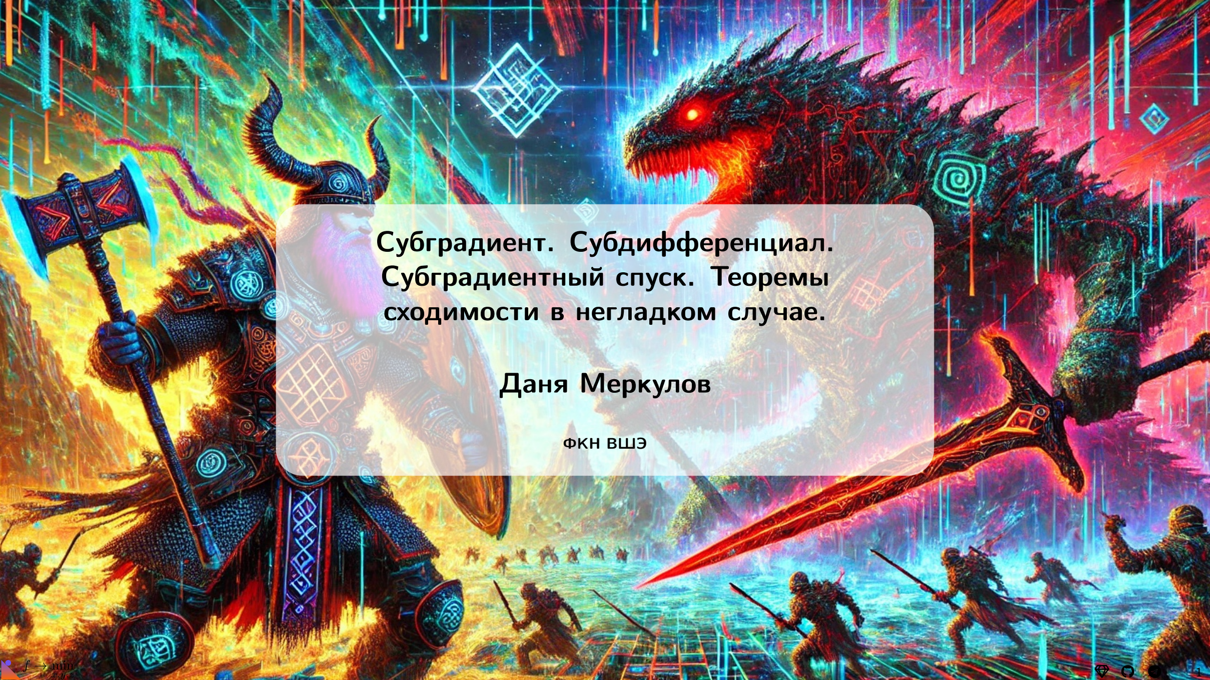


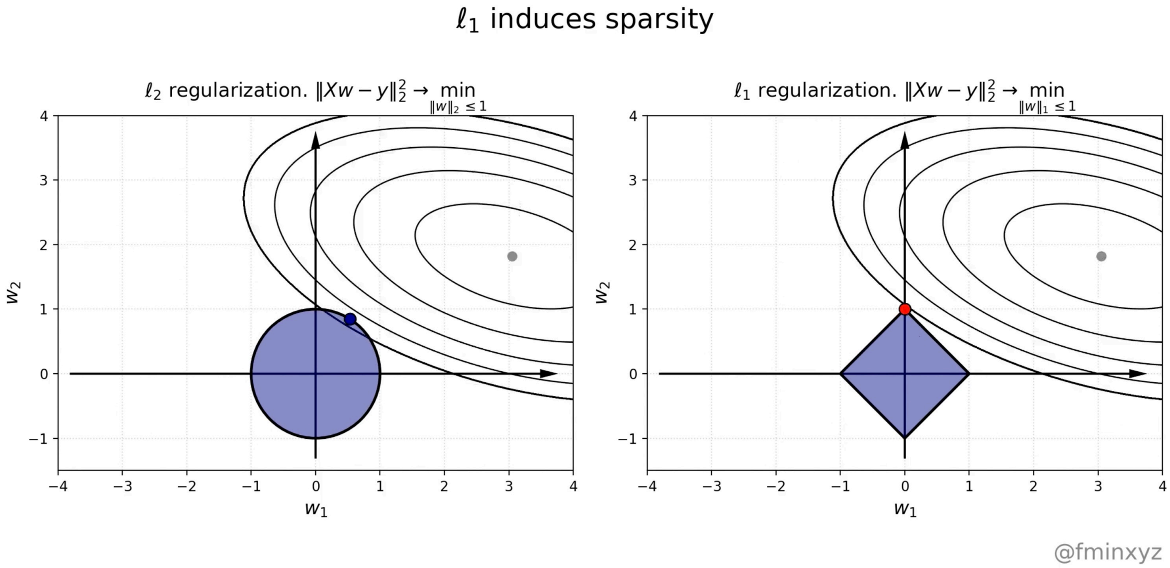
Субградиент. Субдифференциал. Субградиентный спуск. Теоремы сходимости в негладком случае.

Даня Меркулов

ФКН ВШЭ

Негладкие задачи

Линейный МНК с ℓ_1-регуляризацией



Нормы не являются гладкими

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

Рассмотрим классическую задачу выпуклой оптимизации. Предположим, что $f(x)$ выпукла, но теперь мы не требуем гладкости.

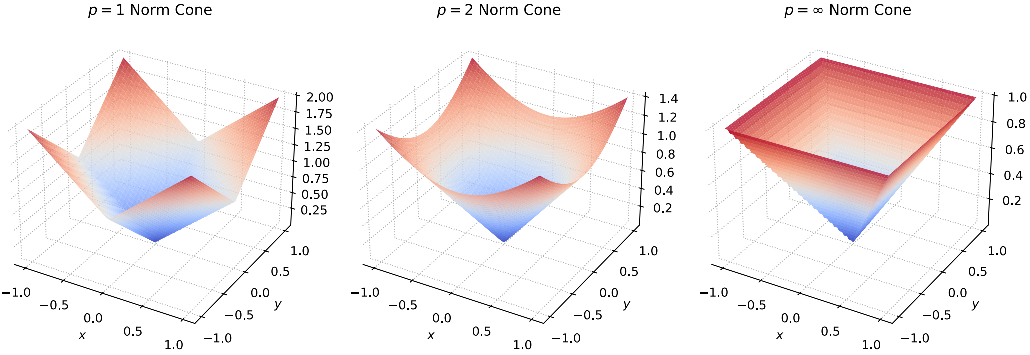


Figure 1: Конусы норм для разных p: нормы негладкие

Пример Вульфа

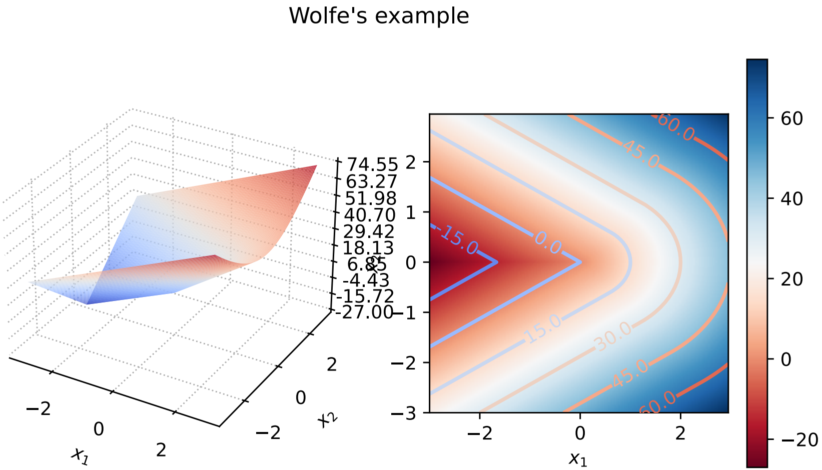


Figure 2: Пример Вульфа

Субградиентное исчисление

Линейная нижняя оценка выпуклой функции

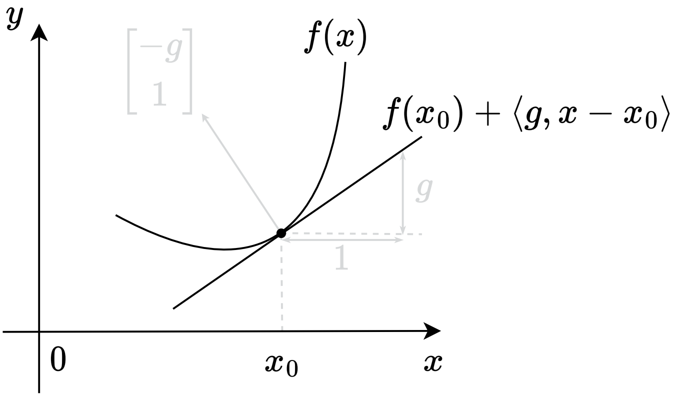


Figure 3: Линейная аппроксимация по Тейлору является глобальной нижней оценкой выпуклой функции

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

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

Линейная нижняя оценка выпуклой функции

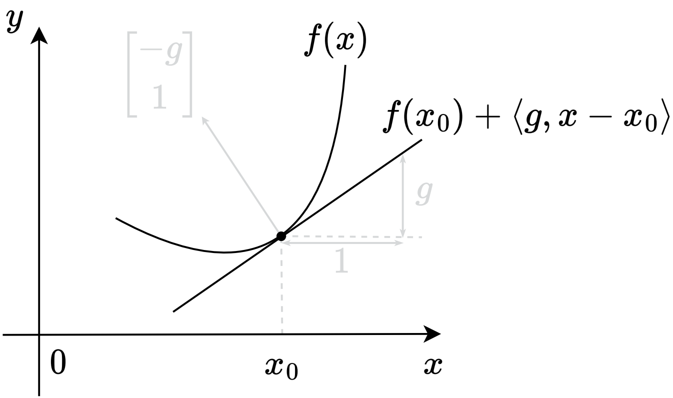


Figure 3: Линейная аппроксимация по Тейлору является глобальной нижней оценкой выпуклой функции

Важное свойство непрерывной выпуклой функции $f(x)$ состоит в том, что в любой выбранной точке 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, то есть касательная к графику функции даёт *глобальную* нижнюю оценку функции.

