

Невыпуклая оптимизация. Нижние оценки.  
Субградиентный спуск.

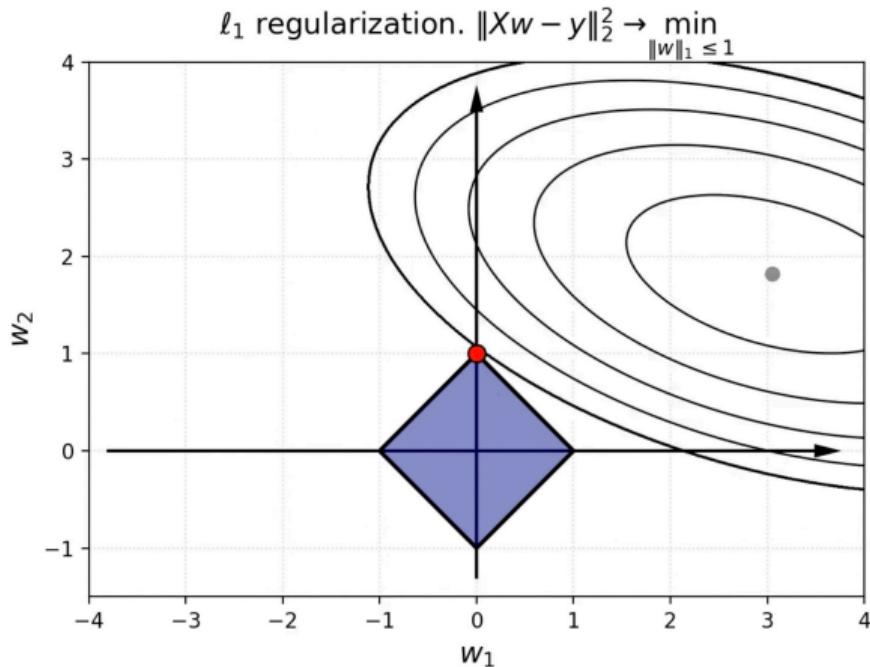
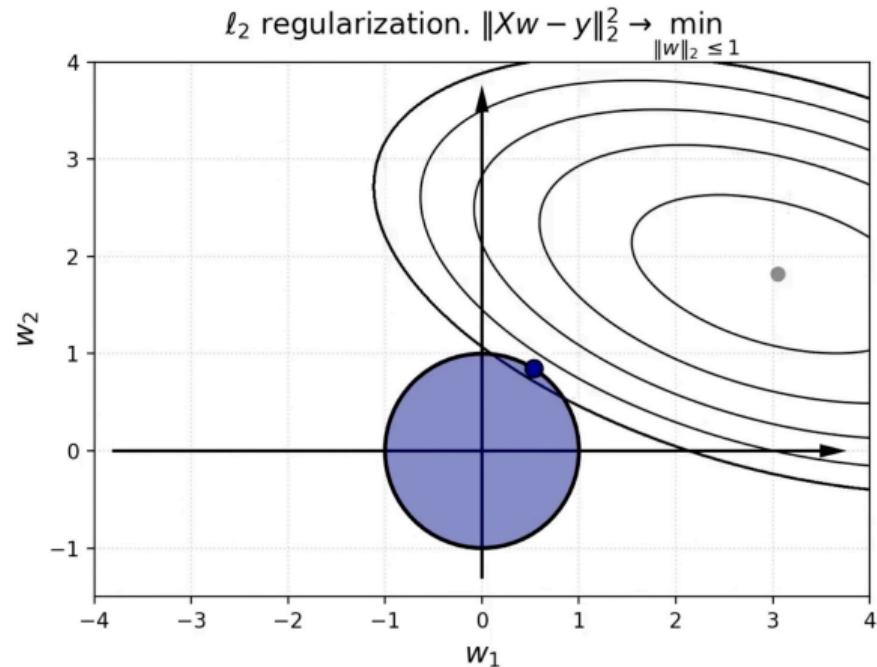
Даня Меркулов

Оптимизация для всех! ЦУ

## Негладкие задачи

## Задача наименьших квадратов с $\ell_1$ -регуляризацией

$\ell_1$  induces sparsity



@fminxyz

## Нормы не являются гладкими

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

Рассмотрим классическую выпуклую задачу оптимизации. Мы предполагаем, что  $f(x)$  является выпуклой функцией, но теперь мы не требуем гладкости.

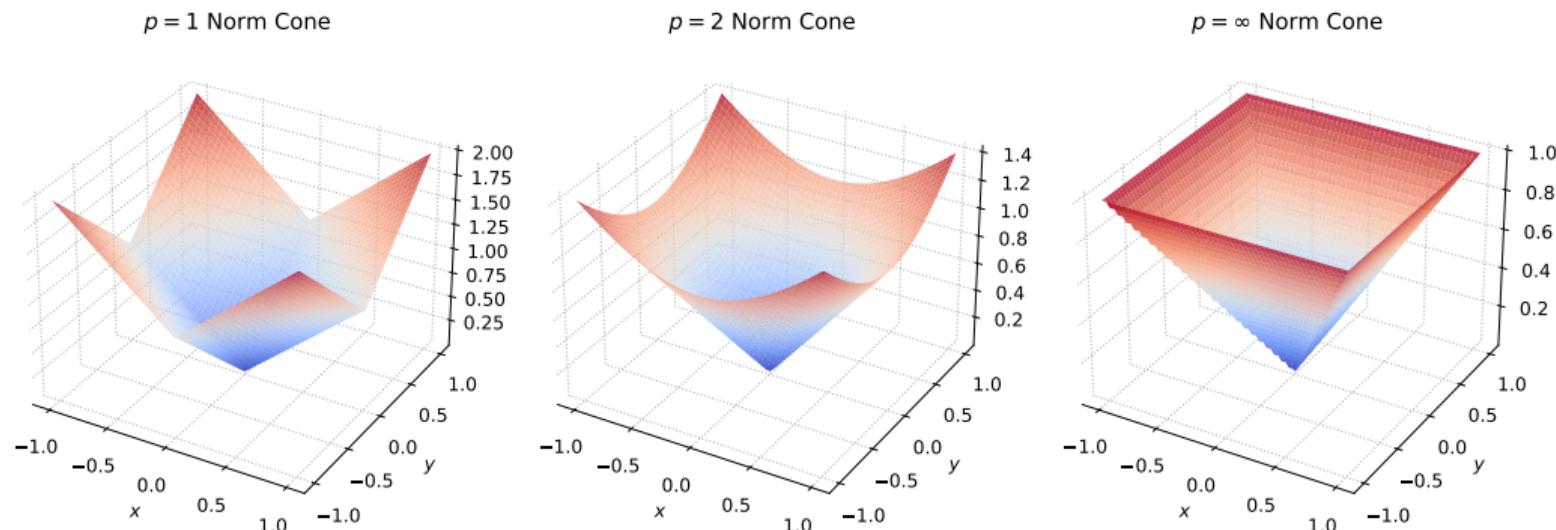


Рис. 1: Нормы конусов для разных  $p$  — нормы не являются гладкими

# Пример Вульфа

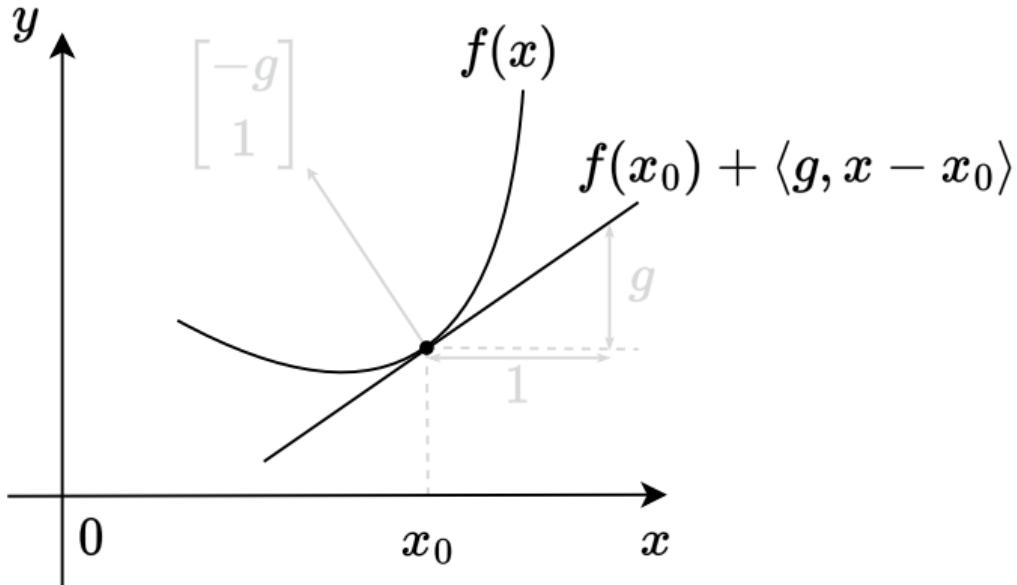
Wolfe's example



Рис. 2: Пример Вульфа. [Открыть в Colab](#)



## Линейная нижняя оценка выпуклых функций

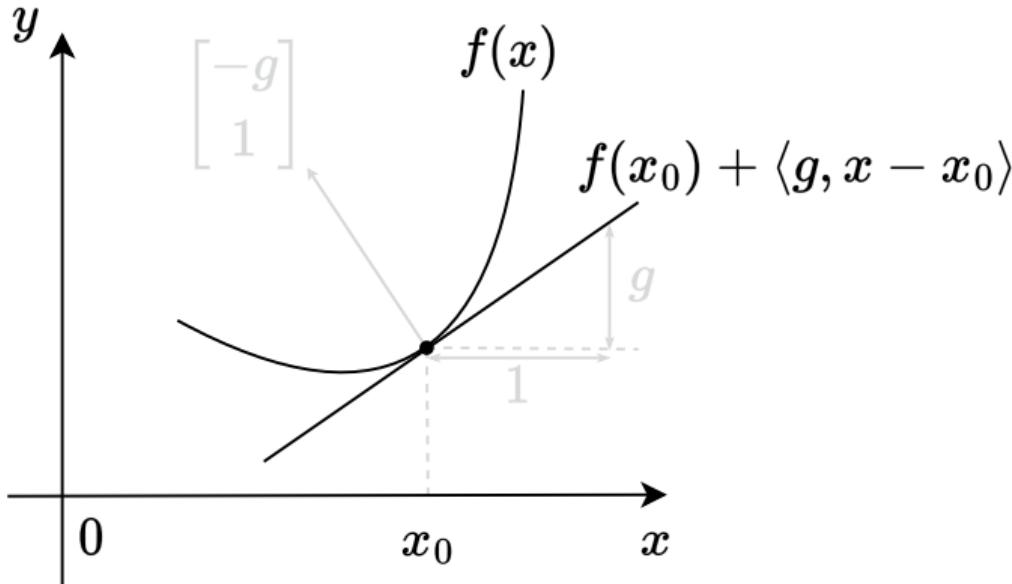


Важное свойство непрерывной выпуклой функции  $f(x)$  заключается в том, что для любой выбранной точки  $x_0$  для всех  $x \in \text{dom } f$  выполняется неравенство:

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

Рис. 3: Линейная аппроксимация Тейлора служит глобальной нижней оценкой для выпуклой функции

## Линейная нижняя оценка выпуклых функций



Важное свойство непрерывной выпуклой функции  $f(x)$  заключается в том, что для любой выбранной точки  $x_0$  для всех  $x \in \text{dom } f$  выполняется неравенство:

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

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

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

Рис. 3: Линейная аппроксимация Тейлора служит глобальной нижней оценкой для выпуклой функции

## Линейная нижняя оценка выпуклых функций



Важное свойство непрерывной выпуклой функции  $f(x)$  заключается в том, что для любой выбранной точки  $x_0$  для всех  $x \in \text{dom } f$  выполняется неравенство:

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

для некоторого вектора  $g$ , т.е. 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.

