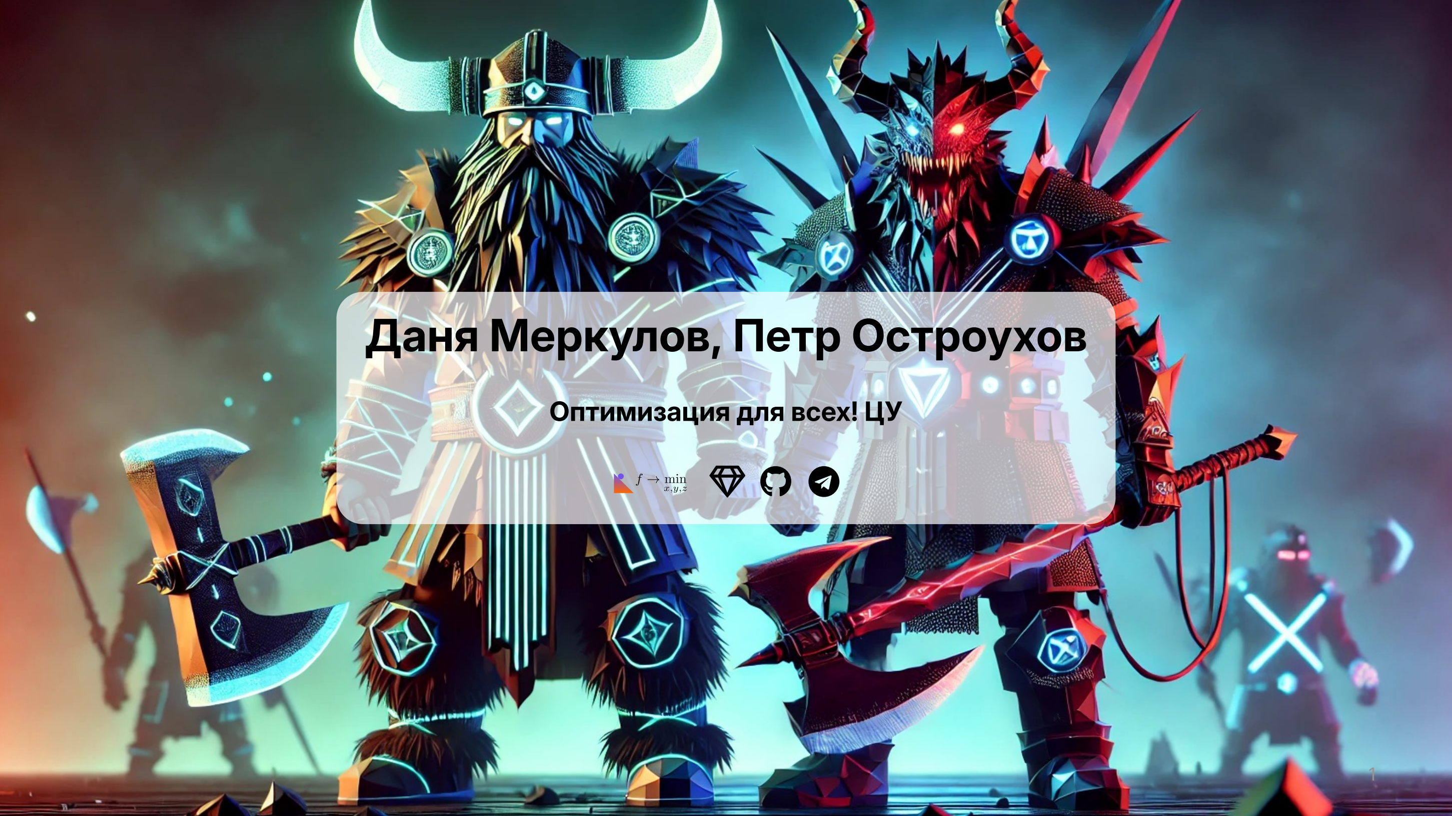
ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

# Субградиенты. Негладкие задачи

МЕТОДЫ ВЫПУКЛОЙ ОПТИМИЗАЦИИ

НЕДЕЛЯ 13

Даня Меркулов
Пётр Остроухов



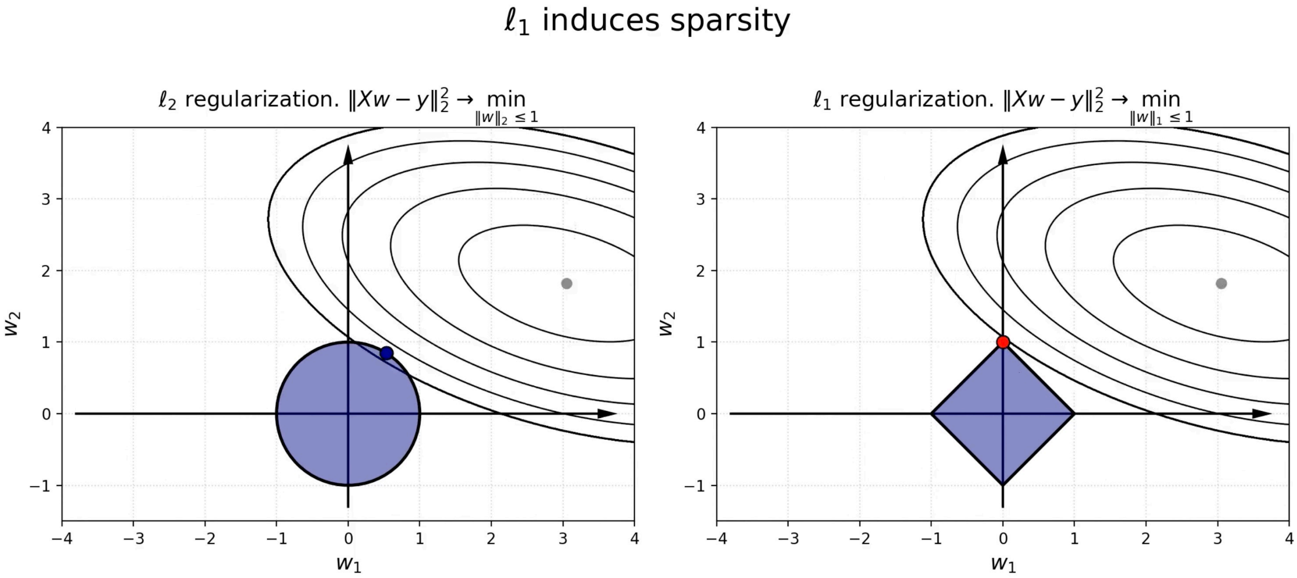
# Даня Меркулов, Петр Остроухов

## Оптимизация для всех! ЦУ

$f \to \min\limits_{x,y,z}$

# Негладкие задачи

# Задача наименьших квадратов с $\ell_1$-регуляризацией




@fminxyz

# Нормы не являются гладкими

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

Рассмотрим задачу выпуклой оптимизации. Будем считать $f(x)$ выпуклой, но не обязательно гладкой.

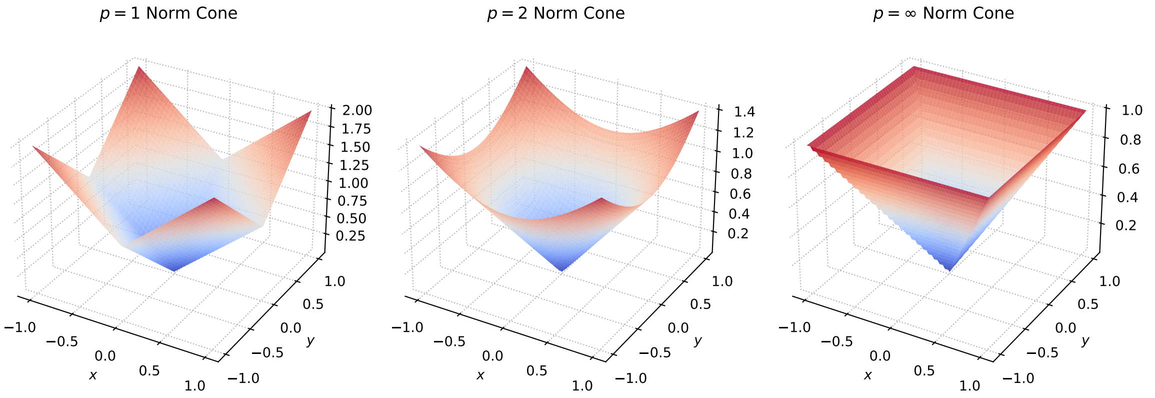



Рисунок 1. Конусы норм для разных $p$-норм не являются гладкими

# Пример Вульфа

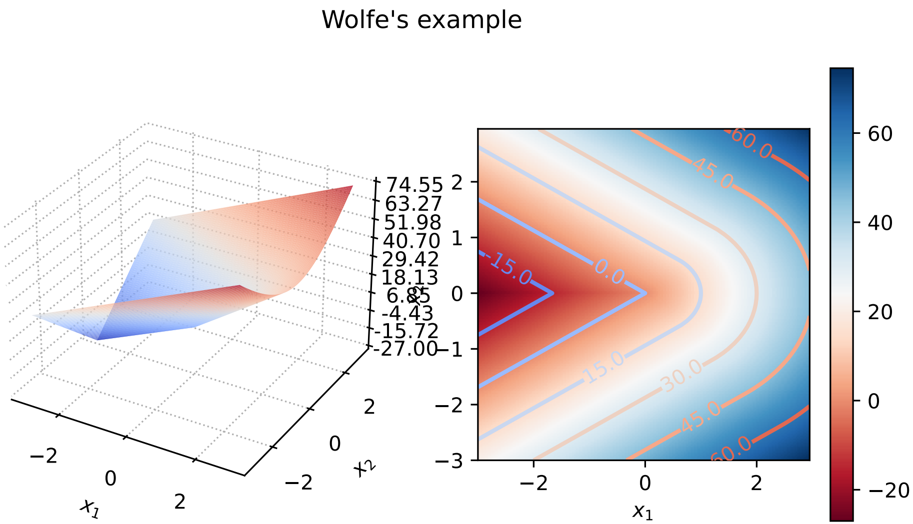



Рисунок 2. Пример Вульфа. Открыть в Colab

# Вычисление субградиента

# Линейная нижняя оценка для выпуклых функций

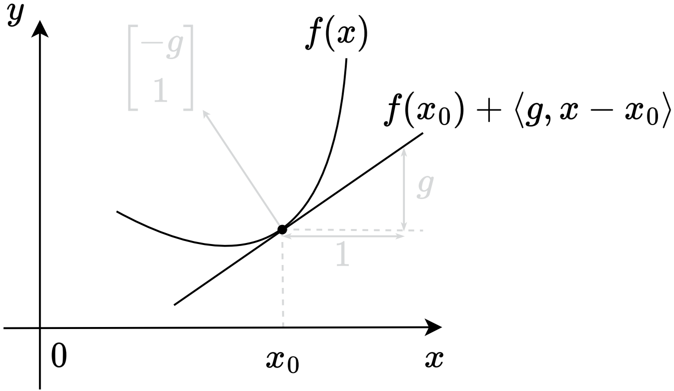



Рисунок 3. Аппроксимация Тейлора первого порядка служит глобальной нижней оценкой для выпуклой функции

Важное свойство выпуклой функции $f(x)$: для любой точки $x_0$ и для всех $x \in \mathrm{dom}\, f$ выполняется неравенство:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

# Линейная нижняя оценка для выпуклых функций

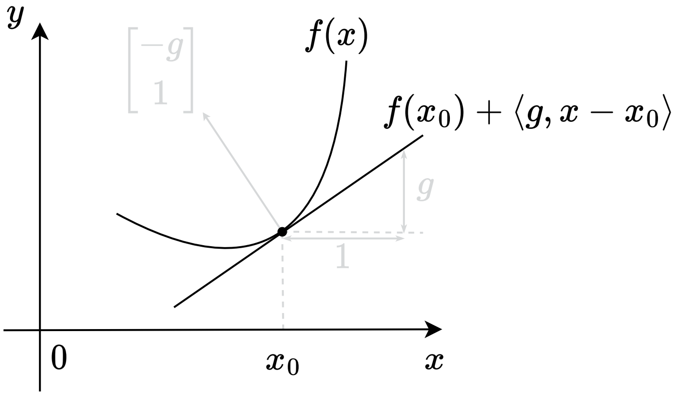



Рисунок 3. Аппроксимация Тейлора первого порядка служит глобальной нижней оценкой для выпуклой функции

Важное свойство выпуклой функции $f(x)$: для любой точки $x_0$ и для всех $x \in \text{dom } f$ выполняется неравенство:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

для некоторого вектора $g$, то есть касательная к графику функции является *глобальной* нижней оценкой для функции.

- Если $f(x)$ дифференцируема, то $g = \nabla f(x_0)$.

# Линейная нижняя оценка для выпуклых функций

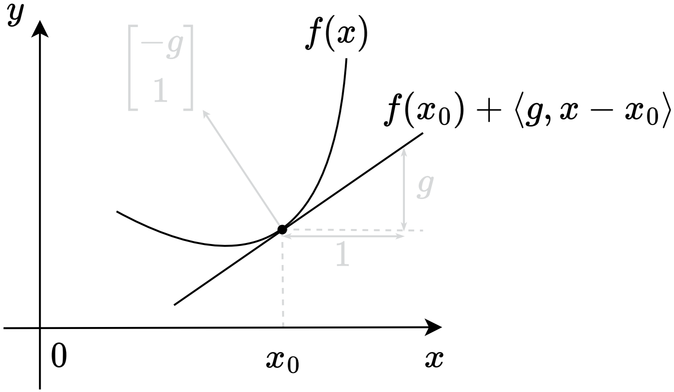



Рисунок 3. Аппроксимация Тейлора первого порядка служит глобальной нижней оценкой для выпуклой функции

Важное свойство выпуклой функции $f(x)$: для любой точки $x_0$ и для всех $x \in \text{dom } f$ выполняется неравенство:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

для некоторого вектора $g$, то есть касательная к графику функции является *глобальной* нижней оценкой для функции.

- Если $f(x)$ дифференцируема, то $g = \nabla f(x_0)$.
- Не все непрерывные выпуклые функции дифференцируемы.

# Линейная нижняя оценка для выпуклых функций

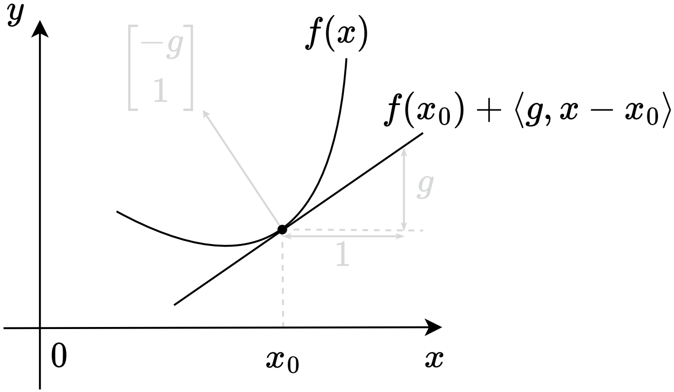



Рисунок 3. Аппроксимация Тейлора первого порядка служит глобальной нижней оценкой для выпуклой функции

Важное свойство выпуклой функции $f(x)$: для любой точки $x_0$ и для всех $x \in \text{dom}\, f$ выполняется неравенство:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

для некоторого вектора $g$, то есть касательная к графику функции является *глобальной* нижней оценкой для функции.