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

Линейная нижняя оценка выпуклой функции

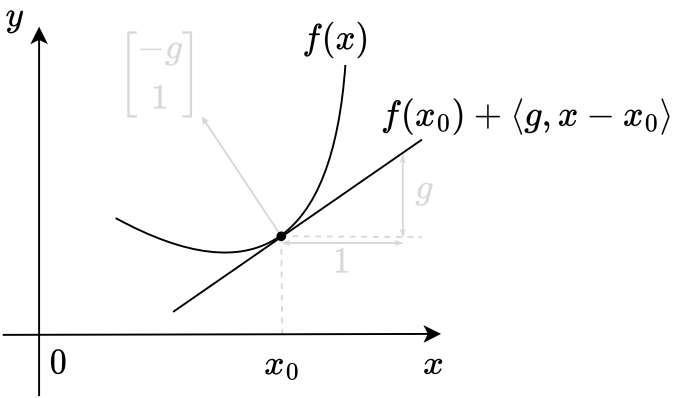


Figure 3: Линейная аппроксимация по Тейлору является глобальной нижней оценкой выпуклой функции

Важное свойство непрерывной выпуклой функции $f(x)$ состоит в том, что в любой выбранной точке 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, то есть касательная к графику функции даёт *глобальную* нижнюю оценку функции.

- Если $f(x)$ дифференцируема, то $g = \nabla f(x_0)$
- Не всякая непрерывная выпуклая функция дифференцируема.

Линейная нижняя оценка выпуклой функции

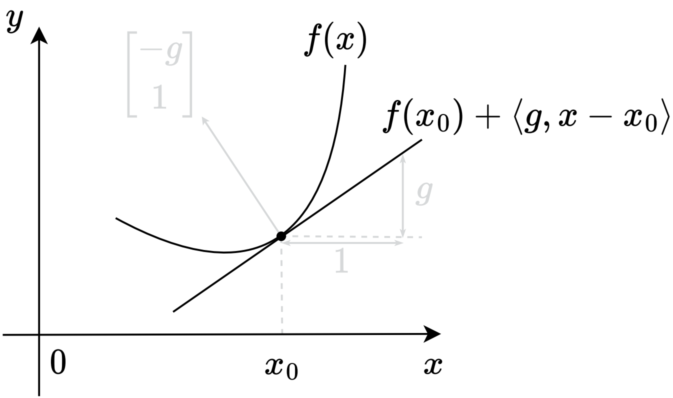


Figure 3: Линейная аппроксимация по Тейлору является глобальной нижней оценкой выпуклой функции

Важное свойство непрерывной выпуклой функции $f(x)$ состоит в том, что в любой выбранной точке 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, то есть касательная к графику функции даёт *глобальную* нижнюю оценку функции.

- Если $f(x)$ дифференцируема, то $g = \nabla f(x_0)$
- Не всякая непрерывная выпуклая функция дифференцируема.

Линейная нижняя оценка выпуклой функции

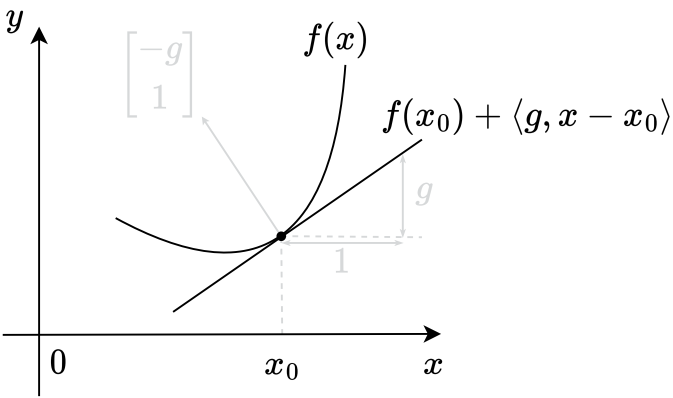


Figure 3: Линейная аппроксимация по Тейлору является глобальной нижней оценкой выпуклой функции

Важное свойство непрерывной выпуклой функции $f(x)$ состоит в том, что в любой выбранной точке 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, то есть касательная к графику функции даёт *глобальную* нижнюю оценку функции.

- Если $f(x)$ дифференцируема, то $g = \nabla f(x_0)$
- Не всякая непрерывная выпуклая функция дифференцируема.

Нам не хочется терять это прекрасное свойство.

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** 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** f в точке x_0 и обозначается $\partial f(x_0)$.

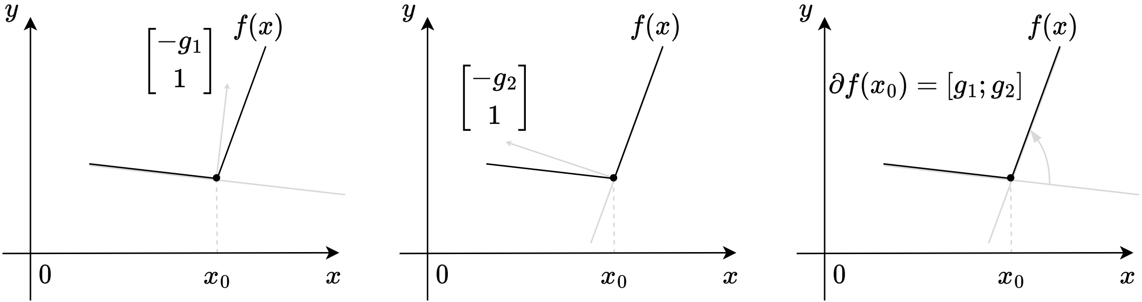


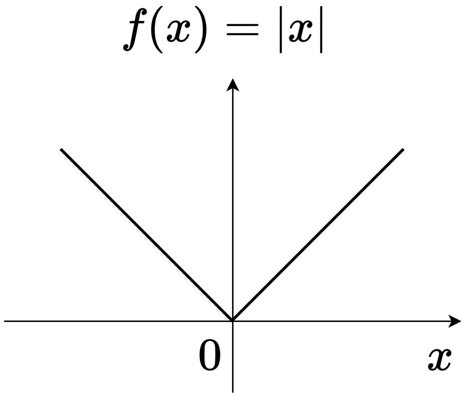
Figure 4: Субдифференциал — множество всех возможных субградиентов

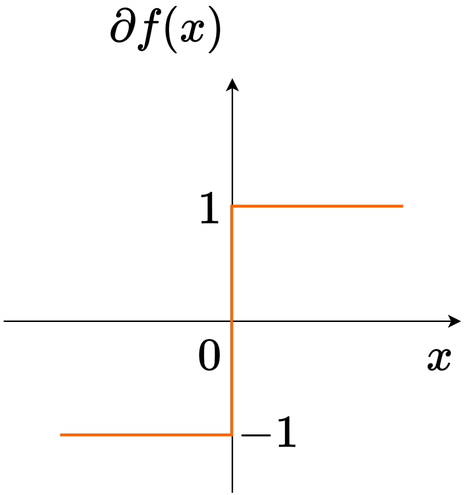
Субградиент и субдифференциал

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

Субградиент и субдифференциал

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





Свойства субдифференциала

- Если $x_0 \in \mathbf{ri}(S)$, то $\partial f(x_0)$ — выпуклое компактное множество.

Свойства субдифференциала

- Если $x_0 \in \mathbf{ri}(S)$, то $\partial f(x_0)$ — выпуклое компактное множество.
- Если выпуклая функция $f(x)$ дифференцируема в точке x_0, то $\partial f(x_0) = \{\nabla f(x_0)\}$.

Свойства субдифференциала

- Если $x_0 \in \mathbf{ri}(S)$, то $\partial f(x_0)$ — выпуклое компактное множество.
- Если выпуклая функция $f(x)$ дифференцируема в точке x_0, то $\partial f(x_0) = \{\nabla f(x_0)\}$.
- Если $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, то $f(x)$ выпукла на S.

Свойства субдифференциала

- Если $x_0 \in \mathbf{ri}(S)$, то $\partial f(x_0)$ — выпуклое компактное множество.
- Если выпуклая функция $f(x)$ дифференцируема в точке x_0, то $\partial f(x_0) = \{\nabla f(x_0)\}$.
- Если $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, то $f(x)$ выпукла на S.

Свойства субдифференциала

- Если $x_0 \in \mathbf{ri}(S)$, то $\partial f(x_0)$ — выпуклое компактное множество.
- Если выпуклая функция $f(x)$ дифференцируема в точке x_0, то $\partial f(x_0) = \{\nabla f(x_0)\}$.
- Если $\partial f(x_0) \neq \emptyset \quad \forall x_0 \in S$, то $f(x)$ выпукла на S.

Субдифференциал дифференцируемой функции

Пусть $f : S \to \mathbb{R}$, $x_0 \in \mathbf{ri}(S)$. Если f дифференцируема в точке x_0, то либо $\partial f(x_0) = \emptyset$, либо $\partial f(x_0) = \{\nabla f(x_0)\}$. Если f выпукла, то $\partial f(x_0) = \{\nabla f(x_0)\}$.

Свойства субдифференциала (доказательство)

1. Пусть $s \in \partial f(x_0)$ и v — единичный вектор. Для достаточно малых $t > 0$ выполняется

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

   Делим на t и переходим к пределу при $t \to 0^+$:

$$\langle \nabla f(x_0), v\rangle \geq \langle s, v\rangle.$$

Свойства субдифференциала (доказательство)

1. Пусть $s \in \partial f(x_0)$ и v — единичный вектор. Для достаточно малых $t > 0$ выполняется

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

Делим на t и переходим к пределу при $t \to 0^+$:

$$\langle \nabla f(x_0), v\rangle \geq \langle s, v\rangle.$$

Свойства субдифференциала (доказательство)

1. Пусть $s \in \partial f(x_0)$ и v — единичный вектор. Для достаточно малых $t > 0$ выполняется

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

   Делим на t и переходим к пределу при $t \to 0^+$:

$$\langle \nabla f(x_0), v\rangle \geq \langle s, v\rangle.$$

2. Значит $\langle \nabla f(x_0) - s, v\rangle \geq 0$ для всех единичных v; берём

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

   и получаем $s = \nabla f(x_0)$.