Рис. 3: Линейная аппроксимация Тейлора служит глобальной нижней оценкой для выпуклой функции

# Линейная нижняя оценка выпуклых функций



Важное свойство непрерывной выпуклой функции  $f(x)$  заключается в том, что для любой выбранной точки  $x_0$  для всех  $x \in \text{dom } f$  выполняется неравенство:

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

для некоторого вектора  $g$ , т.е. 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.

Рис. 3: Линейная аппроксимация Тейлора служит глобальной нижней оценкой для выпуклой функции

## Линейная нижняя оценка выпуклых функций



Важное свойство непрерывной выпуклой функции  $f(x)$  заключается в том, что для любой выбранной точки  $x_0$  для всех  $x \in \text{dom } f$  выполняется неравенство:

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

для некоторого вектора  $g$ , т.е. 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 не хотим потерять такое удобное свойство.

Рис. 3: Линейная аппроксимация Тейлора служит глобальной нижней оценкой для выпуклой функции

## Субградиент и субдифференциал

Вектор  $g$  называется **субградиентом** функции  $f(x) : S \rightarrow \mathbb{R}$  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  $f(x) : S \rightarrow \mathbb{R}$  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  $f(x) : S \rightarrow \mathbb{R}$  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  $\partial f(x_0)$ .

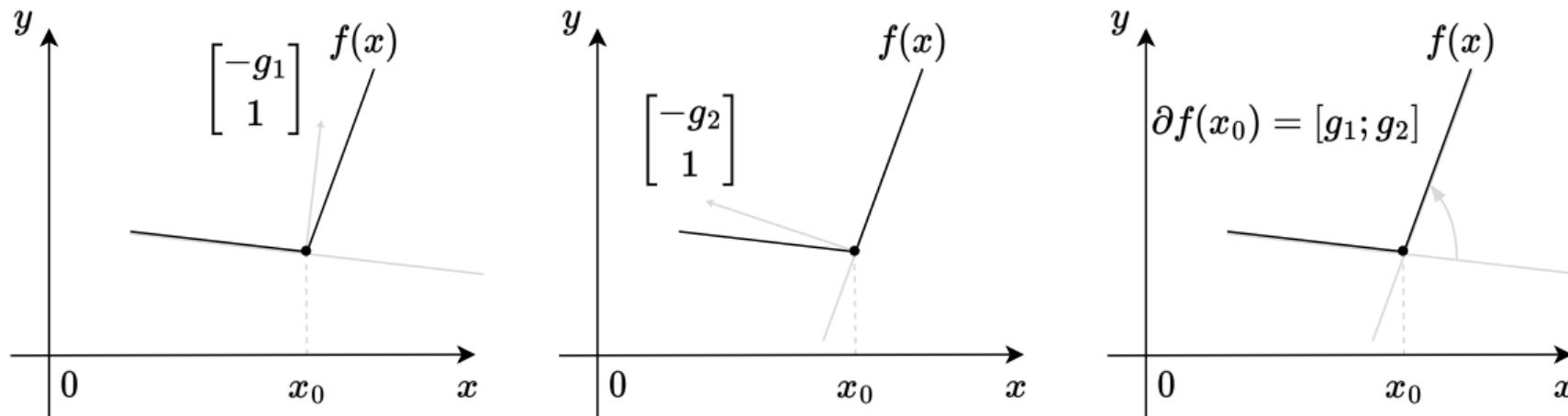


Рис. 4: Субдифференциал — это множество всех возможных субградиентов

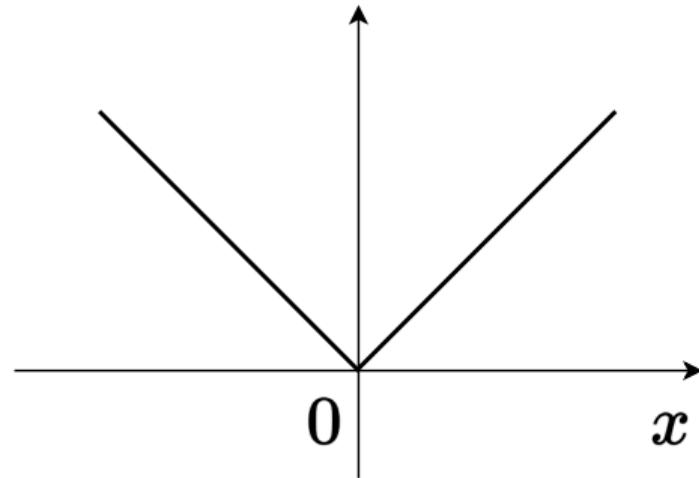
## Субградиент и субдифференциал

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

## Субградиент и субдифференциал

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

$$f(x) = |x|$$



$$\partial f(x)$$



## 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  $x_0 \in \text{ri}(S)$ , 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  $x_0 \in \text{ri}(S)$ , 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  $S$ .

**i** Субдифференциал дифференцируемой функции

Пусть  $f : S \rightarrow \mathbb{R}$  — 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  $S$ .

**i** Субдифференциал дифференцируемой функции

Пусть  $f : S \rightarrow \mathbb{R}$  — 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  $x_0$  является внутренней точкой множества  $S$ , существует  $\delta > 0$  такое, что  $x_0 + tv \in S$  для всех  $0 < t < \delta$ . По определению субградиента:

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

# 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  $S$ .

**i** Субдифференциал дифференцируемой функции

Пусть  $f : S \rightarrow \mathbb{R}$  — 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  $x_0$  является внутренней точкой множества  $S$ , существует  $\delta > 0$  такое, что  $x_0 + tv \in S$  для всех  $0 < t < \delta$ . По определению субградиента:

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

# 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  $S$ .

**i** Субдифференциал дифференцируемой функции

Пусть  $f : S \rightarrow \mathbb{R}$  — 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  $x_0$  является внутренней точкой множества  $S$ , 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  $t \rightarrow 0$  и используя определение градиента, получаем:

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

## Свойства субдифференциала

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

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

которое приводит к  $s = \nabla f(x_0)$ .

## Свойства субдифференциала

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

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

которое приводит к  $s = \nabla f(x_0)$ .

3. Более того, если функция  $f$  выпукла, то согласно дифференциальному условию выпуклости  $f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle$  для всех  $x \in S$ . Но по определению это означает, что  $\nabla f(x_0) \in \partial f(x_0)$ .

# Вычисление субдифференциалов

■ Теорема Моро — Роккафеллара  
(субдифференциал линейной комбинации)

Пусть  $f_i(x)$  — выпуклые функции на выпуклых множествах  $S_i$ ,  $i = \overline{1, n}$ . Тогда если  $\bigcap_{i=1}^n \text{ri}(S_i) \neq$

$\emptyset$ , то функция  $f(x) = \sum_{i=1}^n a_i f_i(x)$ ,  $a_i > 0$

имеет субдифференциал  $\partial_S f(x)$  на множестве

$S = \bigcap_{i=1}^n S_i$  и

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

# Вычисление субдифференциалов

## ■ Теорема Моро — Роккафеллара (субдифференциал линейной комбинации)

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

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

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

## Вычисление субдифференциала

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

## Вычисление субдифференциала

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

## Вычисление субдифференциала

- $\partial(\alpha f)(x) = \alpha \partial f(x)$ , 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

## Вычисление субдифференциала

- $\partial(\alpha f)(x) = \alpha \partial f(x)$ , 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
- $z \in \partial f(x)$  тогда и только тогда, когда  $x \in \partial f^*(z)$ .

## Субградиентный метод

## Алгоритм

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

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

Идея очень проста: заменим градиент  $\nabla f(x_k)$  в методе градиентного спуска субградиентом  $g_k$  в точке  $x_k$ :

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

где  $g_k$  — произвольный субградиент функции  $f(x)$  в точке  $x_k$ ,  $g_k \in \partial f(x_k)$

## Алгоритм

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

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

Идея очень проста: заменим градиент  $\nabla f(x_k)$  в методе градиентного спуска субградиентом  $g_k$  в точке  $x_k$ :

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

где  $g_k$  — произвольный субградиент функции  $f(x)$  в точке  $x_k$ ,  $g_k \in \partial f(x_k)$

Заметьте, что метод субградиента не гарантирует убывание: отрицательный субградиент может не быть направлением убывания, а выбор длины шага может привести к тому, что  $f(x_{k+1}) > f(x_k)$ .

Поэтому мы обычно отслеживаем лучшее значение целевой функции

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

## Сходимость

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

## Сходимость

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

## Сходимость

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

## Сходимость

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

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

## Сходимость

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

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

## Сходимость

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

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

Просуммируем полученное неравенство для  $k = 0, \dots, T - 1$ :

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

## Сходимость

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

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

Просуммируем полученное неравенство для  $k = 0, \dots, T-1$ :

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

## Сходимость

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

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

Просуммируем полученное неравенство для  $k = 0, \dots, T - 1$ :

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

# Сходимость

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

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

Просуммируем полученное неравенство для  $k = 0, \dots, T-1$ :

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

- Запишем, насколько близко мы подошли к оптимуму  $x^* = \arg \min_{x \in \mathbb{R}^n} f(x) = \arg f^*$  на последней итерации:

# Сходимость

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

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

Просуммируем полученное неравенство для  $k = 0, \dots, T-1$ :

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

- Запишем, насколько близко мы подошли к оптимуму  $x^* = \arg \min_{x \in \mathbb{R}^n} f(x) = \arg f^*$  на последней итерации:
- Для субградиента:  $\langle g_k, x^* - x_k \rangle \leq f(x^*) - f(x_k)$ .