- Если $f(x)$ дифференцируема, то $g = \nabla f(x_0)$.
- Не все непрерывные выпуклые функции дифференцируемы.

# Линейная нижняя оценка для выпуклых функций

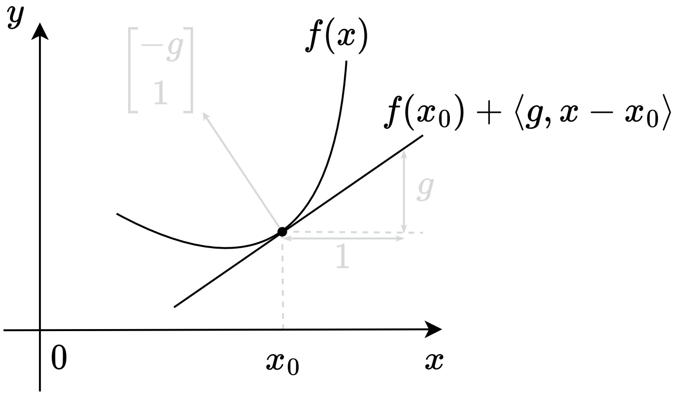



Рисунок 3. Аппроксимация Тейлора первого порядка служит глобальной нижней оценкой для выпуклой функции

Важное свойство выпуклой функции $f(x)$: для любой точки $x_0$ и для всех $x \in \text{dom } f$ выполняется неравенство:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

для некоторого вектора $g$, то есть касательная к графику функции является *глобальной* нижней оценкой для функции.

- Если $f(x)$ дифференцируема, то $g = \nabla f(x_0)$.
- Не все непрерывные выпуклые функции дифференцируемы.

Мы хотим сохранить это полезное свойство и для негладких функций.

# Субградиент и субдифференциал

Вектор $g$ называется **субградиентом** функции $f(x) : S \to \mathbb{R}$ в точке $x_0$, если $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

# Субградиент и субдифференциал

Вектор $g$ называется **субградиентом** функции $f(x) : S \to \mathbb{R}$ в точке $x_0$, если $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

Множество всех субградиентов функции $f(x)$ в точке $x_0$ называется **субдифференциалом** функции $f$ в точке $x_0$ и обозначается $\partial f(x_0)$.

# Субградиент и субдифференциал

Вектор $g$ называется **субградиентом** функции $f(x) : S \to \mathbb{R}$ в точке $x_0$, если $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

Множество всех субградиентов функции $f(x)$ в точке $x_0$ называется **субдифференциалом** функции $f$ в точке $x_0$ и обозначается $\partial f(x_0)$.

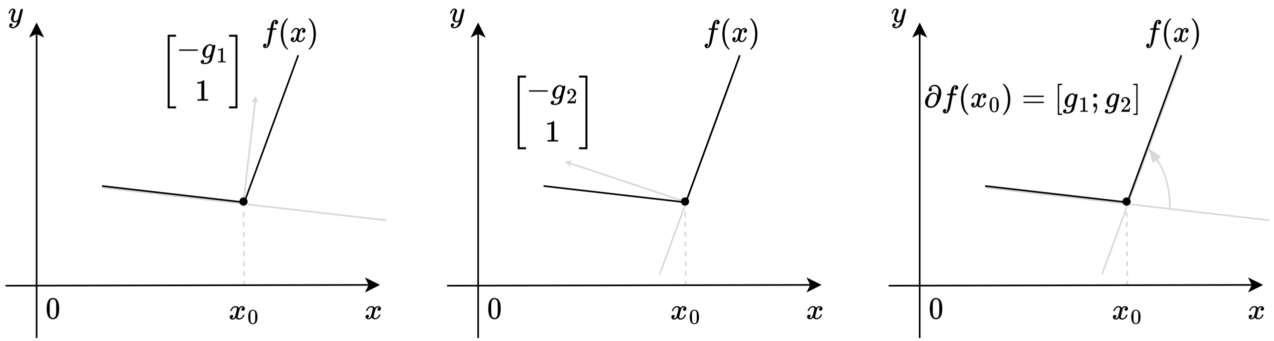


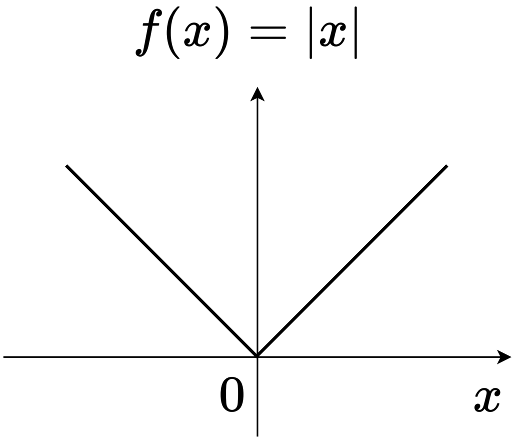

Рисунок 4. Субдифференциал — это множество всех возможных субградиентов

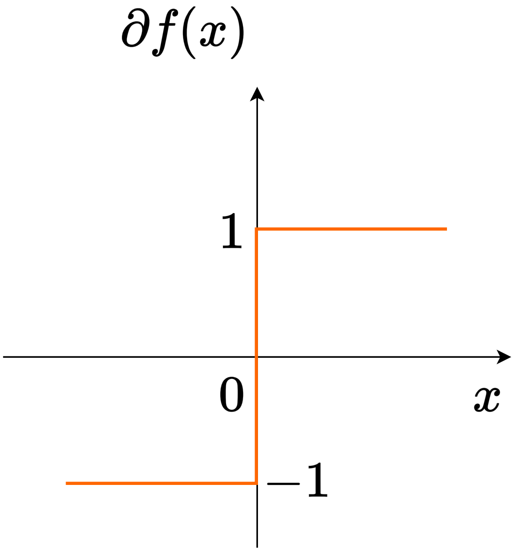
# Субградиент и субдифференциал

Найдите $\partial f(x)$, если $f(x) = |x|$

# Субградиент и субдифференциал

Найдите $\partial f(x)$, если $f(x) = |x|$



# Свойства субдифференциала

- Если $x_0 \in \text{ri}(S)$, то $\partial f(x_0)$ является выпуклым компактным множеством.

# Свойства субдифференциала

- Если $x_0 \in \text{ri}(S)$, то $\partial f(x_0)$ является выпуклым компактным множеством.
- Выпуклая функция $f(x)$ дифференцируема в точке $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.

# Свойства субдифференциала

- Если $x_0 \in \text{ri}(S)$, то $\partial f(x_0)$ является выпуклым компактным множеством.
- Выпуклая функция $f(x)$ дифференцируема в точке $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- Если $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, то $f(x)$ выпукла на $S$.

# Свойства субдифференциала

- Если $x_0 \in \mathrm{ri}(S)$, то $\partial f(x_0)$ является выпуклым компактным множеством.
- Выпуклая функция $f(x)$ дифференцируема в точке $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- Если $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, то $f(x)$ выпукла на $S$.

# Свойства субдифференциала

- Если $x_0 \in \text{ri}(S)$, то $\partial f(x_0)$ является выпуклым компактным множеством.
- Выпуклая функция $f(x)$ дифференцируема в точке $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- Если $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, то $f(x)$ выпукла на $S$.

**Субдифференциал дифференцируемой функции**

Пусть $f : S \to \mathbb{R}$ — функция, определенная на множестве $S$ в евклидовом пространстве $\mathbb{R}^n$. Если $x_0 \in \text{ri}(S)$ и $f$ дифференцируема в точке $x_0$, то либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$. Более того, если функция $f$ выпукла, то первая ситуация невозможна.

# Свойства субдифференциала

- Если $x_0 \in \text{ri}(S)$, то $\partial f(x_0)$ является выпуклым компактным множеством.
- Выпуклая функция $f(x)$ дифференцируема в точке $x_0 \Rightarrow \partial f(x_0) = \{\nabla f(x_0)\}$.
- Если $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, то $f(x)$ выпукла на $S$.

**Субдифференциал дифференцируемой функции**

Пусть $f : S \to \mathbb{R}$ — функция, определенная на множестве $S$ в евклидовом пространстве $\mathbb{R}^n$. Если $x_0 \in \text{ri}(S)$ и $f$ дифференцируема в точке $x_0$, то либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$. Более того, если функция $f$ выпукла, то первая ситуация невозможна.

# Вычисление субдифференциалов

**Теорема Моро — Рокафеллара (субдифференциал линейной комбинации)**

Пусть $f_i(x)$ — выпуклые функции на выпуклых множествах $S_i, \;\; i = 1, \ldots, n$. Тогда если $\bigcap\limits_{i=1}^{n} \mathrm{ri}(S_i) \neq \emptyset$, то функция $f(x) = \sum\limits_{i=1}^{n} a_i f_i(x), \;\; a_i > 0$ имеет субдифференциал $\partial_S f(x)$ на множестве $S = \bigcap\limits_{i=1}^{n} S_i$ и

$$
\partial_S f(x) = \sum_{i=1}^{n} a_i \partial_{S_i} f_i(x).
$$

# Вычисление субдифференциалов

**Теорема Моро — Рокафеллара (субдифференциал линейной комбинации)**

Пусть $f_i(x)$ — выпуклые функции на выпуклых множествах $S_i, \ i = 1, \dots, n$. Тогда если $\bigcap\limits_{i=1}^{n} \text{ri}(S_i) \neq \emptyset$, то функция $f(x) = \sum\limits_{i=1}^{n} a_i f_i(x), \ a_i > 0$ имеет субдифференциал $\partial_S f(x)$ на множестве $S = \bigcap\limits_{i=1}^{n} S_i$ и

$$
\partial_S f(x) = \sum_{i=1}^{n} a_i \partial_{S_i} f_i(x).
$$

**Теорема Дубовицкого — Милютина (субдифференциал поточечного максимума)**

Пусть $f_i(x)$ — выпуклые функции на открытом выпуклом множестве $S \subseteq \mathbb{R}^n, \ x_0 \in S$, и поточечный максимум определяется как $f(x) = \max\limits_i f_i(x)$. Тогда:

$$
\partial_S f(x_0) = \textbf{conv} \left\{ \bigcup_{i \in I(x_0)} \partial_S f_i(x_0) \right\},
$$

$$
I(x) = \{ i \in \{1, \dots, m\} : f_i(x) = f(x) \}
$$

# Вычисление субдифференциалов

- $\partial(\alpha f)(x) = \alpha \partial f(x)$ для $\alpha \geq 0$

# Вычисление субдифференциалов

- $\partial(\alpha f)(x) = \alpha \partial f(x)$ для $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, $f_i$ — выпуклые функции

# Вычисление субдифференциалов

- $\partial(\alpha f)(x) = \alpha \partial f(x)$ для $\alpha \geq 0$
- $\partial(\sum f_i)(x) = \sum \partial f_i(x)$, $f_i$ — выпуклые функции
- $\partial(f(Ax+b))(x) = A^T \partial f(Ax+b)$, $f$ — выпуклая функция

# Субградиентный метод

# Алгоритм

Вектор $g$ называется **субградиентом** функции $f(x) : S \to \mathbb{R}$ в точке $x_0$, если $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

# Алгоритм

Вектор $g$ называется **субградиентом** функции $f(x) : S \to \mathbb{R}$ в точке $x_0$, если $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

Идея метода: заменить градиент $\nabla f(x_k)$ на произвольный субградиент $g_k \in \partial f(x_k)$:

$$x_{k+1} = x_k - \alpha_k g_k,$$

где $g_k$ — произвольный субградиент функции $f(x)$ в точке $x_k$.

# Алгоритм

Вектор $g$ называется **субградиентом** функции $f(x) : S \to \mathbb{R}$ в точке $x_0$, если $\forall x \in S$:

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

Идея метода: заменить градиент $\nabla f(x_k)$ на произвольный субградиент $g_k \in \partial f(x_k)$:

$$x_{k+1} = x_k - \alpha_k g_k,$$

где $g_k$ — произвольный субградиент функции $f(x)$ в точке $x_k$.

Субградиентный метод не является методом спуска: значение функции может расти ($f(x_{k+1}) > f(x_k)$), так как антисубградиент не обязательно является направлением убывания.

Поэтому в качестве приближения решения берут лучшее найденное значение:

$$f_k^{\text{best}} = \min_{i=1,\dots,k} f(x_i).$$

# Сходимость: выпуклый случай

Theorem

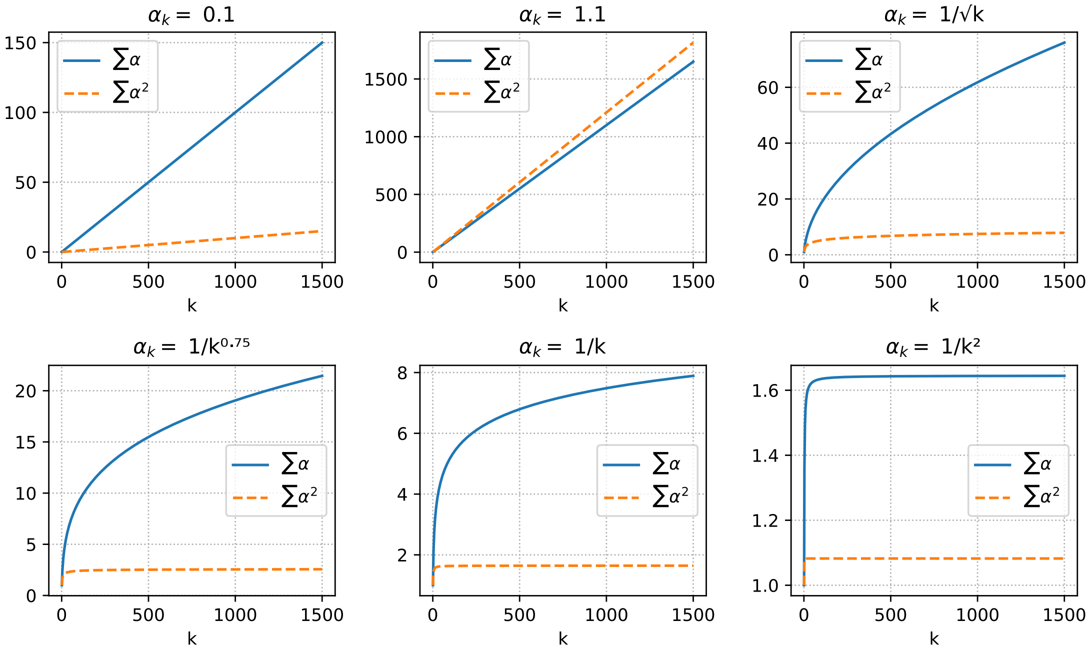
Пусть $f$ — выпуклая функция, $x^*$ — точка минимума. Предположим, что $\|g_k\| \leq G$ для всех $k$. Тогда:

$$f_T^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum\limits_{k=0}^{T-1} \alpha_k^2}{2 \sum\limits_{k=0}^{T-1} \alpha_k}$$

