Градиентные методы в условных задачах оптимизации - метод проекции градиента. Метод Франк - Вульфа.

Даня Меркулов, Петр Остроухов

Оптимизация для всех! ЦУ

Методы оптимизации в условных задачах

Условная оптимизация

Безусловная оптимизация

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

- Любая точка $x_0 \in \mathbb{R}^n$ является допустимой и может быть решением.

Условная оптимизация

Безусловная оптимизация

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

- Любая точка $x_0 \in \mathbb{R}^n$ является допустимой и может быть решением.

Условная оптимизация

Безусловная оптимизация

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

- Любая точка $x_0 \in \mathbb{R}^n$ является допустимой и может быть решением.

Условная оптимизация

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

- Не все точки $x \in \mathbb{R}^n$ являются допустимыми и могут быть решением.

Условная оптимизация

Безусловная оптимизация

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

- Любая точка $x_0 \in \mathbb{R}^n$ является допустимой и может быть решением.

Условная оптимизация

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

- Не все точки $x \in \mathbb{R}^n$ являются допустимыми и могут быть решением.
- Решение должно быть внутри множества S.

Условная оптимизация

Безусловная оптимизация

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

- Любая точка $x_0 \in \mathbb{R}^n$ является допустимой и может быть решением.

Условная оптимизация

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

- Не все точки $x \in \mathbb{R}^n$ являются допустимыми и могут быть решением.
- Решение должно быть внутри множества S.
- Пример:

$$\frac{1}{2}\|Ax - b\|_2^2 \to \min_{\|x\|_2^2 \leq 1}$$

Условная оптимизация

Безусловная оптимизация

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

- Любая точка $x_0 \in \mathbb{R}^n$ является допустимой и может быть решением.

Условная оптимизация

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

- Не все точки $x \in \mathbb{R}^n$ являются допустимыми и могут быть решением.
- Решение должно быть внутри множества S.
- Пример:

$$\frac{1}{2}\|Ax - b\|_2^2 \to \min_{\|x\|_2^2 \leq 1}$$

Условная оптимизация

Безусловная оптимизация

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

- Любая точка $x_0 \in \mathbb{R}^n$ является допустимой и может быть решением.

Условная оптимизация

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

- Не все точки $x \in \mathbb{R}^n$ являются допустимыми и могут быть решением.
- Решение должно быть внутри множества S.
- Пример:

$$\frac{1}{2}\|Ax - b\|_2^2 \to \min_{\|x\|_2^2 \leq 1}$$

Градиентный спуск - хороший способ решать безусловные задачи

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) \tag{GD}$$

Можно ли его скорректировать для решения условных задач?

Условная оптимизация

Безусловная оптимизация

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

- Любая точка $x_0 \in \mathbb{R}^n$ является допустимой и может быть решением.

Условная оптимизация

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

- Не все точки $x \in \mathbb{R}^n$ являются допустимыми и могут быть решением.
- Решение должно быть внутри множества S.
- Пример:

$$
\frac{1}{2}\|Ax - b\|_2^2 \to \min_{\|x\|_2^2 \leq 1}
$$

Градиентный спуск - хороший способ решать безусловные задачи

$$
x_{k+1} = x_k - \alpha_k \nabla f(x_k) \tag{GD}
$$

Можно ли его скорректировать для решения условных задач?

Да. Нужно использовать проекции, чтобы обеспечить допустимость на каждой итерации.

Идея метода проекции градиента

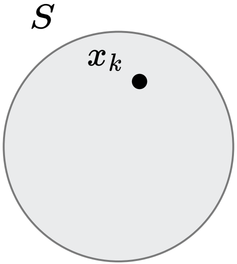


Рис. 1: Предположим, что мы начинаем с точки x_k.

Идея метода проекции градиента

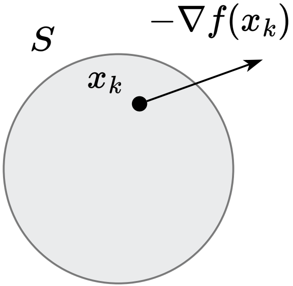


Рис. 2: И идем в направлении $-\nabla f(x_k)$.

Идея метода проекции градиента

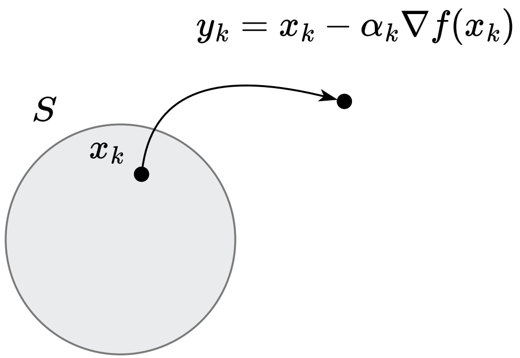


Рис. 3: Иногда мы можем оказаться вне допустимого множества.

Идея метода проекции градиента

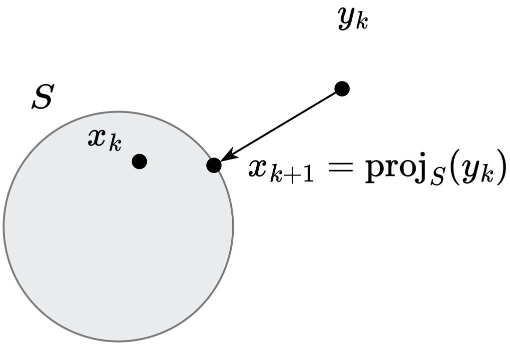


Рис. 4: Решим эту небольшую проблемку с помощью проекции!

Идея метода проекции градиента

$$
x_{k+1} = \text{proj}_S \left(x_k - \alpha_k \nabla f(x_k)\right) \qquad \Leftrightarrow \qquad \begin{aligned} y_k &= x_k - \alpha_k \nabla f(x_k) \\ x_{k+1} &= \text{proj}_S \left(y_k\right) \end{aligned}
$$

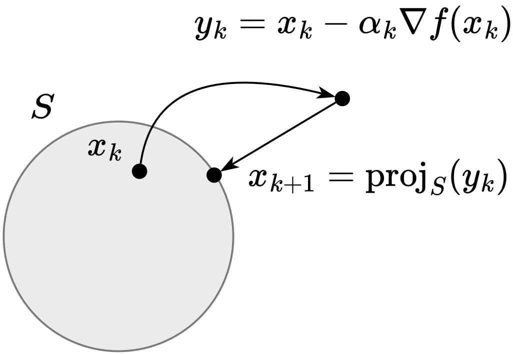


Рис. 5: Иллюстрация алгоритма проекции градиента

Пример: White-box Adversarial Attacks

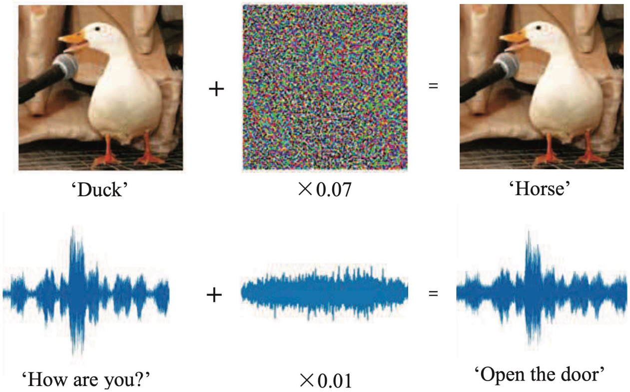


Рис. 6: Источник

- Математически, нейронная сеть - это функция $f(w; x)$

Пример: White-box Adversarial Attacks

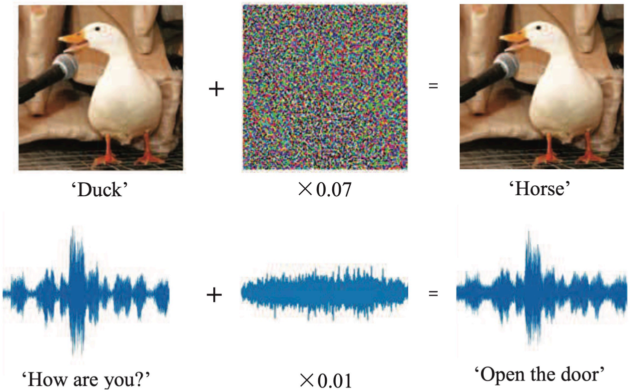


Рис. 6: Источник

- Математически, нейронная сеть - это функция $f(w; x)$
- Обычно, вход x задается, а оптимизируются веса сети w

Пример: White-box Adversarial Attacks

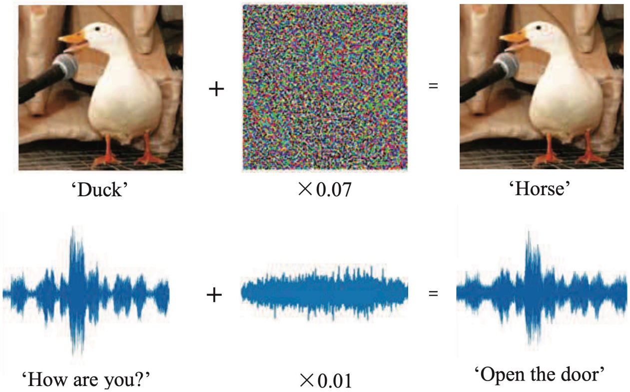


Рис. 6: Источник

- Математически, нейронная сеть - это функция $f(w; x)$
- Обычно, вход x задается, а оптимизируются веса сети w
- Можно наоборот: зафиксировать веса w и оптимизировать вход x враждебным образом:

$$\min_{\delta} \text{size}(\delta) \quad \text{s.t.} \quad \text{pred}[f(w; x+\delta)] \neq y$$

или

$$\max_{\delta} l(w; x+\delta, y) \text{ s.t. } \text{size}(\delta) \leq \epsilon, \; 0 \leq x+\delta \leq 1$$

Проекция

Проекция

Расстояние d от точки $y \in \mathbb{R}^n$ до замкнутого множества $S \subset \mathbb{R}^n$:

$$d(y, S, \|\cdot\|) = \inf\{\|x - y\| \mid x \in S\}$$

Проекция

Расстояние d от точки $y \in \mathbb{R}^n$ до замкнутого множества $S \subset \mathbb{R}^n$:

$$d(y, S, \| \cdot \|) = \inf\{\|x - y\| \mid x \in S\}$$

Мы будем фокусироваться на евклидовой проекции (возможны другие варианты) точки $y \in \mathbb{R}^n$ на множество $S \subseteq \mathbb{R}^n$ - это точка $\text{proj}_S(y) \in S$:

$$\text{proj}_S(y) = \Pi_S(y) = \underset{\mathbf{x} \in S}{\text{argmin}} \frac{1}{2} \|x - y\|_2^2$$

Проекция

Расстояние d от точки $y \in \mathbb{R}^n$ до замкнутого множества $S \subset \mathbb{R}^n$:

$$d(y, S, \|\cdot\|) = \inf\{\|x - y\| \mid x \in S\}$$

Мы будем фокусироваться на евклидовой проекции (возможны другие варианты) точки $y \in \mathbb{R}^n$ на множество $S \subseteq \mathbb{R}^n$ - это точка $\text{proj}_S(y) \in S$:

$$\text{proj}_S(y) = \Pi_S(y) = \underset{\mathbf{x} \in S}{\operatorname{argmin}} \frac{1}{2}\|x - y\|_2^2$$

- **Достаточное условия существования проекции**. Если $S \subseteq \mathbb{R}^n$ - замкнутое множество, то проекция на множество S существует для любой точки.

Проекция

Расстояние d от точки $y \in \mathbb{R}^n$ до замкнутого множества $S \subset \mathbb{R}^n$:

$$
d(y, S, \|\cdot\|) = \inf\{\|x - y\| \mid x \in S\}
$$

Мы будем фокусироваться на евклидовой проекции (возможны другие варианты) точки $y \in \mathbb{R}^n$ на множество $S \subseteq \mathbb{R}^n$ - это точка $\text{proj}_S(y) \in S$:

$$
\text{proj}_S(y) = \Pi_S(y) = \underset{\mathbf{x} \in S}{\text{argmin}} \frac{1}{2}\|x - y\|_2^2
$$

- **Достаточное условия существования проекции**. Если $S \subseteq \mathbb{R}^n$ - замкнутое множество, то проекция на множество S существует для любой точки.
- **Достаточное условия единственности проекции**. Если $S \subseteq \mathbb{R}^n$ - замкнутое выпуклое множество, то проекция на множество S единственна для любой точки.

Проекция

Расстояние d от точки $y \in \mathbb{R}^n$ до замкнутого множества $S \subset \mathbb{R}^n$:

$$d(y, S, \|\cdot\|) = \inf\{\|x - y\| \mid x \in S\}$$

Мы будем фокусироваться на евклидовой проекции (возможны другие варианты) точки $y \in \mathbb{R}^n$ на множество $S \subseteq \mathbb{R}^n$ - это точка $\text{proj}_S(y) \in S$:

$$\text{proj}_S(y) = \Pi_S(y) = \underset{\mathbf{x} \in S}{\text{argmin}} \frac{1}{2}\|x - y\|_2^2$$

- **Достаточное условия существования проекции**. Если $S \subseteq \mathbb{R}^n$ - замкнутое множество, то проекция на множество S существует для любой точки.
- **Достаточное условия единственности проекции**. Если $S \subseteq \mathbb{R}^n$ - замкнутое выпуклое множество, то проекция на множество S единственна для любой точки.
- Если множество открыто, и точка вне этого множества, то ее проекция на это множество может не существовать.

Проекция

Расстояние d от точки $y \in \mathbb{R}^n$ до замкнутого множества $S \subset \mathbb{R}^n$:

$$d(y, S, \|\cdot\|) = \inf\{\|x - y\| \mid x \in S\}$$

Мы будем фокусироваться на евклидовой проекции (возможны другие варианты) точки $y \in \mathbb{R}^n$ на множество $S \subseteq \mathbb{R}^n$ - это точка $\text{proj}_S(y) \in S$:

$$\text{proj}_S(y) = \Pi_S(y) = \underset{\mathbf{x} \in S}{\operatorname{argmin}} \frac{1}{2}\|x - y\|_2^2$$

- **Достаточное условия существования проекции**. Если $S \subseteq \mathbb{R}^n$ - замкнутое множество, то проекция на множество S существует для любой точки.
- **Достаточное условия единственности проекции**. Если $S \subseteq \mathbb{R}^n$ - замкнутое выпуклое множество, то проекция на множество S единственна для любой точки.
- Если множество открыто, и точка вне этого множества, то ее проекция на это множество может не существовать.
- Если точка лежит внутри множества, то ее проекция - это сама точка.