# Сходимость

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

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

Просуммируем полученное неравенство для  $k = 0, \dots, T-1$ :

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

- Запишем, насколько близко мы подошли к оптимуму  $x^* = \arg \min_{x \in \mathbb{R}^n} f(x) = \arg f^*$  на последней итерации:
- Для субградиента:  $\langle g_k, x^* - x_k \rangle \leq f(x^*) - f(x_k)$ .
- Дополнительно предположим, что  $\|g_k\|^2 \leq G^2$

# Сходимость

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

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

Просуммируем полученное неравенство для  $k = 0, \dots, T-1$ :

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

- Запишем, насколько близко мы подошли к оптимуму  $x^* = \arg \min_{x \in \mathbb{R}^n} f(x) = \arg f^*$  на последней итерации:
- Для субградиента:  $\langle g_k, x^* - x_k \rangle \leq f(x^*) - f(x_k)$ .
- Дополнительно предположим, что  $\|g_k\|^2 \leq G^2$
- Используем обозначение  $R = \|x_0 - x^*\|_2$

# Сходимость

- Наконец, заметим:

$$\sum_{k=0}^{T-1} 2\alpha_k(f(x_k) - f(x^*)) \geq \sum_{k=0}^{T-1} 2\alpha_k(f_k^{\text{best}} - f(x^*)) = (f_k^{\text{best}} - f(x^*)) \sum_{k=0}^{T-1} 2\alpha_k$$

## Сходимость

- Наконец, заметим:

$$\sum_{k=0}^{T-1} 2\alpha_k(f(x_k) - f(x^*)) \geq \sum_{k=0}^{T-1} 2\alpha_k(f_k^{\text{best}} - f(x^*)) = (f_k^{\text{best}} - f(x^*)) \sum_{k=0}^{T-1} 2\alpha_k$$

- Которое приводит к основному неравенству:

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

## Сходимость

- Наконец, заметим:

$$\sum_{k=0}^{T-1} 2\alpha_k(f(x_k) - f(x^*)) \geq \sum_{k=0}^{T-1} 2\alpha_k(f_k^{\text{best}} - f(x^*)) = (f_k^{\text{best}} - f(x^*)) \sum_{k=0}^{T-1} 2\alpha_k$$

- Которое приводит к основному неравенству:

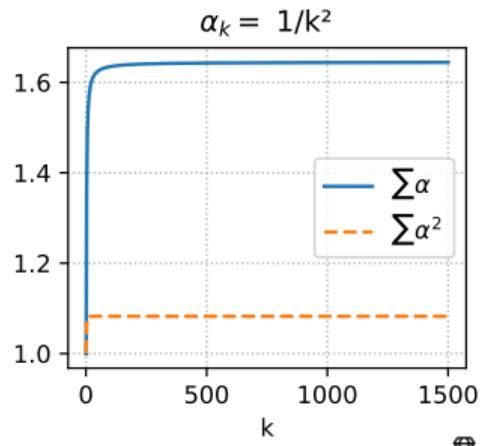
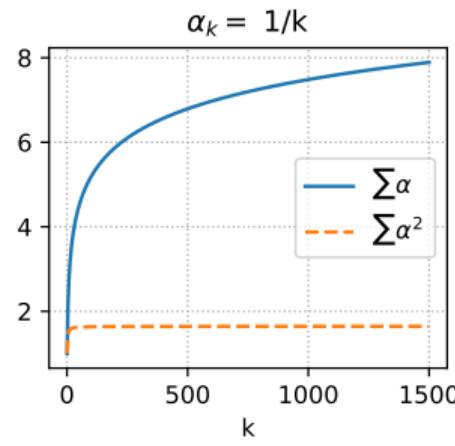
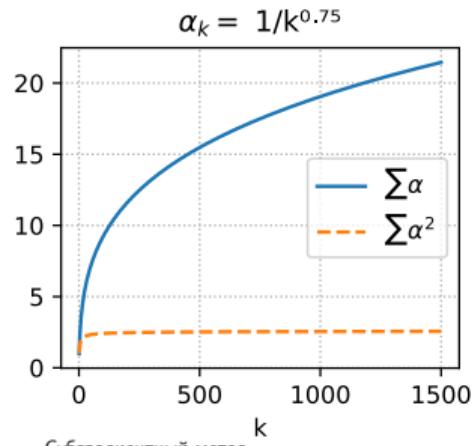
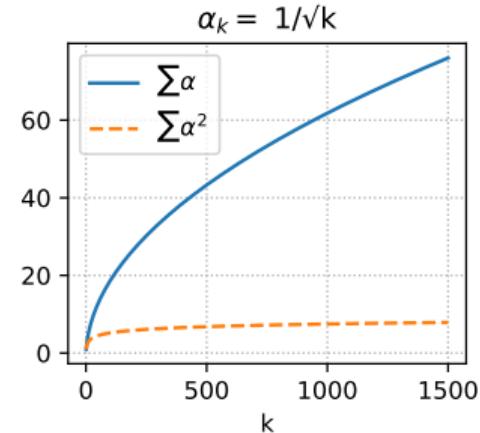
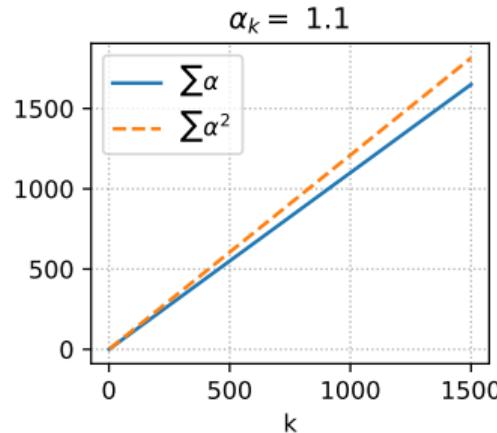
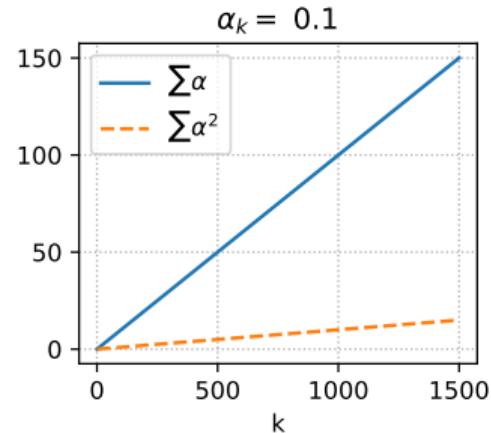
$$f_k^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2}{2 \sum_{k=0}^{T-1} \alpha_k}$$

- Из этого мы можем видеть, что если стратегия шага такая, что

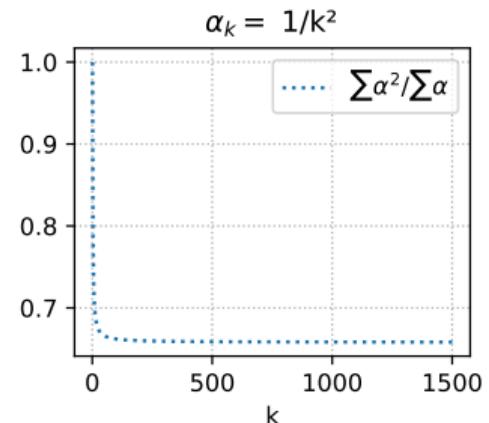
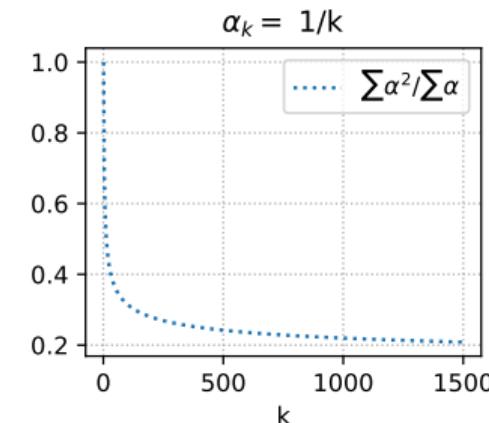
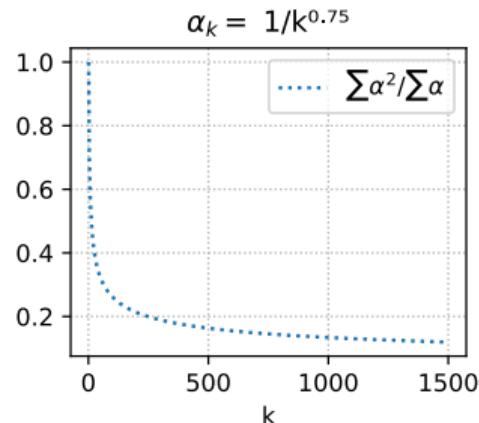
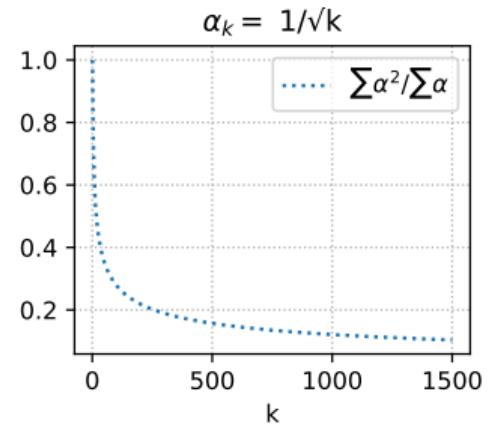
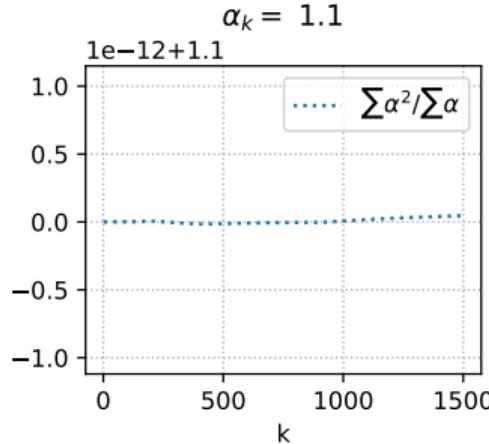
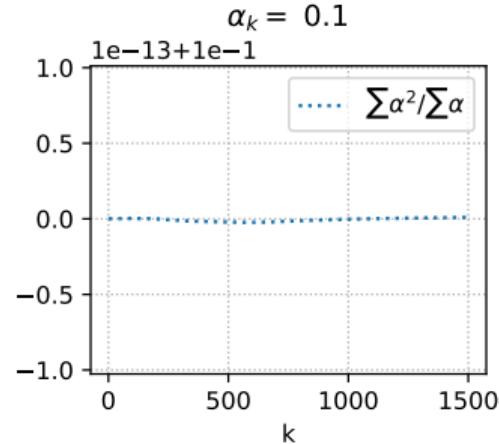
$$\sum_{k=0}^{T-1} \alpha_k^2 < \infty, \quad \sum_{k=0}^{T-1} \alpha_k = \infty,$$

то метод субградиента сходится (шаг должен быть убывающим, но не слишком быстрым).

## Разные стратегии шага



## Разные стратегии шага



## Оценки сходимости. Негладкий выпуклый случай. Постоянный шаг

### Theorem

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

## Оценки сходимости. Негладкий выпуклый случай. Постоянный шаг

### Theorem

Пусть  $f$  — выпуклая  $G$ -липшицева функция и  $R = \|x_0 - x^*\|_2$ . Для постоянного шага  $\alpha$ , метод субградиента 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.
- Некоторые варианты метода субградиента (например, убывающие, но несуммируемые длины шагов) работают даже когда предположение  $\|g_k\|_2 \leq G$  не выполняется; см. <sup>1</sup> или <sup>2</sup>.

<sup>1</sup>B. Polyak. Introduction to Optimization. Optimization Software, Inc., 1987.

<sup>2</sup>N. Shor. Minimization Methods for Non-differentiable Functions. Springer Series in Computational Mathematics. Springer, 1985.