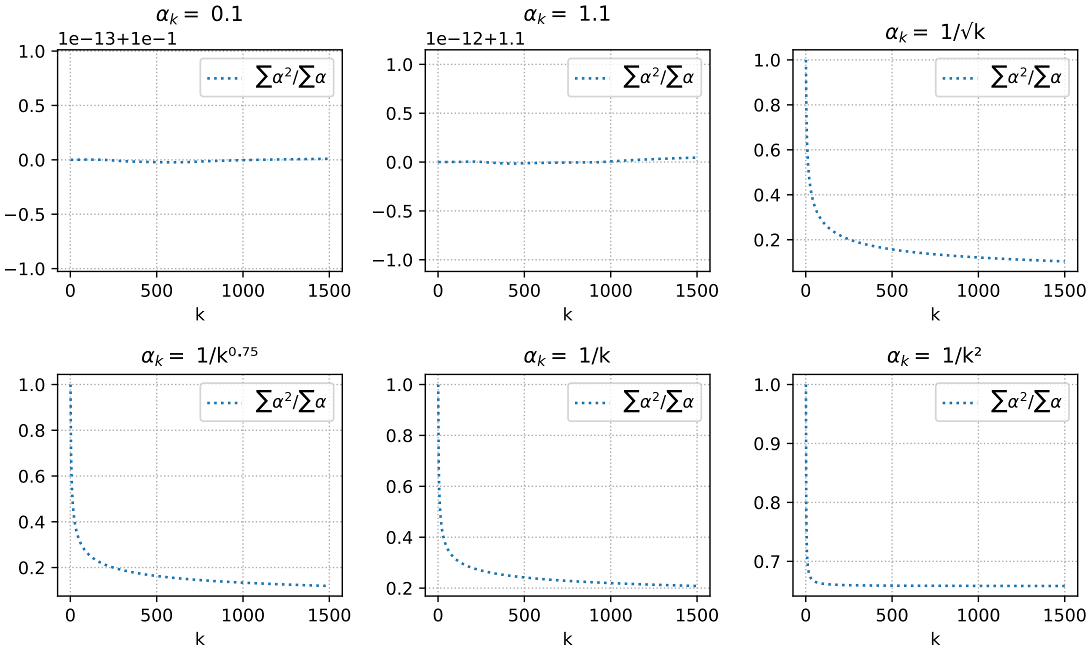
где $R = \|x_0 - x^*\|_2$.
Метод сходится, если:

$$\sum_{k=0}^{T-1} \alpha_k^2 < \infty, \quad \sum_{k=0}^{T-1} \alpha_k = \infty$$

# Различные стратегии выбора шага



# Различные стратегии выбора шага



# Сходимость: выпуклый случай, постоянный шаг (1)

**Theorem**

Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для постоянного шага $\alpha$, субградиентный метод удовлетворяет

$$f_k^{\text{best}} - f(x^*) \leq \frac{R^2}{2\alpha k} + \frac{\alpha}{2}G^2$$

- Заметим, что с любым постоянным шагом первое слагаемое правой части убывает, но второе остается постоянным.

# Сходимость: выпуклый случай, постоянный шаг (1)

**Theorem**

Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для постоянного шага $\alpha$, субградиентный метод удовлетворяет

$$f_k^{\text{best}} - f(x^*) \leq \frac{R^2}{2\alpha k} + \frac{\alpha}{2}G^2$$

- Заметим, что с любым постоянным шагом первое слагаемое правой части убывает, но второе остается постоянным.
- Замечание: существуют вариации метода, работающие и без предположения об ограниченности субградиентов (например, с нормировкой шага). См. [1] или [2].

[1] B. Polyak. Introduction to Optimization. Optimization Software, Inc., 1987.

[2] N. Shor. Minimization Methods for Non-differentiable Functions. Springer Series in Computational Mathematics. Springer, 1985.

# Сходимость: выпуклый случай, постоянный шаг (1)

**Theorem**

Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для постоянного шага $\alpha$, субградиентный метод удовлетворяет

$$
f_k^{\text{best}} - f(x^*) \leq \frac{R^2}{2\alpha k} + \frac{\alpha}{2} G^2
$$

- Заметим, что с любым постоянным шагом первое слагаемое правой части убывает, но второе остается постоянным.
- Замечание: существуют вариации метода, работающие и без предположения об ограниченности субградиентов (например, с нормировкой шага). См. [1] или [2].
- Найдем оптимальный шаг $\alpha$, который минимизирует правую часть неравенства.

[1] B. Polyak. Introduction to Optimization. Optimization Software, Inc., 1987.

[2] N. Shor. Minimization Methods for Non-di erentiable Functions. Springer Series in Computational Mathematics. Springer, 1985.

# Сходимость: выпуклый случай, постоянный шаг (2)

**Theorem**

Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для постоянного шага $\alpha = \frac{R}{G}\sqrt{\frac{1}{k}}$, субградиентный метод удовлетворяет

$$
f_k^{\text{best}} - f(x^*) \leq \frac{GR}{\sqrt{k}}
$$

- Эта версия требует знания числа итераций заранее, что не всегда практично.

# Сходимость: выпуклый случай, постоянный шаг (2)

**Theorem**

Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для постоянного шага $\alpha = \frac{R}{G}\sqrt{\frac{1}{k}}$, субградиентный метод удовлетворяет

$$f_k^{\text{best}} - f(x^*) \leq \frac{GR}{\sqrt{k}}$$

- Эта версия требует знания числа итераций заранее, что не всегда практично.
- Любопытно, что, если вы будете искать оптимальный шаг на каждой итерации: $\alpha_0, \ldots, \alpha_{k-1}$, то получите ту же оценку.

# Сходимость: выпуклый случай, постоянный шаг (2)

**Theorem**

Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для постоянного шага $\alpha = \frac{R}{G}\sqrt{\frac{1}{k}}$, субградиентный метод удовлетворяет

$$f_k^{\text{best}} - f(x^*) \leq \frac{GR}{\sqrt{k}}$$

- Эта версия требует знания числа итераций заранее, что не всегда практично.
- Любопытно, что, если вы будете искать оптимальный шаг на каждой итерации: $\alpha_0, \ldots, \alpha_{k-1}$, то получите ту же оценку.
- Это связано с симметрией и выпуклостью правой части относительно $\alpha_i$.

# Сходимость: выпуклый случай, постоянный шаг (3)

> **Theorem**
>
> Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для постоянного шага $\gamma = \alpha_k \|g_k\|_2$, т.е. $\alpha_k = \frac{\gamma}{\|g_k\|_2}$, субградиентный метод удовлетворяет
>
> $$f_k^{\text{best}} - f(x^*) \leq \frac{GR^2}{2\gamma k} + \frac{G\gamma}{2}$$

- Критерий остановки: норма субградиента не подходит (пример $f(x) = |x|$). Обычно используют фиксированное число итераций.

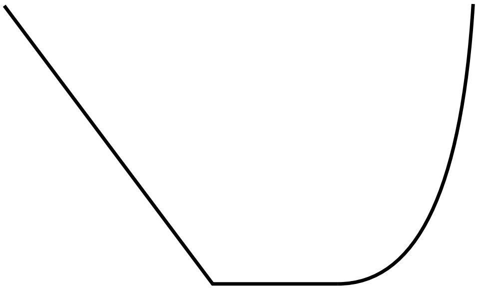
# Сходимость: выпуклый случай, практическая стратегия

Theorem

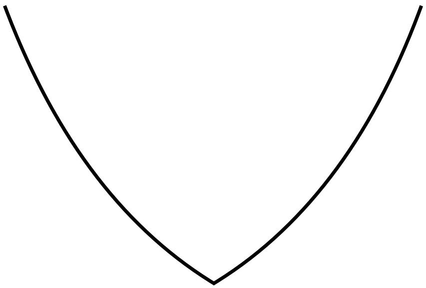
Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для убывающей стратегии шага $\alpha_k = \frac{R}{G\sqrt{k+1}}$, субградиентный метод удовлетворяет

$$f_k^{\text{best}} - f(x^*) \leq \frac{GR(2 + \ln k)}{4\sqrt{k+1}}$$

# Разные комбинации (сильной) выпуклости и (не)гладкости

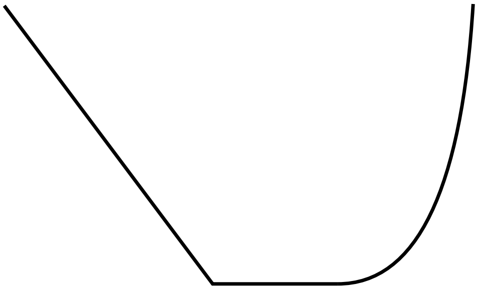


Негладкая
Выпуклая



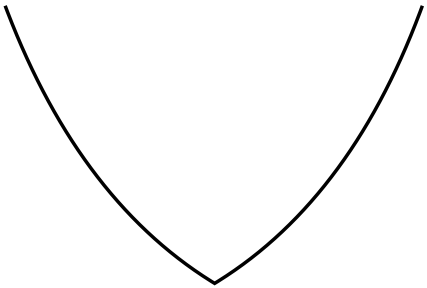
Негладкая
$\mu$ - сильно выпуклая

# Разные комбинации (сильной) выпуклости и (не)гладкости



Негладкая
Выпуклая

$$\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$$



Негладкая
$\mu$ - сильно выпуклая

$$\mathcal{O}\left(\frac{1}{k}\right)$$

# Сильно выпуклая негладкая функция

Theorem

Пусть $f$ — $\mu$-сильно выпуклая на выпуклом множестве и $x, y$ — произвольные точки. Тогда для любого $g \in \partial f(x)$,

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

# Сходимость: сильно выпуклый случай

**Theorem**

Пусть $f$ — $\mu$-сильно выпуклая функция (возможно негладкая) с минимумом $x^*$. Предположим, что субградиенты ограничены на траектории: $\|g_k\| \leq G$. Используя шаг $\alpha_k = \frac{2}{\mu(k+1)}$, субградиентный метод гарантирует для $k > 0$ что:

$$
f_k^{\text{best}} - f(x^*) \leq \frac{2G^2}{\mu k}.
$$

# Summary. Субградиентный метод

| Тип задачи | Стратегия шага | Скорость сходимости | Сложность итераций |
|---|---|---|---|
| Выпуклые липшицевы функции | $\alpha \sim \dfrac{1}{\sqrt{k}}$ | $\mathcal{O}\left(\dfrac{1}{\sqrt{k}}\right)$ | $\mathcal{O}\left(\dfrac{1}{\varepsilon^2}\right)$ |
| Сильно выпуклые функции | $\alpha \sim \dfrac{1}{k}$ | $\mathcal{O}\left(\dfrac{1}{k}\right)$ | $\mathcal{O}\left(\dfrac{1}{\varepsilon}\right)$ |

# Нижние оценки

| выпуклые (негладкие) [3] | гладкие (невыпуклые)[4] | гладкие и выпуклые[5] | гладкие и сильно выпуклые (или PL)[1] |
|---|---|---|---|
| $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$ | $\mathcal{O}\left(\frac{1}{k^2}\right)$ | $\mathcal{O}\left(\frac{1}{k^2}\right)$ | $\mathcal{O}\left(\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k\right)$ |
| $k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$ | $k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right)$ | $k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right)$ | $k_\varepsilon \sim \mathcal{O}\left(\sqrt{\kappa}\log\frac{1}{\varepsilon}\right)$ |

[3] Nesterov, Lectures on Convex Optimization

[4] Carmon, Duchi, Hinder, Sidford, 2017

[5] Nemirovski, Yudin, 1979

# Численные эксперименты

$$f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \lambda\|x\|_1 \to \min_{x \in \mathbb{R}^n}, \qquad A \in \mathbb{R}^{m \times n}, \quad \lambda\left(\tfrac{1}{m}A^TA\right) \in [\mu; L].$$

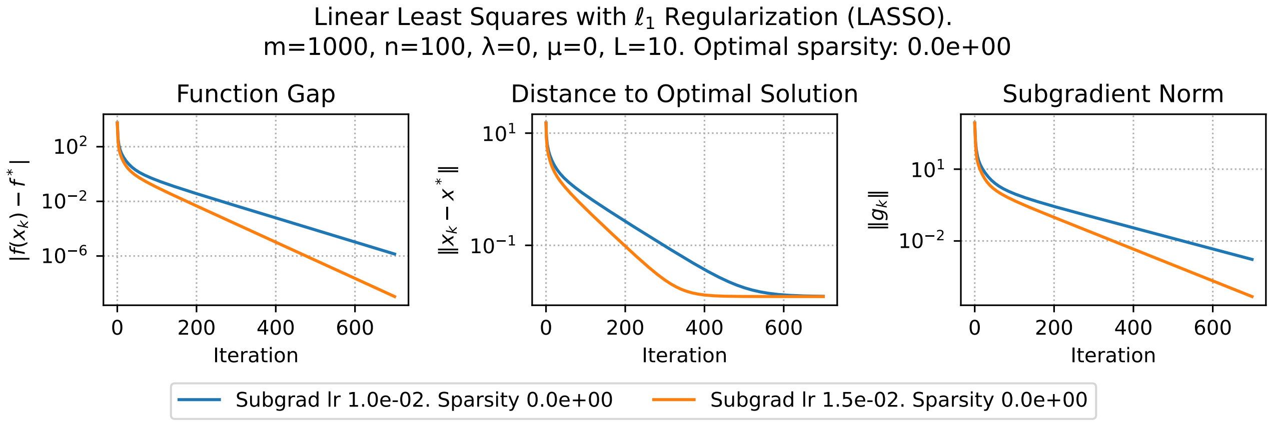



Рисунок 6. Гладкий выпуклый случай. Сублинейная сходимость, не сходится в области определения

# Численные эксперименты

$$f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \lambda\|x\|_1 \to \min_{x \in \mathbb{R}^n}, \qquad A \in \mathbb{R}^{m \times n}, \quad \lambda\left(\tfrac{1}{m}A^T A\right) \in [\mu; L].$$

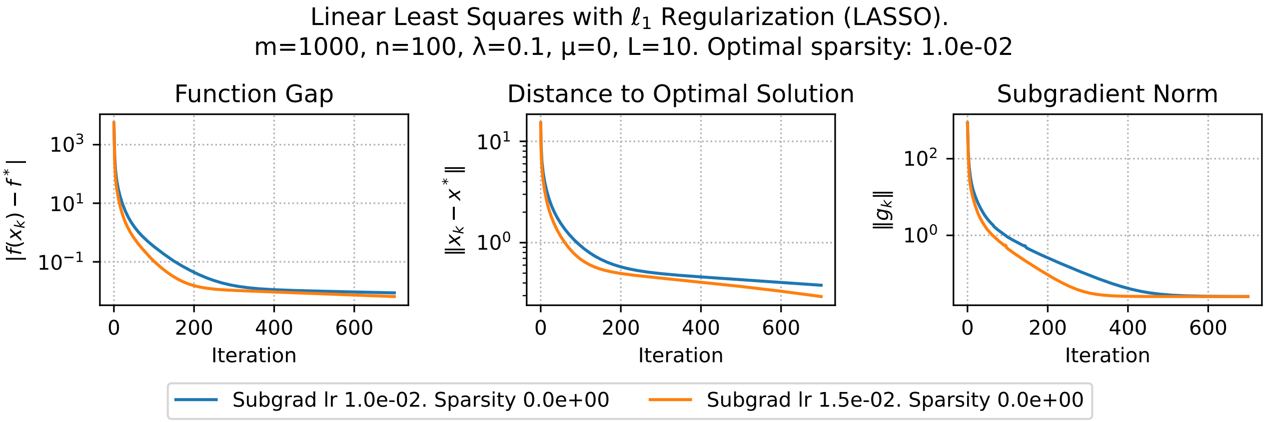



Рисунок 7. Негладкий выпуклый случай. Маленькое значение $\lambda$ приводит к негладкости. Не сходится с постоянным шагом

# Численные эксперименты

$$
f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \lambda\|x\|_1 \to \min_{x \in \mathbb{R}^n}, \qquad A \in \mathbb{R}^{m \times n}, \quad \lambda\left(\frac{1}{m}A^TA\right) \in [\mu; L].
$$

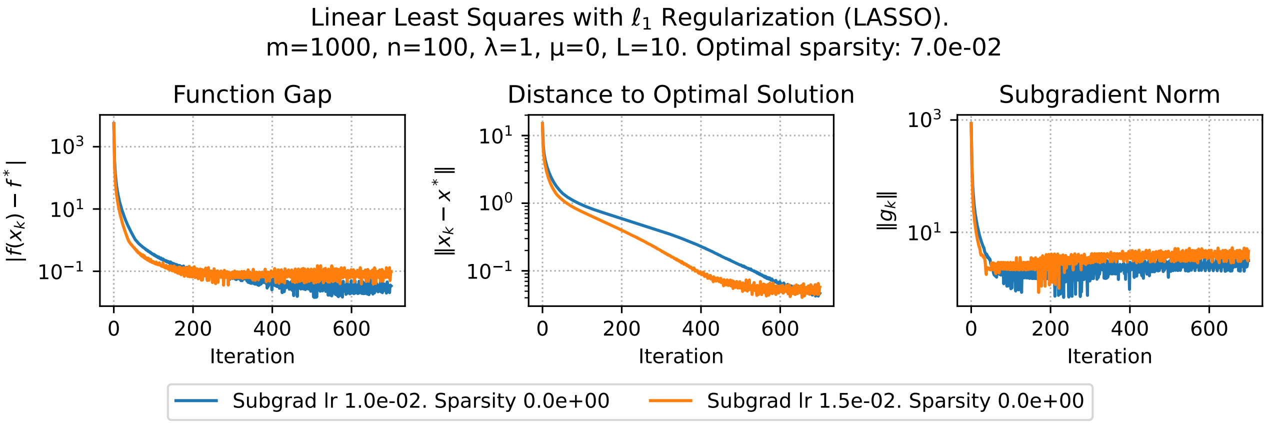



Рисунок 8. Негладкий выпуклый случай. При большем значении $\lambda$ проявляется немонотонность $f(x_k)$. Видно, что меньшая постоянная длина шага приводит к более низкому стационарному уровню функции.

