлична от нуля только внутри куба $B' = \{x \mid \|x - x^*\| \leq \frac{\varepsilon}{M}\}$, который лежит внутри куба B, т.к. $2p \leq \frac{M}{\varepsilon}$. Следовательно, рассмотренный метод на данной функции не может найти ε -решение. Противоречие.

Итак, в указанном классе у любого метода оценки на скорость сходимости весьма пессимистичные. Возникает вопрос: какие свойства нужно потребовать от класса оптимизируемых функций, чтобы оценки стали более оптимистичными? Вернемся к этим вопросам на следующей лекции.

• Сублинейная:

$$\|x^k - x^*\| \le \frac{C}{k^{\alpha}},$$

1
poly(h)

1 K 1/2

где $\alpha > 0$, C > 0 – константа.

• Линейная:

$$\|x^k - x^*\| \le Cq^k$$

где $q \in (0;1)$, C > 0 – константа.

• Сверхлинейная:

$$\|x^k - x^*\| \le Cq^{k\theta},$$

где $q \in (0;1)$, p > 1, C > 0 – константа.

• Квадратичная:

$$||x^k - x^*|| \le Cq^{2^k}$$
 или $||x^k - x^*|| \le C||x^k - x^*||^2$,

где $q \in (0;1)$, C > 0 – константа.