Свойства субдифференциала (доказательство)

1. Пусть $s \in \partial f(x_0)$ и v — единичный вектор. Для достаточно малых $t > 0$ выполняется

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

Делим на t и переходим к пределу при $t \to 0^+$:

$$\langle \nabla f(x_0), v\rangle \geq \langle s, v\rangle.$$

2. Значит $\langle \nabla f(x_0) - s, v\rangle \geq 0$ для всех единичных v; берём

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

и получаем $s = \nabla f(x_0)$.

Свойства субдифференциала (доказательство)

1. Пусть $s \in \partial f(x_0)$ и v — единичный вектор. Для достаточно малых $t > 0$ выполняется

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

   Делим на t и переходим к пределу при $t \to 0^+$:

$$\langle \nabla f(x_0), v\rangle \geq \langle s, v\rangle.$$

2. Значит $\langle \nabla f(x_0) - s, v\rangle \geq 0$ для всех единичных v; берём

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

   и получаем $s = \nabla f(x_0)$.

3. Если f выпукла, то $f(x) \geq f(x_0) + \langle \nabla f(x_0), x - x_0\rangle$ для всех $x \in S$, то есть $\nabla f(x_0) \in \partial f(x_0)$ и $\partial f(x_0) = \{\nabla f(x_0)\}$.

Вычисление субдифференциалов

Теорема Моро — Рокафеллара (субдифференциал линейной комбинации)

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

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

Вычисление субдифференциалов

Теорема Моро — Рокафеллара (субдифференциал линейной комбинации)

Пусть $f_i(x)$ — выпуклые функции на выпуклых множествах S_i, $i = \overline{1,n}$. Тогда если $\bigcap\limits_{i=1}^n \mathbf{ri}(S_i) \neq \emptyset$, 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 $S = \bigcap\limits_{i=1}^n S_i$, причём

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

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 задан как $f(x) = \max\limits_i f_i(x)$. Тогда:

$$
\partial_S f(x_0) = \mathbf{conv}\left\{ \bigcup_{i \in I(x_0)} \partial_S f_i(x_0) \right\}, \quad I(x) = \{i \in [1
$$

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

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

Субградиентный метод

Алгоритм

Вектор g называется **субградиентом** функции $f(x) : S \to \mathbb{R}$ 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 очень простая: заменим градиент $\nabla f(x_k)$ в алгоритме градиентного спуска на субградиент 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 $f(x) : S \to \mathbb{R}$ в точке x_0, если $\forall x \in S$:

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

Идея очень простая: заменим градиент $\nabla f(x_k)$ в алгоритме градиентного спуска на субградиент 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, $g_k \in \partial f(x_k)$

Заметим, что **субградиентный метод не обязан быть методом спуска**: отрицательный субградиент может не быть направлением спуска, а выбор шага может привести к $f(x_{k+1}) > f(x_k)$.

Поэтому обычно отслеживают лучшее достигнутое значение целевой функции

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

Оценка сходимости

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

Оценка сходимости

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

Оценка сходимости

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

Оценка сходимости

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

Оценка сходимости

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

Оценка сходимости

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

Просуммируем полученное неравенство по $k = 0, \ldots, T-1$:

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

Оценка сходимости

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

Просуммируем полученное неравенство по $k = 0, \ldots, T-1$:

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

Оценка сходимости

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

Просуммируем полученное неравенство по $k = 0, \ldots, T-1$:

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

Оценка сходимости

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

Просуммируем полученное неравенство по $k = 0, \ldots, T-1$:

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

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

Оценка сходимости

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

Просуммируем полученное неравенство по $k = 0, \ldots, T-1$:

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

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

Оценка сходимости

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

Просуммируем полученное неравенство по $k = 0, \ldots, T-1$:

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

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

Оценка сходимости

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

Просуммируем полученное неравенство по $k = 0, \ldots, T-1$:

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

- Запишем, насколько близко мы подошли к оптимуму $x^* = \text{arg} \min_{x \in \mathbb{R}^n} f(x) = \text{arg} f^*$ 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 $R = \|x_0 - x^*\|_2$

Оценка сходимости

- Наконец, заметим:

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

Оценка сходимости

- Наконец, заметим:

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

- что приводит к базовому неравенству:

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

Оценка сходимости

- Наконец, заметим:

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

- что приводит к базовому неравенству:

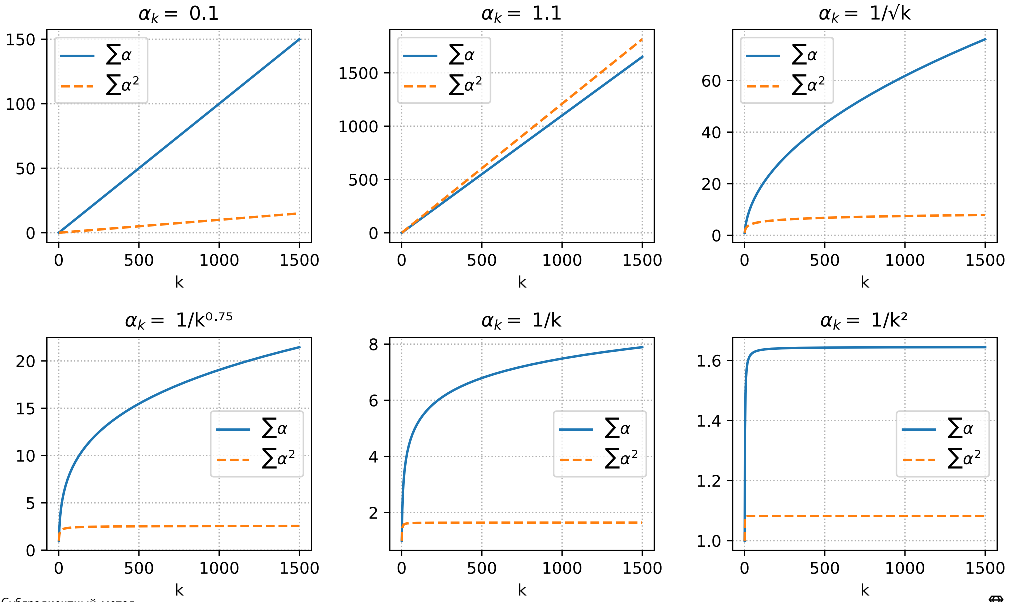
$$\boxed{f_k^{\text{best}} - f(x^*) \leq \frac{R^2 + G^2 \sum\limits_{k=0}^{T-1} \alpha_k^2}{2 \sum\limits_{k=0}^{T-1} \alpha_k}}$$

- Отсюда видно, что если стратегия выбора шага такова, что

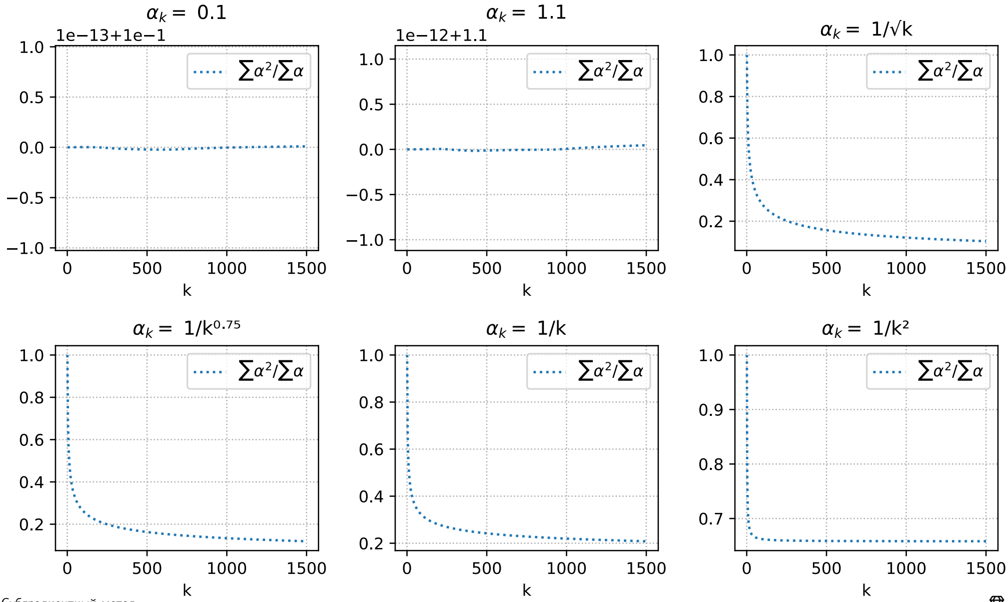
$$\sum_{k=0}^{T-1} \alpha_k^2 < \infty, \quad \sum_{k=0}^{T-1} \alpha_k = \infty,$$

то субградиентный метод сходится (шаг должен убывать, но не слишком быстро).

Различные стратегии выбора шага



Различные стратегии выбора шага



Оценка сходимости. Негладкий выпуклый случай. Постоянный шаг

Theorem

Пусть f — выпуклая G-липшицева функция и $R = \|x_0 - x^*\|_2$. Тогда при фиксированном шаге α субградиентный метод удовлетворяет оценке

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

- Заметим, что при любом постоянном шаге первый член правой части убывает, но второй остаётся постоянным.

Оценка сходимости. Негладкий выпуклый случай. Постоянный шаг

Theorem

Пусть f — выпуклая G-липшицева функция и $R = \|x_0 - x^*\|_2$. Тогда при фиксированном шаге α субградиентный метод удовлетворяет оценке

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

- Заметим, что при любом постоянном шаге первый член правой части убывает, но второй остаётся постоянным.
- Найдём оптимальный шаг α, минимизирующий правую часть неравенства.

Оценка сходимости. Негладкий выпуклый случай. Постоянный шаг

Theorem

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

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

- Эта версия требует заранее знать число итераций, что обычно непрактично.

Оценка сходимости. Негладкий выпуклый случай. Постоянный шаг

Theorem

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

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

- Эта версия требует заранее знать число итераций, что обычно непрактично.
- Интересно, что если искать оптимальные шаги для всей последовательности $\alpha_0, \alpha_1, \ldots, \alpha_{k-1}$, получится тот же результат.

Оценка сходимости. Негладкий выпуклый случай. Постоянный шаг

Theorem

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

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

- Эта версия требует заранее знать число итераций, что обычно непрактично.
- Интересно, что если искать оптимальные шаги для всей последовательности $\alpha_0, \alpha_1, \ldots, \alpha_{k-1}$, получится тот же результат.
- Почему? Потому что правая часть является выпуклой и **симметричной** функцией от $\alpha_0, \alpha_1, \ldots, \alpha_{k-1}$.