# Численные эксперименты

$$f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \lambda\|x\|_1 \to \min_{x \in \mathbb{R}^n}, \qquad A \in \mathbb{R}^{m \times n}, \quad \lambda\left(\tfrac{1}{m}A^TA\right) \in [\mu; L].$$

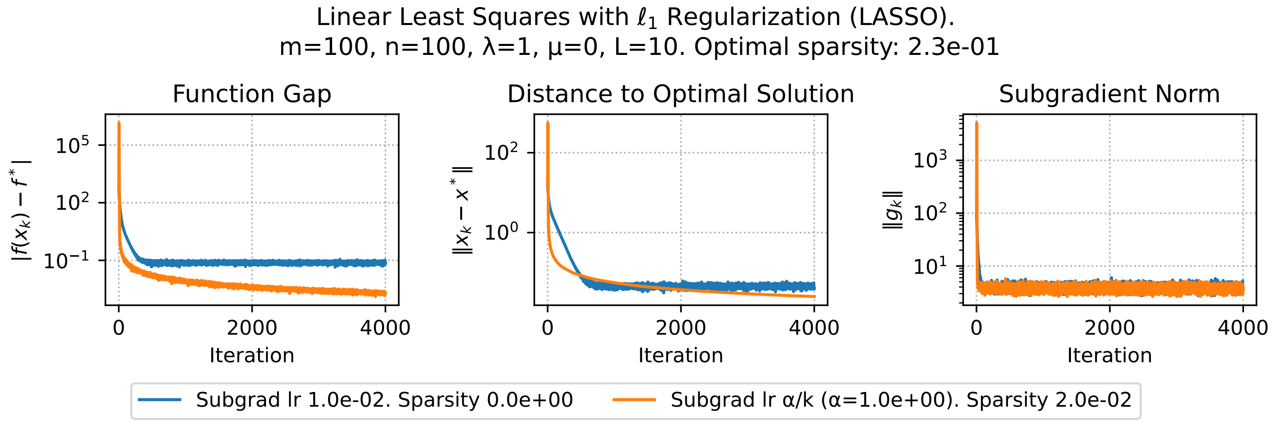



Рисунок 9. Негладкий выпуклый случай. Убывающая длина шага приводит к сходимости для $f_k^{\text{best}}$

# Численные эксперименты

$$f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \lambda\|x\|_1 \to \min_{x\in\mathbb{R}^n}, \qquad A \in \mathbb{R}^{m\times n}, \quad \lambda\left(\tfrac{1}{m}A^TA\right) \in [\mu; L].$$

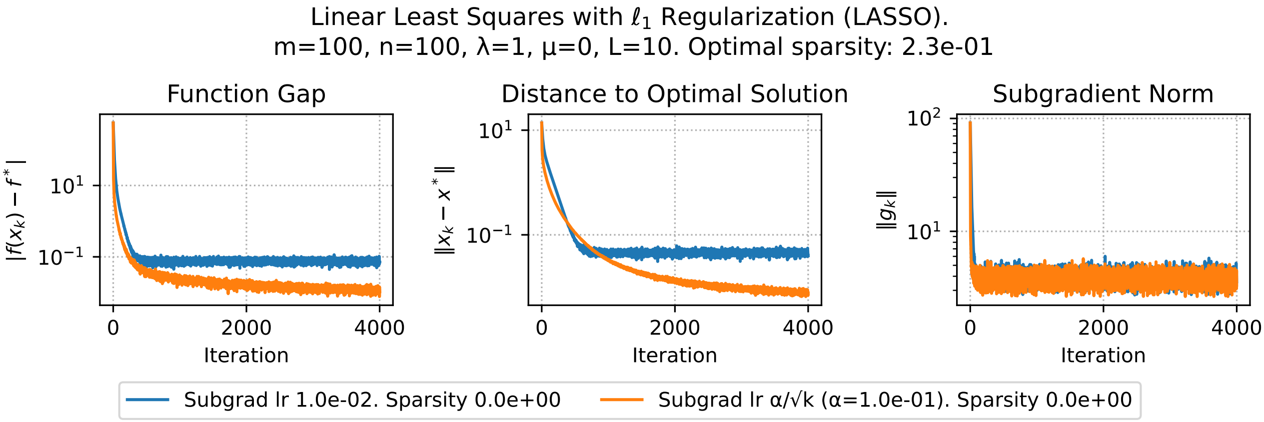



Рисунок 10. Негладкий выпуклый случай. $\frac{\alpha_0}{\sqrt{k}}$ шаг приводит к сходимости для $f_k^{\text{best}}$

# Численные эксперименты

$$f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \lambda\|x\|_1 \to \min_{x \in \mathbb{R}^n}, \qquad A \in \mathbb{R}^{m \times n}, \quad \lambda\left(\tfrac{1}{m}A^TA\right) \in [\mu; L].$$

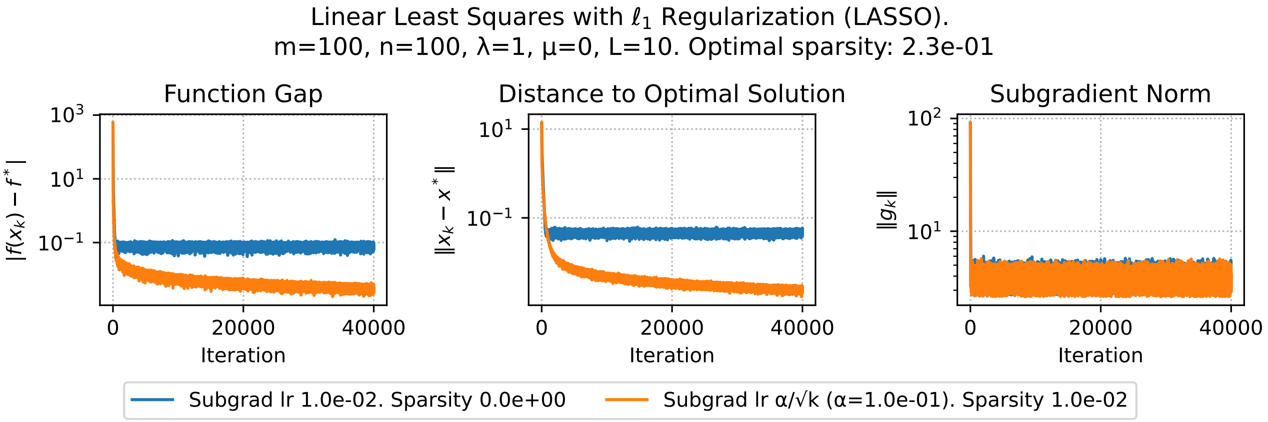



Рисунок 11. Негладкий выпуклый случай. $\frac{\alpha_0}{\sqrt{k}}$ шаг приводит к сходимости для $f_k^{\text{best}}$

# Численные эксперименты

$$f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \lambda\|x\|_1 \to \min_{x \in \mathbb{R}^n}, \qquad A \in \mathbb{R}^{m \times n}, \quad \lambda\left(\tfrac{1}{m}A^TA\right) \in [\mu; L].$$

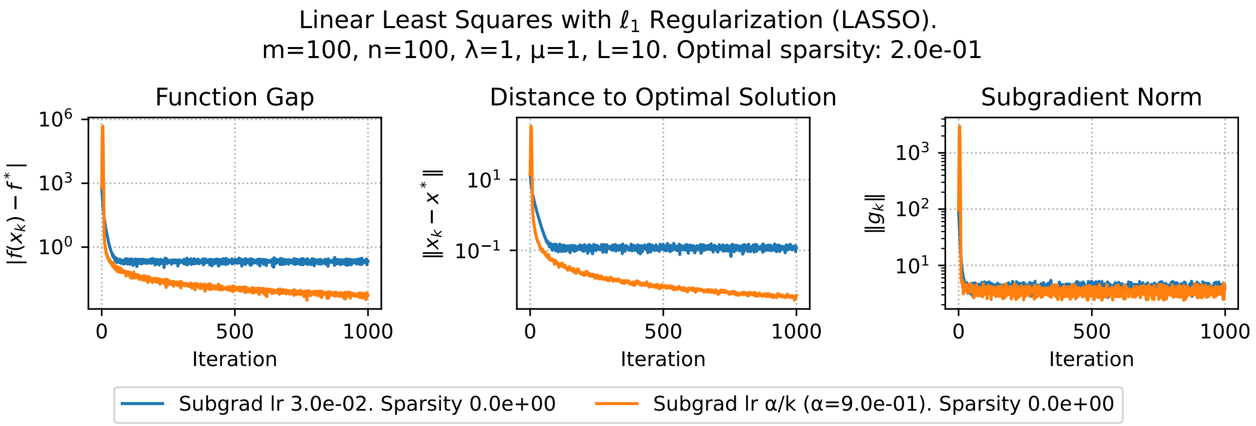



Рисунок 12. Негладкий сильно выпуклый случай. $\frac{\alpha_0}{k}$ шаг приводит к сходимости для $f_k^{\text{best}}$

# Численные эксперименты

$$f(x) = \frac{1}{2m}\|Ax - b\|_2^2 + \lambda\|x\|_1 \to \min_{x \in \mathbb{R}^n}, \qquad A \in \mathbb{R}^{m \times n}, \quad \lambda\left(\tfrac{1}{m}A^T A\right) \in [\mu; L].$$

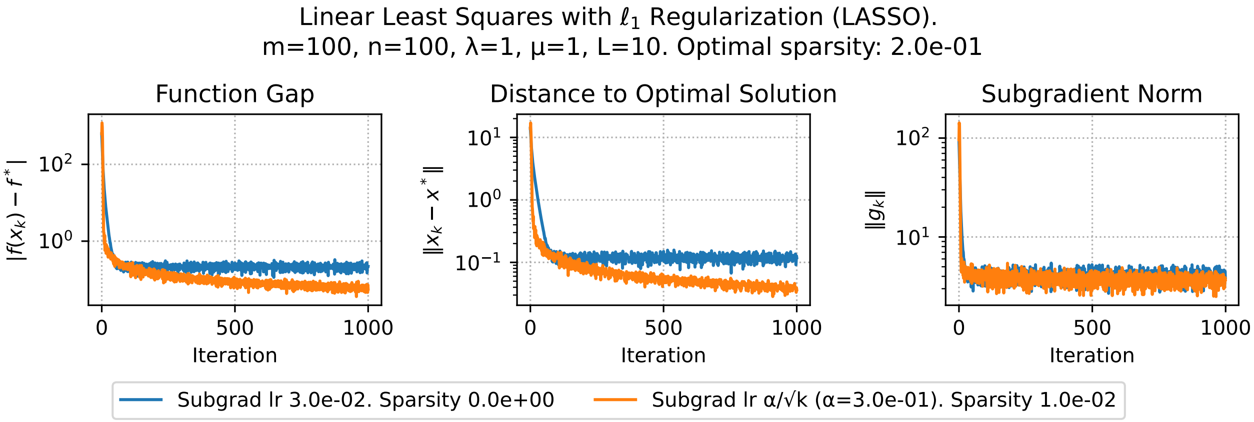



Рисунок 13. Негладкий сильно выпуклый случай. Шаг $\frac{\alpha_0}{\sqrt{k}}$ работает хуже

# Численные эксперименты

$$
f(x) = \frac{1}{m}\sum_{i=1}^{m} \text{log}(1 + \text{exp}(-b_i(A_i x))) + \lambda\|x\|_1 \to \min_{x\in\mathbb{R}^n}, \qquad A_i \in \mathbb{R}^n, \quad b_i \in \{-1, 1\}
$$

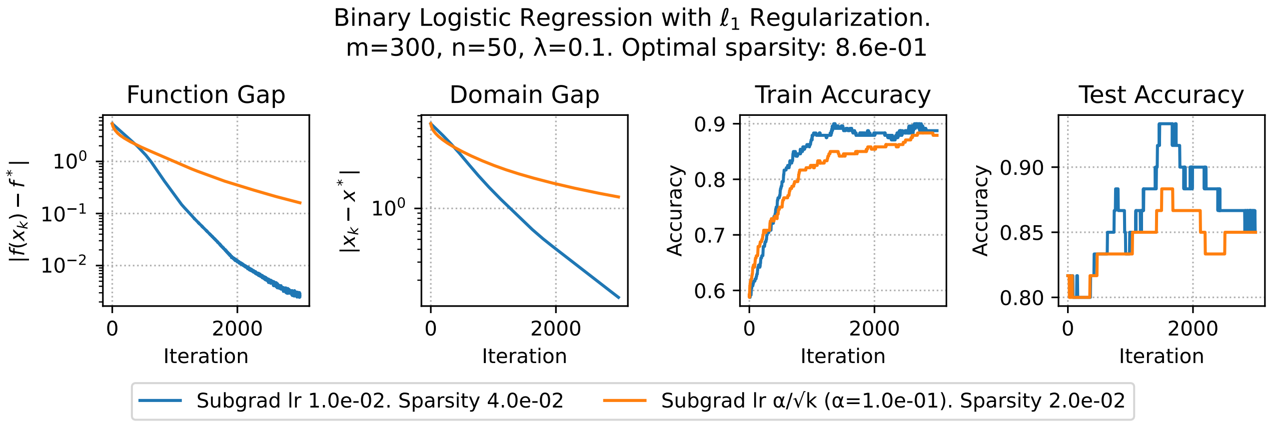



Рисунок 14. Логистическая регрессия с $\ell_1$ регуляризацией

# Численные эксперименты

$$f(x) = \frac{1}{m}\sum_{i=1}^{m} \log(1 + \exp(-b_i(A_i x))) + \lambda\|x\|_1 \to \min_{x\in\mathbb{R}^n}, \qquad A_i \in \mathbb{R}^n, \quad b_i \in \{-1, 1\}$$