## Оценки сходимости. Негладкий выпуклый случай. Постоянный шаг

### Theorem

Пусть  $f$  — выпуклая  $G$ -липшицева функция и  $R = \|x_0 - x^*\|_2$ . Для постоянного шага  $\alpha$ , метод субградиента 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.
- Некоторые варианты метода субградиента (например, убывающие, но несуммируемые длины шагов) работают даже когда предположение  $\|g_k\|_2 \leq G$  не выполняется; см. <sup>1</sup> или <sup>2</sup>.
- Найдем оптимальный шаг  $\alpha$  который минимизирует правую часть неравенства.

<sup>1</sup>B. Polyak. Introduction to Optimization. Optimization Software, Inc., 1987.

<sup>2</sup>N. Shor. Minimization Methods for Non-differentiable Functions. Springer Series in Computational Mathematics. Springer, 1985.

## Theorem

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

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

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

## Theorem

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

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

- Эта версия требует знания числа итераций заранее, что не всегда практично.
- Интересно отметить, что если мы хотим найти оптимальные шаги для всей последовательности  $\alpha_0, \alpha_1, \dots, \alpha_{k-1}$ , мы получим тот же результат.

## Оценки сходимости. Негладкий выпуклый случай. Постоянный шаг

### Theorem

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

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

- Эта версия требует знания числа итераций заранее, что не всегда практично.
- Интересно отметить, что если мы хотим найти оптимальные шаги для всей последовательности  $\alpha_0, \alpha_1, \dots, \alpha_{k-1}$ , мы получим тот же результат.
- Почему? Потому что правая часть является выпуклой и **симметричной** функцией  $\alpha_0, \alpha_1, \dots, \alpha_{k-1}$ .

## Theorem

Пусть  $f$  — выпуклая  $G$ -липшицева функция и  $R = \|x_0 - x^*\|_2$ . Для постоянного шага  $\gamma = \alpha_k \|g_k\|_2$ , т.е.  $\alpha_k = \frac{\gamma}{\|g_k\|_2}$ , метод субградиента удовлетворяет

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

- Заметим, что для метода субградиента, мы обычно не можем использовать норму субградиента как критерий остановки (представьте  $f(x) = |x|$ ). Существуют более продвинутые варианты критериев остановки, но из-за очень медленной сходимости обычно просто задают максимальное число итераций.

## Theorem

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

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

### 1. Оценка сумм:

## Theorem

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

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

### 1. Оценка сумм:

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

Theorem

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

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

1. Оценка сумм:

$$\sum_{k=0}^{T-1} \alpha_k^2 = \frac{R^2}{G^2} \sum_{k=1}^T \frac{1}{k} \leq \frac{R^2}{G^2} (1 + \ln T); \quad \sum_{k=0}^{T-1} \alpha_k = \frac{R}{G} \sum_{k=1}^T \frac{1}{\sqrt{k}} \geq \frac{R}{G} \int_1^{T+1} \frac{1}{\sqrt{t}} dt = \frac{2R}{G} (\sqrt{T+1} - 1).$$

Theorem

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

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

1. Оценка сумм:

$$\sum_{k=0}^{T-1} \alpha_k^2 = \frac{R^2}{G^2} \sum_{k=1}^T \frac{1}{k} \leq \frac{R^2}{G^2} (1 + \ln T); \quad \sum_{k=0}^{T-1} \alpha_k = \frac{R}{G} \sum_{k=1}^T \frac{1}{\sqrt{k}} \geq \frac{R}{G} \int_1^{T+1} \frac{1}{\sqrt{t}} dt = \frac{2R}{G} (\sqrt{T+1} - 1).$$

2. Мы отбрасываем последний  $-1$  в верхней оценке выше и используем основное неравенство:

Theorem

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

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

1. Оценка сумм:

$$\sum_{k=0}^{T-1} \alpha_k^2 = \frac{R^2}{G^2} \sum_{k=1}^T \frac{1}{k} \leq \frac{R^2}{G^2} (1 + \ln T); \quad \sum_{k=0}^{T-1} \alpha_k = \frac{R}{G} \sum_{k=1}^T \frac{1}{\sqrt{k}} \geq \frac{R}{G} \int_1^{T+1} \frac{1}{\sqrt{t}} dt = \frac{2R}{G} (\sqrt{T+1} - 1).$$

2. Мы отбрасываем последний  $-1$  в верхней оценке выше и используем основное неравенство:

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

Theorem

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

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

1. Оценка сумм:

$$\sum_{k=0}^{T-1} \alpha_k^2 = \frac{R^2}{G^2} \sum_{k=1}^T \frac{1}{k} \leq \frac{R^2}{G^2} (1 + \ln T); \quad \sum_{k=0}^{T-1} \alpha_k = \frac{R}{G} \sum_{k=1}^T \frac{1}{\sqrt{k}} \geq \frac{R}{G} \int_1^{T+1} \frac{1}{\sqrt{t}} dt = \frac{2R}{G} (\sqrt{T+1} - 1).$$

2. Мы отбрасываем последний  $-1$  в верхней оценке выше и используем основное неравенство:

$$f_T^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2}{2 \sum_{k=0}^{T-1} \alpha_k} \leq \frac{R^2 + R^2(1 + \ln T)}{4 \frac{R}{G} (\sqrt{T+1})}$$

Theorem

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

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

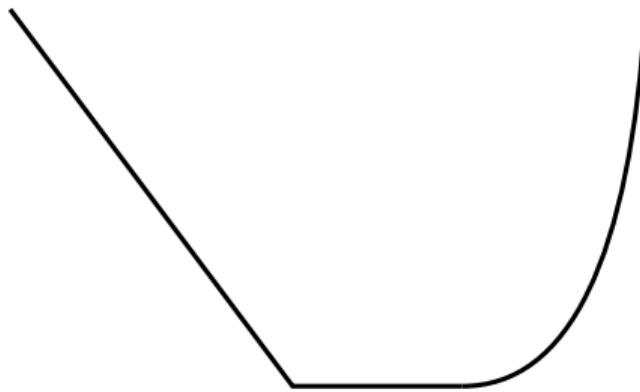
1. Оценка сумм:

$$\sum_{k=0}^{T-1} \alpha_k^2 = \frac{R^2}{G^2} \sum_{k=1}^T \frac{1}{k} \leq \frac{R^2}{G^2} (1 + \ln T); \quad \sum_{k=0}^{T-1} \alpha_k = \frac{R}{G} \sum_{k=1}^T \frac{1}{\sqrt{k}} \geq \frac{R}{G} \int_1^{T+1} \frac{1}{\sqrt{t}} dt = \frac{2R}{G} (\sqrt{T+1} - 1).$$

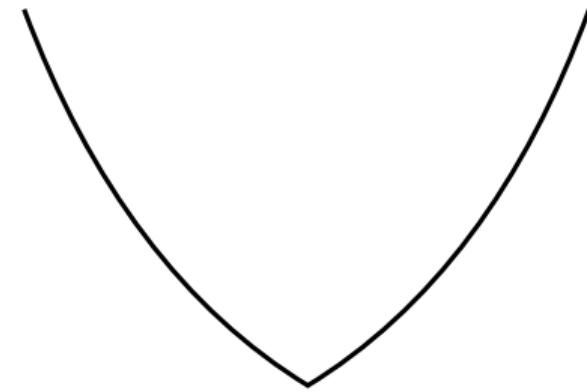
2. Мы отбрасываем последний  $-1$  в верхней оценке выше и используем основное неравенство:

$$f_T^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum_{k=0}^{T-1} \alpha_k^2}{2 \sum_{k=0}^{T-1} \alpha_k} \leq \frac{R^2 + R^2(1 + \ln T)}{4 \frac{R}{G} (\sqrt{T+1})} = \frac{GR(2 + \ln T)}{4\sqrt{T+1}}$$

## Негладкий сильно выпуклый случай

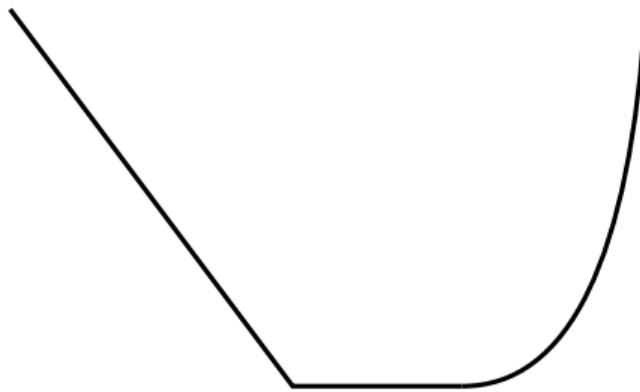


Негладкая  
Выпуклая



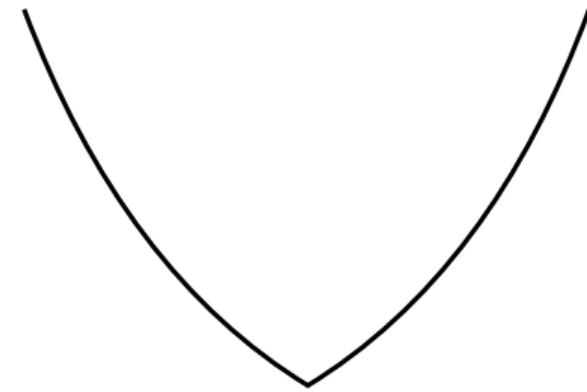
Негладкая  
 $\mu$  - сильно выпуклая

## Негладкий сильно выпуклый случай



Негладкая  
Выпуклая

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



Негладкая  
 $\mu$  - сильно выпуклая

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

## Негладкий сильно выпуклый случай

### Theorem

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

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

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

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

## Негладкий сильно выпуклый случай

### Theorem

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

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

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

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

## Негладкий сильно выпуклый случай

### Theorem

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

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

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

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

3. Таким образом,

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

## Негладкий сильно выпуклый случай

### Theorem

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

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

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

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

3. Таким образом,

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

## Оценки сходимости. Негладкий сильно выпуклый случай.

### Theorem

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

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

1. Начнем с формулировки метода как раньше:

## Оценки сходимости. Негладкий сильно выпуклый случай.

### Theorem

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

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

1. Начнем с формулировки метода как раньше:

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

## Оценки сходимости. Негладкий сильно выпуклый случай.

### Theorem

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

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

1. Начнем с формулировки метода как раньше:

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

## Оценки сходимости. Негладкий сильно выпуклый случай.

### Theorem

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

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

1. Начнем с формулировки метода как раньше:

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

## Оценки сходимости. Негладкий сильно выпуклый случай.

### Theorem

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

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

1. Начнем с формулировки метода как раньше:

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

## Оценки сходимости. Негладкий сильно выпуклый случай.

### Theorem

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

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

1. Начнем с формулировки метода как раньше:

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай.