Оценка сходимости. Негладкий выпуклый случай. Постоянная длина шага

Theorem

Пусть f — выпуклая G-липшицева функция и $R = \|x_0 - x^*\|_2$. Тогда для фиксированной длины шага $\gamma = \alpha_k \|g_k\|_2$, т.е. $\alpha_k = \frac{\gamma}{\|g_k\|_2}$, субградиентный метод удовлетворяет оценке

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

- Заметим, что в субградиентном методе обычно нельзя использовать норму субградиента как критерий остановки (представьте $f(x) = |x|$). Существуют более продвинутые критерии, но сходимость и так очень медленная, поэтому обычно просто задают максимальное число итераций.

Оценка сходимости. Негладкий выпуклый случай. Практическая стратегия

Theorem

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

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

1. Оценим суммы:

Оценка сходимости. Негладкий выпуклый случай. Практическая стратегия

Theorem

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

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

1. Оценим суммы:

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

Оценка сходимости. Негладкий выпуклый случай. Практическая стратегия

Theorem

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

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

1. Оценим суммы:

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

Оценка сходимости. Негладкий выпуклый случай. Практическая стратегия

Theorem

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

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

1. Оценим суммы:

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

2. В верхней оценке выше отбросим последнее -1 и воспользуемся базовым неравенством:

Оценка сходимости. Негладкий выпуклый случай. Практическая стратегия

Theorem

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

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

1. Оценим суммы:

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

2. В верхней оценке выше отбросим последнее -1 и воспользуемся базовым неравенством:

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

Оценка сходимости. Негладкий выпуклый случай. Практическая стратегия

Theorem

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

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

1. Оценим суммы:

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

2. В верхней оценке выше отбросим последнее -1 и воспользуемся базовым неравенством:

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

Оценка сходимости. Негладкий выпуклый случай. Практическая стратегия

Theorem

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

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

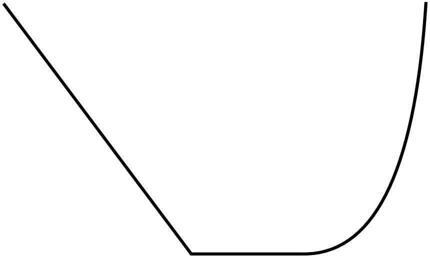
1. Оценим суммы:

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

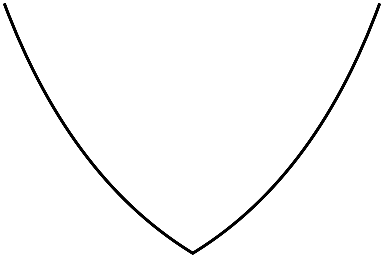
2. В верхней оценке выше отбросим последнее -1 и воспользуемся базовым неравенством:

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

Негладкий сильно выпуклый случай

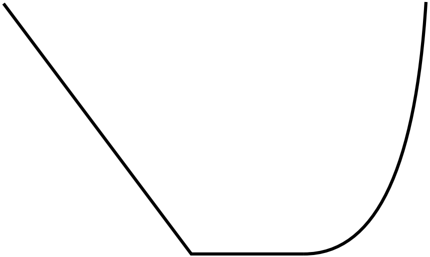


Негладкая
Выпуклая



Негладкая
μ - сильно выпуклая

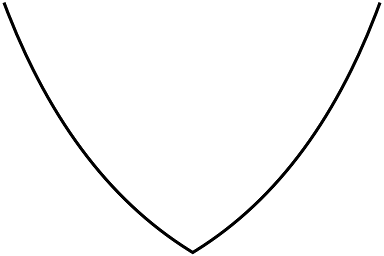
Негладкий сильно выпуклый случай



Негладкая

Выпуклая

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



Негладкая

μ - сильно выпуклая

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

Негладкий сильно выпуклый случай

Theorem

Пусть f μ-сильно выпукла на выпуклом множестве, а 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 μ-сильной выпуклости следует

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

Негладкий сильно выпуклый случай

Theorem

Пусть f μ-сильно выпукла на выпуклом множестве, а 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 μ-сильной выпуклости следует

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

2. Из субградиентного неравенства в точке x получаем

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

Негладкий сильно выпуклый случай

Theorem

Пусть f μ-сильно выпукла на выпуклом множестве, а 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 μ-сильной выпуклости следует
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x - y\|^2.$$
2. Из субградиентного неравенства в точке x получаем
$$f(\lambda x + (1-\lambda)y) \geq f(x) + \langle g, \lambda x + (1-\lambda)y - x \rangle \quad \to \quad f(\lambda x + (1-\lambda)y) \geq f(x) - (1-\lambda)\langle g, x - y \rangle.$$
3. Следовательно,
$$f(x) - (1-\lambda)\langle g, x - y \rangle \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x - y\|^2$$
$$(1-\lambda)f(x) \leq (1-\lambda)f(y) + (1-\lambda)\langle g, x - y \rangle - \frac{\mu}{2}\lambda(1-\lambda)\|x - y\|^2$$
$$f(x) \leq f(y) + \langle g, x - y \rangle - \frac{\mu}{2}\lambda\|x - y\|^2$$

Негладкий сильно выпуклый случай

Theorem

Пусть f μ-сильно выпукла на выпуклом множестве, а 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 μ-сильной выпуклости следует

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

2. Из субградиентного неравенства в точке x 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. Следовательно,

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

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

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

4. Переходя к пределу при $\lambda \to 1^-$, получаем

Оценка сходимости. Негладкий сильно выпуклый случай.

Theorem

Пусть f — μ-сильно выпуклая (возможно, негладкая) функция с минимизатором x^* и ограниченными субградиентами $\|g_k\| \leq G$. При выборе шага $\alpha_k = \frac{2}{\mu(k+1)}$ субградиентный метод гарантирует для $k > 0$:

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

1. Начинаем с той же записи метода, что и ранее:

Оценка сходимости. Негладкий сильно выпуклый случай.

Theorem

Пусть f — μ-сильно выпуклая (возможно, негладкая) функция с минимизатором x^* и ограниченными субградиентами $\|g_k\| \leq G$. При выборе шага $\alpha_k = \frac{2}{\mu(k+1)}$ субградиентный метод гарантирует для $k > 0$:

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

1. Начинаем с той же записи метода, что и ранее:

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

Оценка сходимости. Негладкий сильно выпуклый случай.

Theorem

Пусть f — μ-сильно выпуклая (возможно, негладкая) функция с минимизатором x^* и ограниченными субградиентами $\|g_k\| \leq G$. При выборе шага $\alpha_k = \frac{2}{\mu(k+1)}$ субградиентный метод гарантирует для $k > 0$:

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

1. Начинаем с той же записи метода, что и ранее:

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

Оценка сходимости. Негладкий сильно выпуклый случай.

Theorem

Пусть f — μ-сильно выпуклая (возможно, негладкая) функция с минимизатором x^* и ограниченными субградиентами $\|g_k\| \leq G$. При выборе шага $\alpha_k = \frac{2}{\mu(k+1)}$ субградиентный метод гарантирует для $k > 0$:

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

1. Начинаем с той же записи метода, что и ранее:

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

Оценка сходимости. Негладкий сильно выпуклый случай.

Theorem

Пусть f — μ-сильно выпуклая (возможно, негладкая) функция с минимизатором x^* и ограниченными субградиентами $\|g_k\| \leq G$. При выборе шага $\alpha_k = \frac{2}{\mu(k+1)}$ субградиентный метод гарантирует для $k > 0$:

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

1. Начинаем с той же записи метода, что и ранее:

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

Оценка сходимости. Негладкий сильно выпуклый случай.

Theorem

Пусть f — μ-сильно выпуклая (возможно, негладкая) функция с минимизатором x^* и ограниченными субградиентами $\|g_k\| \leq G$. При выборе шага $\alpha_k = \frac{2}{\mu(k+1)}$ субградиентный метод гарантирует для $k > 0$:

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

1. Начинаем с той же записи метода, что и ранее:

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

Оценка сходимости. Негладкий сильно выпуклый случай.

Theorem

Пусть f — μ-сильно выпуклая (возможно, негладкая) функция с минимизатором x^* и ограниченными субградиентами $\|g_k\| \leq G$. При выборе шага $\alpha_k = \frac{2}{\mu(k+1)}$ субградиентный метод гарантирует для $k > 0$:

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

1. Начинаем с той же записи метода, что и ранее:

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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. Просуммировав неравенства по всем $k = 0, 1, \ldots, T-1$, получаем:

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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. Просуммировав неравенства по всем $k = 0, 1, \ldots, T-1$, получаем:

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

3. Просуммировав неравенства по всем $k = 0, 1, \ldots, T-1$, получаем:

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

3. Просуммировав неравенства по всем $k = 0, 1, \ldots, T-1$, получаем:

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

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

3. Просуммировав неравенства по всем $k = 0, 1, \ldots, T-1$, получаем:

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

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

3. Просуммировав неравенства по всем $k = 0, 1, \ldots, T-1$, получаем:

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

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

3. Просуммировав неравенства по всем $k = 0, 1, \ldots, T-1$, получаем:

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

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

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

3. Просуммировав неравенства по всем $k = 0, 1, \ldots, T-1$, получаем:

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

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

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

Оценка сходимости. Негладкий сильно выпуклый случай. Доказательство

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

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

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

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

3. Просуммировав неравенства по всем $k = 0, 1, \ldots, T-1$, получаем:

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

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

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

Резюме. Субградиентный метод

Тип задачи	Правило выбора шага	Скорость сходимости	Итерационная сложность
Выпуклые липшицевы задачи	$\alpha \sim \dfrac{1}{\sqrt{k}}$	$\mathcal{O}\left(\dfrac{1}{\sqrt{k}}\right)$	$\mathcal{O}\left(\dfrac{1}{\varepsilon^2}\right)$
Сильно выпуклые липшицевы задачи	$\alpha \sim \dfrac{1}{k}$	$\mathcal{O}\left(\dfrac{1}{k}\right)$	$\mathcal{O}\left(\dfrac{1}{\varepsilon}\right)$