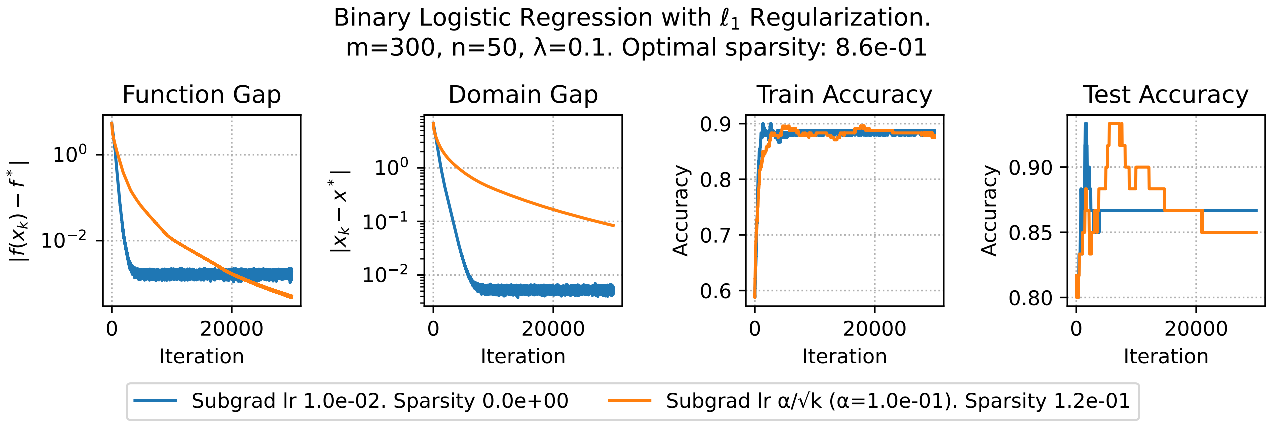



Рисунок 15. Логистическая регрессия с $\ell_1$ регуляризацией

# Численные эксперименты

$$f(x) = \frac{1}{m}\sum_{i=1}^{m} \log(1+\exp(-b_i(A_i x))) + \lambda\|x\|_1 \to \min_{x\in\mathbb{R}^n}, \qquad A_i \in \mathbb{R}^n, \quad b_i \in \{-1,1\}$$

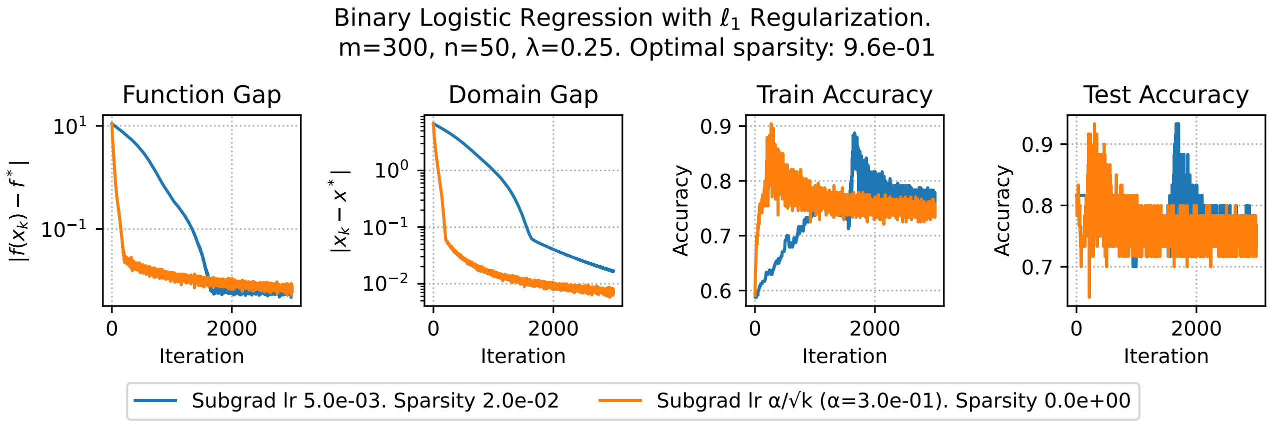



Рисунок 16. Логистическая регрессия с $\ell_1$ регуляризацией

# Численные эксперименты

$$f(x) = \frac{1}{m}\sum_{i=1}^{m} \log(1 + \exp(-b_i(A_i x))) + \lambda \|x\|_1 \to \min_{x \in \mathbb{R}^n}, \qquad A_i \in \mathbb{R}^n, \quad b_i \in \{-1, 1\}$$

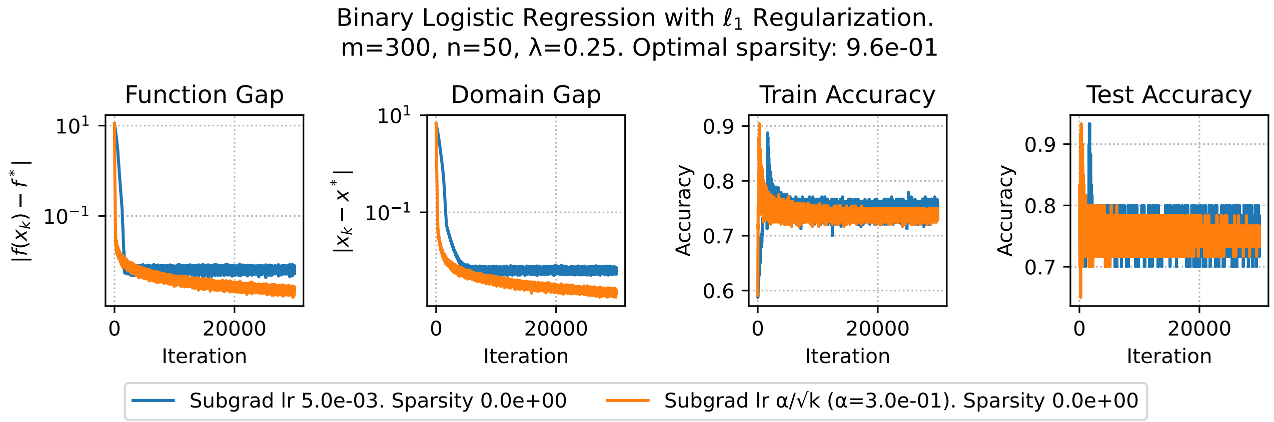



Рисунок 17. Логистическая регрессия с $\ell_1$ регуляризацией

# Численные эксперименты

$$f(x) = \frac{1}{m}\sum_{i=1}^{m} \log(1+\exp(-b_i(A_i x))) + \lambda\|x\|_1 \to \min_{x\in\mathbb{R}^n}, \qquad A_i \in \mathbb{R}^n, \quad b_i \in \{-1, 1\}$$

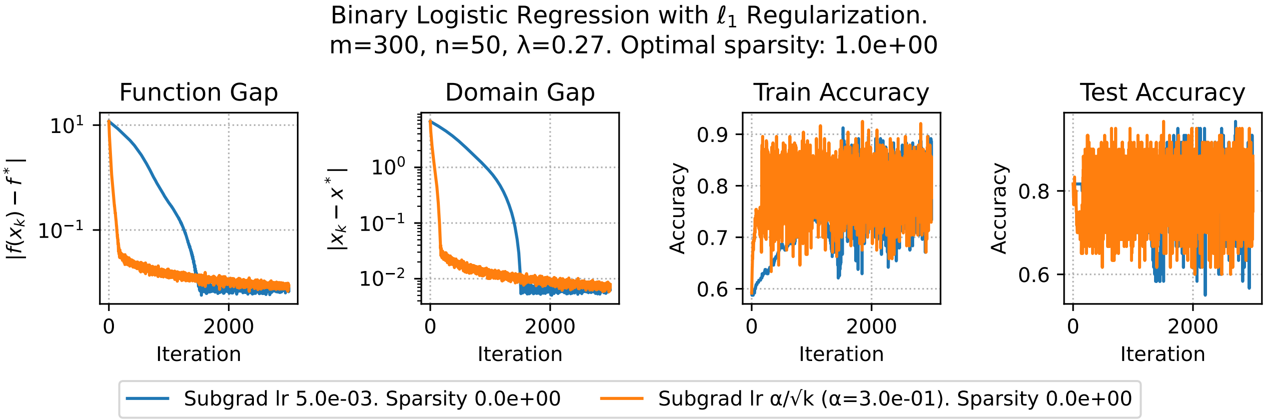



Рисунок 18. Логистическая регрессия с $\ell_1$ регуляризацией

# Численные эксперименты

$$f(x) = \frac{1}{m} \sum_{i=1}^{m} \text{log}(1 + \text{exp}(-b_i(A_i x))) + \lambda \|x\|_1 \to \min_{x \in \mathbb{R}^n}, \qquad A_i \in \mathbb{R}^n, \quad b_i \in \{-1, 1\}$$

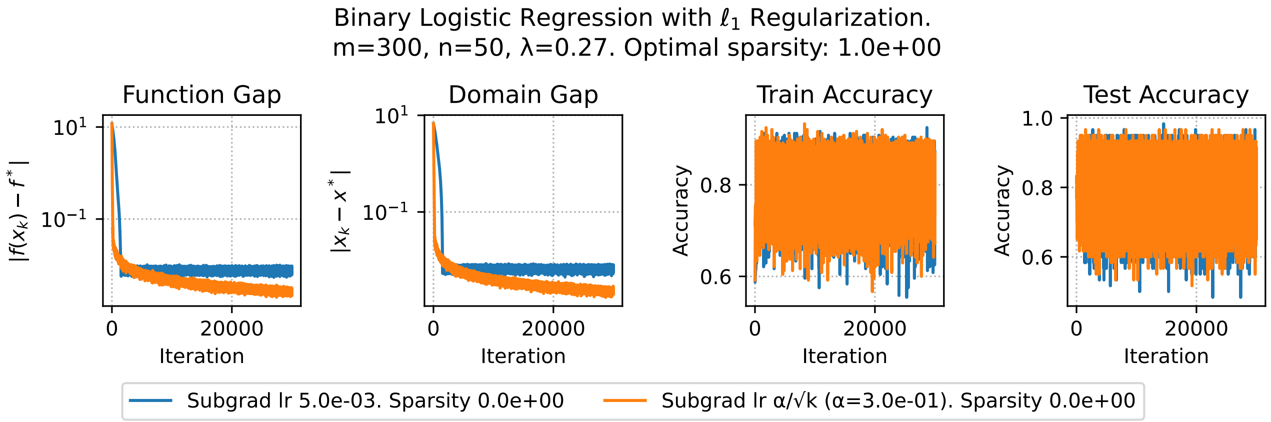



Рисунок 19. Логистическая регрессия с $\ell_1$ регуляризацией

# Бонус: доказательства

# Субдифференциал дифференцируемой функции

> **Субдифференциал дифференцируемой функции**
>
> Пусть $f : S \to \mathbb{R}$ — функция, определенная на множестве $S$ в евклидовом пространстве $\mathbb{R}^n$. Если $x_0 \in \mathrm{ri}(S)$ и $f$ дифференцируема в точке $x_0$, то либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$. Более того, если функция $f$ выпукла, то первая ситуация невозможна.

**Доказательство**

1. Пусть $s \in \partial f(x_0)$ для некоторого $s \in \mathbb{R}^n$, отличного от $\nabla f(x_0)$. Пусть $v \in \mathbb{R}^n$ — единичный вектор. Поскольку $x_0$ — внутренняя точка $S$, существует $\delta > 0$, такое что $x_0 + tv \in S$ для всех $0 < t < \delta$. По определению субградиента:
$$f(x_0 + tv) \geq f(x_0) + t\langle s, v \rangle$$

# Субдифференциал дифференцируемой функции

> **Субдифференциал дифференцируемой функции**
>
> Пусть $f : S \to \mathbb{R}$ — функция, определенная на множестве $S$ в евклидовом пространстве $\mathbb{R}^n$. Если $x_0 \in \text{ri}(S)$ и $f$ дифференцируема в точке $x_0$, то либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$. Более того, если функция $f$ выпукла, то первая ситуация невозможна.

**Доказательство**

1. Пусть $s \in \partial f(x_0)$ для некоторого $s \in \mathbb{R}^n$, отличного от $\nabla f(x_0)$. Пусть $v \in \mathbb{R}^n$ — единичный вектор. Поскольку $x_0$ — внутренняя точка $S$, существует $\delta > 0$, такое что $x_0 + tv \in S$ для всех $0 < t < \delta$. По определению субградиента:
$$f(x_0 + tv) \geq f(x_0) + t\langle s, v\rangle$$

# Субдифференциал дифференцируемой функции

**Субдифференциал дифференцируемой функции**

Пусть $f : S \to \mathbb{R}$ — функция, определенная на множестве $S$ в евклидовом пространстве $\mathbb{R}^n$. Если $x_0 \in \mathbf{ri}(S)$ и $f$ дифференцируема в точке $x_0$, то либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$. Более того, если функция $f$ выпукла, то первая ситуация невозможна.

**Доказательство**

1. Пусть $s \in \partial f(x_0)$ для некоторого $s \in \mathbb{R}^n$, отличного от $\nabla f(x_0)$. Пусть $v \in \mathbb{R}^n$ — единичный вектор. Поскольку $x_0$ — внутренняя точка $S$, существует $\delta > 0$, такое что $x_0 + tv \in S$ для всех $0 < t < \delta$. По определению субградиента:

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v \rangle$$

что влечёт:

$$\frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v \rangle$$

для всех $0 < t < \delta$. Переходя к пределу при $t \to 0$ и используя определение градиента, получаем:

$$\langle \nabla f(x_0), v \rangle = \lim_{t \to 0; 0 < t < \delta} \frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v \rangle$$

2. Отсюда $\langle s - \nabla f(x_0), v \rangle \leq 0$. В силу произвольности $v$ можно выбрать

$$v = \frac{s - \nabla f(x_0)}{\|s - \nabla f(x_0)\|},$$

# Субдифференциал дифференцируемой функции

**Субдифференциал дифференцируемой функции**

Пусть $f : S \to \mathbb{R}$ — функция, определенная на множестве $S$ в евклидовом пространстве $\mathbb{R}^n$. Если $x_0 \in \text{ri}(S)$ и $f$ дифференцируема в точке $x_0$, то либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$. Более того, если функция $f$ выпукла, то первая ситуация невозможна.