### Theorem

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

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

1. Начнем с формулировки метода как раньше:

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

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

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

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

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

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

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

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

## Оценки сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

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

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

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

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

## Summary. Метод субградиента

Тип задачи	Стратегия шага	Скорость сходимости	Сложность итераций
Выпуклые и липшицевые задачи	$\alpha \sim \frac{1}{\sqrt{k}}$	$\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$	$\mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$
Сильно выпуклые и липшицевые задачи	$\alpha \sim \frac{1}{k}$	$\mathcal{O}\left(\frac{1}{k}\right)$	$\mathcal{O}\left(\frac{1}{\varepsilon}\right)$

## Численные эксперименты

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

Linear Least Squares with  $\ell_1$  Regularization (LASSO).  
m=1000, n=100,  $\lambda=0$ ,  $\mu=0$ ,  $L=10$ . Optimal sparsity: 0.0e+00

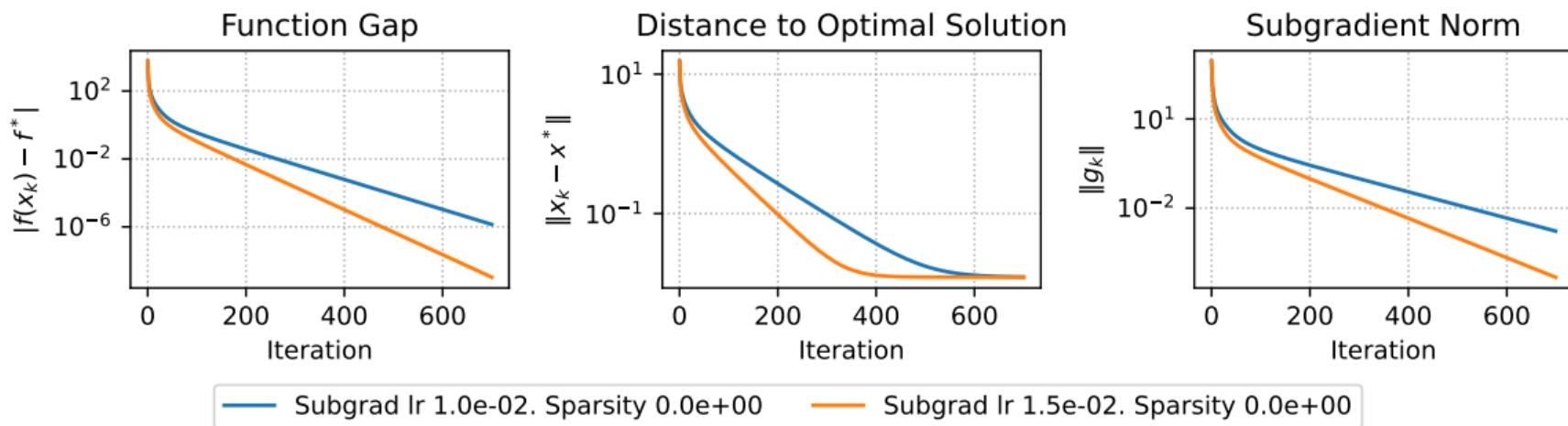


Рис. 6: Гладкий выпуклый случай. Сублинейная сходимость, не сходится в области определения

## Численные эксперименты

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

Linear Least Squares with  $\ell_1$  Regularization (LASSO).  
 $m=1000, n=100, \lambda=0.1, \mu=0, L=10$ . Optimal sparsity: 1.0e-02

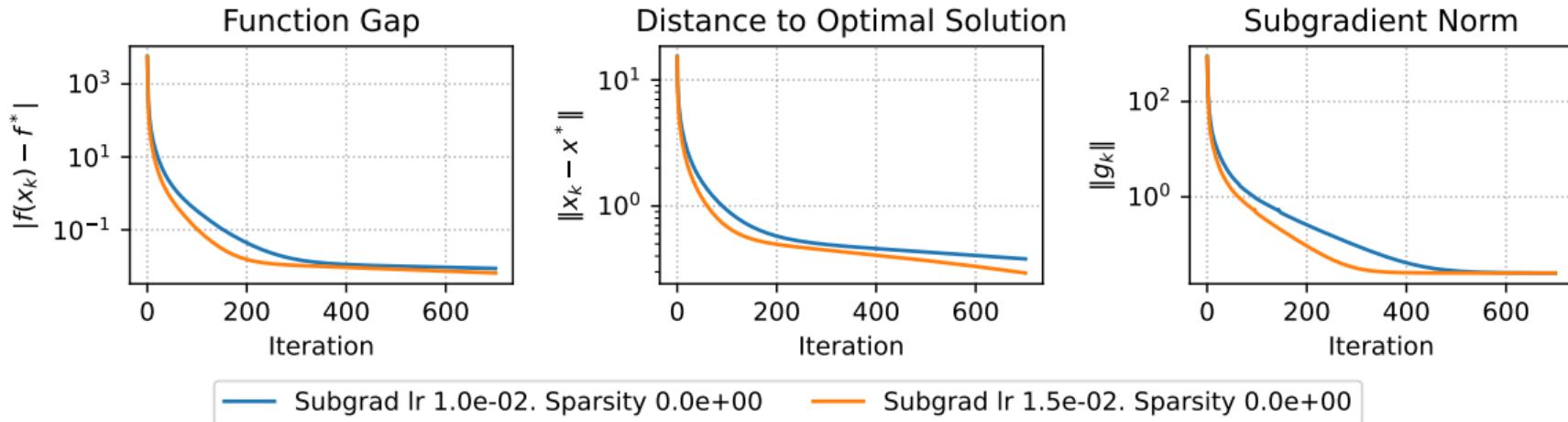


Рис. 7: Негладкий выпуклый случай. Маленькое значение  $\lambda$  приводит к негладкости. Не сходится с постоянным шагом

## Численные эксперименты

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

Linear Least Squares with  $\ell_1$  Regularization (LASSO).  
 $m=1000, n=100, \lambda=1, \mu=0, L=10$ . Optimal sparsity: 7.0e-02

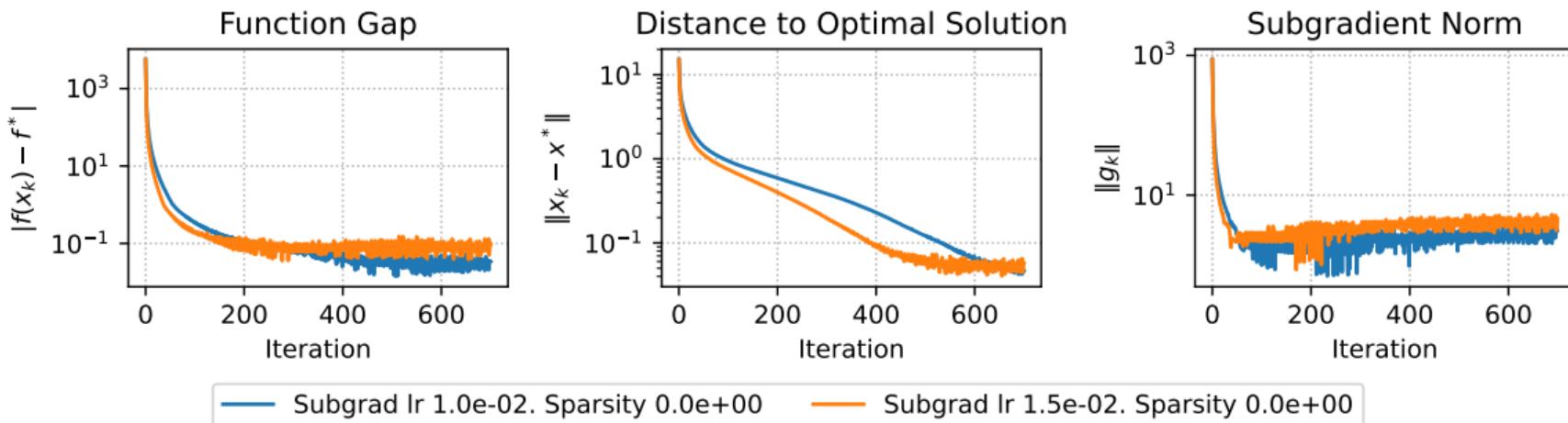


Рис. 8: Негладкий выпуклый случай. При большем значении  $\lambda$  проявляется немонотонность  $f(x_k)$ . Видно, что меньшая постоянная длина шага приводит к более низкому стационарному уровню функции.

## Численные эксперименты

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

Linear Least Squares with  $\ell_1$  Regularization (LASSO).  
 $m=100, n=100, \lambda=1, \mu=0, L=10$ . Optimal sparsity:  $2.3e-01$

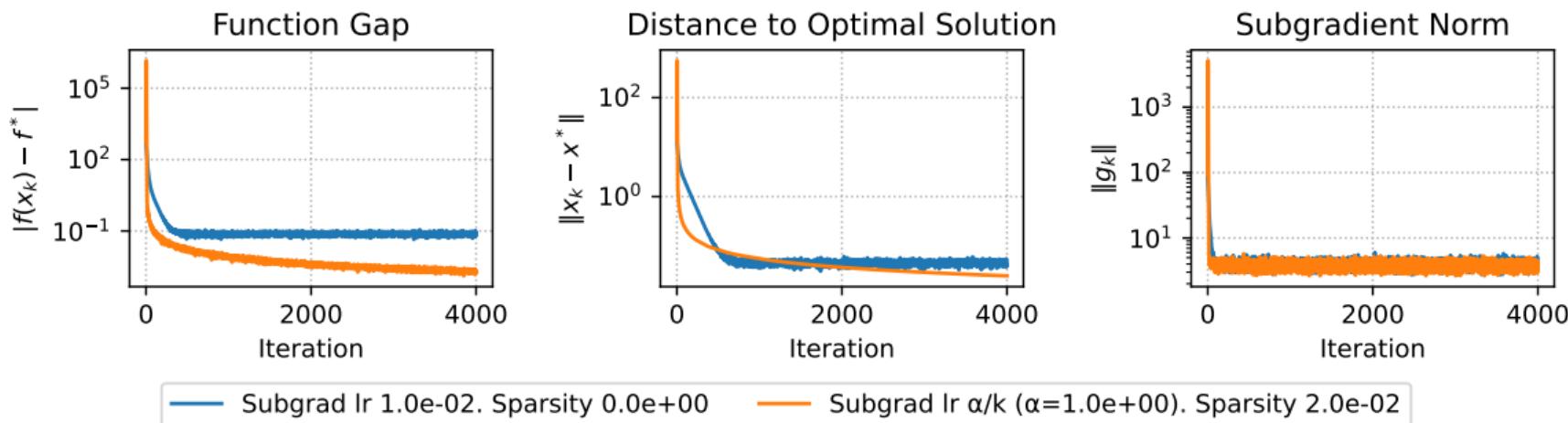


Рис. 9: Негладкий выпуклый случай. Убывающая длина шага приводит к сходимости для  $f_k^{\text{best}}$

## Численные эксперименты

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

Linear Least Squares with  $\ell_1$  Regularization (LASSO).  
 $m=100, n=100, \lambda=1, \mu=0, L=10$ . Optimal sparsity: 2.3e-01