Численные эксперименты

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

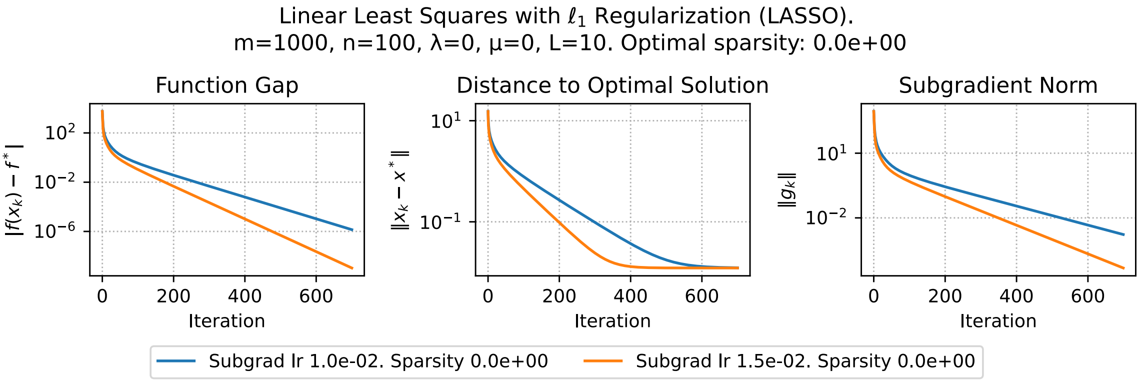


Figure 6: Гладкий выпуклый случай. Сублинейная сходимость, нет сходимости по аргументу

Численные эксперименты

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

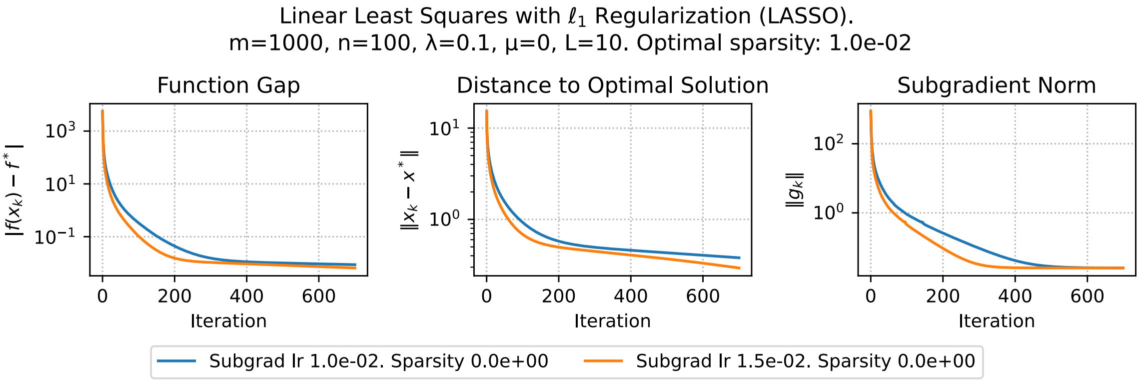


Figure 7: Негладкий выпуклый случай. Малое значение λ делает задачу негладкой. Нет сходимости при постоянном шаге

Численные эксперименты

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

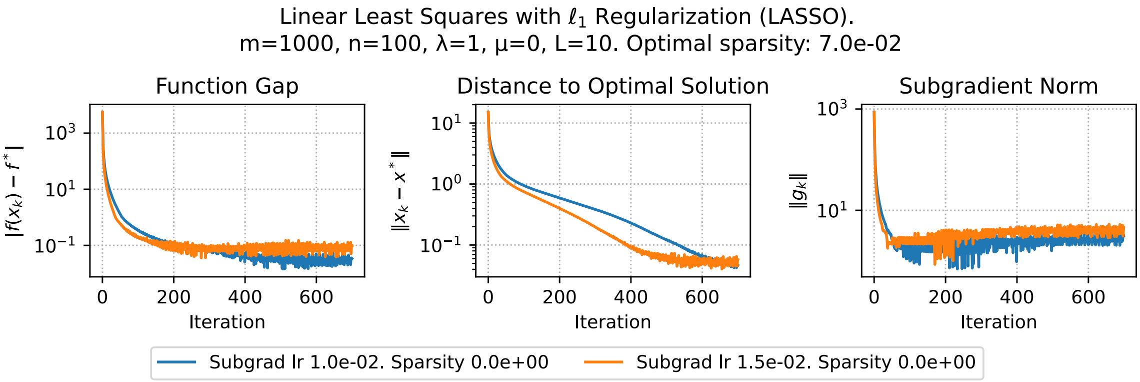


Figure 8: Негладкий выпуклый случай. Большее значение λ проявляет немонотонность $f(x_k)$. Меньший постоянный шаг приводит к более низкому стационарному уровню

Численные эксперименты

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

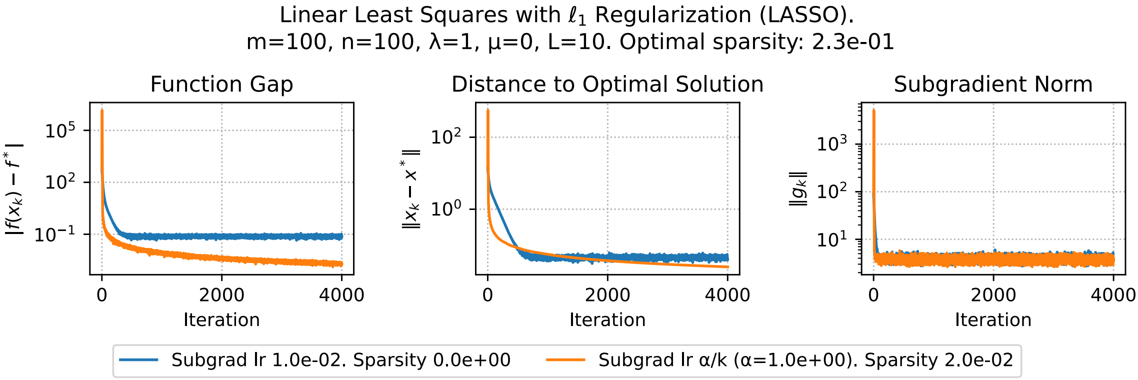


Figure 9: Негладкий выпуклый случай. Убывающий шаг обеспечивает сходимость f_k^{best}

Численные эксперименты

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

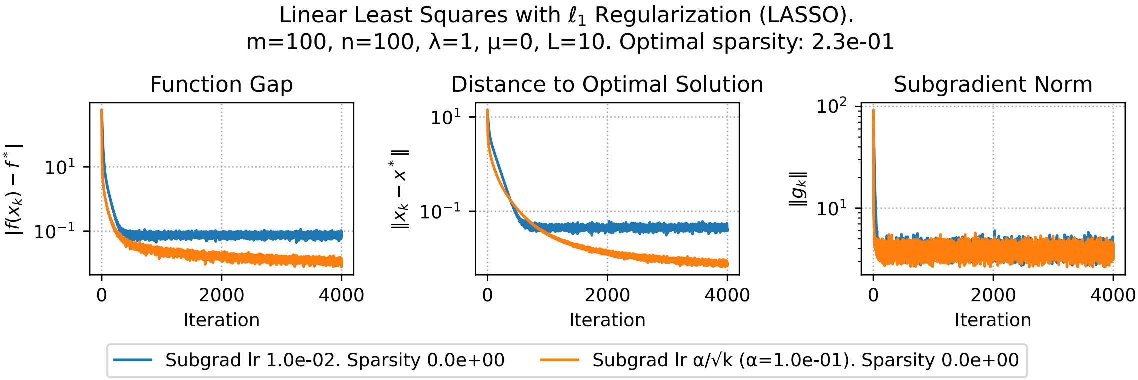


Figure 10: Негладкий выпуклый случай. Шаг $\frac{\alpha_0}{\sqrt{k}}$ обеспечивает сходимость f_k^{best}

Численные эксперименты

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

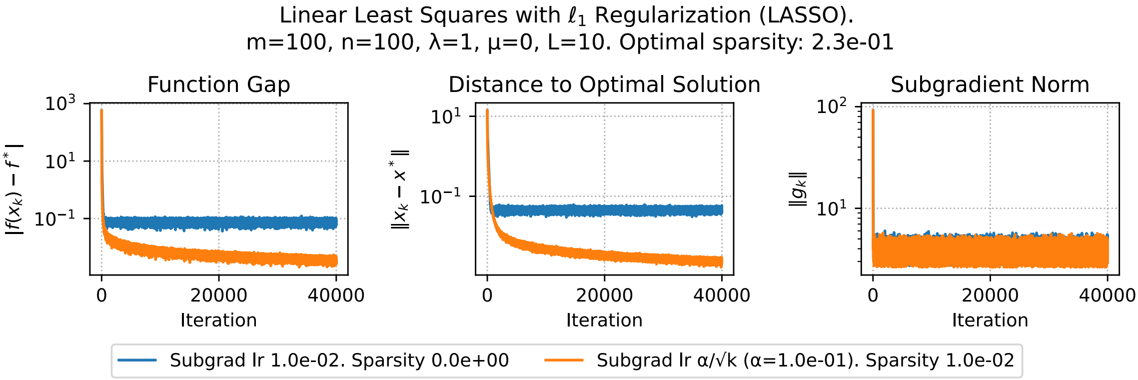


Figure 11: Негладкий выпуклый случай. Шаг $\frac{\alpha_0}{\sqrt{k}}$ обеспечивает сходимость f_k^{best}