Оператор проекции является нерастягивающим

- Функция f называется нерастягивающей, если она L-Липшицева с $L \leq 1$ [1]. То есть, для любых двух точек $x, y \in \text{dom} f$,

$$\|f(x) - f(y)\| \leq L\|x - y\|, \text{ where } L \leq 1.$$

Это означает, что расстояние между образами точек не больше, чем расстояние между исходными точками.

[1] Нерастягивающая становится сжимающей, если $L < 1$.

Оператор проекции является нерастягивающим

- Функция f называется нерастягивающей, если она L-Липшицева с $L \leq 1$ [1]. То есть, для любых двух точек $x, y \in \text{dom} f$,

$$\|f(x) - f(y)\| \leq L\|x - y\|, \text{ where } L \leq 1.$$

  Это означает, что расстояние между образами точек не больше, чем расстояние между исходными точками.

- Оператор проекции является нерастягивающим:

$$\|\text{proj}(x) - \text{proj}(y)\|_2 \leq \|x - y\|_2.$$

[1] Нерастягивающая становится сжимающей, если $L < 1$.

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq R\}$, $y \notin S$

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq R\}$, $y \notin S$

Построим гипотезу из рисунка: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \le R\}$, $y \notin S$

Построим гипотезу из рисунка: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

Проверим неравенство для выпуклого замкнутого множества:
$(\pi - y)^T(x - \pi) \ge 0$

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \le R\}$, $y \notin S$

Построим гипотезу из рисунка: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

Проверим неравенство для выпуклого замкнутого множества:
$(\pi - y)^T (x - \pi) \ge 0$

$$\left(x_0 - y + R\frac{y - x_0}{\|y - x_0\|} \right)^T \left(x - x_0 - R\frac{y - x_0}{\|y - x_0\|} \right) =$$

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq R\}$, $y \notin S$

Построим гипотезу из рисунка: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

Проверим неравенство для выпуклого замкнутого множества:
$(\pi - y)^T(x - \pi) \geq 0$

$$\left(x_0 - y + R\frac{y - x_0}{\|y - x_0\|}\right)^T \left(x - x_0 - R\frac{y - x_0}{\|y - x_0\|}\right) =$$

$$\left(\frac{(y - x_0)(R - \|y - x_0\|)}{\|y - x_0\|}\right)^T \left(\frac{(x - x_0)\|y - x_0\| - R(y - x_0)}{\|y - x_0\|}\right) =$$

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \le R\}$, $y \notin S$

Построим гипотезу из рисунка: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T(x - \pi) \ge 0$

$$\left(x_0 - y + R\frac{y - x_0}{\|y - x_0\|} \right)^T \left(x - x_0 - R\frac{y - x_0}{\|y - x_0\|} \right) =$$

$$\left(\frac{(y - x_0)(R - \|y - x_0\|)}{\|y - x_0\|} \right)^T \left(\frac{(x - x_0)\|y - x_0\| - R(y - x_0)}{\|y - x_0\|} \right) =$$

$$\frac{R - \|y - x_0\|}{\|y - x_0\|^2} (y - x_0)^T \left((x - x_0) \|y - x_0\| - R(y - x_0) \right) =$$

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \le R\}$, $y \notin S$

Построим гипотезу из рисунка: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T(x - \pi) \ge 0$

$$
\left(x_0 - y + R\frac{y - x_0}{\|y - x_0\|} \right)^T \left(x - x_0 - R\frac{y - x_0}{\|y - x_0\|} \right) =
$$

$$
\left(\frac{(y - x_0)(R - \|y - x_0\|)}{\|y - x_0\|} \right)^T \left(\frac{(x - x_0)\|y - x_0\| - R(y - x_0)}{\|y - x_0\|} \right) =
$$

$$
\frac{R - \|y - x_0\|}{\|y - x_0\|^2} (y - x_0)^T \left((x - x_0) \|y - x_0\| - R \left(y - x_0 \right) \right) =
$$

$$
\frac{R - \|y - x_0\|}{\|y - x_0\|} \left((y - x_0)^T (x - x_0) - R\|y - x_0\| \right) =
$$

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq R\}$, $y \notin S$

Построим гипотезу из рисунка: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T(x - \pi) \geq 0$

$$
\begin{aligned}
\left(x_0 - y + R\frac{y - x_0}{\|y - x_0\|} \right)^T \left(x - x_0 - R\frac{y - x_0}{\|y - x_0\|} \right) &= \\
\left(\frac{(y - x_0)(R - \|y - x_0\|)}{\|y - x_0\|} \right)^T \left(\frac{(x - x_0)\|y - x_0\| - R(y - x_0)}{\|y - x_0\|} \right) &= \\
\frac{R - \|y - x_0\|}{\|y - x_0\|^2} (y - x_0)^T \left((x - x_0)\|y - x_0\| - R(y - x_0) \right) &= \\
\frac{R - \|y - x_0\|}{\|y - x_0\|} \left((y - x_0)^T (x - x_0) - R\|y - x_0\| \right) &= \\
(R - \|y - x_0\|) \left(\frac{(y - x_0)^T (x - x_0)}{\|y - x_0\|} - R \right)
\end{aligned}
$$

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq R\}$, $y \notin S$

Построим гипотезу из рисунка: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T(x - \pi) \geq 0$

$$
\left(x_0 - y + R\frac{y - x_0}{\|y - x_0\|} \right)^T \left(x - x_0 - R\frac{y - x_0}{\|y - x_0\|} \right) =
$$

$$
\left(\frac{(y - x_0)(R - \|y - x_0\|)}{\|y - x_0\|} \right)^T \left(\frac{(x - x_0)\|y - x_0\| - R(y - x_0)}{\|y - x_0\|} \right) =
$$

$$
\frac{R - \|y - x_0\|}{\|y - x_0\|^2} (y - x_0)^T \left((x - x_0) \|y - x_0\| - R\left(y - x_0\right) \right) =
$$

$$
\frac{R - \|y - x_0\|}{\|y - x_0\|} \left((y - x_0)^T (x - x_0) - R\|y - x_0\| \right) =
$$

$$
(R - \|y - x_0\|) \left(\frac{(y - x_0)^T(x - x_0)}{\|y - x_0\|} - R \right)
$$

Первый множитель отрицателен для выбора точки y. Второй множитель тоже отрицателен, что следует из неравенства КБШ:

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \le R\}$, $y \notin S$

Построим гипотезу из рисунка: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T(x - \pi) \ge 0$

$$
\left(x_0 - y + R\frac{y - x_0}{\|y - x_0\|}\right)^T \left(x - x_0 - R\frac{y - x_0}{\|y - x_0\|}\right) =
$$

$$
\left(\frac{(y - x_0)(R - \|y - x_0\|)}{\|y - x_0\|}\right)^T \left(\frac{(x - x_0)\|y - x_0\| - R(y - x_0)}{\|y - x_0\|}\right) =
$$

$$
\frac{R - \|y - x_0\|}{\|y - x_0\|^2}(y - x_0)^T\left((x - x_0)\|y - x_0\| - R(y - x_0)\right) =
$$

$$
\frac{R - \|y - x_0\|}{\|y - x_0\|}\left((y - x_0)^T(x - x_0) - R\|y - x_0\|\right) =
$$

$$
(R - \|y - x_0\|)\left(\frac{(y - x_0)^T(x - x_0)}{\|y - x_0\|} - R\right)
$$

Первый множитель отрицателен для выбора точки y. Второй множитель тоже отрицателен, что следует из неравенства КБШ:

$$
(y - x_0)^T(x - x_0) \le \|y - x_0\|\|x - x_0\|
$$

Пример: проекция на шар

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid \|x - x_0\| \le R\}$, $y \notin S$

Построим гипотезу из рисунка: $\pi = x_0 + R \cdot \frac{y - x_0}{\|y - x_0\|}$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T(x - \pi) \ge 0$

$$
\left(x_0 - y + R\frac{y - x_0}{\|y - x_0\|} \right)^T \left(x - x_0 - R\frac{y - x_0}{\|y - x_0\|} \right) =
$$

$$
\left(\frac{(y - x_0)(R - \|y - x_0\|)}{\|y - x_0\|} \right)^T \left(\frac{(x - x_0)\|y - x_0\| - R(y - x_0)}{\|y - x_0\|} \right) =
$$

$$
\frac{R - \|y - x_0\|}{\|y - x_0\|^2} (y - x_0)^T \left((x - x_0) \|y - x_0\| - R \left(y - x_0 \right) \right) =
$$

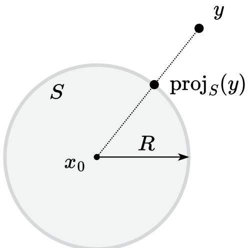
$$
\frac{R - \|y - x_0\|}{\|y - x_0\|} \left((y - x_0)^T (x - x_0) - R\|y - x_0\| \right) =
$$

$$
(R - \|y - x_0\|) \left(\frac{(y - x_0)^T (x - x_0)}{\|y - x_0\|} - R \right)
$$

Первый множитель отрицателен для выбора точки y. Второй множитель тоже отрицателен, что следует из неравенства КБШ:

$$
(y - x_0)^T (x - x_0) \le \|y - x_0\| \|x - x_0\|
$$

$$
\frac{(y - x_0)^T (x - x_0)}{\|y - x_0\|} - R \le \frac{\|y - x_0\| \|x - x_0\|}{\|y - x_0\|} -
$$



Пример: проекция на гиперплоскость

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Построим гипотезу из рисунка: $\pi = y + \alpha c$. Коэффициент α выбирается так, чтобы $\pi \in S$: $c^T \pi = b$, поэтому:

Пример: проекция на гиперплоскость

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Построим гипотезу из рисунка: $\pi = y + \alpha c$. Коэффициент α выбирается так, чтобы $\pi \in S$: $c^T \pi = b$, поэтому:

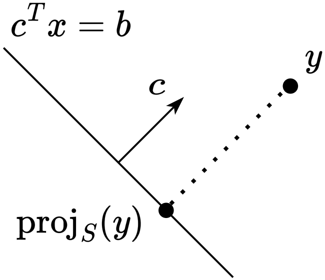


Рис. 8: Гиперплоскость

Пример: проекция на гиперплоскость

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Построим гипотезу из рисунка: $\pi = y + \alpha c$.
Коэффициент α выбирается так, чтобы $\pi \in S$: $c^T \pi = b$, поэтому:

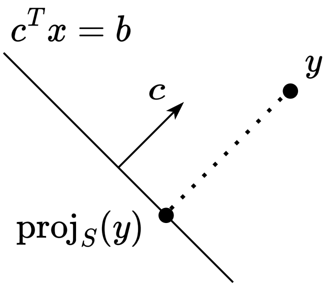


Рис. 8: Гиперплоскость

$$c^T(y + \alpha c) = b$$

Пример: проекция на гиперплоскость

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Построим гипотезу из рисунка: $\pi = y + \alpha c$. Коэффициент α выбирается так, чтобы $\pi \in S$: $c^T \pi = b$, поэтому:

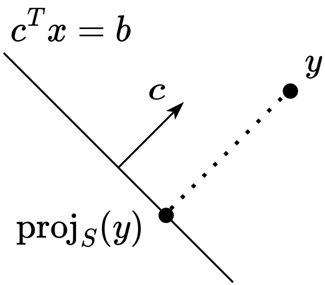


Рис. 8: Гиперплоскость

$$c^T(y + \alpha c) = b$$
$$c^T y + \alpha c^T c = b$$

Пример: проекция на гиперплоскость

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Построим гипотезу из рисунка: $\pi = y + \alpha c$. Коэффициент α выбирается так, чтобы $\pi \in S$: $c^T \pi = b$, поэтому:

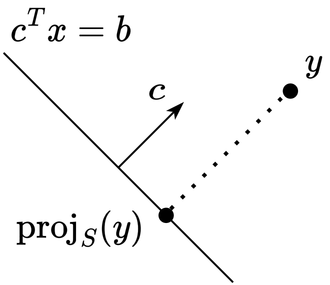


Рис. 8: Гиперплоскость

$$
\begin{aligned}
c^T(y + \alpha c) &= b \\
c^T y + \alpha c^T c &= b \\
c^T y &= b - \alpha c^T c
\end{aligned}
$$

Пример: проекция на гиперплоскость

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Построим гипотезу из рисунка: $\pi = y + \alpha c$. Коэффициент α выбирается так, чтобы $\pi \in S$: $c^T\pi = b$, поэтому:

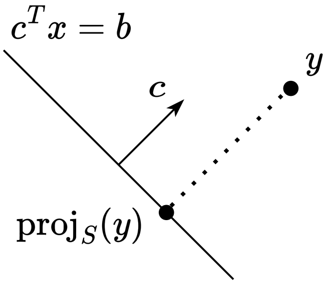


Рис. 8: Гиперплоскость

$$c^T(y + \alpha c) = b$$
$$c^T y + \alpha c^T c = b$$
$$c^T y = b - \alpha c^T c$$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T(x - \pi) \geq 0$

$$(y + \alpha c - y)^T(x - y - \alpha c) =$$

Пример: проекция на гиперплоскость

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Построим гипотезу из рисунка: $\pi = y + \alpha c$. Коэффициент α выбирается так, чтобы $\pi \in S$: $c^T \pi = b$, поэтому:

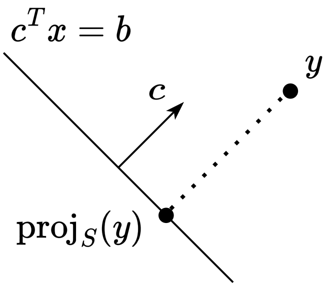


Рис. 8: Гиперплоскость

$$
\begin{aligned}
c^T(y + \alpha c) &= b \\
c^T y + \alpha c^T c &= b \\
c^T y &= b - \alpha c^T c
\end{aligned}
$$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T(x - \pi) \geq 0$

$$
\begin{aligned}
(y + \alpha c - y)^T(x - y - \alpha c) &= \\
\alpha c^T(x - y - \alpha c) &=
\end{aligned}
$$

Пример: проекция на гиперплоскость

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Построим гипотезу из рисунка: $\pi = y + \alpha c$. Коэффициент α выбирается так, чтобы $\pi \in S$: $c^T \pi = b$, поэтому:

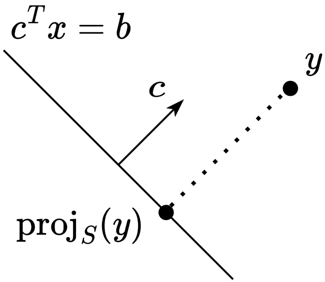


Рис. 8: Гиперплоскость

$$
\begin{aligned}
c^T(y + \alpha c) &= b \\
c^T y + \alpha c^T c &= b \\
c^T y &= b - \alpha c^T c
\end{aligned}
$$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T(x - \pi) \geq 0$

$$
\begin{aligned}
(y + \alpha c - y)^T(x - y - \alpha c) &= \\
\alpha c^T(x - y - \alpha c) &= \\
\alpha(c^T x) - \alpha(c^T y) - \alpha^2(c^T c) &=
\end{aligned}
$$