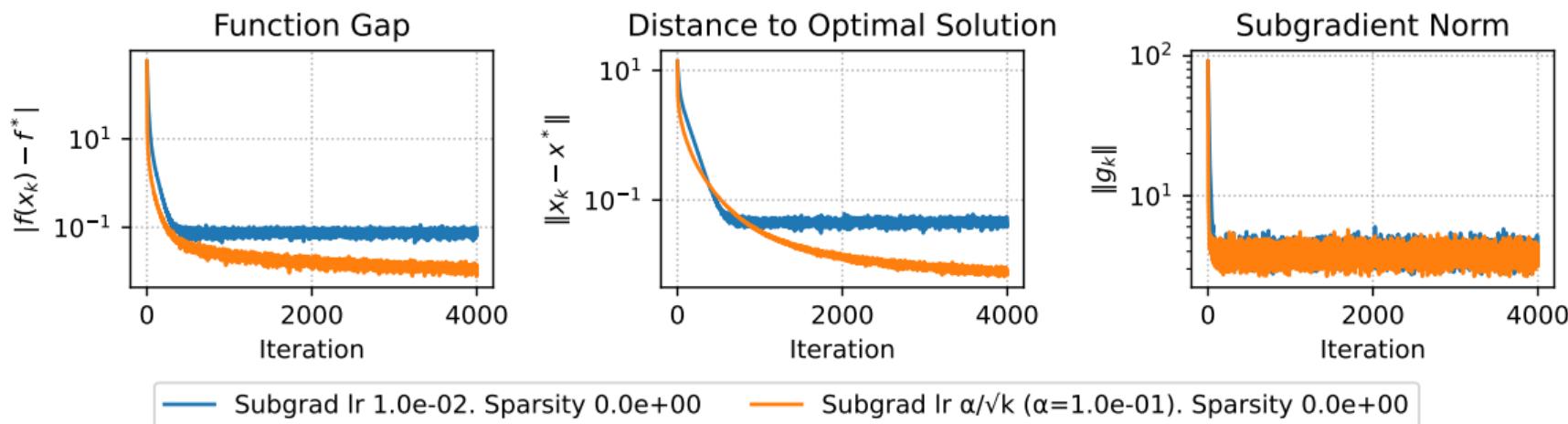


Рис. 10: Негладкий выпуклый случай.  $\frac{\alpha_0}{\sqrt{k}}$  шаг приводит к сходимости для  $f_k^{\text{best}}$

## Численные эксперименты

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

Linear Least Squares with  $\ell_1$  Regularization (LASSO).  
 $m=100, n=100, \lambda=1, \mu=0, L=10$ . Optimal sparsity:  $2.3e-01$

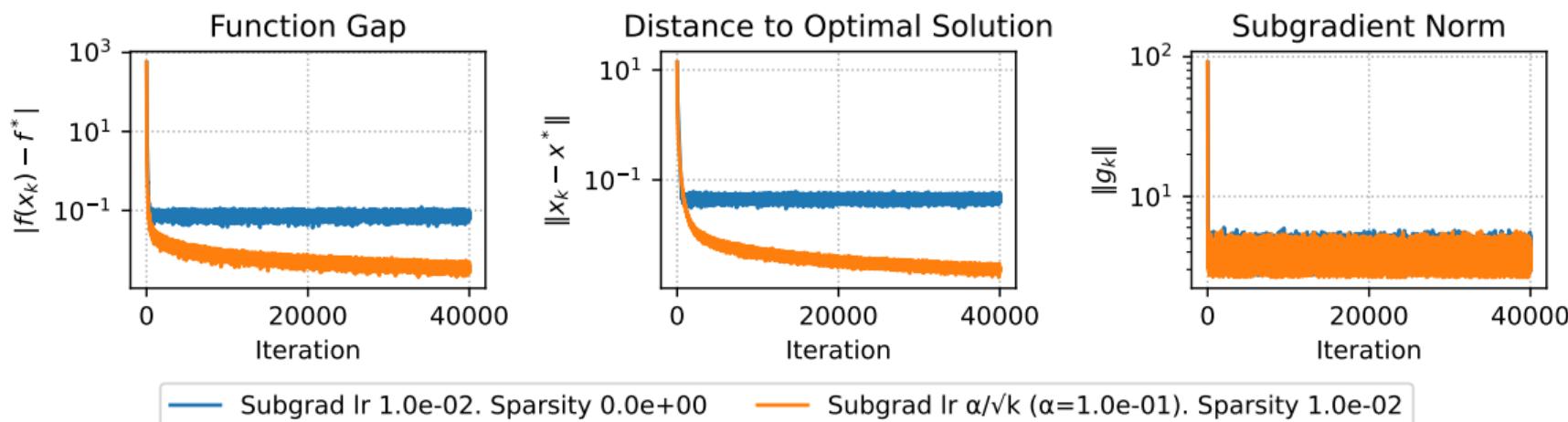


Рис. 11: Негладкий выпуклый случай.  $\frac{\alpha_0}{\sqrt{k}}$  шаг приводит к сходимости для  $f_k^{\text{best}}$

## Численные эксперименты

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

Linear Least Squares with  $\ell_1$  Regularization (LASSO).  
 $m=100, n=100, \lambda=1, \mu=1, L=10$ . Optimal sparsity:  $2.0e-01$

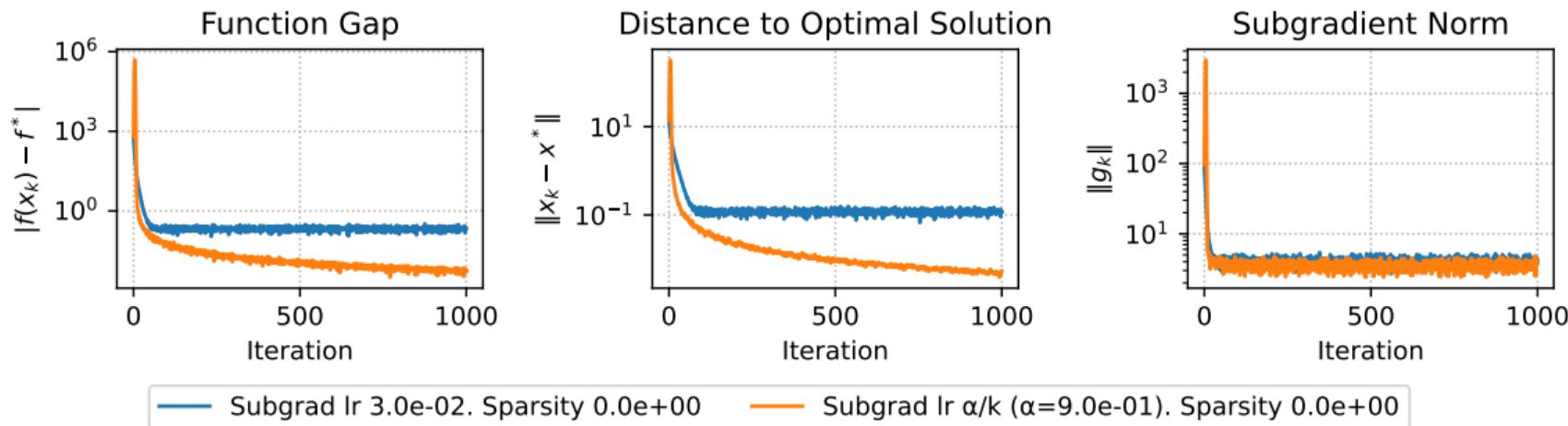


Рис. 12: Негладкий сильно выпуклый случай.  $\frac{\alpha_0}{k}$  шаг приводит к сходимости для  $f_k^{\text{best}}$

## Численные эксперименты

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

Linear Least Squares with  $\ell_1$  Regularization (LASSO).  
 $m=100, n=100, \lambda=1, \mu=1, L=10$ . Optimal sparsity:  $2.0e-01$

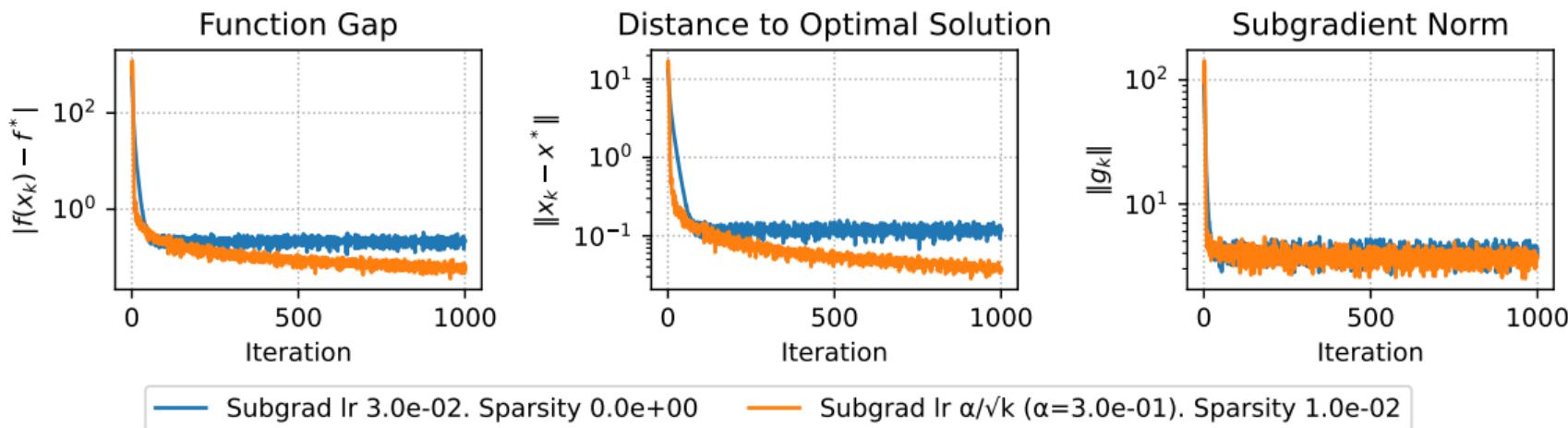


Рис. 13: Негладкий сильно выпуклый случай. Шаг  $\frac{\alpha_0}{\sqrt{k}}$  работает хуже

## Численные эксперименты

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

Binary Logistic Regression with  $\ell_1$  Regularization.  
 $m=300$ ,  $n=50$ ,  $\lambda=0.1$ . Optimal sparsity: 8.6e-01