Численные эксперименты

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

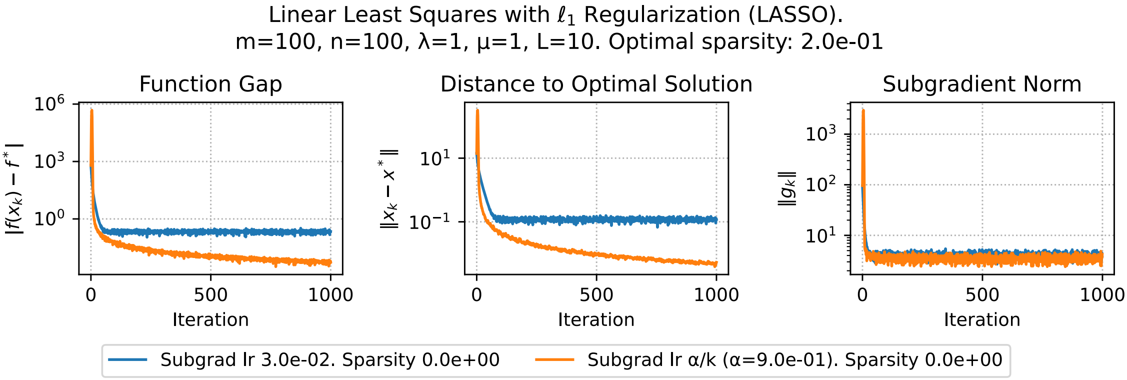


Figure 12: Негладкий сильно выпуклый случай. Шаг $\frac{\alpha_0}{k}$ обеспечивает сходимость f_k^{best}

Численные эксперименты

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

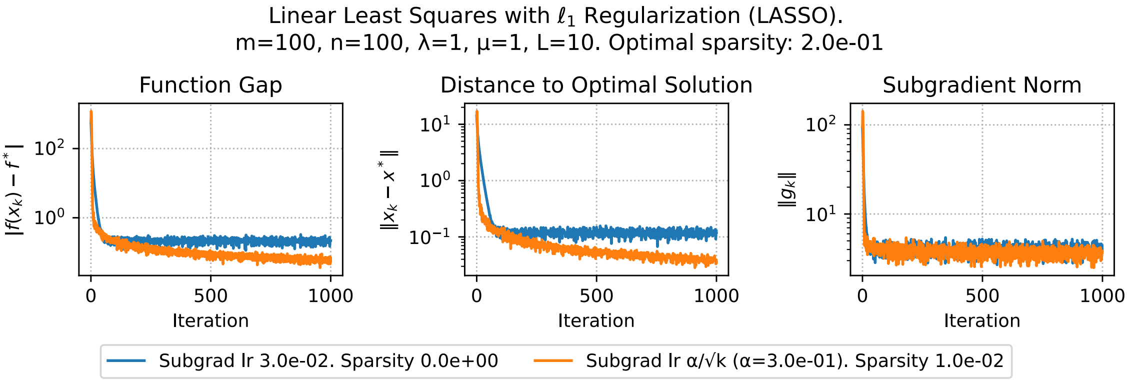


Figure 13: Негладкий сильно выпуклый случай. Шаг $\frac{\alpha_0}{\sqrt{k}}$ работает хуже