Пример: проекция на гиперплоскость

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Построим гипотезу из рисунка: $\pi = y + \alpha c$. Коэффициент α выбирается так, чтобы $\pi \in S$: $c^T \pi = b$, поэтому:

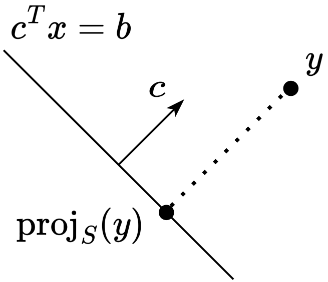


Рис. 8: Гиперплоскость

$$
\begin{aligned}
c^T(y + \alpha c) &= b \\
c^T y + \alpha c^T c &= b \\
c^T y &= b - \alpha c^T c
\end{aligned}
$$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T (x - \pi) \geq 0$

$$
\begin{aligned}
(y + \alpha c - y)^T (x - y - \alpha c) &= \\
\alpha c^T (x - y - \alpha c) &= \\
\alpha (c^T x) - \alpha (c^T y) - \alpha^2 (c^T c) &= \\
\alpha b - \alpha (b - \alpha c^T c) - \alpha^2 c^T c &=
\end{aligned}
$$

Пример: проекция на гиперплоскость

Найти $\pi_S(y) = \pi$, если $S = \{x \in \mathbb{R}^n \mid c^T x = b\}$, $y \notin S$. Построим гипотезу из рисунка: $\pi = y + \alpha c$. Коэффициент α выбирается так, чтобы $\pi \in S$: $c^T \pi = b$, поэтому:

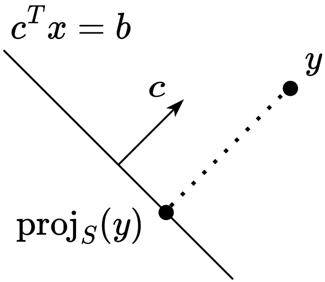


Рис. 8: Гиперплоскость

$$
\begin{aligned}
c^T(y + \alpha c) &= b \\
c^T y + \alpha c^T c &= b \\
c^T y &= b - \alpha c^T c
\end{aligned}
$$

Проверим неравенство для выпуклого замкнутого множества: $(\pi - y)^T (x - \pi) \geq 0$

$$
\begin{aligned}
(y + \alpha c - y)^T (x - y - \alpha c) &= \\
\alpha c^T (x - y - \alpha c) &= \\
\alpha (c^T x) - \alpha (c^T y) - \alpha^2 (c^T c) &= \\
\alpha b - \alpha (b - \alpha c^T c) - \alpha^2 c^T c &= \\
\alpha b - \alpha b + \alpha^2 c^T c - \alpha^2 c^T c &= 0 \geq 0
\end{aligned}
$$

Метод проекции градиента (PGD)

Идея

$$x_{k+1} = \text{proj}_S \left(x_k - \alpha_k \nabla f(x_k)\right) \qquad \Leftrightarrow \qquad \begin{aligned} y_k &= x_k - \alpha_k \nabla f(x_k) \\ x_{k+1} &= \text{proj}_S \left(y_k\right) \end{aligned}$$

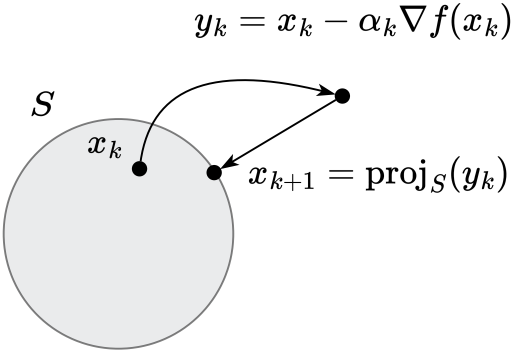


Рис. 9: Иллюстрация алгоритма метода проекции градиента

Скорость сходимости для гладких выпуклых функций

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ выпуклая и дифференцируемая. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть есть минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Алгоритм проекции градиента с шагом $\frac{1}{L}$ достигает следующей сходимости после итерации $k > 0$:

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

Скорость сходимости для гладких сильно выпуклых функций

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ является μ-сильно выпуклой. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть есть минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Алгоритм проекции градиента с шагом $\alpha \leq \frac{1}{L}$ достигает следующей сходимости после $k > 0$:

$$\|x_k - x^*\|_2^2 \leq (1 - \alpha\mu)^k \|x_0 - x^*\|_2^2$$

Результаты сходимости метода проекции градиента для гладких функций

Метод проекции градиента: $\min\limits_{x \in S} f(x)$ $\quad x_{k+1} = \Pi_S(x_k - \alpha_k \nabla f(x_k))$ $\quad \lambda\left(\nabla^2 f(x)\right) \in [\mu, L], \varkappa = \frac{L}{\mu}$

гладкая & выпуклая	гладкая & сильно выпуклая
$f(x_k) - f^* = \mathcal{O}\left(\frac{1}{k}\right)$ $k_\varepsilon = \mathcal{O}\left(\frac{1}{\varepsilon}\right)$	$\\|x_k - x^*\\|^2 = \mathcal{O}\left(\left(1 - \frac{\mu}{L}\right)^k\right)$ $k_\varepsilon = \mathcal{O}\left(\varkappa \log \frac{1}{\varepsilon}\right)$

Результаты сходимости метода проекции градиента для гладких функций

Метод проекции градиента: $\min\limits_{x \in S} f(x) \qquad x_{k+1} = \Pi_S(x_k - \alpha_k \nabla f(x_k)) \quad \lambda\left(\nabla^2 f(x)\right) \in [\mu, L]\,, \varkappa = \frac{L}{\mu}$

гладкая & выпуклая	гладкая & сильно выпуклая
$f(x_k) - f^* = \mathcal{O}\left(\frac{1}{k}\right)$ $k_\varepsilon = \mathcal{O}\left(\frac{1}{\varepsilon}\right)$	$\|x_k - x^*\|^2 = \mathcal{O}\left(\left(1 - \frac{\mu}{L}\right)^k\right)$ $k_\varepsilon = \mathcal{O}\left(\varkappa \log \frac{1}{\varepsilon}\right)$

- Обратите внимание, что характер сходимости метода по итерациям в точности совпадает с безусловным случаем (градиентным спуском).

Результаты сходимости метода проекции градиента для гладких функций

Метод проекции градиента: $\min\limits_{x \in S} f(x)$ $\quad x_{k+1} = \Pi_S(x_k - \alpha_k \nabla f(x_k))$ $\quad \lambda\left(\nabla^2 f(x)\right) \in [\mu, L]\,, \varkappa = \frac{L}{\mu}$

гладкая & выпуклая	гладкая & сильно выпуклая
$f(x_k) - f^* = \mathcal{O}\left(\frac{1}{k}\right)$ $k_\varepsilon = \mathcal{O}\left(\frac{1}{\varepsilon}\right)$	$\|x_k - x^*\|^2 = \mathcal{O}\left(\left(1 - \frac{\mu}{L}\right)^k\right)$ $k_\varepsilon = \mathcal{O}\left(\varkappa \log \frac{1}{\varepsilon}\right)$

- Обратите внимание, что характер сходимости метода по итерациям в точности совпадает с безусловным случаем (градиентным спуском).
- Однако, стоимость итерации может быть **сильно** больше. Иногда, стоимость одной итерации может по порядку совпадать со стоимостью решения всей задачи.

Метод Франк-Вульфа

Рис. 10: Маргарет Штраус Франк (1927-2024)

Рис. 11: Филипп Вульф (1927-2016)

Идея

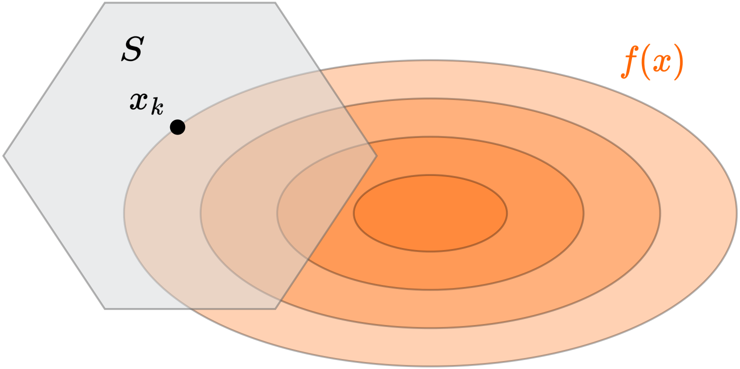


Рис. 12: Иллюстрация метода Франк-Вульфа (метод условного градиента)

Идея

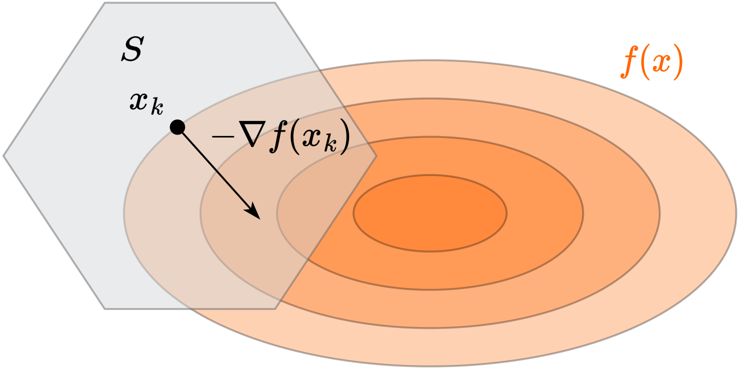


Рис. 13: Иллюстрация метода Франк-Вульфа (метод условного градиента)

Идея

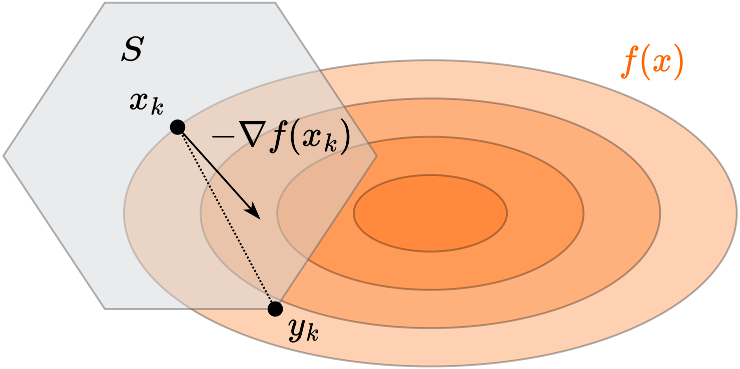


Рис. 14: Иллюстрация метода Франк-Вульфа (метод условного градиента)

Идея

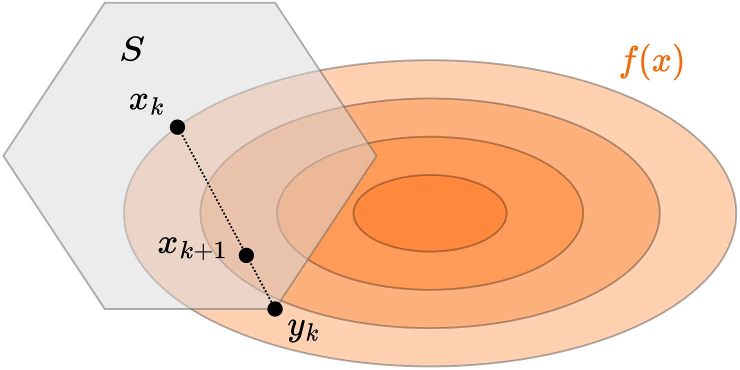


Рис. 15: Иллюстрация метода Франк-Вульфа (метод условного градиента)

Идея

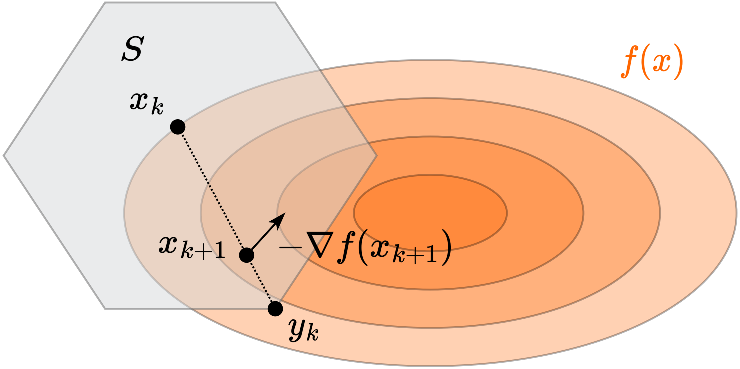


Рис. 16: Иллюстрация метода Франк-Вульфа (метод условного градиента)

Идея

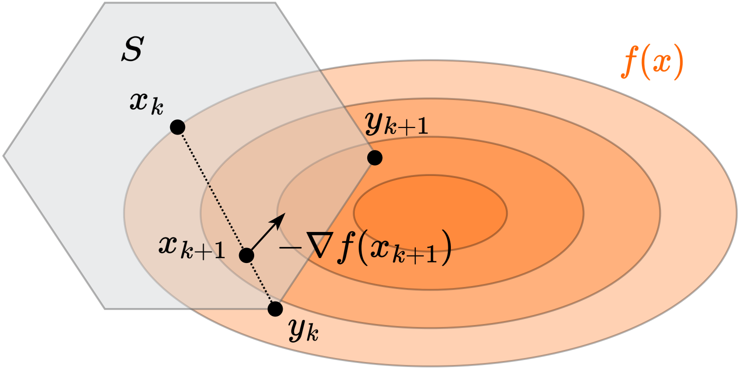


Рис. 17: Иллюстрация метода Франк-Вульфа (метод условного градиента)

Идея

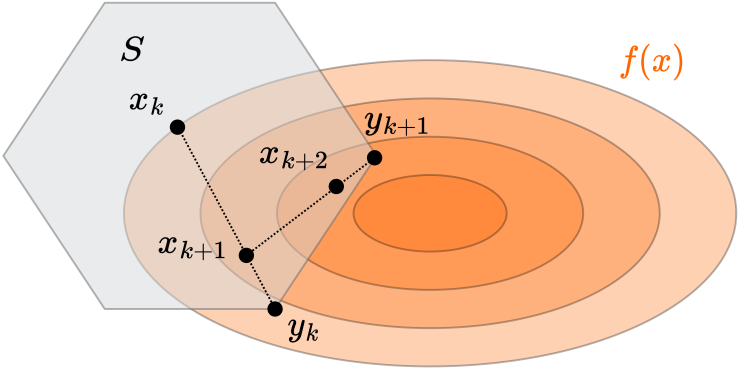


Рис. 18: Иллюстрация метода Франк-Вульфа (метод условного градиента)

Идея

$$
y_k = \arg\min_{x \in S} f^I_{x_k}(x) = \arg\min_{x \in S} \langle \nabla f(x_k), x \rangle
$$

$$
x_{k+1} = \gamma_k x_k + (1 - \gamma_k) y_k
$$

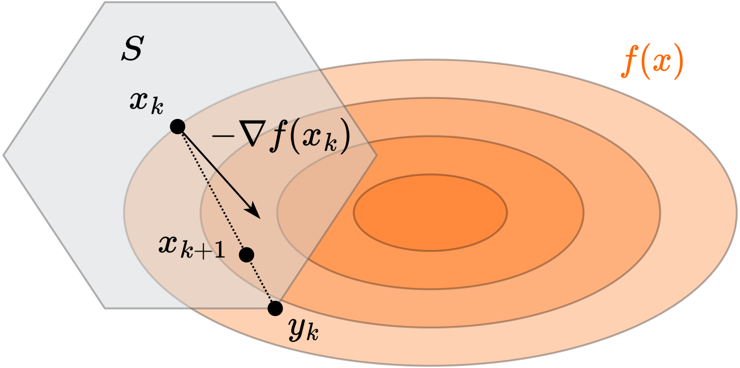


Рис. 19: Иллюстрация метода Франк-Вульфа (метод условного градиента)

Скорость сходимости для гладкой и выпуклой функции

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ является выпуклой и дифференцируемой. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть существует минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Метод Франк-Вульфа с шагом $\gamma_k = \frac{k-1}{k+1}$ достигает следующей сходимости после $k > 0$ итераций:

$$
f(x_k) - f^* \leq \frac{2LR^2}{k+1}
$$

где $R = \max\limits_{x,y \in S} \|x - y\|$ является диаметром множества S.