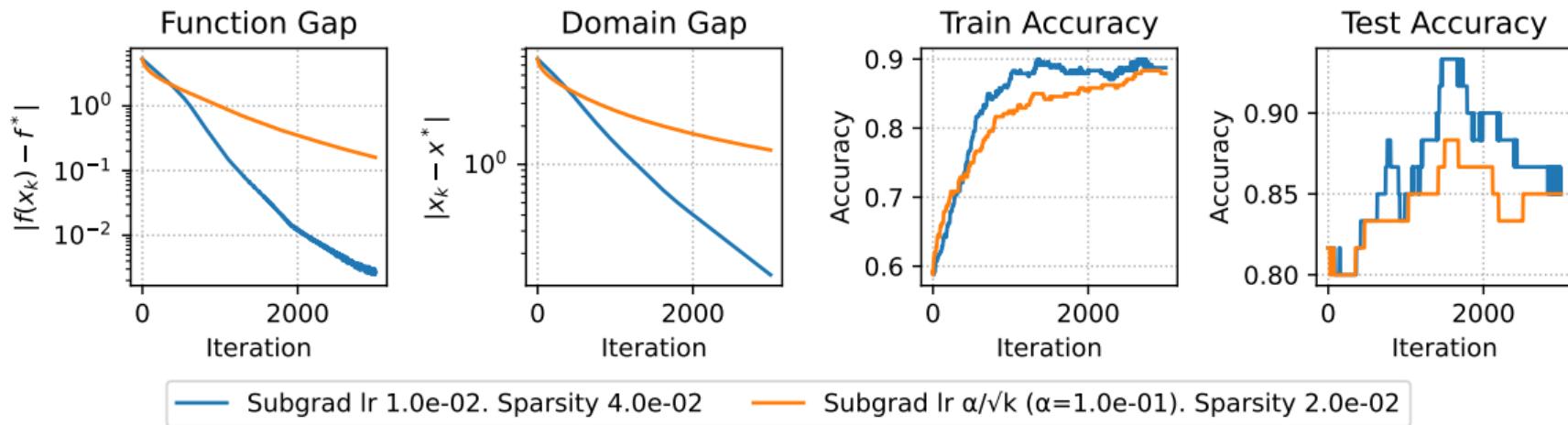


Рис. 14: Логистическая регрессия с  $\ell_1$  регуляризацией

## Численные эксперименты

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

Binary Logistic Regression with  $\ell_1$  Regularization.  
 $m=300$ ,  $n=50$ ,  $\lambda=0.1$ . Optimal sparsity: 8.6e-01

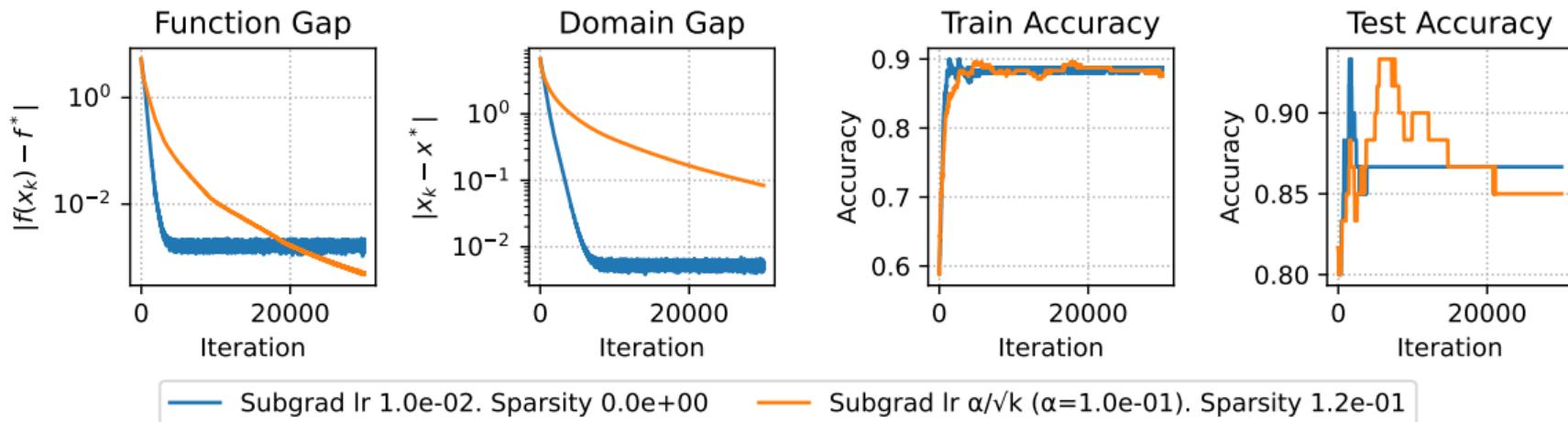


Рис. 15: Логистическая регрессия с  $\ell_1$  регуляризацией

## Численные эксперименты

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

Binary Logistic Regression with  $\ell_1$  Regularization.  
 $m=300$ ,  $n=50$ ,  $\lambda=0.25$ . Optimal sparsity: 9.6e-01

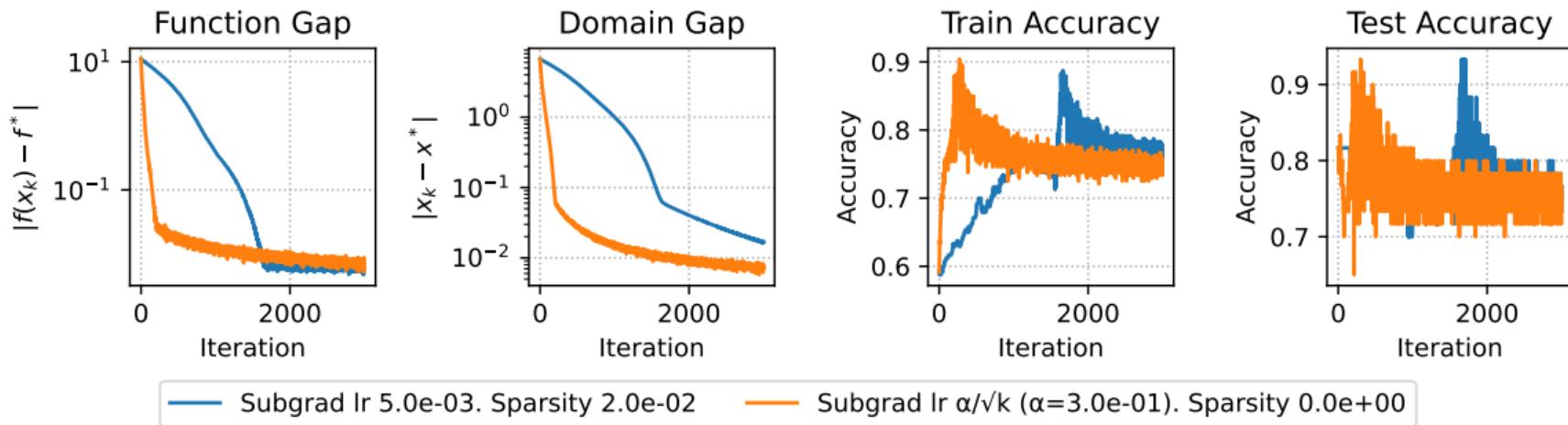


Рис. 16: Логистическая регрессия с  $\ell_1$  регуляризацией

## Численные эксперименты

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

Binary Logistic Regression with  $\ell_1$  Regularization.  
 $m=300$ ,  $n=50$ ,  $\lambda=0.25$ . Optimal sparsity: 9.6e-01

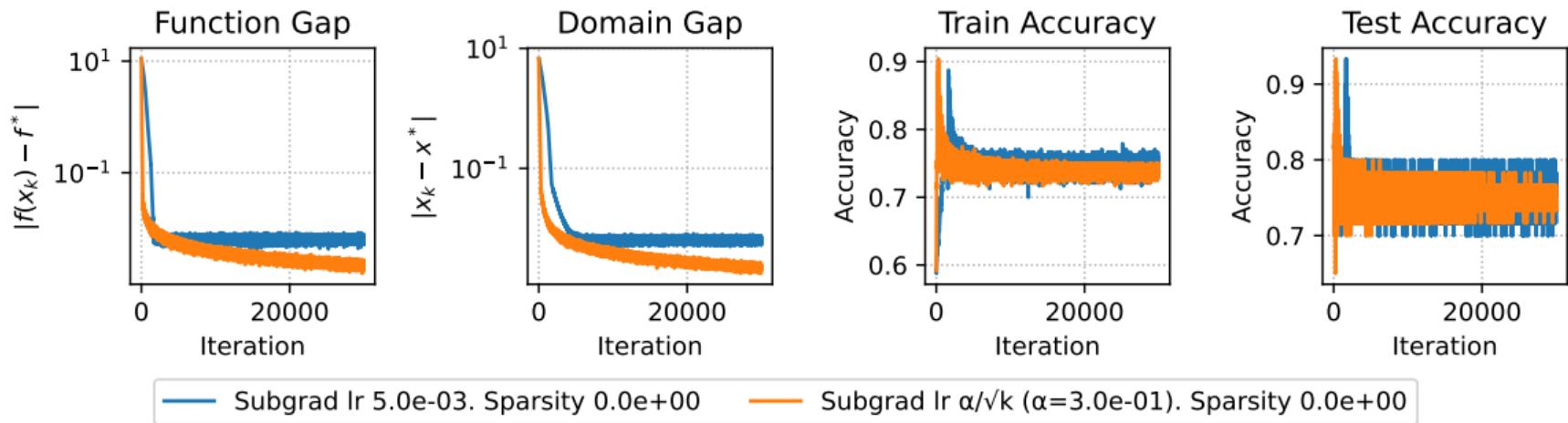


Рис. 17: Логистическая регрессия с  $\ell_1$  регуляризацией

## Численные эксперименты

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

Binary Logistic Regression with  $\ell_1$  Regularization.  
 $m=300$ ,  $n=50$ ,  $\lambda=0.27$ . Optimal sparsity: 1.0e+00

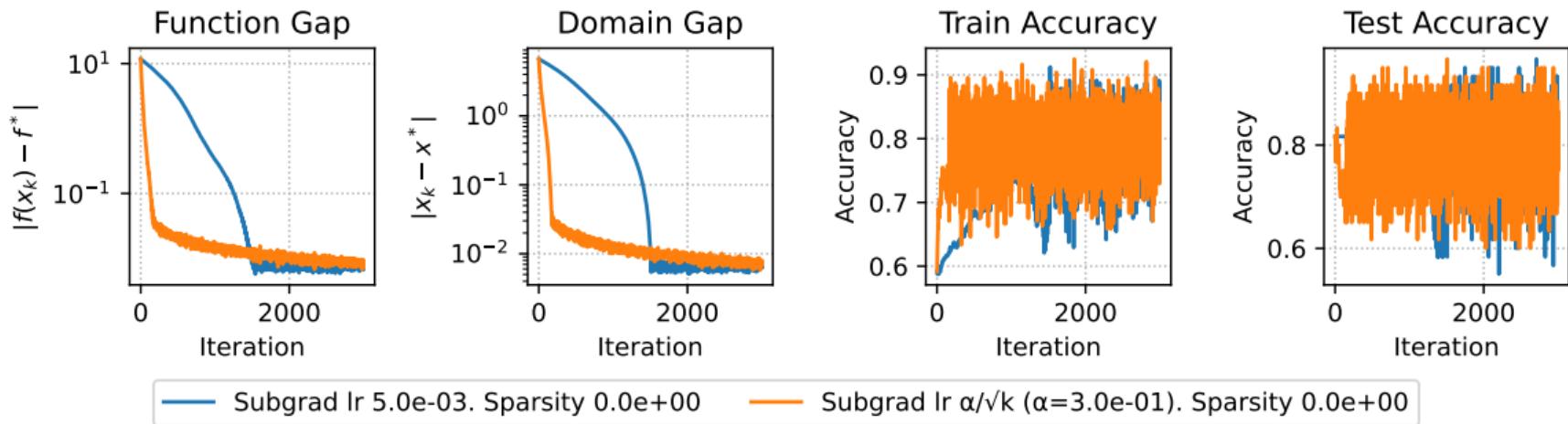


Рис. 18: Логистическая регрессия с  $\ell_1$  регуляризацией

## Численные эксперименты

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

Binary Logistic Regression with  $\ell_1$  Regularization.  
 $m=300$ ,  $n=50$ ,  $\lambda=0.27$ . Optimal sparsity:  $1.0e+00$