Численные эксперименты

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

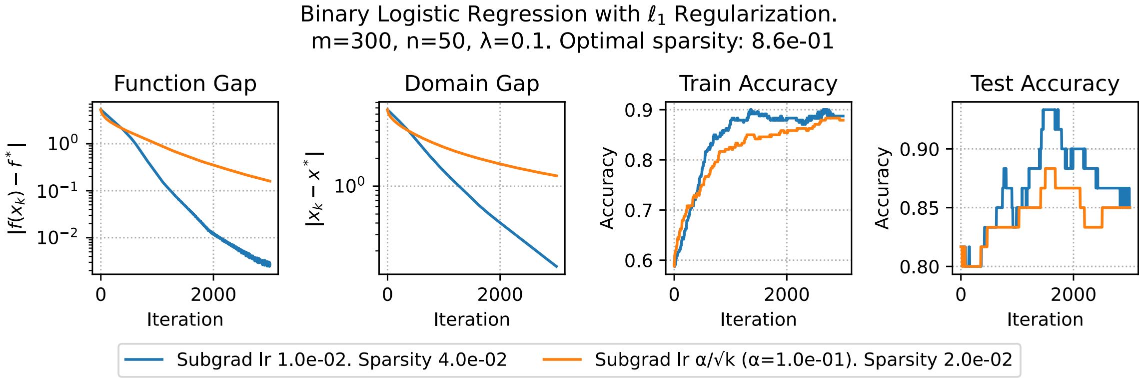


Figure 14: Логистическая регрессия с ℓ_1-регуляризацией

Численные эксперименты

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

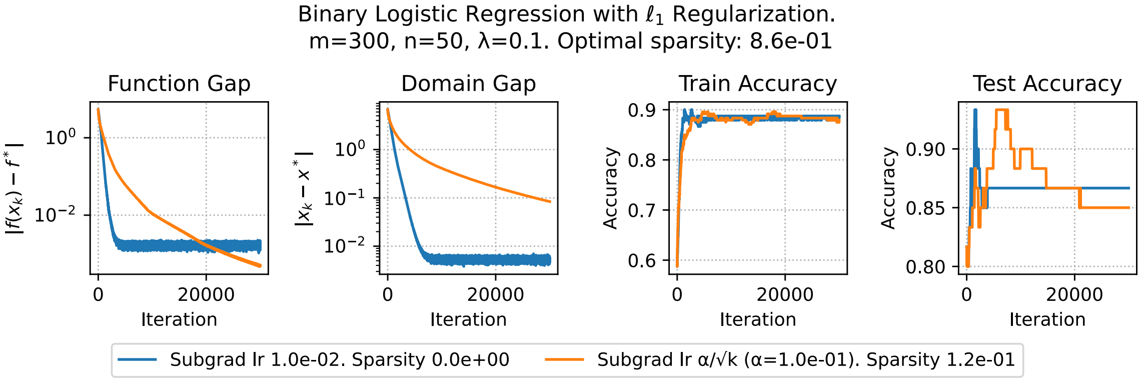


Figure 15: Логистическая регрессия с ℓ_1-регуляризацией

Численные эксперименты

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

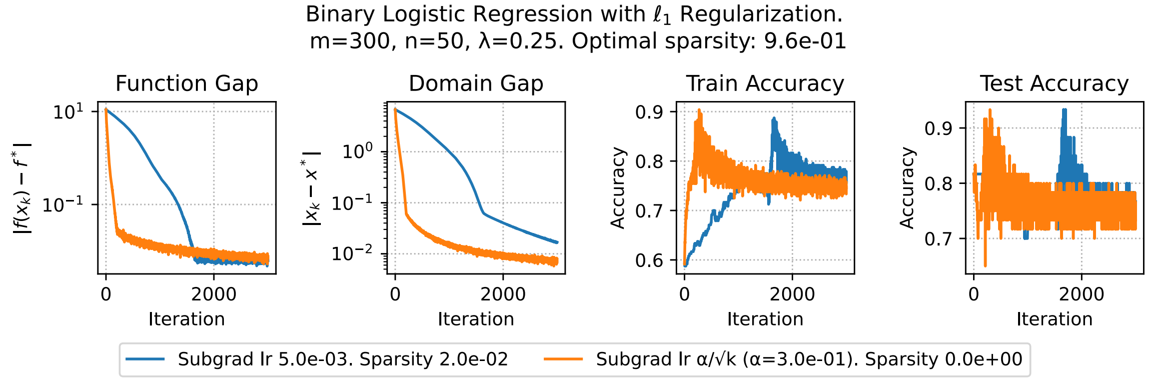


Figure 16: Логистическая регрессия с ℓ_1-регуляризацией

Численные эксперименты

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

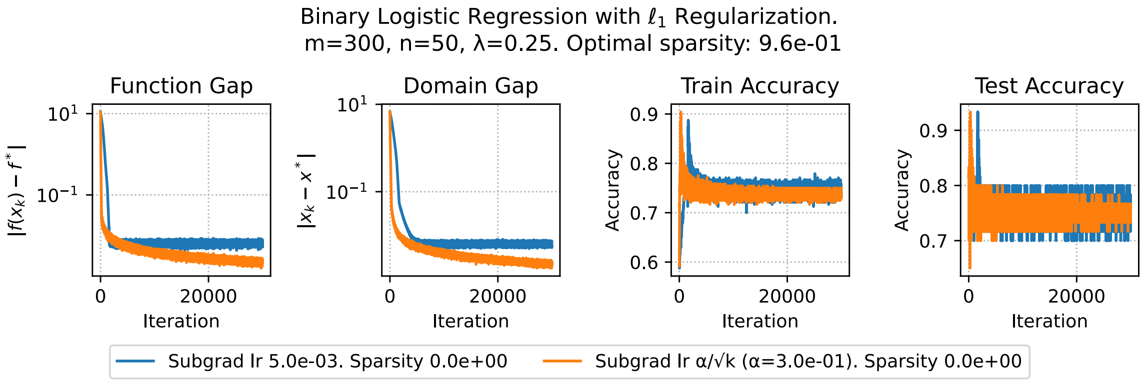


Figure 17: Логистическая регрессия с ℓ_1-регуляризацией

Численные эксперименты

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

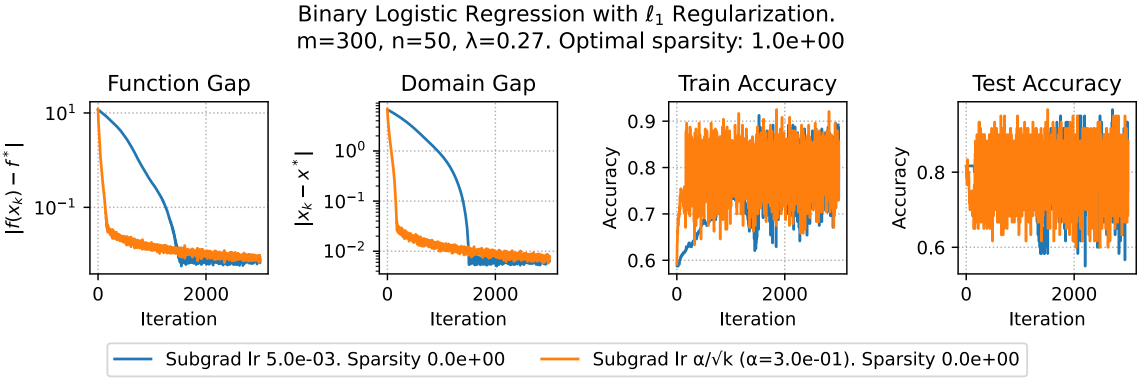


Figure 18: Логистическая регрессия с ℓ_1-регуляризацией

Численные эксперименты

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

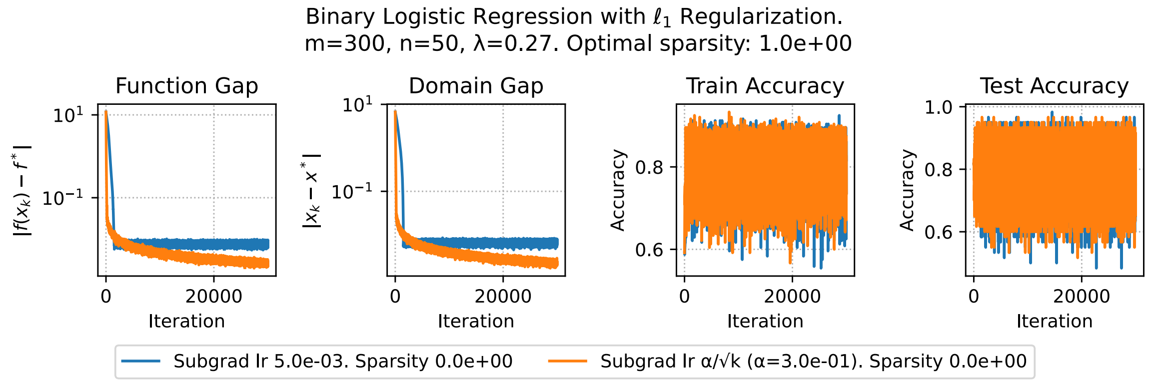


Figure 19: Логистическая регрессия с ℓ_1-регуляризацией

Нижние оценки

Нижние оценки

выпуклая (негладкая)	гладкая (невыпуклая)	гладкая & выпуклая	гладкая & сильно выпуклая (или PL)
$\mathcal{O}\left(\dfrac{1}{\sqrt{k}}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon^2}\right)$	$\mathcal{O}\left(\dfrac{1}{k^2}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\sqrt{\varepsilon}}\right)$	$\mathcal{O}\left(\dfrac{1}{k^2}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\sqrt{\varepsilon}}\right)$	$\mathcal{O}\left(\left(\dfrac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k\right)$ $k_\varepsilon \sim \mathcal{O}\left(\sqrt{\kappa}\log\dfrac{1}{\varepsilon}\right)$

Итерация «чёрного ящика»

Итерация градиентного спуска:

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

Итерация «чёрного ящика»

Итерация градиентного спуска:

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

Рассмотрим семейство методов первого порядка, для которых

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

Итерация «чёрного ящика»

Итерация градиентного спуска:

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

Рассмотрим семейство методов первого порядка, для которых

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

Чтобы построить нижнюю оценку, нужно найти функцию f из соответствующего класса так, чтобы любой метод из семейства 1 работал не быстрее, чем предсказано нижней оценкой.

Негладкий выпуклый случай

Theorem

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

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

где e_i — i-й вектор стандартного базиса. На итерации $k \leq n$ как минимум $n-k$ координат x равны 0. Это помогает получить оценку погрешности.

Негладкий 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, а $x[1:k]$ обозначает первые k компонент вектора x.

Негладкий 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, а $x[1:k]$ обозначает первые k компонент вектора x.

Ключевые свойства:

- Функция $f(x)$ является α-сильно выпуклой благодаря квадратичному слагаемому $\frac{\alpha}{2}\|x\|_2^2$.

Негладкий 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, а $x[1:k]$ обозначает первые k компонент вектора x.

Ключевые свойства:

- Функция $f(x)$ является α-сильно выпуклой благодаря квадратичному слагаемому $\frac{\alpha}{2}\|x\|_2^2$.
- Функция негладкая, потому что первое слагаемое недифференцируемо в точке, где достигается максимум по координатам x.

Негладкий 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, а $x[1:k]$ обозначает первые k компонент вектора x.

Ключевые свойства:

- Функция $f(x)$ является α-сильно выпуклой благодаря квадратичному слагаемому $\frac{\alpha}{2}\|x\|_2^2$.
- Функция негладкая, потому что первое слагаемое недифференцируемо в точке, где достигается максимум по координатам x.

Негладкий 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, а $x[1:k]$ обозначает первые k компонент вектора x.

Ключевые свойства:

- Функция $f(x)$ является α-сильно выпуклой благодаря квадратичному слагаемому $\frac{\alpha}{2}\|x\|_2^2$.
- Функция негладкая, потому что первое слагаемое недифференцируемо в точке, где достигается максимум по координатам x.

Рассмотрим субдифференциал $f(x)$ в точке 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 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, а $x[1:k]$ обозначает первые k компонент вектора x.

Ключевые свойства:

- Функция $f(x)$ является α-сильно выпуклой благодаря квадратичному слагаемому $\frac{\alpha}{2}\|x\|_2^2$.
- Функция негладкая, потому что первое слагаемое недифференцируемо в точке, где достигается максимум по координатам x.

Рассмотрим субдифференциал $f(x)$ в точке 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

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

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

Негладкий случай (доказательство)

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

$$\alpha x + \beta e_i,$$

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

- Гарантируем $\|x^0\| \leq R$, начиная с $x^0 = 0$.

Негладкий случай (доказательство)

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

$$\alpha x + \beta e_i,$$

где i — *первая* координата, для которой $x[i] = \max_{1 \le j \le k} x[j]$.

- Гарантируем $\|x^0\| \le R$, начиная с $x^0 = 0$.
- При запросе в точке $x^0 = 0$ оракул возвращает e_1. Следовательно, x^1 лежит на прямой, порождённой e_1.

Негладкий случай (доказательство)

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

$$\alpha x + \beta e_i,$$

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

- Гарантируем $\|x^0\| \leq R$, начиная с $x^0 = 0$.
- При запросе в точке $x^0 = 0$ оракул возвращает e_1. Следовательно, x^1 лежит на прямой, порождённой e_1.
- По индукции можно показать, что для всех i итерат x^i лежит в линейной оболочке $\{e_1, \dots, e_i\}$. В частности, при $i \leq k$ координата $k+1$ у x_i равна нулю, и из структуры $f(x)$ следует:

$$f(x^i) \geq 0.$$

Негладкий случай (доказательство)

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

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

Негладкий случай (доказательство)

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

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

- 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 случай (доказательство)

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

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

- Заметим, что $0 \in \partial f(y)$:

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

- 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 случай (доказательство)

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

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

- Заметим, что $0 \in \partial f(y)$:

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

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

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

- Тогда имеем:

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

Негладкий случай (доказательство)

Мы получили: $f(x^i) - f(x^*) \geq \frac{\beta^2}{2\alpha k}$, но нам нужно доказать, что $\min\limits_{i \in [1,k]} f(x^i) - f(x^*) \geq \frac{GR}{2(1+\sqrt{k})}$.

Негладкий случай (доказательство)

Мы получили: $f(x^i) - f(x^*) \geq \frac{\beta^2}{2\alpha k}$, но нам нужно доказать, что $\min\limits_{i \in [1,k]} f(x^i) - f(x^*) \geq \frac{GR}{2(1+\sqrt{k})}$.

Выпуклый случай

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

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

Заметим, в частности, что $\|y\|_2^2 = \frac{\beta^2}{\alpha^2 k} = R^2$ при этих параметрах

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

Негладкий случай (доказательство)

Мы получили: $f(x^i) - f(x^*) \geq \frac{\beta^2}{2\alpha k}$, но нам нужно доказать, что $\min\limits_{i \in [1,k]} f(x^i) - f(x^*) \geq \frac{GR}{2(1+\sqrt{k})}$.

Выпуклый случай

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

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

Заметим, в частности, что $\|y\|_2^2 = \frac{\beta^2}{\alpha^2 k} = R^2$ при этих параметрах

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

Сильно выпуклый случай

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

Заметим, в частности, что $\|y\|_2^2 = \frac{\beta^2}{\alpha^2 k} = \frac{G^2}{4\alpha^2 k} = R^2$ при этих параметрах

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

Приложения

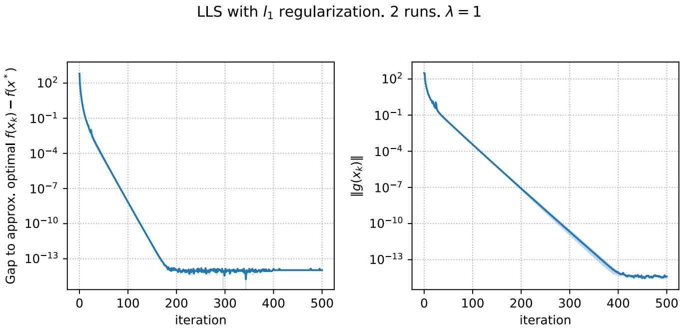
Линейные наименьшие квадраты с l_1-регуляризацией

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \lambda \|x\|_1$$

Алгоритм можно записать так:

$$x_{k+1} = x_k - \alpha_k \left(A^\top (Ax_k - b) + \lambda \text{sign}(x_k)\right),$$

где функция знака применяется поэлементно.



Логистическая регрессия с регуляризацией

Пусть $(x_i, y_i) \in \mathbb{R}^p \times \{0, 1\}$ для $i = 1, \dots, n$. Тогда функция логистической регрессии задаётся как:

$$f(\theta) = \sum_{i=1}^{n} \left(-y_i x_i^T \theta + \log(1 + \exp(x_i^T \theta))\right)$$

Это гладкая выпуклая функция; её градиент равен:

$$\nabla f(\theta) = \sum_{i=1}^{n} \left(y_i - s_i(\theta)\right) x_i$$

где $s_i(\theta) = \frac{\exp(x_i^T \theta)}{1+\exp(x_i^T \theta)}$, $i = 1, \dots, n$. Рассмотрим регуляризованную задачу:

$$f(\theta) + \lambda r(\theta) \to \min_{\theta}$$

где $r(\theta) = \|\theta\|_2^2$ соответствует гребневой (ridge) регуляризации, а $r(\theta) = \|\theta\|_1$ — лассо (lasso).