Нижняя граница для метода Франк-Вульфа [2]

Theorem

Рассмотрим любой алгоритм, который использует только линейный минимизатор (LMO). Пусть диаметр множества S равен R. Существует L-гладкая сильно выпуклая функция $f : \mathbb{R}^n \to \mathbb{R}$ такая, что этот алгоритм требует по крайней мере

$$\min\left(\frac{n}{2}, \frac{LR^2}{16\varepsilon}\right)$$

итераций (т.е. вызовов LMO) для построения точки $\hat{x} \in S$ с $f(\hat{x}) - \min\limits_{x \in S} f(x) \leq \varepsilon$. Нижняя граница применима как для выпуклых, так и для сильно выпуклых функций.

[2] The Complexity of Large-scale Convex Programming under a Linear Optimization Oracle

Summary по методу Франк-Вульфа

- Метод не требует проекций, в некоторых специальных случаях позволяет вычислять итерации аналитически

Summary по методу Франк-Вульфа

- Метод не требует проекций, в некоторых специальных случаях позволяет вычислять итерации аналитически
- Глобальная скорость сходимости равна $O\left(\frac{1}{k}\right)$ для гладких и выпуклых функций. Сильная выпуклость не улучшает скорость - это нижняя граница для LMO

Summary по методу Франк-Вульфа

- Метод не требует проекций, в некоторых специальных случаях позволяет вычислять итерации аналитически
- Глобальная скорость сходимости равна $O\left(\frac{1}{k}\right)$ для гладких и выпуклых функций. Сильная выпуклость не улучшает скорость - это нижняя граница для LMO
- В сравнении с методом проекции градиента, скорость хуже, но итерация может быть дешевле и более разреженной

Summary по методу Франк-Вульфа

- Метод не требует проекций, в некоторых специальных случаях позволяет вычислять итерации аналитически
- Глобальная скорость сходимости равна $O\left(\frac{1}{k}\right)$ для гладких и выпуклых функций. Сильная выпуклость не улучшает скорость - это нижняя граница для LMO
- В сравнении с методом проекции градиента, скорость хуже, но итерация может быть дешевле и более разреженной
- Недавно было показано, что для сильно выпуклых множеств, скорость может быть улучшена до $O\left(\frac{1}{k^2}\right)$ (paper)

Summary по методу Франк-Вульфа

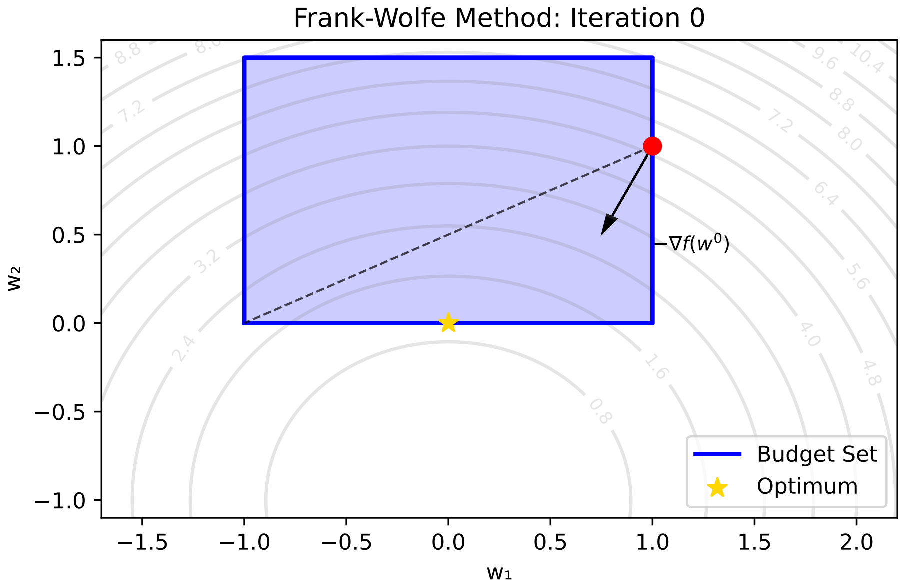
- Метод не требует проекций, в некоторых специальных случаях позволяет вычислять итерации аналитически
- Глобальная скорость сходимости равна $O\left(\frac{1}{k}\right)$ для гладких и выпуклых функций. Сильная выпуклость не улучшает скорость - это нижняя граница для LMO
- В сравнении с методом проекции градиента, скорость хуже, но итерация может быть дешевле и более разреженной
- Недавно было показано, что для сильно выпуклых множеств, скорость может быть улучшена до $O\left(\frac{1}{k^2}\right)$ (paper)
- Если мы позволяем steps away, сходимость становится линейной (paper) в сильно выпуклом случае

Summary по методу Франк-Вульфа

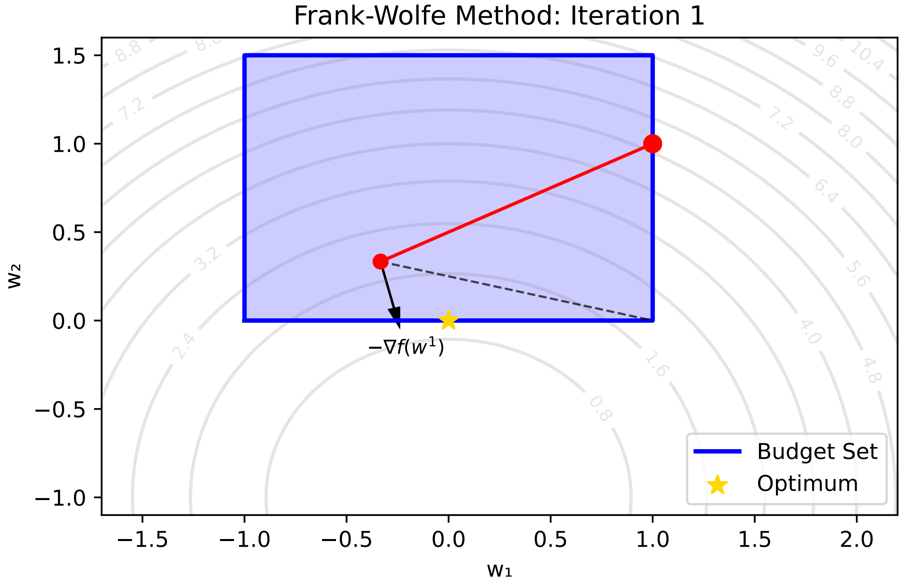
- Метод не требует проекций, в некоторых специальных случаях позволяет вычислять итерации аналитически
- Глобальная скорость сходимости равна $O\left(\frac{1}{k}\right)$ для гладких и выпуклых функций. Сильная выпуклость не улучшает скорость - это нижняя граница для LMO
- В сравнении с методом проекции градиента, скорость хуже, но итерация может быть дешевле и более разреженной
- Недавно было показано, что для сильно выпуклых множеств, скорость может быть улучшена до $O\left(\frac{1}{k^2}\right)$ (paper)
- Если мы позволяем steps away, сходимость становится линейной (paper) в сильно выпуклом случае
- Недавняя работа показала расширение на негладкий случай (paper) с скоростью сходимости $O\left(\frac{1}{\sqrt{k}}\right)$