**Доказательство**

1. Пусть $s \in \partial f(x_0)$ для некоторого $s \in \mathbb{R}^n$, отличного от $\nabla f(x_0)$. Пусть $v \in \mathbb{R}^n$ — единичный вектор. Поскольку $x_0$ — внутренняя точка $S$, существует $\delta > 0$, такое что $x_0 + tv \in S$ для всех $0 < t < \delta$. По определению субградиента:

$$f(x_0 + tv) \geq f(x_0) + t\langle s, v\rangle$$

что влечёт:

$$\frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v\rangle$$

для всех $0 < t < \delta$. Переходя к пределу при $t \to 0$ и используя определение градиента, получаем:

$$\langle \nabla f(x_0), v\rangle = \lim_{t \to 0; 0 < t < \delta} \frac{f(x_0 + tv) - f(x_0)}{t} \geq \langle s, v\rangle$$

2. Отсюда $\langle s - \nabla f(x_0), v\rangle \leq 0$. В силу произвольности $v$ можно выбрать

$$v = \frac{s - \nabla f(x_0)}{\| s - \nabla f(x_0)\|},$$

# Сходимость: выпуклый случай

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ — выпуклая функция, и мы используем субградиентный метод. Пусть $\|g_k\| \leq G$ для всех $k$, и $\|x_0 - x^*\| \leq R$. Тогда для лучшего найденного значения функции $f_T^{\text{best}} = \min_{0 \leq k < T} f(x_k)$ справедливо:

$$
f_T^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2}{2 \sum_{k=0}^{T-1} \alpha_k}.
$$

**Доказательство** Воспользуемся определением шага субградиентного метода $x_{k+1} = x_k - \alpha_k g_k$ и неравенством, определяющим субградиент: $f(x^*) \geq f(x_k) + \langle g_k, x^* - x_k \rangle$:

$$
\|x_{k+1} - x^*\|^2 = \|x_k - x^* - \alpha_k g_k\|^2 =
$$

# Сходимость: выпуклый случай

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ — выпуклая функция, и мы используем субградиентный метод. Пусть $\|g_k\| \leq G$ для всех $k$, и $\|x_0 - x^*\| \leq R$. Тогда для лучшего найденного значения функции $f_T^{\text{best}} = \min_{0 \leq k < T} f(x_k)$ справедливо:

$$
f_T^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2}{2 \sum_{k=0}^{T-1} \alpha_k}.
$$

**Доказательство** Воспользуемся определением шага субградиентного метода $x_{k+1} = x_k - \alpha_k g_k$ и неравенством, определяющим субградиент: $f(x^*) \geq f(x_k) + \langle g_k, x^* - x_k \rangle$:

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - x^* - \alpha_k g_k\|^2 = \\
&= \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k \langle g_k, x_k - x^* \rangle
\end{aligned}
$$

# Сходимость: выпуклый случай

**Theorem**

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ — выпуклая функция, и мы используем субградиентный метод. Пусть $\|g_k\| \leq G$ для всех $k$, и $\|x_0 - x^*\| \leq R$. Тогда для лучшего найденного значения функции $f_T^{\text{best}} = \min_{0 \leq k < T} f(x_k)$ справедливо:

$$
f_T^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2}{2 \sum_{k=0}^{T-1} \alpha_k}.
$$

**Доказательство** Воспользуемся определением шага субградиентного метода $x_{k+1} = x_k - \alpha_k g_k$ и неравенством, определяющим субградиент: $f(x^*) \geq f(x_k) + \langle g_k, x^* - x_k \rangle$:

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - x^* - \alpha_k g_k\|^2 = \\
&= \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k \langle g_k, x_k - x^* \rangle \\
&\leq \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k (f(x_k) - f(x^*))
\end{aligned}
$$

# Сходимость: выпуклый случай

> **Theorem**
>
> Пусть $f : \mathbb{R}^n \to \mathbb{R}$ — выпуклая функция, и мы используем субградиентный метод. Пусть $\|g_k\| \leq G$ для всех $k$, и $\|x_0 - x^*\| \leq R$. Тогда для лучшего найденного значения функции $f_T^{\text{best}} = \min_{0 \leq k < T} f(x_k)$ справедливо:
>
> $$f_T^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2}{2 \sum_{k=0}^{T-1} \alpha_k}.$$

**Доказательство** Воспользуемся определением шага субградиентного метода $x_{k+1} = x_k - \alpha_k g_k$ и неравенством, определяющим субградиент: $f(x^*) \geq f(x_k) + \langle g_k, x^* - x_k \rangle$:

$$\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - x^* - \alpha_k g_k\|^2 = \\
&= \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k \langle g_k, x_k - x^* \rangle \\
&\leq \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k (f(x_k) - f(x^*)) \\
2\alpha_k (f(x_k) - f(x^*)) &\leq \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 + \alpha_k^2 \|g_k\|^2
\end{aligned}$$

# Сходимость: выпуклый случай (доказательство 2)

Просуммируем неравенство по $k$ от $0$ до $T-1$ и воспользуемся предположением на ограниченность нормы субградиента: $\|g_k\| \leq G$:

$$\sum_{k=0}^{T-1} 2\alpha_k(f(x_k) - f(x^*)) \leq \|x_0 - x^*\|^2 - \|x_T - x^*\|^2 + \sum_{k=0}^{T-1} \alpha_k^2 \|g_k\|^2$$

# Сходимость: выпуклый случай (доказательство 2)

Просуммируем неравенство по $k$ от $0$ до $T-1$ и воспользуемся предположением на ограниченность нормы субградиента: $\|g_k\| \leq G$:

$$
\begin{aligned}
\sum_{k=0}^{T-1} 2\alpha_k (f(x_k) - f(x^*)) &\leq \|x_0 - x^*\|^2 - \|x_T - x^*\|^2 + \sum_{k=0}^{T-1} \alpha_k^2 \|g_k\|^2 \\
&\leq \|x_0 - x^*\|^2 + \sum_{k=0}^{T-1} \alpha_k^2 \|g_k\|^2
\end{aligned}
$$

# Сходимость: выпуклый случай (доказательство 2)

Просуммируем неравенство по $k$ от $0$ до $T-1$ и воспользуемся предположением на ограниченность нормы субградиента: $\|g_k\| \leq G$:

$$
\begin{aligned}
\sum_{k=0}^{T-1} 2\alpha_k(f(x_k) - f(x^*)) &\leq \|x_0 - x^*\|^2 - \|x_T - x^*\|^2 + \sum_{k=0}^{T-1} \alpha_k^2 \|g_k\|^2 \\
&\leq \|x_0 - x^*\|^2 + \sum_{k=0}^{T-1} \alpha_k^2 \|g_k\|^2 \\
&\leq R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2
\end{aligned}
$$

# Сходимость: выпуклый случай (доказательство 2)

Просуммируем неравенство по $k$ от $0$ до $T-1$ и воспользуемся предположением на ограниченность нормы субградиента: $\|g_k\| \leq G$:

$$
\begin{aligned}
\sum_{k=0}^{T-1} 2\alpha_k (f(x_k) - f(x^*)) &\leq \|x_0 - x^*\|^2 - \|x_T - x^*\|^2 + \sum_{k=0}^{T-1} \alpha_k^2 \|g_k\|^2 \\
&\leq \|x_0 - x^*\|^2 + \sum_{k=0}^{T-1} \alpha_k^2 \|g_k\|^2 \\
&\leq R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2
\end{aligned}
$$

$$
(f_T^{\text{best}} - f(x^*)) \sum_{k=0}^{T-1} 2\alpha_k \leq \sum_{k=0}^{T-1} 2\alpha_k (f(x_k) - f(x^*)) \leq R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2
$$

# Сходимость: выпуклый случай (доказательство 2)

Просуммируем неравенство по $k$ от $0$ до $T-1$ и воспользуемся предположением на ограниченность нормы субградиента: $\|g_k\| \leq G$:

$$
\begin{aligned}
\sum_{k=0}^{T-1} 2\alpha_k (f(x_k) - f(x^*)) &\leq \|x_0 - x^*\|^2 - \|x_T - x^*\|^2 + \sum_{k=0}^{T-1} \alpha_k^2 \|g_k\|^2 \\
&\leq \|x_0 - x^*\|^2 + \sum_{k=0}^{T-1} \alpha_k^2 \|g_k\|^2 \\
&\leq R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2
\end{aligned}
$$

$$
(f_T^{\text{best}} - f(x^*)) \sum_{k=0}^{T-1} 2\alpha_k \leq \sum_{k=0}^{T-1} 2\alpha_k (f(x_k) - f(x^*)) \leq R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2
$$

Это дает оценку

$$
f_T^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum \alpha_k^2}{2 \sum \alpha_k}.
$$

Отсюда следует сходимость при $\sum \alpha_k^2 < \infty, \sum \alpha_k = \infty$ (шаг убывает, но не слишком быстро).

# Сходимость: выпуклый случай, практическая стратегия

Theorem

Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для убывающей стратегии шага $\alpha_k = \frac{R}{G\sqrt{k+1}}$, субградиентный метод удовлетворяет

$$f_k^{\text{best}} - f(x^*) \leq \frac{GR(2 + \ln k)}{4\sqrt{k+1}}$$

Оценка сумм:

$$\sum_{i=0}^{k-1} \alpha_i^2 = \frac{R^2}{G^2} \sum_{i=1}^{k} \frac{1}{i} \leq \frac{R^2}{G^2}(1 + \ln k);$$

# Сходимость: выпуклый случай, практическая стратегия

> **Theorem**
>
> Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для убывающей стратегии шага $\alpha_k = \frac{R}{G\sqrt{k+1}}$, субградиентный метод удовлетворяет
>
> $$f_k^{\text{best}} - f(x^*) \leq \frac{GR(2 + \ln k)}{4\sqrt{k+1}}$$

Оценка сумм:

$$\sum_{i=0}^{k-1} \alpha_i^2 = \frac{R^2}{G^2} \sum_{i=1}^{k} \frac{1}{i} \leq \frac{R^2}{G^2}(1 + \ln k); \qquad \sum_{i=0}^{k-1} \alpha_i = \frac{R}{G} \sum_{i=1}^{k} \frac{1}{\sqrt{i}} \geq \frac{R}{G} \int_1^{k+1} \frac{1}{\sqrt{t}}\, dt = \frac{2R}{G}(\sqrt{k+1} - 1).$$

# Сходимость: выпуклый случай, практическая стратегия

**Theorem**

Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для убывающей стратегии шага $\alpha_k = \frac{R}{G\sqrt{k+1}}$, субградиентный метод удовлетворяет

$$
f_k^{\text{best}} - f(x^*) \leq \frac{GR(2 + \ln k)}{4\sqrt{k+1}}
$$

Оценка сумм:

$$
\sum_{i=0}^{k-1} \alpha_i^2 = \frac{R^2}{G^2} \sum_{i=1}^{k} \frac{1}{i} \leq \frac{R^2}{G^2}(1 + \ln k); \qquad \sum_{i=0}^{k-1} \alpha_i = \frac{R}{G} \sum_{i=1}^{k} \frac{1}{\sqrt{i}} \geq \frac{R}{G} \int_1^{k+1} \frac{1}{\sqrt{t}}\, dt = \frac{2R}{G}(\sqrt{k+1} - 1).
$$

$$
f_k^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum\limits_{i=0}^{k-1} \alpha_i^2}{2 \sum\limits_{i=0}^{k-1} \alpha_i}
$$

# Сходимость: выпуклый случай, практическая стратегия

Theorem

Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для убывающей стратегии шага $\alpha_k = \frac{R}{G\sqrt{k+1}}$, субградиентный метод удовлетворяет

$$
f_k^{\text{best}} - f(x^*) \leq \frac{GR(2 + \ln k)}{4\sqrt{k+1}}
$$

Оценка сумм:

$$
\sum_{i=0}^{k-1} \alpha_i^2 = \frac{R^2}{G^2} \sum_{i=1}^{k} \frac{1}{i} \leq \frac{R^2}{G^2}(1 + \ln k); \qquad \sum_{i=0}^{k-1} \alpha_i = \frac{R}{G} \sum_{i=1}^{k} \frac{1}{\sqrt{i}} \geq \frac{R}{G} \int_1^{k+1} \frac{1}{\sqrt{t}}\, dt = \frac{2R}{G}(\sqrt{k+1} - 1).
$$

$$
f_k^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum\limits_{i=0}^{k-1} \alpha_i^2}{2 \sum\limits_{i=0}^{k-1} \alpha_i} \leq \frac{R^2 + R^2(1 + \ln k)}{4\frac{R}{G}(\sqrt{k+1})}
$$

# Сходимость: выпуклый случай, практическая стратегия

> **Theorem**
>
> Пусть $f$ — выпуклая $G$-липшицева функция и $R = \|x_0 - x^*\|_2$. Для убывающей стратегии шага $\alpha_k = \frac{R}{G\sqrt{k+1}}$, субградиентный метод удовлетворяет
>
> $$f_k^{\text{best}} - f(x^*) \leq \frac{GR(2 + \ln k)}{4\sqrt{k+1}}$$

Оценка сумм:

$$\sum_{i=0}^{k-1} \alpha_i^2 = \frac{R^2}{G^2} \sum_{i=1}^{k} \frac{1}{i} \leq \frac{R^2}{G^2}(1 + \ln k); \qquad \sum_{i=0}^{k-1} \alpha_i = \frac{R}{G} \sum_{i=1}^{k} \frac{1}{\sqrt{i}} \geq \frac{R}{G} \int_1^{k+1} \frac{1}{\sqrt{t}}\, dt = \frac{2R}{G}(\sqrt{k+1} - 1).$$

$$f_k^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum\limits_{i=0}^{k-1} \alpha_i^2}{2 \sum\limits_{i=0}^{k-1} \alpha_i} \leq \frac{R^2 + R^2(1 + \ln k)}{4\frac{R}{G}(\sqrt{k+1})} = \frac{GR(2 + \ln k)}{4\sqrt{k+1}}$$

# Сильно выпуклая негладкая функция

> **Theorem**
>
> Пусть $f$ — $\mu$-сильно выпуклая на выпуклом множестве и $x, y$ — произвольные точки. Тогда для любого $g \in \partial f(x)$,
>
> $$\langle g, x - y \rangle \geq f(x) - f(y) + \frac{\mu}{2}\|x - y\|^2.$$

**Доказательство**

1. Для любого $\lambda \in [0, 1)$, из $\mu$-сильной выпуклости,

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

# Сильно выпуклая негладкая функция

**Theorem**

Пусть $f$ — $\mu$-сильно выпуклая на выпуклом множестве и $x, y$ — произвольные точки. Тогда для любого $g \in \partial f(x)$,

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

**Доказательство**

1. Для любого $\lambda \in [0, 1)$, из $\mu$-сильной выпуклости,

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

2. Из неравенства субградиента в $x$, мы получаем

$$f(\lambda x + (1-\lambda)y) \geq f(x) + \langle g, \lambda x + (1-\lambda)y - x\rangle \quad \to \quad f(\lambda x + (1-\lambda)y) \geq f(x) - (1-\lambda)\langle g, x - y\rangle.$$

# Сильно выпуклая негладкая функция

3. Таким образом,

$$f(x) - (1-\lambda)\langle g, x-y\rangle \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2$$

$$(1-\lambda) f(x) \leq (1-\lambda) f(y) + (1-\lambda)\langle g, x-y\rangle - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2$$

$$f(x) \leq f(y) + \langle g, x-y\rangle - \frac{\mu}{2}\lambda\|x-y\|^2$$

# Сильно выпуклая негладкая функция

3. Таким образом,

$$f(x) - (1-\lambda)\langle g, x-y\rangle \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2$$

$$(1-\lambda) f(x) \leq (1-\lambda) f(y) + (1-\lambda)\langle g, x-y\rangle - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2$$

$$f(x) \leq f(y) + \langle g, x-y\rangle - \frac{\mu}{2}\lambda\|x-y\|^2$$

4. Переходя к пределу при $\lambda \to 1^-$ получаем $f(x) \leq f(y) + \langle g, x-y\rangle - \frac{\mu}{2}\|x-y\|^2 \to \langle g, x-y\rangle \geq f(x) - f(y) + \frac{\mu}{2}\|x-y\|^2$.