Метод опорных векторов

Пусть $D = \{(x_i, y_i) \mid x_i \in \mathbb{R}^n, y_i \in \{\pm 1\}\}$

Нужно найти $\theta \in \mathbb{R}^n$ и $b \in \mathbb{R}$ такие, что

$$\min_{\theta \in \mathbb{R}^n, b \in \mathbb{R}} \frac{1}{2}\|\theta\|_2^2 + C\sum_{i=1}^m \text{max}[0, 1 - y_i(\theta^\top x_i + b)]$$

Метод опорных векторов. Субградиент

Субградиент hinge loss $h(\theta) = \max[0, 1 - y_i(\theta^\top x_i + b)]$:

$$
\partial_\theta h = \begin{cases} 0, & \text{если } y_i(\theta^\top x_i + b) > 1 \\ -y_i x_i, & \text{если } y_i(\theta^\top x_i + b) < 1 \\ [-y_i x_i, 0], & \text{если } y_i(\theta^\top x_i + b) = 1 \end{cases}
$$

Метод опорных векторов. Субградиент

Субградиент hinge loss $h(\theta) = \max[0, 1 - y_i(\theta^\top x_i + b)]$:

$$\partial_\theta h = \begin{cases} 0, & \text{если } y_i(\theta^\top x_i + b) > 1 \\ -y_i x_i, & \text{если } y_i(\theta^\top x_i + b) < 1 \\ [-y_i x_i, 0], & \text{если } y_i(\theta^\top x_i + b) = 1 \end{cases}$$

Итерация субградиентного метода (для упрощённой SVM без b):

$$\theta_{k+1} = \theta_k - \alpha_k \left(\theta_k - C \frac{1}{m} \sum_{i: y_i \theta_k^\top x_i < 1} y_i x_i \right)$$

Метод опорных векторов. Численные эксперименты

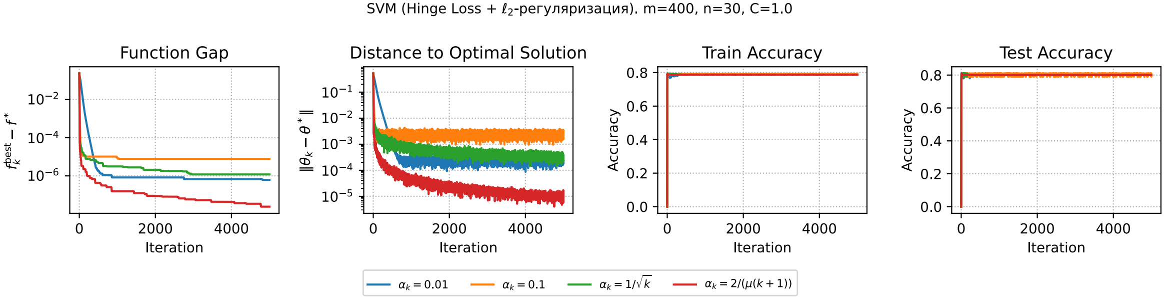


Figure 21: Субградиентный метод для SVM. Сравнение различных стратегий выбора шага

Метод опорных векторов. Немонотонность

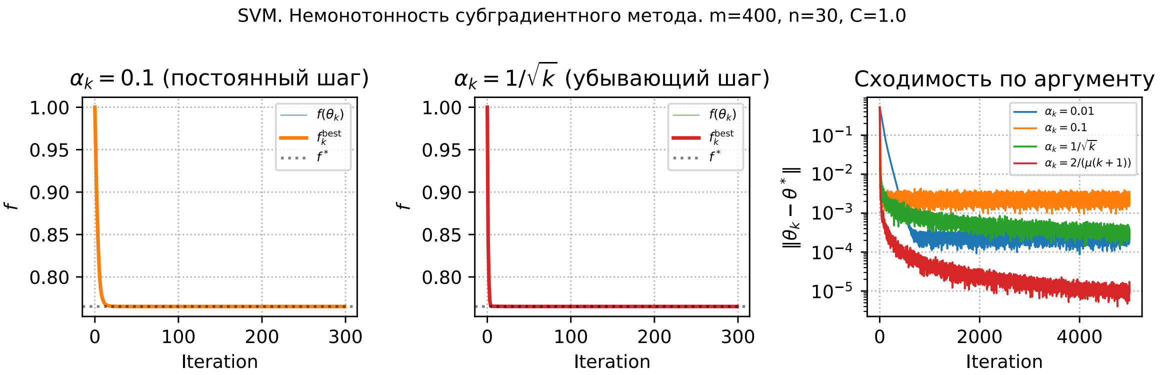
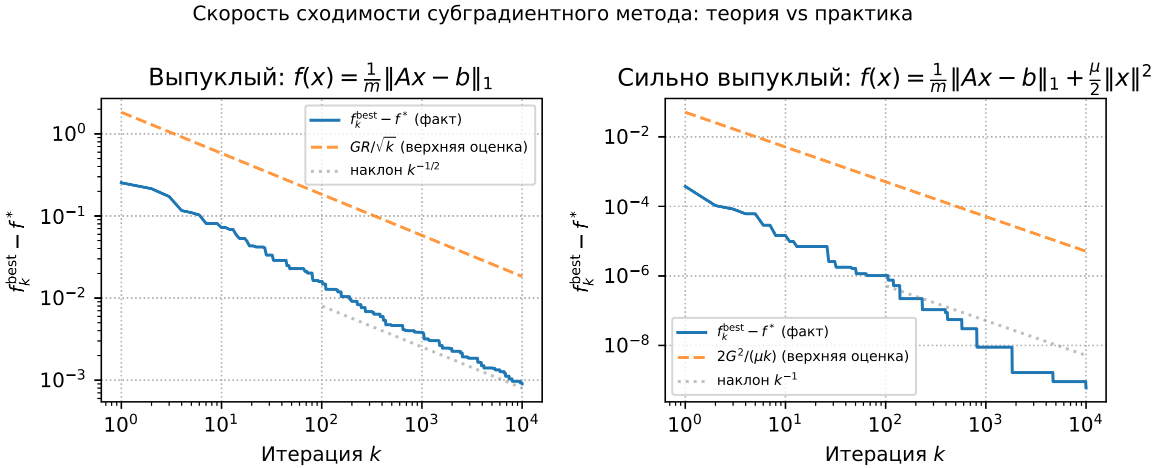


Figure 22: Немонотонность $f(\theta_k)$ при постоянном шаге. Убывающий шаг уменьшает осцилляции. Сходимость по аргументу зависит от стратегии

Дополнительные эксперименты

Скорость сходимости: теория vs практика

$$f(x) = \frac{1}{m}\|Ax - b\|_1$$



Шаг Поляка

$$
\alpha_k = \frac{f(x_k) - f^*}{\|g_k\|^2}
$$

Если известно оптимальное значение f^*, шаг Поляка значительно ускоряет сходимость субградиентного метода.

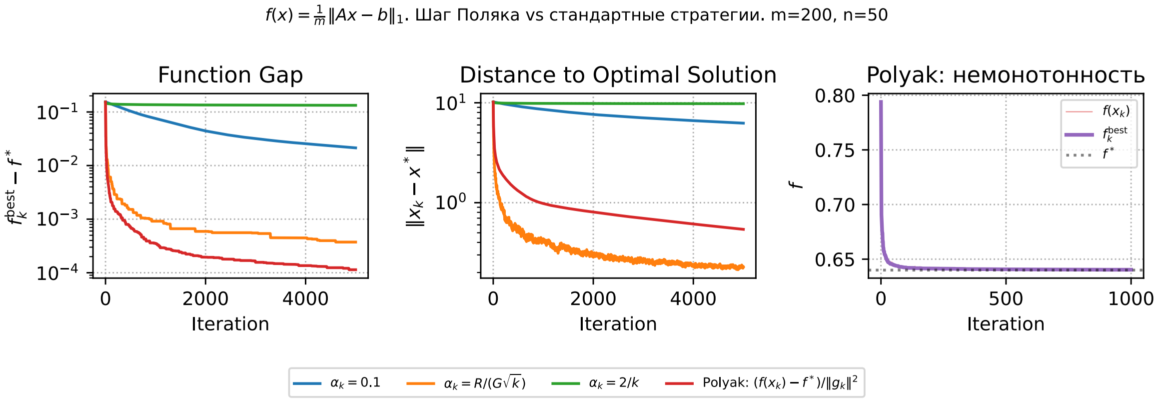
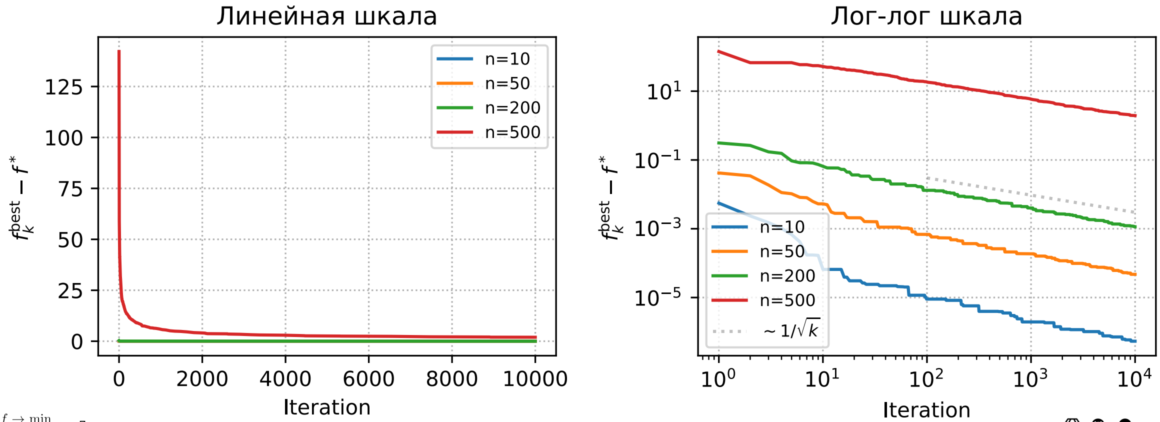


Figure 24: Шаг Поляка существенно превосходит стандартные стратегии и обеспечивает сходимость по аргументу. Однако

Влияние размерности задачи

$$f(x) = \frac{1}{m}\|Ax - b\|_1, \quad \alpha_k = \frac{R}{G\sqrt{k}}$$

Влияние размерности n на сходимость. $f(x) = \frac{1}{m}\|Ax - b\|_1$, $m = 500$, $\alpha_k = R/(G\sqrt{k})$



Литература

- Subgradient Methods Stephen Boyd (with help from Jaehyun Park) *hse26.fmin.xyz* | Оптимизация в ML | ФКН ВШЭ 2026