Численные эксперименты

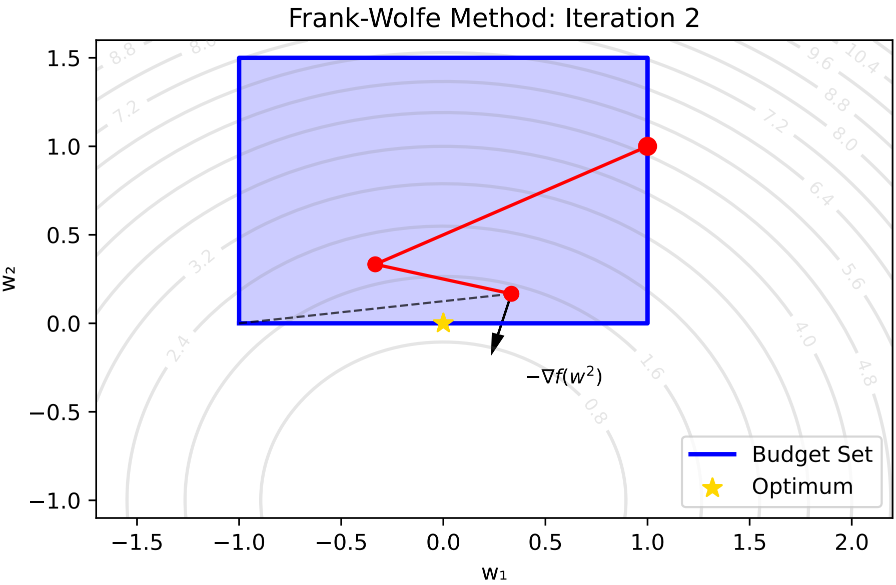
2d пример. Метод Франк-Вульфа



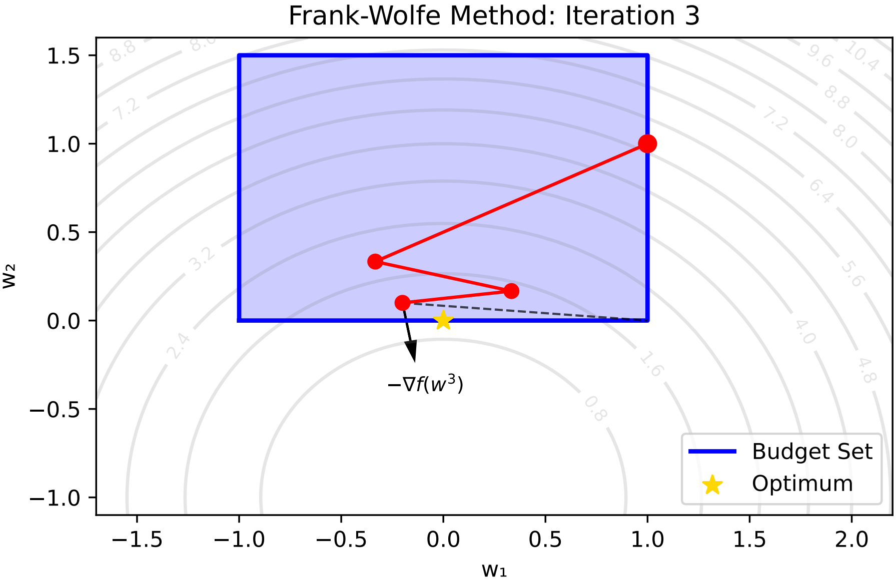
2d пример. Метод Франк-Вульфа



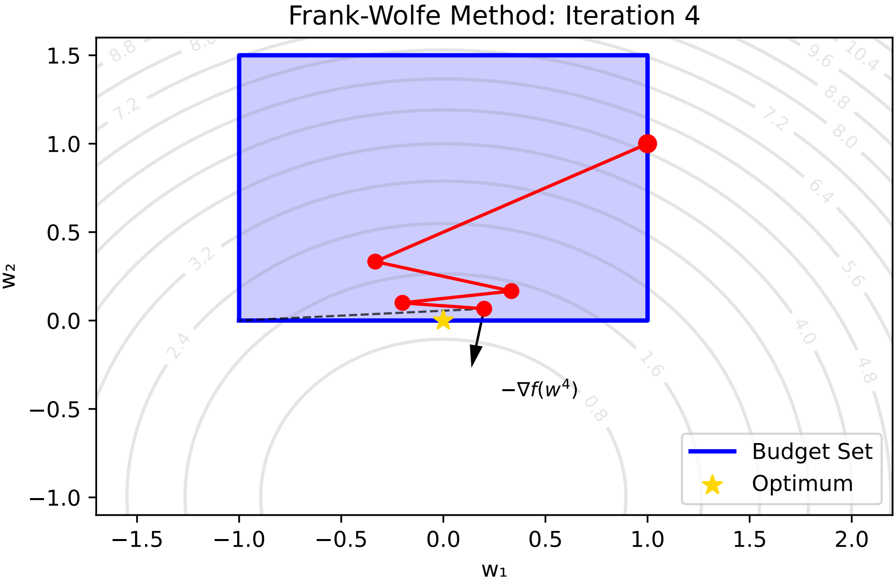
2d пример. Метод Франк-Вульфа



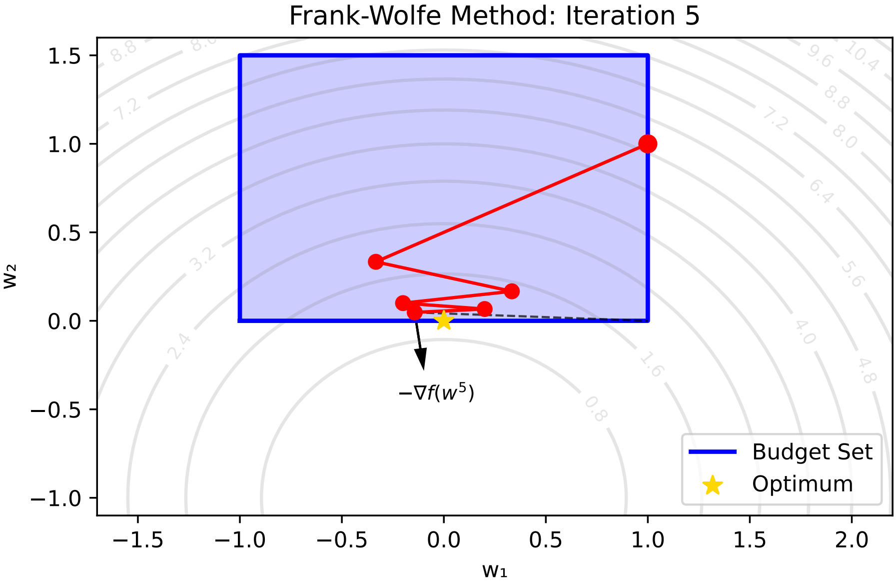
2d пример. Метод Франк-Вульфа



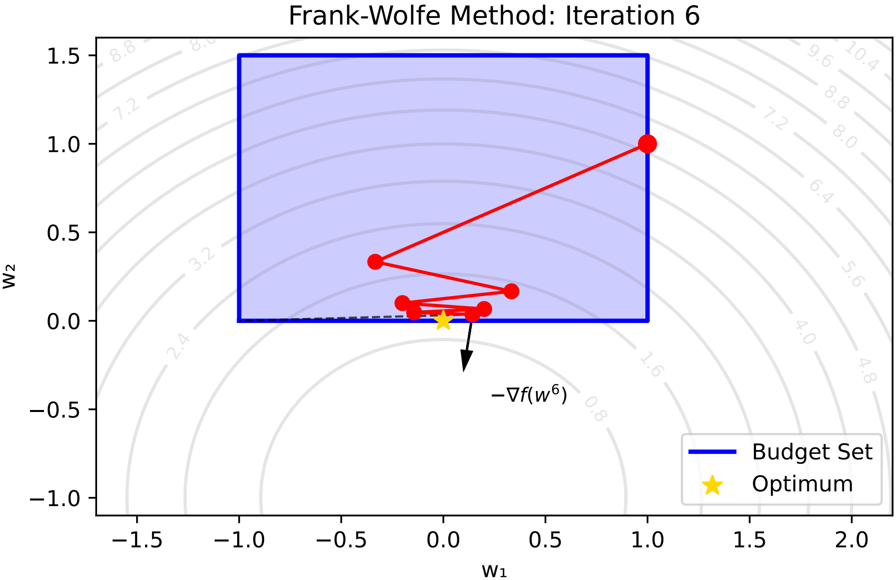
2d пример. Метод Франк-Вульфа



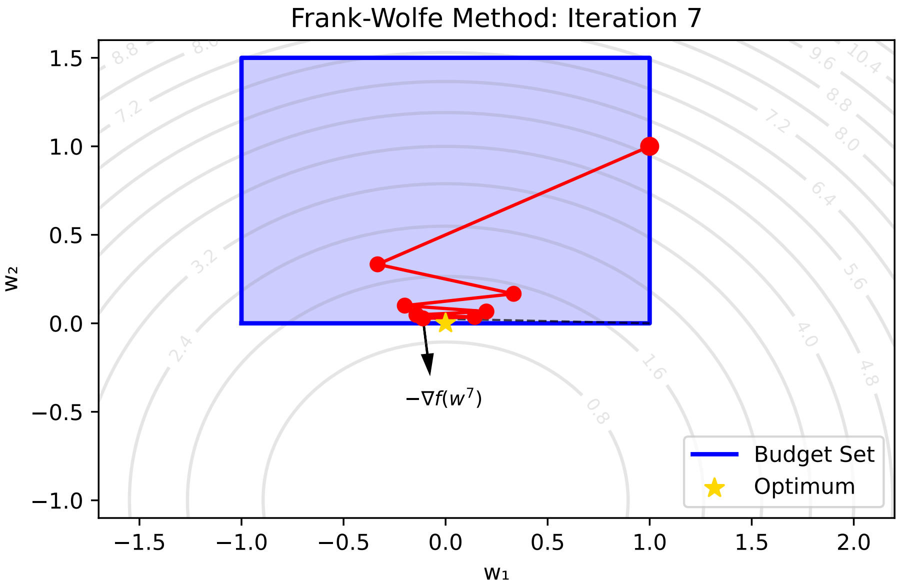
2d пример. Метод Франк-Вульфа



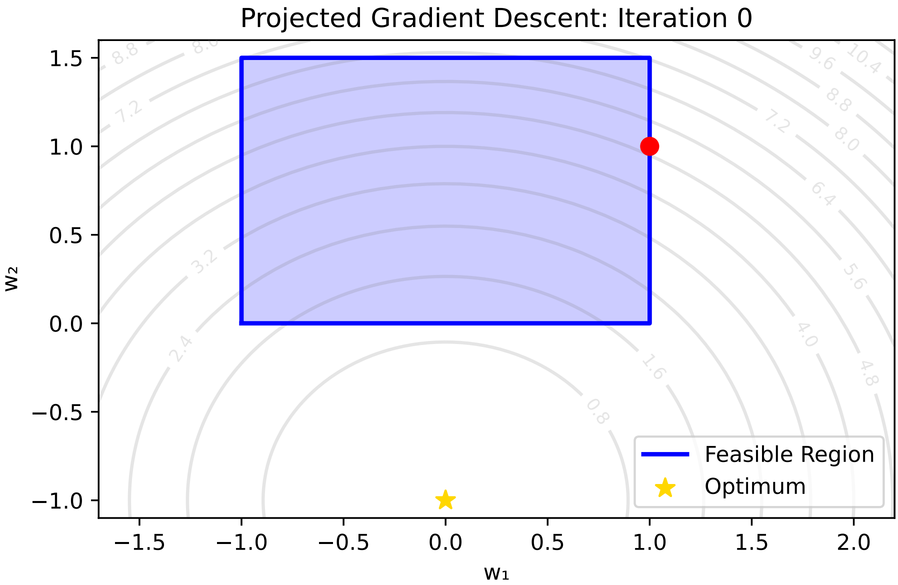
2d пример. Метод Франк-Вульфа



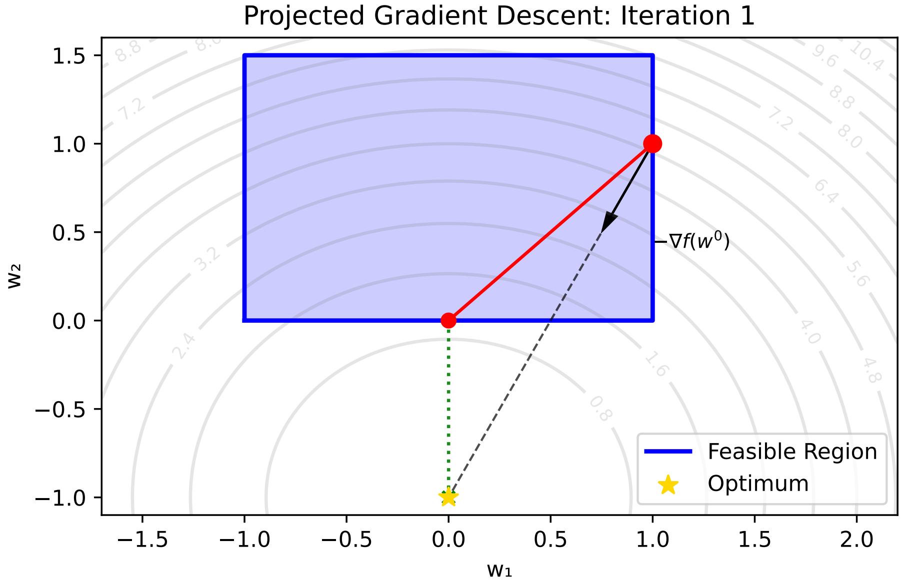
2d пример. Метод Франк-Вульфа



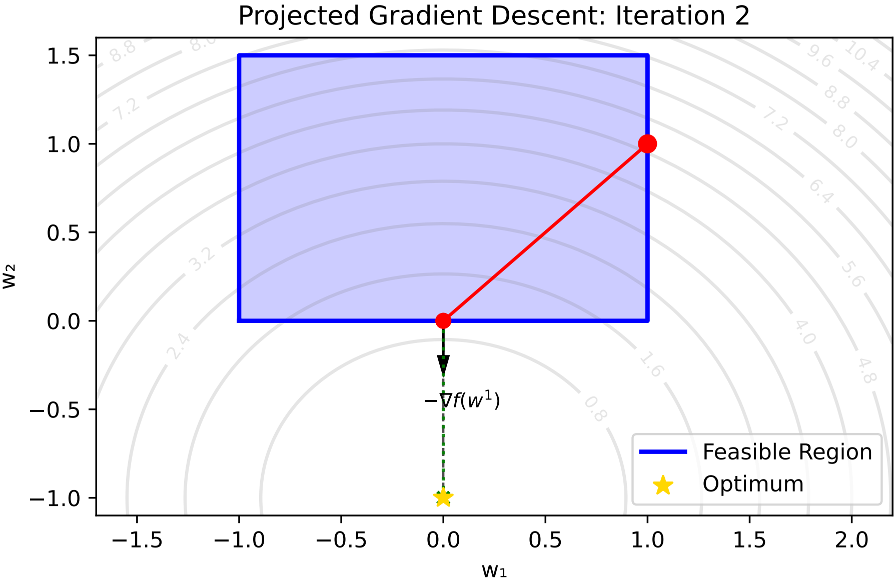
2d пример. Метод проекции градиента



2d пример. Метод проекции градиента



2d пример. Метод проекции градиента



Квадратичная функция на кубе

$$\min_{\substack{x \in \mathbb{R}^n \\ -\mathbf{1} \preceq x \preceq \mathbf{1}}} \frac{1}{2} x^\top A x - b^\top x,$$

$$A \in \mathbb{R}^{n \times n}, \quad \lambda(A) \in [\mu; L].$$

Проекция простая:

$$\pi_S(x) = \text{clip}(x, -\mathbf{1}, \mathbf{1}).$$

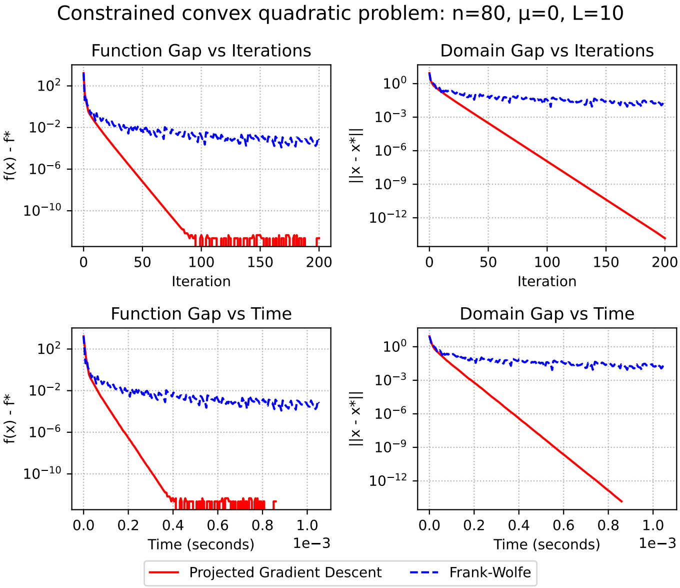
или

$$\pi_S(x) = \max\left(-\mathbf{1}, \min(\mathbf{1}, x)\right).$$

Линейный минимизатор (LMO) равен $y = \underset{z \in S}{\operatorname{argmin}} \langle g, z \rangle$.

Поскольку множество допустимых значений разделяется по координатам, решение вычисляется покоординатно как

$$y_i = \begin{cases} -1, & \text{if } g_i > 0, \\ 1, & \text{if } g_i \leq 0. \end{cases}$$



Квадратичная функция на кубе

$$
\min_{\substack{x \in \mathbb{R}^n \\ -\mathbf{1} \preceq x \preceq \mathbf{1}}} \frac{1}{2} x^\top A x - b^\top x,
$$

$$
A \in \mathbb{R}^{n \times n}, \quad \lambda(A) \in [\mu; L].
$$

Проекция простая:

$$
\pi_S(x) = \text{clip}(x, -\mathbf{1}, \mathbf{1}).
$$

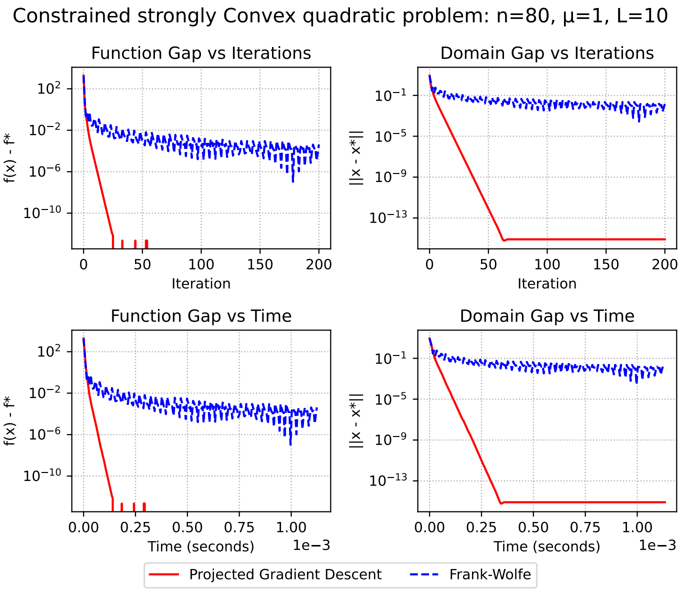
или

$$
\pi_S(x) = \max\left(-\mathbf{1}, \min(\mathbf{1}, x)\right).
$$

Линейный минимизатор (LMO) равен $y = \underset{z \in S}{\text{argmin}} \langle g, z \rangle$.