# Сходимость: сильно выпуклый случай

**Theorem**

Пусть $f$ — $\mu$-сильно выпуклая функция (возможно негладкая) с минимумом $x^*$. Предположим, что субградиенты ограничены на траектории: $\|g_k\| \leq G$. Используя шаг $\alpha_k = \frac{2}{\mu(k+1)}$, субградиентный метод гарантирует для $k > 0$ что:

$$
f_k^{\text{best}} - f(x^*) \leq \frac{2G^2}{\mu k}.
$$

**Доказательство**

1. Начнем с формулировки метода как раньше:

# Сходимость: сильно выпуклый случай

**Theorem**

Пусть $f$ — $\mu$-сильно выпуклая функция (возможно негладкая) с минимумом $x^*$. Предположим, что субградиенты ограничены на траектории: $\|g_k\| \leq G$. Используя шаг $\alpha_k = \frac{2}{\mu(k+1)}$, субградиентный метод гарантирует для $k > 0$ что:

$$
f_k^{\text{best}} - f(x^*) \leq \frac{2G^2}{\mu k}.
$$

**Доказательство**

1. Начнем с формулировки метода как раньше:

$$
\|x_{k+1} - x^*\|^2 = \|x_k - x^* - \alpha_k g_k\|^2 =
$$

# Сходимость: сильно выпуклый случай

**Theorem**

Пусть $f$ — $\mu$-сильно выпуклая функция (возможно негладкая) с минимумом $x^*$. Предположим, что субградиенты ограничены на траектории: $\|g_k\| \leq G$. Используя шаг $\alpha_k = \frac{2}{\mu(k+1)}$, субградиентный метод гарантирует для $k > 0$ что:

$$
f_k^{\text{best}} - f(x^*) \leq \frac{2G^2}{\mu k}.
$$

**Доказательство**

1. Начнем с формулировки метода как раньше:

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 = \|x_k - x^* - \alpha_k g_k\|^2 = \\
= \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k \langle g_k, x_k - x^* \rangle
\end{aligned}
$$

# Сходимость: сильно выпуклый случай

Theorem

Пусть $f$ — $\mu$-сильно выпуклая функция (возможно негладкая) с минимумом $x^*$. Предположим, что субградиенты ограничены на траектории: $\|g_k\| \leq G$. Используя шаг $\alpha_k = \frac{2}{\mu(k+1)}$, субградиентный метод гарантирует для $k > 0$ что:

$$
f_k^{\text{best}} - f(x^*) \leq \frac{2G^2}{\mu k}.
$$

**Доказательство**

1. Начнем с формулировки метода как раньше:

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 = \|x_k - x^* - \alpha_k g_k\|^2 = \\
&= \|x_k - x^*\|^2 + \alpha_k^2\|g_k\|^2 - 2\alpha_k\langle g_k, x_k - x^*\rangle \\
&\leq \|x_k - x^*\|^2 + \alpha_k^2\|g_k\|^2 - 2\alpha_k(f(x_k) - f(x^*)) - \alpha_k\mu\|x_k - x^*\|^2
\end{aligned}
$$

# Сходимость: сильно выпуклый случай

**Theorem**

Пусть $f$ — $\mu$-сильно выпуклая функция (возможно негладкая) с минимумом $x^*$. Предположим, что субградиенты ограничены на траектории: $\|g_k\| \leq G$. Используя шаг $\alpha_k = \frac{2}{\mu(k+1)}$, субградиентный метод гарантирует для $k > 0$ что:

$$
f_k^{\text{best}} - f(x^*) \leq \frac{2G^2}{\mu k}.
$$

**Доказательство**

1. Начнем с формулировки метода как раньше:

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - x^* - \alpha_k g_k\|^2 = \\
&= \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k \langle g_k, x_k - x^* \rangle \\
&\leq \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k (f(x_k) - f(x^*)) - \alpha_k \mu \|x_k - x^*\|^2 \\
&= (1 - \mu\alpha_k)\|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k \left(f(x_k) - f(x^*)\right)
\end{aligned}
$$

# Сходимость: сильно выпуклый случай

**Theorem**

Пусть $f$ — $\mu$-сильно выпуклая функция (возможно негладкая) с минимумом $x^*$. Предположим, что субградиенты ограничены на траектории: $\|g_k\| \leq G$. Используя шаг $\alpha_k = \frac{2}{\mu(k+1)}$, субградиентный метод гарантирует для $k > 0$ что:

$$
f_k^{\text{best}} - f(x^*) \leq \frac{2G^2}{\mu k}.
$$

**Доказательство**

1. Начнем с формулировки метода как раньше:

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - x^* - \alpha_k g_k\|^2 = \\
&= \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k \langle g_k, x_k - x^* \rangle \\
&\leq \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k (f(x_k) - f(x^*)) - \alpha_k \mu \|x_k - x^*\|^2 \\
&= (1 - \mu\alpha_k)\|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k \left(f(x_k) - f(x^*)\right) \\
2\alpha_k \left(f(x_k) - f(x^*)\right) &\leq (1 - \mu\alpha_k)\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 + \alpha_k^2 \|g_k\|^2
\end{aligned}
$$

# Сходимость: сильно выпуклый случай

**Theorem**

Пусть $f$ — $\mu$-сильно выпуклая функция (возможно негладкая) с минимумом $x^*$. Предположим, что субградиенты ограничены на траектории: $\|g_k\| \leq G$. Используя шаг $\alpha_k = \frac{2}{\mu(k+1)}$, субградиентный метод гарантирует для $k > 0$ что:

$$
f_k^{\text{best}} - f(x^*) \leq \frac{2G^2}{\mu k}.
$$

**Доказательство**

1. Начнем с формулировки метода как раньше:

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - x^* - \alpha_k g_k\|^2 = \\
&= \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k \langle g_k, x_k - x^* \rangle \\
&\leq \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k (f(x_k) - f(x^*)) - \alpha_k \mu \|x_k - x^*\|^2 \\
&= (1 - \mu\alpha_k) \|x_k - x^*\|^2 + \alpha_k^2 \|g_k\|^2 - 2\alpha_k \left(f(x_k) - f(x^*)\right) \\
2\alpha_k \left(f(x_k) - f(x^*)\right) &\leq (1 - \mu\alpha_k) \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 + \alpha_k^2 \|g_k\|^2 \\
f(x_k) - f(x^*) &\leq \frac{1 - \mu\alpha_k}{2\alpha_k} \|x_k - x^*\|^2 - \frac{1}{2\alpha_k} \|x_{k+1} - x^*\|^2 + \frac{\alpha_k}{2} \|g_k\|^2
\end{aligned}
$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$
f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2
$$

$$
f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2
$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$
f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2
$$

$$
f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2
$$

$$
k\left(f(x_k) - f(x^*)\right) \leq \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2
$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2$$

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2$$

$$k\left(f(x_k) - f(x^*)\right) \leq \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2$$

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2$$

$$k\left(f(x_k) - f(x^*)\right) \leq \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2$$

3. Суммируя неравенства для всех $k = 0, 1, \ldots, T-1$, мы получаем:

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2$$

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2$$

$$k\left(f(x_k) - f(x^*)\right) \leq \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2$$

3. Суммируя неравенства для всех $k = 0, 1, \ldots, T-1$, мы получаем:

$$\sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq 0 - \frac{\mu(T-1)T}{4}\|x_T - x^*\|^2 + \frac{1}{\mu}\sum_{k=0}^{T-1}\|g_k\|^2$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$
\begin{aligned}
f(x_k) - f(x^*) &\leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2 \\
f(x_k) - f(x^*) &\leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2 \\
k\left(f(x_k) - f(x^*)\right) &\leq \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2
\end{aligned}
$$

3. Суммируя неравенства для всех $k = 0, 1, \ldots, T-1$, мы получаем:

$$
\sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq 0 - \frac{\mu(T-1)T}{4}\|x_T - x^*\|^2 + \frac{1}{\mu}\sum_{k=0}^{T-1}\|g_k\|^2 \leq \frac{G^2 T}{\mu}
$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2$$

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2$$

$$k\left(f(x_k) - f(x^*)\right) \leq \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2$$

3. Суммируя неравенства для всех $k = 0, 1, \ldots, T-1$, мы получаем:

$$\sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq 0 - \frac{\mu(T-1)T}{4}\|x_T - x^*\|^2 + \frac{1}{\mu}\sum_{k=0}^{T-1}\|g_k\|^2 \leq \frac{G^2 T}{\mu}$$

$$\left(f_{T-1}^{\text{best}} - f(x^*)\right)\sum_{k=0}^{T-1} k = \sum_{k=0}^{T-1} k\left(f_{T-1}^{\text{best}} - f(x^*)\right)$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$f(x_k) - f(x^*) \le \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2$$

$$f(x_k) - f(x^*) \le \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2$$

$$k\left(f(x_k) - f(x^*)\right) \le \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2$$

3. Суммируя неравенства для всех $k = 0, 1, \ldots, T-1$, мы получаем:

$$\sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \le 0 - \frac{\mu(T-1)T}{4}\|x_T - x^*\|^2 + \frac{1}{\mu}\sum_{k=0}^{T-1}\|g_k\|^2 \le \frac{G^2 T}{\mu}$$

$$\left(f_{T-1}^{\text{best}} - f(x^*)\right)\sum_{k=0}^{T-1} k = \sum_{k=0}^{T-1} k\left(f_{T-1}^{\text{best}} - f(x^*)\right) \le \sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right)$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2$$

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2$$

$$k\left(f(x_k) - f(x^*)\right) \leq \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2$$

3. Суммируя неравенства для всех $k = 0, 1, \ldots, T-1$, мы получаем:

$$\sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq 0 - \frac{\mu(T-1)T}{4}\|x_T - x^*\|^2 + \frac{1}{\mu}\sum_{k=0}^{T-1}\|g_k\|^2 \leq \frac{G^2 T}{\mu}$$

$$\left(f_{T-1}^{\text{best}} - f(x^*)\right)\sum_{k=0}^{T-1} k = \sum_{k=0}^{T-1} k\left(f_{T-1}^{\text{best}} - f(x^*)\right) \leq \sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq \frac{G^2 T}{\mu}$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2$$

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2$$

$$k\left(f(x_k) - f(x^*)\right) \leq \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2$$

3. Суммируя неравенства для всех $k = 0, 1, \ldots, T-1$, мы получаем:

$$\sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq 0 - \frac{\mu(T-1)T}{4}\|x_T - x^*\|^2 + \frac{1}{\mu}\sum_{k=0}^{T-1}\|g_k\|^2 \leq \frac{G^2 T}{\mu}$$

$$\left(f_{T-1}^{\text{best}} - f(x^*)\right)\sum_{k=0}^{T-1} k = \sum_{k=0}^{T-1} k\left(f_{T-1}^{\text{best}} - f(x^*)\right) \leq \sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq \frac{G^2 T}{\mu}$$

$$f_{T-1}^{\text{best}} - f(x^*) \leq \frac{G^2 T}{\mu \sum_{k=0}^{T-1} k}$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$
f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2
$$

$$
f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2
$$

$$
k\left(f(x_k) - f(x^*)\right) \leq \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2
$$

3. Суммируя неравенства для всех $k = 0, 1, \ldots, T-1$, мы получаем:

$$
\sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq 0 - \frac{\mu(T-1)T}{4}\|x_T - x^*\|^2 + \frac{1}{\mu}\sum_{k=0}^{T-1}\|g_k\|^2 \leq \frac{G^2 T}{\mu}
$$

$$
\left(f_{T-1}^{\text{best}} - f(x^*)\right)\sum_{k=0}^{T-1} k = \sum_{k=0}^{T-1} k\left(f_{T-1}^{\text{best}} - f(x^*)\right) \leq \sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq \frac{G^2 T}{\mu}
$$

$$
f_{T-1}^{\text{best}} - f(x^*) \leq \frac{G^2 T}{\mu \sum_{k=0}^{T-1} k} = \frac{2G^2 T}{\mu T(T-1)}
$$

# Сходимость: сильно выпуклый случай

2. Подставим шаг $\alpha_k = \frac{2}{\mu(k+1)}$ в неравенство:

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu(k+1)}\|g_k\|^2$$

$$f(x_k) - f(x^*) \leq \frac{\mu(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu k}\|g_k\|^2$$

$$k\left(f(x_k) - f(x^*)\right) \leq \frac{\mu k(k-1)}{4}\|x_k - x^*\|^2 - \frac{\mu k(k+1)}{4}\|x_{k+1} - x^*\|^2 + \frac{1}{\mu}\|g_k\|^2$$

3. Суммируя неравенства для всех $k = 0, 1, \ldots, T-1$, мы получаем:

$$\sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq 0 - \frac{\mu(T-1)T}{4}\|x_T - x^*\|^2 + \frac{1}{\mu}\sum_{k=0}^{T-1}\|g_k\|^2 \leq \frac{G^2 T}{\mu}$$

$$\left(f_{T-1}^{\text{best}} - f(x^*)\right)\sum_{k=0}^{T-1} k = \sum_{k=0}^{T-1} k\left(f_{T-1}^{\text{best}} - f(x^*)\right) \leq \sum_{k=0}^{T-1} k\left(f(x_k) - f(x^*)\right) \leq \frac{G^2 T}{\mu}$$

$$f_{T-1}^{\text{best}} - f(x^*) \leq \frac{G^2 T}{\mu\sum_{k=0}^{T-1} k} = \frac{2G^2 T}{\mu T(T-1)} \qquad f_k^{\text{best}} - f(x^*) \leq \frac{2G^2}{\mu k}.$$

# Двойной бонус: Нижние оценки

# Нижние оценки

| выпуклые (негладкие) [6] | гладкие (невыпуклые)[7] | гладкие и выпуклые[8] | гладкие и сильно выпуклые (или PL)[1] |
|---|---|---|---|
| $\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$ <br> $k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$ | $\mathcal{O}\left(\frac{1}{k^2}\right)$ <br> $k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right)$ | $\mathcal{O}\left(\frac{1}{k^2}\right)$ <br> $k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\sqrt{\varepsilon}}\right)$ | $\mathcal{O}\left(\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k\right)$ <br> $k_\varepsilon \sim \mathcal{O}\left(\sqrt{\kappa}\log\frac{1}{\varepsilon}\right)$ |

[6] Nesterov, Lectures on Convex Optimization

[7] Carmon, Duchi, Hinder, Sidford, 2017

[8] Nemirovski, Yudin, 1979

# Модель черного ящика

Итерация градиентного спуска:

$$
\begin{aligned}
x_{k+1} &= x_k - \alpha_k \nabla f(x_k) \\
&= x_{k-1} - \alpha_{k-1} \nabla f(x_{k-1}) - \alpha_k \nabla f(x_k) \\
&\ \ \vdots \\
&= x_0 - \sum_{i=0}^{k} \alpha_{k-i} \nabla f(x_{k-i})
\end{aligned}
$$

# Модель черного ящика

Итерация градиентного спуска:

$$
\begin{aligned}
x_{k+1} &= x_k - \alpha_k \nabla f(x_k) \\
&= x_{k-1} - \alpha_{k-1} \nabla f(x_{k-1}) - \alpha_k \nabla f(x_k) \\
&\ \vdots \\
&= x_0 - \sum_{i=0}^{k} \alpha_{k-i} \nabla f(x_{k-i})
\end{aligned}
$$

Рассмотрим семейство методов первого порядка, где

$$
\begin{aligned}
&x_{k+1} \in x_0 + \text{span}\left\{\nabla f(x_0), \nabla f(x_1), \ldots, \nabla f(x_k)\right\} && f \text{ - smooth} \\
&x_{k+1} \in x_0 + \text{span}\left\{g_0, g_1, \ldots, g_k\right\}\text{, where } g_i \in \partial f(x_i) && f \text{ - non-smooth}
\end{aligned}
\tag{1}
$$

# Модель черного ящика

Итерация градиентного спуска:

$$
\begin{aligned}
x_{k+1} &= x_k - \alpha_k \nabla f(x_k) \\
&= x_{k-1} - \alpha_{k-1} \nabla f(x_{k-1}) - \alpha_k \nabla f(x_k) \\
&\ \vdots \\
&= x_0 - \sum_{i=0}^{k} \alpha_{k-i} \nabla f(x_{k-i})
\end{aligned}
$$

Рассмотрим семейство методов первого порядка, где

$$
\begin{aligned}
&x_{k+1} \in x_0 + \text{span}\left\{\nabla f(x_0), \nabla f(x_1), \ldots, \nabla f(x_k)\right\} && f \text{ - smooth} \\
&x_{k+1} \in x_0 + \text{span}\left\{g_0, g_1, \ldots, g_k\right\} \text{, where } g_i \in \partial f(x_i) && f \text{ - non-smooth}
\end{aligned} \tag{1}
$$

Для получения нижней оценки построим «худшую» функцию $f$ из соответствующего класса, на которой любой метод из семейства Уравнение 1 сходится медленно.