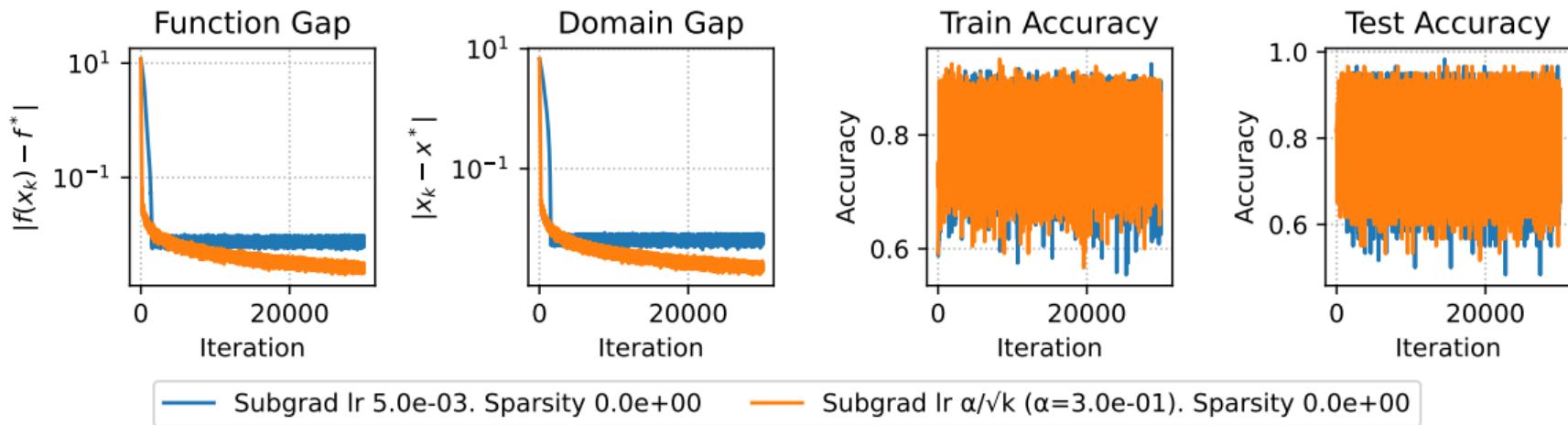


Рис. 19: Логистическая регрессия с  $\ell_1$  регуляризацией

## Нижние оценки

## Нижние оценки

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

<sup>3</sup>Nesterov, Lectures on Convex Optimization

<sup>4</sup>Carmon, Duchi, Hinder, Sidford, 2017

<sup>5</sup>Nemirovski, Yudin, 1979

# Итерация «чёрного ящика»

Итерация градиентного спуска:

$$\begin{aligned}x^{k+1} &= x^k - \alpha^k \nabla f(x^k) \\&= x^{k-1} - \alpha^{k-1} \nabla f(x^{k-1}) - \alpha^k \nabla f(x^k) \\&\quad \vdots \\&= x^0 - \sum_{i=0}^k \alpha^{k-i} \nabla f(x^{k-i})\end{aligned}$$

## Итерация «чёрного ящика»

Итерация градиентного спуска:

$$\begin{aligned}x^{k+1} &= x^k - \alpha^k \nabla f(x^k) \\&= x^{k-1} - \alpha^{k-1} \nabla f(x^{k-1}) - \alpha^k \nabla f(x^k) \\&\quad \vdots \\&= x^0 - \sum_{i=0}^k \alpha^{k-i} \nabla f(x^{k-i})\end{aligned}$$

Рассмотрим семейство методов первого порядка, где

$$\begin{aligned}x^{k+1} &\in x^0 + \text{span} \{ \nabla f(x^0), \nabla f(x^1), \dots, \nabla f(x^k) \} && f - \text{smooth} \\x^{k+1} &\in x^0 + \text{span} \{ g_0, g_1, \dots, g_k \}, \text{ where } g_i \in \partial f(x^i) && f - \text{non-smooth}\end{aligned}\tag{1}$$

## Итерация «чёрного ящика»

Итерация градиентного спуска:

$$\begin{aligned}x^{k+1} &= x^k - \alpha^k \nabla f(x^k) \\&= x^{k-1} - \alpha^{k-1} \nabla f(x^{k-1}) - \alpha^k \nabla f(x^k) \\&\quad \vdots \\&= x^0 - \sum_{i=0}^k \alpha^{k-i} \nabla f(x^{k-i})\end{aligned}$$

Рассмотрим семейство методов первого порядка, где

$$\begin{aligned}x^{k+1} &\in x^0 + \text{span} \{ \nabla f(x^0), \nabla f(x^1), \dots, \nabla f(x^k) \} && f - \text{smooth} \\x^{k+1} &\in x^0 + \text{span} \{ g_0, g_1, \dots, g_k \}, \text{ where } g_i \in \partial f(x^i) && f - \text{non-smooth}\end{aligned}\tag{1}$$

Чтобы построить нижнюю оценку, нужно найти функцию  $f$  из соответствующего класса так, чтобы любой метод из семейства Уравнение 1 сходился не быстрее этой нижней оценки.

## Негладкий выпуклый случай

### Theorem

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

$$\min_{i \in [1, k]} f(x^i) - \min_{x \in \mathbb{B}(R)} f(x) \geq \frac{GR}{2(1 + \sqrt{k})}$$

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

$$\min_{i \in [1, k]} f(x^i) - \min_{x \in \mathbb{B}(R)} f(x) \geq \frac{GR}{2(1 + \sqrt{k})}$$

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

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

$$\text{span} \{g_0, g_1, \dots, g_k\} \subset \text{span} \{e_1, e_2, \dots, e_i\}$$

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:

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

где  $\alpha, \beta \in \mathbb{R}$  — параметры, и  $x[1 : k]$  — первые  $k$  компонент  $x$ .

## Негладкий выпуклый случай (доказательство)

Рассмотрим функцию:

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

где  $\alpha, \beta \in \mathbb{R}$  — параметры, и  $x[1 : k]$  — первые  $k$  компонент  $x$ .

**Основные свойства:**

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

## Негладкий выпуклый случай (доказательство)

Рассмотрим функцию:

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

где  $\alpha, \beta \in \mathbb{R}$  — параметры, и  $x[1 : k]$  — первые  $k$  компонент  $x$ .

**Основные 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:

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

где  $\alpha, \beta \in \mathbb{R}$  — параметры, и  $x[1 : k]$  — первые  $k$  компонент  $x$ .

**Основные 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:

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

где  $\alpha, \beta \in \mathbb{R}$  — параметры, и  $x[1 : k]$  — первые  $k$  компонент  $x$ .

**Основные 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  $x$ :

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

## Негладкий выпуклый случай (доказательство)

Рассмотрим функцию:

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

где  $\alpha, \beta \in \mathbb{R}$  — параметры, и  $x[1 : k]$  — первые  $k$  компонент  $x$ .

**Основные 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  $x$ :

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

Легко видеть, что если  $g \in \partial f(x)$  и  $\|x\| \leq R$ , то  
 $\|g\| \leq \alpha R + \beta$

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  $i$  — первая координата для которой  $x[i] = \max_{1 \leq j \leq k} x[j]$ .

- Мы обеспечиваем  $\|x^0\| \leq R$  начиная с  $x^0 = 0$ .

## 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  $i$  — первая координата для которой  $x[i] = \max_{1 \leq j \leq k} x[j]$ .

- Мы обеспечиваем  $\|x^0\| \leq R$  начиная с  $x^0 = 0$ .
- При запросе оракула в  $x^0 = 0$ , он возвращает  $e_1$ . Следовательно,  $x^1$  должен лежать 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  $i$  — первая координата для которой  $x[i] = \max_{1 \leq j \leq k} x[j]$ .

- Мы обеспечиваем  $\|x^0\| \leq R$  начиная с  $x^0 = 0$ .
- При запросе оракула в  $x^0 = 0$ , он возвращает  $e_1$ . Следовательно,  $x^1$  должен лежать на прямой, порождённой  $e_1$ .
- По индукции, показывается, что для всех  $i$ , итерация  $x^i$  лежит в линейной оболочке  $\{e_1, \dots, e_i\}$ . 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)

- Остаётся вычислить минимальное значение  $f$ . Определим точку  $y \in \mathbb{R}^n$  как

$$y[i] = -\frac{\beta}{\alpha k} \quad \text{for } 1 \leq i \leq k, \quad y[i] = 0 \quad \text{for } k+1 \leq i \leq n.$$

## Негладкий выпуклый случай (доказательство)

- Остаётся вычислить минимальное значение  $f$ . Определим точку  $y \in \mathbb{R}^n$  как

$$y[i] = -\frac{\beta}{\alpha k} \quad \text{for } 1 \leq i \leq k, \quad y[i] = 0 \quad \text{for } k+1 \leq i \leq n.$$

- Заметим, что  $0 \in \partial f(y)$ :

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

## Негладкий выпуклый случай (доказательство)

- Остаётся вычислить минимальное значение  $f$ . Определим точку  $y \in \mathbb{R}^n$  как

$$y[i] = -\frac{\beta}{\alpha k} \quad \text{for } 1 \leq i \leq k, \quad y[i] = 0 \quad \text{for } k+1 \leq i \leq n.$$

- Заметим, что  $0 \in \partial f(y)$ :

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

- 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)

- Остаётся вычислить минимальное значение  $f$ . Определим точку  $y \in \mathbb{R}^n$  как

$$y[i] = -\frac{\beta}{\alpha k} \quad \text{for } 1 \leq i \leq k, \quad y[i] = 0 \quad \text{for } k+1 \leq i \leq n.$$

- Заметим, что  $0 \in \partial f(y)$ :

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

- 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:

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

## Негладкий выпуклый случай (доказательство)

У нас есть:  $f(x^i) - f(x^*) \geq \frac{\beta^2}{2\alpha k}$ , в то время как мы должны доказать, что  $\min_{i \in [1, k]} f(x^i) - f(x^*) \geq \frac{GR}{2(1+\sqrt{k})}$ .

## Негладкий выпуклый случай (доказательство)

У нас есть:  $f(x^i) - f(x^*) \geq \frac{\beta^2}{2\alpha k}$ , в то время как мы должны доказать, что  $\min_{i \in [1, k]} f(x^i) - f(x^*) \geq \frac{GR}{2(1 + \sqrt{k})}$ .

### 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

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

## Негладкий выпуклый случай (доказательство)

У нас есть:  $f(x^i) - f(x^*) \geq \frac{\beta^2}{2\alpha k}$ , в то время как мы должны доказать, что  $\min_{i \in [1, k]} f(x^i) - f(x^*) \geq \frac{GR}{2(1 + \sqrt{k})}$ .

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

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

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

$$\min_{i \in [1, k]} f(x^i) - f(x^*) \geq \frac{G^2}{8\alpha k}$$

# Ссылки

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