Поскольку множество допустимых значений разделяется по координатам, решение вычисляется покоординатно как

$$
y_i = \begin{cases} -1, & \text{if } g_i > 0, \\ 1, & \text{if } g_i \leq 0. \end{cases}
$$



Квадратичная функция на симплексе (Диагональная матрица)

$$\min_{\substack{x \in \mathbb{R}^n \\ x \geq 0, \mathbf{1}^T x = 1}} \frac{1}{2} x^T A x,$$

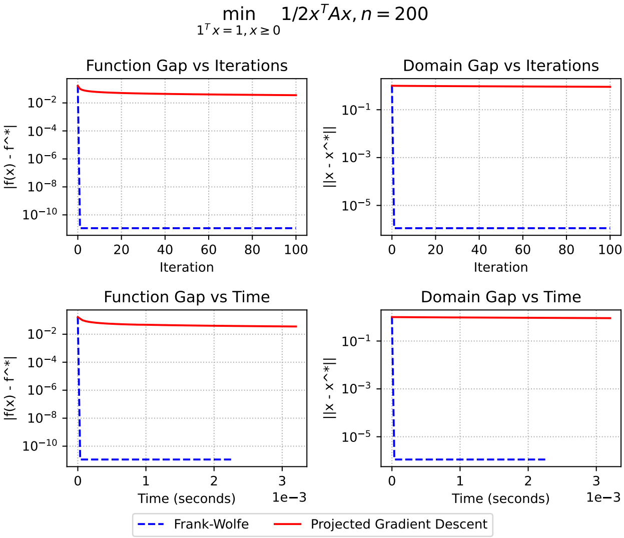
$$A \in \mathbb{R}^{n \times n}, \quad \lambda(A) \in [0; 100].$$

Метод	Обновление, мс	LMO/проекция, мс
PGD	0.0069	0.0167
FW	0.0070	0.0066

Проекция на единичный симплекс $\pi_S(x)$ может быть выполнена за $\mathcal{O}(n \log n)$ или в среднем за $\mathcal{O}(n)$ операций. [3]

Линейный минимизатор (LMO) равен $y = \underset{z \in S}{\operatorname{argmin}} \langle g, z \rangle$. Решение соответствует вершине симплекса:

$$y = e_j \quad \text{where} \quad j = \underset{i}{\operatorname{argmin}}\, g_i.$$



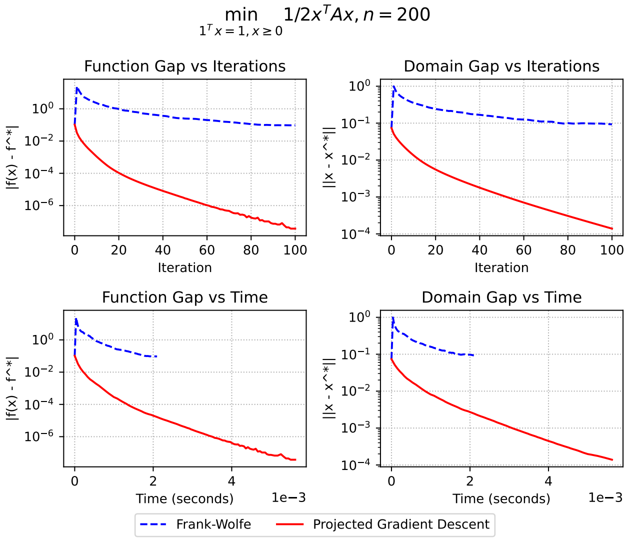
[3] Efficient Projections onto the ℓ_1-Ball for Learning in High Dimensions

Квадратичная функция на симплексе

$$\min_{\substack{x \in \mathbb{R}^n \\ x \geq 0, \mathbf{1}^T x = 1}} \frac{1}{2} x^T A x,$$

$$A \in \mathbb{R}^{n \times n}, \quad \lambda(A) \in [0; 100].$$

Метод	Обновление, мс	LMO/проекция, мс
PGD	0.0069	0.0420
FW	0.0069	0.0066

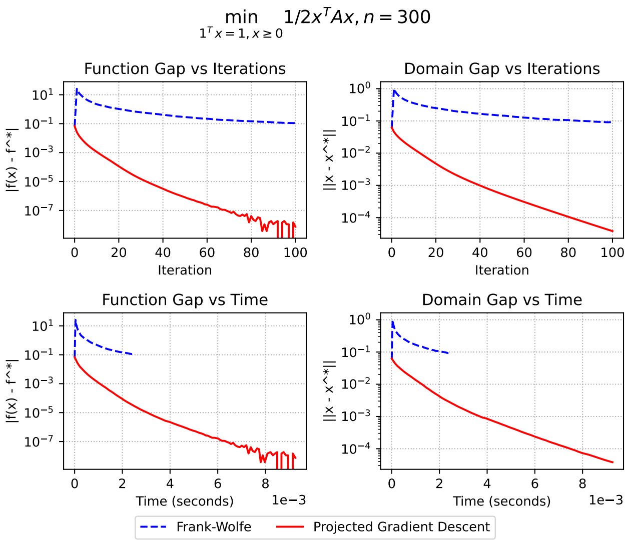


Квадратичная функция на симплексе

$$\min_{\substack{x \in \mathbb{R}^n \\ x \geq 0, \mathbf{1}^T x = 1}} \frac{1}{2} x^T A x,$$

$$A \in \mathbb{R}^{n \times n}, \quad \lambda(A) \in [0; 100].$$

Метод	Обновление, мс	LMO/проекция, мс
PGD	0.0068	0.0761
FW	0.0069	0.0070

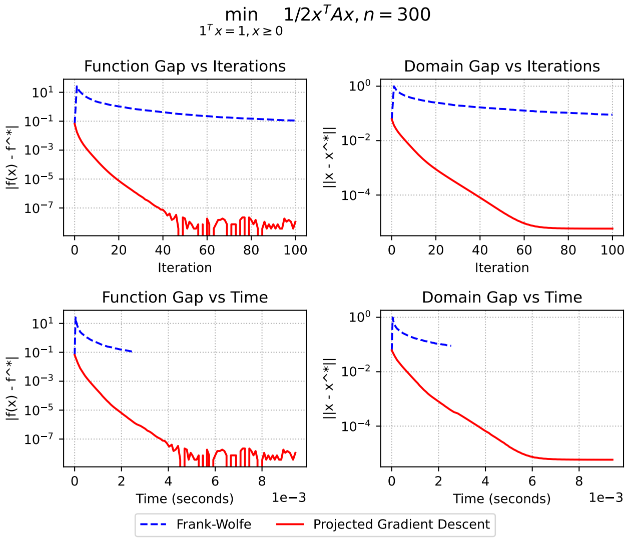


Квадратичная функция на симплексе

$$\min_{\substack{x \in \mathbb{R}^n \\ x \geq 0, \mathbf{1}^T x = 1}} \frac{1}{2} x^T A x,$$

$$A \in \mathbb{R}^{n \times n}, \quad \lambda(A) \in [1; 100].$$

Метод	Обновление, мс	LMO/проекция, мс
PGD	0.0068	0.0752
FW	0.0067	0.0068



Метод проекции градиента vs метод Франк-Вульфа

Основное отличие между методами заключается в том, что метод проекции градиента требует проекцию, в то время как метод Франк-Вульфа требует только линейный минимизатор (LMO).

В недавней книге авторы представили следующую таблицу сравнения сложности линейных минимизаций и проекций на некоторые выпуклые множества с точностью до аддитивной ошибки ϵ в евклидовой норме.

Множество	**Линейный минимизатор**	**Проекция**
n-мерный ℓ_p-шар, $p \neq 1, 2, \infty$	$\mathcal{O}(n)$	$\tilde{\mathcal{O}}(\frac{n}{\epsilon^2})$
Ядро нормы матрицы $n \times m$	$\mathcal{O}\left(\nu \ln(m+n) \frac{\sqrt{\sigma_1}}{\sqrt{\epsilon}}\right)$	$\mathcal{O}(m\, n\, \min\{m, n\})$
Потоковый многогранник на графе с m вершинами и n ребрами (ограничение на пропускную способность ребер)	$\mathcal{O}\Big((n \log m)(n + m \log m)\Big)$	$\tilde{\mathcal{O}}(\frac{n}{\epsilon^2})$ or $\mathcal{O}(n^4 \log n)$
Многогранник Биркгофа ($n \times n$ дважды стохастических матриц)	$\mathcal{O}(n^3)$	$\tilde{\mathcal{O}}(\frac{n^2}{\epsilon^2})$

Когда ϵ отсутствует, нет аддитивной ошибки. $\tilde{\mathcal{O}}$ скрывает полилогарифмические факторы в размерностях и полиномиальные факторы в константах, связанных с расстоянием до оптимума. Для ядерной нормы шара, ν обозначает количество ненулевых элементов, а σ_1 обозначает наибольшее сингулярное значение проекции матрицы.

Бонус: Доказательства

Критерий проекции (неравенство Бурбаки-Чейни-Гольдштейна)

Theorem

Пусть $S \subseteq \mathbb{R}^n$ - замкнутое и выпуклое, $\forall x \in S, y \in \mathbb{R}^n$. Тогда

$$\langle y - \text{proj}_S(y), x - \text{proj}_S(y)\rangle \leq 0 \tag{1}$$

$$\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2 \tag{2}$$

1. $\text{proj}_S(y)$ - минимизатор дифференцируемой выпуклой функции $d(y, S, \|\cdot\|) = \|x - y\|^2$ на S. По условию оптимальности первого порядка:

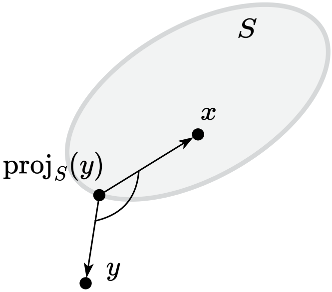


Рис. 20: Геометрическая иллюстрация первого утверждения. Для любой точки $x \in S$ угол между векторами $x - \text{proj}_S(y)$ и $y - \text{proj}_S(y)$ должен быть не острый

Критерий проекции (неравенство Бурбаки-Чейни-Гольдштейна)

Theorem

Пусть $S \subseteq \mathbb{R}^n$ - замкнутое и выпуклое, $\forall x \in S, y \in \mathbb{R}^n$. Тогда

$$\langle y - \text{proj}_S(y), x - \text{proj}_S(y)\rangle \leq 0 \tag{1}$$

$$\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2 \tag{2}$$

1. $\text{proj}_S(y)$ - минимизатор дифференцируемой выпуклой функции $d(y, S, \|\cdot\|) = \|x - y\|^2$ на S. По условию оптимальности первого порядка:

$$\nabla d(\text{proj}_S(y))^T(x - \text{proj}_S(y)) \geq 0$$

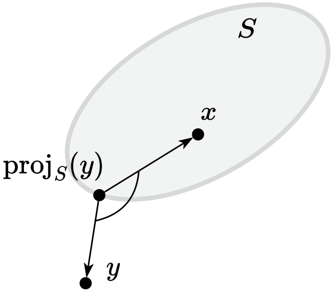


Рис. 20: Геометрическая иллюстрация первого утверждения. Для любой точки $x \in S$ угол между векторами $x - \text{proj}_S(y)$ и $y - \text{proj}_S(y)$ должен быть не острый

Критерий проекции (неравенство Бурбаки-Чейни-Гольдштейна)

Theorem

Пусть $S \subseteq \mathbb{R}^n$ - замкнутое и выпуклое, $\forall x \in S, y \in \mathbb{R}^n$. Тогда

$$\langle y - \text{proj}_S(y), x - \text{proj}_S(y)\rangle \leq 0 \tag{1}$$

$$\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2 \tag{2}$$

1. $\text{proj}_S(y)$ - минимизатор дифференцируемой выпуклой функции $d(y, S, \|\cdot\|) = \|x - y\|^2$ на S. По условию оптимальности первого порядка:

$$\nabla d(\text{proj}_S(y))^T(x - \text{proj}_S(y)) \geq 0$$

$$2\left(\text{proj}_S(y) - y\right)^T(x - \text{proj}_S(y)) \geq 0$$

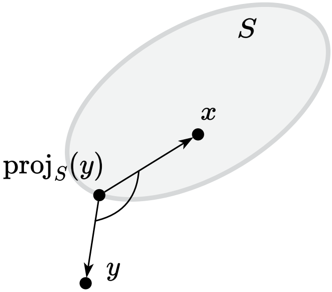


Рис. 20: Геометрическая иллюстрация первого утверждения. Для любой точки $x \in S$ угол между векторами $x - \text{proj}_S(y)$ и $y - \text{proj}_S(y)$ должен быть не острый

Критерий проекции (неравенство Бурбаки-Чейни-Гольдштейна)

Theorem

Пусть $S \subseteq \mathbb{R}^n$ - замкнутое и выпуклое, $\forall x \in S, y \in \mathbb{R}^n$. Тогда

$$\langle y - \text{proj}_S(y), x - \text{proj}_S(y)\rangle \leq 0 \tag{1}$$

$$\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2 \tag{2}$$

1. $\text{proj}_S(y)$ - минимизатор дифференцируемой выпуклой функции $d(y, S, \|\cdot\|) = \|x - y\|^2$ на S. По условию оптимальности первого порядка:

$$\nabla d(\text{proj}_S(y))^T(x - \text{proj}_S(y)) \geq 0$$

$$2\left(\text{proj}_S(y) - y\right)^T(x - \text{proj}_S(y)) \geq 0$$

$$\left(y - \text{proj}_S(y)\right)^T(x - \text{proj}_S(y)) \leq 0$$

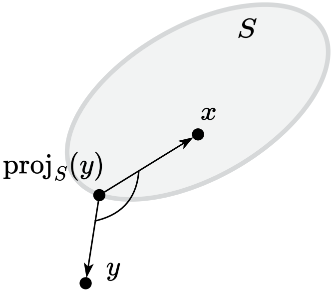


Рис. 20: Геометрическая иллюстрация первого утверждения. Для любой точки $x \in S$ угол между векторами $x - \text{proj}_S(y)$ и $y - \text{proj}_S(y)$ должен быть не острый

Критерий проекции (неравенство Бурбаки-Чейни-Гольдштейна)

Theorem

Пусть $S \subseteq \mathbb{R}^n$ - замкнутое и выпуклое, $\forall x \in S, y \in \mathbb{R}^n$. Тогда

$$\langle y - \text{proj}_S(y), x - \text{proj}_S(y)\rangle \leq 0 \tag{1}$$

$$\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2 \tag{2}$$

1. $\text{proj}_S(y)$ - минимизатор дифференцируемой выпуклой функции $d(y, S, \|\cdot\|) = \|x - y\|^2$ на S. По условию оптимальности первого порядка:

$$\nabla d(\text{proj}_S(y))^T(x - \text{proj}_S(y)) \geq 0$$

$$2\left(\text{proj}_S(y) - y\right)^T(x - \text{proj}_S(y)) \geq 0$$

$$\left(y - \text{proj}_S(y)\right)^T(x - \text{proj}_S(y)) \leq 0$$

2. Используем правило косинусов $2x^Ty = \|x\|^2 + \|y\|^2 - \|x - y\|^2$ с $x = x - \text{proj}_S(y)$ и $y = y - \text{proj}_S(y)$. По первому свойству теоремы:

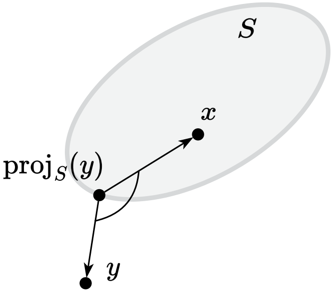


Рис. 20: Геометрическая иллюстрация первого утверждения. Для любой точки $x \in S$ угол между векторами $x - \text{proj}_S(y)$ и $y - \text{proj}_S(y)$ должен быть не острый

Критерий проекции (неравенство Бурбаки-Чейни-Гольдштейна)