# Негладкий выпуклый случай

**Theorem**

Существует функция $f$, которая является $G$-липшицевой и выпуклой, такая, что любой метод Уравнение 1 удовлетворяет

$$\min_{1 \leq i \leq k} f(x_i) - \min_{x \in \mathbb{B}(R)} f(x) \geq \frac{GR}{2(1 + \sqrt{k})}$$

для $R > 0$ и $k \leq n$, где $n$ — размерность задачи.

# Негладкий выпуклый случай

**Theorem**

Существует функция $f$, которая является $G$-липшицевой и выпуклой, такая, что любой метод Уравнение 1 удовлетворяет

$$\min_{1\leq i\leq k} f(x_i) - \min_{x\in\mathbb{B}(R)} f(x) \geq \frac{GR}{2(1+\sqrt{k})}$$

для $R > 0$ и $k \leq n$, где $n$ — размерность задачи.

**Идея доказательства:** построить такую функцию $f$, что для любого метода Уравнение 1 получаем

$$x_k \in \text{span}\{e_1, e_2, \ldots, e_k\}$$

где $e_i$ — $i$-й стандартный базисный вектор. На итерации $k \leq n$, есть по крайней мере $n-k$ координат $x$, равных $0$. Это позволяет нам получить оценку на ошибку.

# Негладкий выпуклый случай (доказательство)

Рассмотрим функцию:

$$f(x) = \beta \max_{1 \le i \le k} x^{(i)} + \frac{\alpha}{2}\|x\|_2^2,$$

где $\alpha, \beta \in \mathbb{R}$ — параметры, и $x^{(1)}, \dots, x^{(k)}$ — первые $k$ компонент $x$.

# Негладкий выпуклый случай (доказательство)

Рассмотрим функцию:

$$
f(x) = \beta \max_{1 \le i \le k} x^{(i)} + \frac{\alpha}{2}\|x\|_2^2,
$$

где $\alpha, \beta \in \mathbb{R}$ — параметры, и $x^{(1)}, \dots, x^{(k)}$ — первые $k$ компонент $x$.

**Свойства:**

- Функция $f(x)$ является $\alpha$-сильно выпуклой из-за квадратичного члена $\frac{\alpha}{2}\|x\|_2^2$.

# Негладкий выпуклый случай (доказательство)

Рассмотрим функцию:

$$f(x) = \beta \max_{1 \le i \le k} x^{(i)} + \frac{\alpha}{2}\|x\|_2^2,$$

где $\alpha, \beta \in \mathbb{R}$ — параметры, и $x^{(1)}, \dots, x^{(k)}$ — первые $k$ компонент $x$.

**Свойства:**

- Функция $f(x)$ является $\alpha$-сильно выпуклой из-за квадратичного члена $\frac{\alpha}{2}\|x\|_2^2$.
- Функция негладкая, поскольку первое слагаемое вносит недифференцируемость в точке, соответствующей максимальной координате $x$.

# Негладкий выпуклый случай (доказательство)

Рассмотрим функцию:

$$f(x) = \beta \max_{1 \leq i \leq k} x^{(i)} + \frac{\alpha}{2} \|x\|_2^2,$$

где $\alpha, \beta \in \mathbb{R}$ — параметры, и $x^{(1)}, \ldots, x^{(k)}$ — первые $k$ компонент $x$.

**Свойства:**

- Функция $f(x)$ является $\alpha$-сильно выпуклой из-за квадратичного члена $\frac{\alpha}{2}\|x\|_2^2$.
- Функция негладкая, поскольку первое слагаемое вносит недифференцируемость в точке, соответствующей максимальной координате $x$.

# Негладкий выпуклый случай (доказательство)

Рассмотрим функцию:

$$f(x) = \beta \max_{1\le i\le k} x^{(i)} + \frac{\alpha}{2}\|x\|_2^2,$$

где $\alpha, \beta \in \mathbb{R}$ — параметры, и $x^{(1)}, \dots, x^{(k)}$ — первые $k$ компонент $x$.

**Свойства:**

- Функция $f(x)$ является $\alpha$-сильно выпуклой из-за квадратичного члена $\frac{\alpha}{2}\|x\|_2^2$.
- Функция негладкая, поскольку первое слагаемое вносит недифференцируемость в точке, соответствующей максимальной координате $x$.

Рассмотрим субдифференциал $f(x)$ в $x$:

$$\begin{aligned}
\partial f(x) &= \partial\left(\beta \max_{1\le i\le k} x^{(i)}\right) + \partial\left(\frac{\alpha}{2}\|x\|_2^2\right) \\
&= \beta\partial\left(\max_{1\le i\le k} x^{(i)}\right) + \alpha x \\
&= \beta \,\mathrm{conv}\left\{e_i \mid i : x^{(i)} = \max_j x^{(j)}\right\} + \alpha x
\end{aligned}$$

# Негладкий выпуклый случай (доказательство)

Рассмотрим функцию:

$$f(x) = \beta \max_{1 \le i \le k} x^{(i)} + \frac{\alpha}{2}\|x\|_2^2,$$

где $\alpha, \beta \in \mathbb{R}$ — параметры, и $x^{(1)}, \ldots, x^{(k)}$ — первые $k$ компонент $x$.

**Свойства:**

- Функция $f(x)$ является $\alpha$-сильно выпуклой из-за квадратичного члена $\frac{\alpha}{2}\|x\|_2^2$.
- Функция негладкая, поскольку первое слагаемое вносит недифференцируемость в точке, соответствующей максимальной координате $x$.

Рассмотрим субдифференциал $f(x)$ в $x$:

$$\begin{aligned}
\partial f(x) &= \partial\left(\beta \max_{1 \le i \le k} x^{(i)}\right) + \partial\left(\frac{\alpha}{2}\|x\|_2^2\right) \\
&= \beta\partial\left(\max_{1 \le i \le k} x^{(i)}\right) + \alpha x \\
&= \beta \operatorname{conv}\left\{e_i \mid i : x^{(i)} = \max_j x^{(j)}\right\} + \alpha x
\end{aligned}$$

Легко видеть, что если $g \in \partial f(x)$ и $\|x\| \le R$, то

$$\|g\| \le \alpha R + \beta$$

Таким образом, $f$ является $\alpha R + \beta$-липшицевой на $B(R)$.

# Негладкий выпуклый случай (доказательство)

Далее опишем оракул первого порядка для этой функции. При запросе субградиента в точке $x$ оракул возвращает:

$$\alpha x + \gamma e_i,$$

где $i$ — *наименьший* индекс, на котором достигается максимум $x^{(i)} = \max_{1 \le j \le k} x^{(j)}$.

- Пусть $x_0 = 0$.

# Негладкий выпуклый случай (доказательство)

Далее опишем оракул первого порядка для этой функции. При запросе субградиента в точке $x$ оракул возвращает:

$$\alpha x + \gamma e_i,$$

где $i$ — *наименьший* индекс, на котором достигается максимум $x^{(i)} = \max_{1 \leq j \leq k} x^{(j)}$.

- Пусть $x_0 = 0$.
- При запросе оракула в $x_0 = 0$, он возвращает $e_1$. Следовательно, $x_1$ лежит на прямой, порождённой $e_1$.

# Негладкий выпуклый случай (доказательство)

Далее опишем оракул первого порядка для этой функции. При запросе субградиента в точке $x$ оракул возвращает:

$$\alpha x + \gamma e_i,$$

где $i$ — *наименьший* индекс, на котором достигается максимум $x^{(i)} = \max_{1 \leq j \leq k} x^{(j)}$.

- Пусть $x_0 = 0$.
- При запросе оракула в $x_0 = 0$, он возвращает $e_1$. Следовательно, $x_1$ лежит на прямой, порождённой $e_1$.
- Индукцией по $i$ можно показать, что $x_i \in \text{span}\{e_1, \dots, e_i\}$. В частности, для $i \leq k$, $k+1$-я координата $x_i$ равна нулю и вследствие структуры $f(x)$:

$$f(x_i) \geq 0.$$

# Негладкий выпуклый случай (доказательство)

- Найдем минимум функции $f$. Определим точку $y \in \mathbb{R}^n$:

$$y^{(i)} = -\frac{\beta}{\alpha k} \quad \text{for } 1 \leq i \leq k, \qquad y^{(i)} = 0 \quad \text{for } k+1 \leq i \leq n.$$

# Негладкий выпуклый случай (доказательство)

- Найдем минимум функции $f$. Определим точку $y \in \mathbb{R}^n$:

$$y^{(i)} = -\frac{\beta}{\alpha k} \quad \text{for } 1 \leq i \leq k, \qquad y^{(i)} = 0 \quad \text{for } k+1 \leq i \leq n.$$

- Заметим, что $0 \in \partial f(y)$:

$$\begin{aligned}
\partial f(y) &= \alpha y + \beta \,\text{conv}\left\{e_i \mid i : y^{(i)} = \max_j y^{(j)}\right\} \\
&= \alpha y + \beta \,\text{conv}\left\{e_i \mid i : y^{(i)} = 0\right\} \\
0 &\in \partial f(y).
\end{aligned}$$

# Негладкий выпуклый случай (доказательство)

- Найдем минимум функции $f$. Определим точку $y \in \mathbb{R}^n$:

$$y^{(i)} = -\frac{\beta}{\alpha k} \quad \text{for } 1 \le i \le k, \qquad y^{(i)} = 0 \quad \text{for } k+1 \le i \le n.$$

- Заметим, что $0 \in \partial f(y)$:

$$\begin{aligned}
\partial f(y) &= \alpha y + \beta \,\text{conv} \left\{ e_i \mid i : y^{(i)} = \max_j y^{(j)} \right\} \\
&= \alpha y + \beta \,\text{conv} \left\{ e_i \mid i : y^{(i)} = 0 \right\} \\
0 &\in \partial f(y).
\end{aligned}$$

- Следовательно, минимальное значение $f = f(y) = f(x^*)$ равно

$$f(y) = -\frac{\beta^2}{\alpha k} + \frac{\alpha}{2} \cdot \frac{\beta^2}{\alpha^2 k} = -\frac{\beta^2}{2\alpha k}.$$

# Негладкий выпуклый случай (доказательство)

- Найдем минимум функции $f$. Определим точку $y \in \mathbb{R}^n$:

$$y^{(i)} = -\frac{\beta}{\alpha k} \quad \text{for } 1 \le i \le k, \qquad y^{(i)} = 0 \quad \text{for } k+1 \le i \le n.$$

- Заметим, что $0 \in \partial f(y)$:

$$\begin{aligned} \partial f(y) &= \alpha y + \beta \operatorname{conv}\left\{ e_i \mid i : y^{(i)} = \max_j y^{(j)} \right\} \\ &= \alpha y + \beta \operatorname{conv}\left\{ e_i \mid i : y^{(i)} = 0 \right\} \\ 0 &\in \partial f(y). \end{aligned}$$

- Следовательно, минимальное значение $f = f(y) = f(x^*)$ равно

$$f(y) = -\frac{\beta^2}{\alpha k} + \frac{\alpha}{2} \cdot \frac{\beta^2}{\alpha^2 k} = -\frac{\beta^2}{2\alpha k}.$$

- Теперь мы получаем:

$$f(x_i) - f(x^*) \ge 0 - \left( -\frac{\beta^2}{2\alpha k} \right) \ge \frac{\beta^2}{2\alpha k}.$$

# Негладкий выпуклый случай (доказательство)

Имеем оценку снизу $f(x_i) - f(x^*) \geq \frac{\beta^2}{2\alpha k}$, что (при правильном выборе параметров) совпадает с требуемой $\frac{GR}{2(1+\sqrt{k})}$.

# Негладкий выпуклый случай (доказательство)

Имеем оценку снизу $f(x_i) - f(x^*) \geq \frac{\beta^2}{2\alpha k}$, что (при правильном выборе параметров) совпадает с требуемой $\frac{GR}{2(1+\sqrt{k})}$.

Выпуклый случай

$$\alpha = \frac{G}{R}\frac{1}{1+\sqrt{k}} \quad \beta = \frac{\sqrt{k}}{1+\sqrt{k}}$$

$$\frac{\beta^2}{2\alpha} = \frac{GRk}{2(1+\sqrt{k})}$$

Заметим, в частности, что $\|y\|_2^2 = \frac{\beta^2}{\alpha^2 k} = R^2$ с этими параметрами

$$\min_{1\leq i\leq k} f(x_i) - f(x^*) \geq \frac{\beta^2}{2\alpha k} = \frac{GR}{2(1+\sqrt{k})}$$

# Негладкий выпуклый случай (доказательство)

Имеем оценку снизу $f(x_i) - f(x^*) \geq \frac{\beta^2}{2\alpha k}$, что (при правильном выборе параметров) совпадает с требуемой $\frac{GR}{2(1+\sqrt{k})}$.

## Выпуклый случай

$$\alpha = \frac{G}{R}\frac{1}{1+\sqrt{k}} \quad \beta = \frac{\sqrt{k}}{1+\sqrt{k}}$$

$$\frac{\beta^2}{2\alpha} = \frac{GRk}{2(1+\sqrt{k})}$$

Заметим, в частности, что $\|y\|_2^2 = \frac{\beta^2}{\alpha^2 k} = R^2$ с этими параметрами

$$\min_{1\leq i\leq k} f(x_i) - f(x^*) \geq \frac{\beta^2}{2\alpha k} = \frac{GR}{2(1+\sqrt{k})}$$

## Сильно выпуклый случай

$$\alpha = \frac{G}{2R} \quad \beta = \frac{G}{2}$$

Заметим, в частности, что $\|y\|_2^2 = \frac{\beta^2}{\alpha^2 k} = \frac{G^2}{4\alpha^2 k} = R^2$ с этими параметрами

$$\min_{1\leq i\leq k} f(x_i) - f(x^*) \geq \frac{G^2}{8\alpha k}$$

# Ссылки


- Subgradient Methods Stephen Boyd (with help from Jaehyun Park)