Theorem

Пусть $S \subseteq \mathbb{R}^n$ - замкнутое и выпуклое, $\forall x \in S, y \in \mathbb{R}^n$. Тогда

$$\langle y - \text{proj}_S(y), x - \text{proj}_S(y)\rangle \leq 0 \tag{1}$$

$$\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2 \tag{2}$$

1. $\text{proj}_S(y)$ - минимизатор дифференцируемой выпуклой функции $d(y, S, \|\cdot\|) = \|x - y\|^2$ на S. По условию оптимальности первого порядка:

$$\nabla d(\text{proj}_S(y))^T(x - \text{proj}_S(y)) \geq 0$$

$$2\left(\text{proj}_S(y) - y\right)^T (x - \text{proj}_S(y)) \geq 0$$

$$\left(y - \text{proj}_S(y)\right)^T (x - \text{proj}_S(y)) \leq 0$$

2. Используем правило косинусов $2x^Ty = \|x\|^2 + \|y\|^2 - \|x - y\|^2$ с $x = x - \text{proj}_S(y)$ и $y = y - \text{proj}_S(y)$. По первому свойству теоремы:

$$0 \geq 2x^Ty = \|x - \text{proj}_S(y)\|^2 + \|y + \text{proj}_S(y)\|^2 - \|x - y\|^2$$

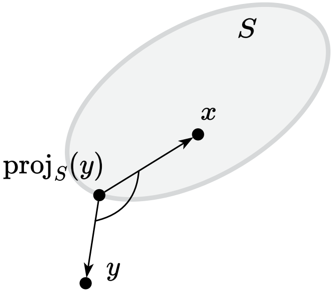


Рис. 20: Геометрическая иллюстрация первого утверждения. Для любой точки $x \in S$ угол между векторами $x - \text{proj}_S(y)$ и $y - \text{proj}_S(y)$ должен быть не острый

Критерий проекции (неравенство Бурбаки-Чейни-Гольдштейна)

Theorem

Пусть $S \subseteq \mathbb{R}^n$ - замкнутое и выпуклое, $\forall x \in S, y \in \mathbb{R}^n$. Тогда

$$\langle y - \text{proj}_S(y), x - \text{proj}_S(y)\rangle \leq 0 \tag{1}$$

$$\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2 \tag{2}$$

1. $\text{proj}_S(y)$ - минимизатор дифференцируемой выпуклой функции $d(y, S, \|\cdot\|) = \|x - y\|^2$ на S. По условию оптимальности первого порядка:

$$\nabla d(\text{proj}_S(y))^T(x - \text{proj}_S(y)) \geq 0$$

$$2\left(\text{proj}_S(y) - y\right)^T (x - \text{proj}_S(y)) \geq 0$$

$$\left(y - \text{proj}_S(y)\right)^T (x - \text{proj}_S(y)) \leq 0$$

2. Используем правило косинусов $2x^Ty = \|x\|^2 + \|y\|^2 - \|x - y\|^2$ с $x = x - \text{proj}_S(y)$ и $y = y - \text{proj}_S(y)$. По первому свойству теоремы:

$$0 \geq 2x^Ty = \|x - \text{proj}_S(y)\|^2 + \|y + \text{proj}_S(y)\|^2 - \|x - y\|^2$$

$$\|x - \text{proj}_S(y)\|^2 + \|y + \text{proj}_S(y)\|^2 \leq \|x - y\|^2$$

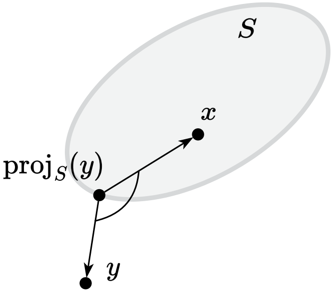


Рис. 20: Геометрическая иллюстрация первого утверждения. Для любой точки $x \in S$ угол между векторами $x - \text{proj}_S(y)$ и $y - \text{proj}_S(y)$ должен быть не острый

Оператор проекции является нерастягивающим

Сокращенная запись: пусть $\pi = \text{proj}$ и $\pi(x)$ обозначает $\text{proj}(x)$.

Начнём с важного свойства проекции:

$$\langle y - \pi(y), x - \pi(y) \rangle \leq 0 \quad \forall x \in S. \tag{3}$$

Оператор проекции является нерастягивающим

Сокращенная запись: пусть $\pi = \text{proj}$ и $\pi(x)$ обозначает $\text{proj}(x)$.

Начнём с важного свойства проекции:

$$
\langle y - \pi(y), x - \pi(y) \rangle \leq 0 \quad \forall x \in S. \tag{3}
$$

Заменим x на $\pi(x)$ в (3)

$$
\langle y - \pi(y), \pi(x) - \pi(y) \rangle \leq 0. \tag{4}
$$

Оператор проекции является нерастягивающим

Сокращенная запись: пусть $\pi = \text{proj}$ и $\pi(x)$ обозначает $\text{proj}(x)$.

Начнём с важного свойства проекции:

$$\langle y - \pi(y), x - \pi(y)\rangle \leq 0 \quad \forall x \in S. \tag{3}$$

Заменим x на $\pi(x)$ в (3)

$$\langle y - \pi(y), \pi(x) - \pi(y)\rangle \leq 0. \tag{4}$$

Заменим y на x и x на $\pi(y)$ в (3)

$$\langle x - \pi(x), \pi(y) - \pi(x)\rangle \leq 0. \tag{5}$$

Оператор проекции является нерастягивающим

Сокращенная запись: пусть $\pi = \text{proj}$ и $\pi(x)$ обозначает $\text{proj}(x)$.

Начнём с важного свойства проекции:

$$\langle y - \pi(y), x - \pi(y)\rangle \leq 0 \quad \forall x \in S. \tag{3}$$

Заменим x на $\pi(x)$ в (3)

$$\langle y - \pi(y), \pi(x) - \pi(y)\rangle \leq 0. \tag{4}$$

Заменим y на x и x на $\pi(y)$ в (3)

$$\langle x - \pi(x), \pi(y) - \pi(x)\rangle \leq 0. \tag{5}$$

(4)+(5) сократило $\pi(y) - \pi(x)$, что не хорошо. Поэтому перевернем знак в (5), что даст нам

$$\langle \pi(x) - x, \pi(x) - \pi(y)\rangle \leq 0. \tag{6}$$

Оператор проекции является нерастягивающим

Сокращенная запись: пусть $\pi = \text{proj}$ и $\pi(x)$ обозначает $\text{proj}(x)$.

Начнём с важного свойства проекции:

$$\langle y - \pi(y), x - \pi(y)\rangle \leq 0 \quad \forall x \in S. \tag{3}$$

Заменим x на $\pi(x)$ в (3)

$$\langle y - \pi(y), \pi(x) - \pi(y)\rangle \leq 0. \tag{4}$$

Заменим y на x и x на $\pi(y)$ в (3)

$$\langle x - \pi(x), \pi(y) - \pi(x)\rangle \leq 0. \tag{5}$$

(4)+(5) сократило $\pi(y) - \pi(x)$, что не хорошо. Поэтому перевернем знак в (5), что даст нам

$$\langle \pi(x) - x, \pi(x) - \pi(y)\rangle \leq 0. \tag{6}$$

$$\langle y - \pi(y) + \pi(x) - x, \pi(x) - \pi(y)\rangle \leq 0$$

Оператор проекции является нерастягивающим

Сокращенная запись: пусть $\pi = \text{proj}$ и $\pi(x)$ обозначает $\text{proj}(x)$.

Начнём с важного свойства проекции:

$$\langle y - \pi(y), x - \pi(y)\rangle \leq 0 \quad \forall x \in S. \tag{3}$$

Заменим x на $\pi(x)$ в (3)

$$\langle y - \pi(y), \pi(x) - \pi(y)\rangle \leq 0. \tag{4}$$

Заменим y на x и x на $\pi(y)$ в (3)

$$\langle x - \pi(x), \pi(y) - \pi(x)\rangle \leq 0. \tag{5}$$

(4)+(5) сократило $\pi(y) - \pi(x)$, что не хорошо. Поэтому перевернем знак в (5), что даст нам

$$\langle \pi(x) - x, \pi(x) - \pi(y)\rangle \leq 0. \tag{6}$$

$$\begin{aligned}\langle y - \pi(y) + \pi(x) - x, \pi(x) - \pi(y)\rangle &\leq 0 \\ \langle y - x, \pi(x) - \pi(y)\rangle &\leq -\langle \pi(x) - \pi(y), \pi(x) - \pi(y)\rangle\end{aligned}$$

Оператор проекции является нерастягивающим

Сокращенная запись: пусть $\pi = \text{proj}$ и $\pi(x)$ обозначает $\text{proj}(x)$.

Начнём с важного свойства проекции:

$$\langle y - \pi(y), x - \pi(y)\rangle \leq 0 \quad \forall x \in S. \tag{3}$$

Заменим x на $\pi(x)$ в (3)

$$\langle y - \pi(y), \pi(x) - \pi(y)\rangle \leq 0. \tag{4}$$

Заменим y на x и x на $\pi(y)$ в (3)

$$\langle x - \pi(x), \pi(y) - \pi(x)\rangle \leq 0. \tag{5}$$

(4)+(5) сократило $\pi(y) - \pi(x)$, что не хорошо. Поэтому перевернем знак в (5), что даст нам

$$\langle \pi(x) - x, \pi(x) - \pi(y)\rangle \leq 0. \tag{6}$$

$$\begin{aligned}
\langle y - \pi(y) + \pi(x) - x, \pi(x) - \pi(y)\rangle &\leq 0 \\
\langle y - x, \pi(x) - \pi(y)\rangle &\leq -\langle \pi(x) - \pi(y), \pi(x) - \pi(y)\rangle \\
\langle y - x, \pi(y) - \pi(x)\rangle &\geq \|\pi(x) - \pi(y)\|_2^2
\end{aligned}$$

Оператор проекции является нерастягивающим

Сокращенная запись: пусть $\pi = \text{proj}$ и $\pi(x)$ обозначает $\text{proj}(x)$.

Начнём с важного свойства проекции:

$$\langle y - \pi(y), x - \pi(y)\rangle \leq 0 \quad \forall x \in S. \tag{3}$$

Заменим x на $\pi(x)$ в (3)

$$\langle y - \pi(y), \pi(x) - \pi(y)\rangle \leq 0. \tag{4}$$

Заменим y на x и x на $\pi(y)$ в (3)

$$\langle x - \pi(x), \pi(y) - \pi(x)\rangle \leq 0. \tag{5}$$

(4)+(5) сократило $\pi(y) - \pi(x)$, что не хорошо. Поэтому перевернем знак в (5), что даст нам

$$\langle \pi(x) - x, \pi(x) - \pi(y)\rangle \leq 0. \tag{6}$$

$$\begin{aligned}
\langle y - \pi(y) + \pi(x) - x, \pi(x) - \pi(y)\rangle &\leq 0 \\
\langle y - x, \pi(x) - \pi(y)\rangle &\leq -\langle \pi(x) - \pi(y), \pi(x) - \pi(y)\rangle \\
\langle y - x, \pi(y) - \pi(x)\rangle &\geq \|\pi(x) - \pi(y)\|_2^2 \\
\|(y - x)^\top(\pi(y) - \pi(x))\|_2 &\geq \|\pi(x) - \pi(y)\|_2^2
\end{aligned}$$

Оператор проекции является нерастягивающим

Сокращенная запись: пусть $\pi = \text{proj}$ и $\pi(x)$ обозначает $\text{proj}(x)$.

Начнём с важного свойства проекции:

$$\langle y - \pi(y), x - \pi(y)\rangle \leq 0 \quad \forall x \in S. \tag{3}$$

Заменим x на $\pi(x)$ в (3)

$$\langle y - \pi(y), \pi(x) - \pi(y)\rangle \leq 0. \tag{4}$$

Заменим y на x и x на $\pi(y)$ в (3)

$$\langle x - \pi(x), \pi(y) - \pi(x)\rangle \leq 0. \tag{5}$$

(4)+(5) сократило $\pi(y) - \pi(x)$, что не хорошо. Поэтому перевернем знак в (5), что даст нам

$$\langle \pi(x) - x, \pi(x) - \pi(y)\rangle \leq 0. \tag{6}$$

$$\begin{aligned}
\langle y - \pi(y) + \pi(x) - x, \pi(x) - \pi(y)\rangle &\leq 0 \\
\langle y - x, \pi(x) - \pi(y)\rangle &\leq -\langle \pi(x) - \pi(y), \pi(x) - \pi(y)\rangle \\
\langle y - x, \pi(y) - \pi(x)\rangle &\geq \|\pi(x) - \pi(y)\|_2^2 \\
\|(y - x)^\top(\pi(y) - \pi(x))\|_2 &\geq \|\pi(x) - \pi(y)\|_2^2
\end{aligned}$$

По неравенству КБШ, левая часть ограничена сверху $\|y - x\|_2 \|\pi(y) - \pi(x)\|_2$, получаем $\|y - x\|_2 \|\pi(y) - \pi(x)\|_2 \geq \|\pi(x) - \pi(y)\|_2^2$. Деление на $\|\pi(x) - \pi(y)\|_2$ даёт требуемое неравенство.

Инструменты сходимости 💎💎💎💎

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ - L-гладкая выпуклая функция. Тогда, для любых $x, y \in \mathbb{R}^n$, выполняется следующее неравенство:

$$
f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y) \text{ or, equivalently,}
$$
$$
\|\nabla f(y) - \nabla f(x)\|_2^2 = \|\nabla f(x) - \nabla f(y)\|_2^2 \leq 2L \left(f(x) - f(y) - \langle \nabla f(y), x - y \rangle \right)
$$

Инструменты сходимости

Доказательство

1. Для доказательства этого рассмотрим другую функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$. Она очевидно выпукла (как сумма выпуклых функций). И легко проверить, что она L-гладкая по определению, так как $\nabla \varphi(y) = \nabla f(y) - \nabla f(x)$ и $\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leq L\|y_1 - y_2\|$.

Инструменты сходимости

Доказательство

1. Для доказательства этого рассмотрим другую функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$. Она очевидно выпукла (как сумма выпуклых функций). И легко проверить, что она L-гладкая по определению, так как $\nabla \varphi(y) = \nabla f(y) - \nabla f(x)$ и $\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leq L\|y_1 - y_2\|$.
2. Теперь рассмотрим свойство гладкости параболы для функции $\varphi(y)$:

Инструменты сходимости

Доказательство

1. Для доказательства этого рассмотрим другую функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$. Она очевидно выпукла (как сумма выпуклых функций). И легко проверить, что она L-гладкая по определению, так как $\nabla \varphi(y) = \nabla f(y) - \nabla f(x)$ и $\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leq L\|y_1 - y_2\|$.
2. Теперь рассмотрим свойство гладкости параболы для функции $\varphi(y)$:

$$\varphi(y) \leq \varphi(x) + \langle \nabla \varphi(x), y - x \rangle + \frac{L}{2}\|y - x\|_2^2$$

Инструменты сходимости 💎💎💎💎

Доказательство

1. Для доказательства этого рассмотрим другую функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$. Она очевидно выпукла (как сумма выпуклых функций). И легко проверить, что она L-гладкая по определению, так как $\nabla \varphi(y) = \nabla f(y) - \nabla f(x)$ и $\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leq L\|y_1 - y_2\|$.
2. Теперь рассмотрим свойство гладкости параболы для функции $\varphi(y)$:

$$
\begin{aligned}
&\varphi(y) \leq \varphi(x) + \langle \nabla \varphi(x), y - x \rangle + \frac{L}{2}\|y - x\|_2^2 \\
\stackrel{x := y, y := y - \frac{1}{L}\nabla \varphi(y)}{} \quad &\varphi\left(y - \frac{1}{L}\nabla \varphi(y)\right) \leq \varphi(y) + \left\langle \nabla \varphi(y), -\frac{1}{L}\nabla \varphi(y) \right\rangle + \frac{1}{2L}\|\nabla \varphi(y)\|_2^2
\end{aligned}
$$

Инструменты сходимости

Доказательство

1. Для доказательства этого рассмотрим другую функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$. Она очевидно выпукла (как сумма выпуклых функций). И легко проверить, что она L-гладкая по определению, так как $\nabla \varphi(y) = \nabla f(y) - \nabla f(x)$ и $\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leq L\|y_1 - y_2\|$.
2. Теперь рассмотрим свойство гладкости параболы для функции $\varphi(y)$:

$$
\begin{aligned}
&\varphi(y) \leq \varphi(x) + \langle \nabla \varphi(x), y - x \rangle + \frac{L}{2}\|y - x\|_2^2 \\
{\scriptstyle x := y,\, y := y - \frac{1}{L}\nabla\varphi(y)} \quad &\varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) + \left\langle \nabla \varphi(y), -\frac{1}{L}\nabla\varphi(y) \right\rangle + \frac{1}{2L}\|\nabla\varphi(y)\|_2^2 \\
&\varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2
\end{aligned}
$$

Инструменты сходимости 💎💎💎💎

3. Из условий оптимальности первого порядка для выпуклой функции $\nabla \varphi(y) = \nabla f(y) - \nabla f(x) = 0$. Мы можем заключить, что для любого x, минимум функции $\varphi(y)$ находится в точке $y = x$. Следовательно:

$$\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla \varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla \varphi(y)\|_2^2$$

Инструменты сходимости

3. Из условий оптимальности первого порядка для выпуклой функции $\nabla \varphi(y) = \nabla f(y) - \nabla f(x) = 0$. Мы можем заключить, что для любого x, минимум функции $\varphi(y)$ находится в точке $y = x$. Следовательно:

$$\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$

4. Теперь заменим $\varphi(y) = f(y) - \langle \nabla f(x), y\rangle$:

Инструменты сходимости 💎💎💎💎

3. Из условий оптимальности первого порядка для выпуклой функции $\nabla\varphi(y) = \nabla f(y) - \nabla f(x) = 0$. Мы можем заключить, что для любого x, минимум функции $\varphi(y)$ находится в точке $y = x$. Следовательно:

$$\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$

4. Теперь заменим $\varphi(y) = f(y) - \langle\nabla f(x), y\rangle$:

$$f(x) - \langle\nabla f(x), x\rangle \leq f(y) - \langle\nabla f(x), y\rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2$$

Инструменты сходимости

3. Из условий оптимальности первого порядка для выпуклой функции $\nabla \varphi(y) = \nabla f(y) - \nabla f(x) = 0$. Мы можем заключить, что для любого x, минимум функции $\varphi(y)$ находится в точке $y = x$. Следовательно:

$$\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$

4. Теперь заменим $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$:

$$f(x) - \langle \nabla f(x), x \rangle \leq f(y) - \langle \nabla f(x), y \rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2$$

$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y)$$

Инструменты сходимости 💎💎💎💎

3. Из условий оптимальности первого порядка для выпуклой функции $\nabla\varphi(y) = \nabla f(y) - \nabla f(x) = 0$. Мы можем заключить, что для любого x, минимум функции $\varphi(y)$ находится в точке $y = x$. Следовательно:

$$\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$

4. Теперь заменим $\varphi(y) = f(y) - \langle\nabla f(x), y\rangle$:

$$\begin{aligned}
&f(x) - \langle\nabla f(x), x\rangle \leq f(y) - \langle\nabla f(x), y\rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2 \\
&f(x) + \langle\nabla f(x), y - x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y) \\
&\|\nabla f(y) - \nabla f(x)\|_2^2 \leq 2L\left(f(y) - f(x) - \langle\nabla f(x), y - x\rangle\right)
\end{aligned}$$

Инструменты сходимости 💎💎💎💎

3. Из условий оптимальности первого порядка для выпуклой функции $\nabla\varphi(y) = \nabla f(y) - \nabla f(x) = 0$. Мы можем заключить, что для любого x, минимум функции $\varphi(y)$ находится в точке $y = x$. Следовательно:

$$\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$

4. Теперь заменим $\varphi(y) = f(y) - \langle\nabla f(x), y\rangle$:

$$f(x) - \langle\nabla f(x), x\rangle \leq f(y) - \langle\nabla f(x), y\rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2$$

$$f(x) + \langle\nabla f(x), y - x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y)$$

$$\|\nabla f(y) - \nabla f(x)\|_2^2 \leq 2L\left(f(y) - f(x) - \langle\nabla f(x), y - x\rangle\right)$$

поменять x и y $$\|\nabla f(x) - \nabla f(y)\|_2^2 \leq 2L\left(f(x) - f(y) - \langle\nabla f(y), x - y\rangle\right)$$

Инструменты сходимости 💎💎💎💎

3. Из условий оптимальности первого порядка для выпуклой функции $\nabla\varphi(y) = \nabla f(y) - \nabla f(x) = 0$. Мы можем заключить, что для любого x, минимум функции $\varphi(y)$ находится в точке $y = x$. Следовательно:

$$\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$

4. Теперь заменим $\varphi(y) = f(y) - \langle\nabla f(x), y\rangle$:

$$f(x) - \langle\nabla f(x), x\rangle \leq f(y) - \langle\nabla f(x), y\rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2$$

$$f(x) + \langle\nabla f(x), y - x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y)$$

$$\|\nabla f(y) - \nabla f(x)\|_2^2 \leq 2L\left(f(y) - f(x) - \langle\nabla f(x), y - x\rangle\right)$$

поменять x и y $\quad \|\nabla f(x) - \nabla f(y)\|_2^2 \leq 2L\left(f(x) - f(y) - \langle\nabla f(y), x - y\rangle\right)$

Инструменты сходимости 💎💎💎💎

3. Из условий оптимальности первого порядка для выпуклой функции $\nabla\varphi(y) = \nabla f(y) - \nabla f(x) = 0$. Мы можем заключить, что для любого x, минимум функции $\varphi(y)$ находится в точке $y = x$. Следовательно:

$$\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$

4. Теперь заменим $\varphi(y) = f(y) - \langle \nabla f(x), y\rangle$:

$$f(x) - \langle \nabla f(x), x\rangle \leq f(y) - \langle \nabla f(x), y\rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2$$

$$f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y)$$

$$\|\nabla f(y) - \nabla f(x)\|_2^2 \leq 2L\left(f(y) - f(x) - \langle \nabla f(x), y - x\rangle\right)$$

поменять x и y $$\|\nabla f(x) - \nabla f(y)\|_2^2 \leq 2L\left(f(x) - f(y) - \langle \nabla f(y), x - y\rangle\right)$$

Лемма доказана. С первого взгляда она не имеет много геометрического смысла, но мы будем использовать ее как удобный инструмент для оценки разницы между градиентами.

Инструменты сходимости

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ непрерывно дифференцируема на $\mathbb{R}^n$. Тогда функция f является μ-сильно выпуклой тогда и только тогда, когда для любых $x, y \in \mathbb{R}^d$ выполняется следующее:

$$
\begin{aligned}
\text{Strongly convex case } \mu > 0 \quad & \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2 \\
\text{Convex case } \mu = 0 \quad & \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0
\end{aligned}
$$

Доказательство

1. Мы дадим доказательство только для сильно выпуклого случая, выпуклый случай следует из него с установкой $\mu = 0$. Начнем с необходимости. Для сильно выпуклой функции

$$
\begin{aligned}
& f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2}\|x - y\|_2^2 \\
& f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{\mu}{2}\|x - y\|_2^2 \\
\text{сумма} \quad & \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2
\end{aligned}
$$

Инструменты сходимости 💎💎💎💎

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x-y)), x - y \rangle dt$:

Инструменты сходимости 💎💎💎💎

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x-y)), x - y \rangle dt$:

$$
f(x) - f(y) - \langle \nabla f(y), x - y \rangle = \int_0^1 \langle \nabla f(y + t(x-y)), x - y \rangle dt - \langle \nabla f(y), x - y \rangle
$$

Инструменты сходимости 💎💎💎💎

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x-y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x-y)), x - y \rangle dt$:

$$
\begin{aligned}
f(x) - f(y) - \langle \nabla f(y), x - y \rangle &= \int_0^1 \langle \nabla f(y + t(x-y)), x - y \rangle dt - \langle \nabla f(y), x - y \rangle \\
{\scriptstyle \langle \nabla f(y), x-y \rangle = \int_0^1 \langle \nabla f(y), x-y \rangle dt} \qquad &= \int_0^1 \langle \nabla f(y + t(x-y)) - \nabla f(y), (x - y) \rangle dt
\end{aligned}
$$

Инструменты сходимости 💎💎💎💎

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x - y)), x - y \rangle dt$:

$$
\begin{aligned}
f(x) - f(y) - \langle \nabla f(y), x - y \rangle &= \int_0^1 \langle \nabla f(y + t(x - y)), x - y \rangle dt - \langle \nabla f(y), x - y \rangle \\
\overset{\langle \nabla f(y), x-y\rangle = \int_0^1 \langle \nabla f(y), x-y\rangle dt}{} &= \int_0^1 \langle \nabla f(y + t(x - y)) - \nabla f(y), (x - y) \rangle dt \\
\overset{y + t(x-y) - y = t(x-y)}{} &= \int_0^1 t^{-1} \langle \nabla f(y + t(x - y)) - \nabla f(y), t(x - y) \rangle dt
\end{aligned}
$$

Инструменты сходимости 💎💎💎💎

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu\|x-y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x-y)), x - y\rangle dt$:

$$
\begin{aligned}
f(x) - f(y) - \langle \nabla f(y), x - y\rangle &= \int_0^1 \langle \nabla f(y + t(x-y)), x - y\rangle dt - \langle \nabla f(y), x - y\rangle \\
\overset{\langle \nabla f(y), x-y\rangle = \int_0^1 \langle \nabla f(y), x-y\rangle dt}{} \quad &= \int_0^1 \langle \nabla f(y + t(x-y)) - \nabla f(y), (x - y)\rangle dt \\
\overset{y + t(x-y) - y = t(x-y)}{} \quad &= \int_0^1 t^{-1}\langle \nabla f(y + t(x-y)) - \nabla f(y), t(x - y)\rangle dt \\
&\geq \int_0^1 t^{-1}\mu\|t(x-y)\|^2 dt
\end{aligned}
$$

Инструменты сходимости 💎💎💎💎

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu\|x-y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x-y)), x - y\rangle dt$:

$$
\begin{aligned}
f(x) - f(y) - \langle \nabla f(y), x - y\rangle &= \int_0^1 \langle \nabla f(y + t(x-y)), x - y\rangle dt - \langle \nabla f(y), x - y\rangle \\
\overset{\langle \nabla f(y), x-y\rangle = \int_0^1 \langle \nabla f(y), x-y\rangle dt}{} &= \int_0^1 \langle \nabla f(y + t(x-y)) - \nabla f(y), (x - y)\rangle dt \\
\overset{y + t(x-y) - y = t(x-y)}{} &= \int_0^1 t^{-1}\langle \nabla f(y + t(x-y)) - \nabla f(y), t(x - y)\rangle dt \\
&\geq \int_0^1 t^{-1}\mu\|t(x-y)\|^2 dt = \mu\|x-y\|^2 \int_0^1 t dt
\end{aligned}
$$

Инструменты сходимости 💎💎💎💎

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu \|x-y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x-y)), x-y\rangle dt$:

$$
\begin{aligned}
f(x) - f(y) - \langle \nabla f(y), x - y\rangle &= \int_0^1 \langle \nabla f(y + t(x-y)), x-y\rangle dt - \langle \nabla f(y), x-y\rangle \\
{\scriptstyle \langle \nabla f(y), x-y\rangle = \int_0^1 \langle \nabla f(y), x-y\rangle dt} \quad &= \int_0^1 \langle \nabla f(y + t(x-y)) - \nabla f(y), (x-y)\rangle dt \\
{\scriptstyle y + t(x-y) - y = t(x-y)} \quad &= \int_0^1 t^{-1}\langle \nabla f(y + t(x-y)) - \nabla f(y), t(x-y)\rangle dt \\
&\geq \int_0^1 t^{-1}\mu \|t(x-y)\|^2 dt = \mu\|x-y\|^2 \int_0^1 t dt = \frac{\mu}{2}\|x-y\|_2^2
\end{aligned}
$$

Инструменты сходимости

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x-y)), x - y \rangle dt$:

$$
\begin{aligned}
f(x) - f(y) - \langle \nabla f(y), x - y \rangle &= \int_0^1 \langle \nabla f(y + t(x-y)), x - y \rangle dt - \langle \nabla f(y), x - y \rangle \\
\overset{\langle \nabla f(y), x-y \rangle = \int_0^1 \langle \nabla f(y), x-y \rangle dt}{} &= \int_0^1 \langle \nabla f(y + t(x-y)) - \nabla f(y), (x - y) \rangle dt \\
\overset{y + t(x-y) - y = t(x-y)}{} &= \int_0^1 t^{-1} \langle \nabla f(y + t(x-y)) - \nabla f(y), t(x - y) \rangle dt \\
&\geq \int_0^1 t^{-1} \mu \|t(x-y)\|^2 dt = \mu \|x - y\|^2 \int_0^1 t dt = \frac{\mu}{2} \|x - y\|_2^2
\end{aligned}
$$

Инструменты сходимости

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu\|x-y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x-y)), x-y\rangle dt$:

$$
\begin{aligned}
f(x) - f(y) - \langle \nabla f(y), x - y\rangle &= \int_0^1 \langle \nabla f(y + t(x-y)), x-y\rangle dt - \langle \nabla f(y), x-y\rangle \\
{\scriptstyle \langle \nabla f(y), x-y\rangle = \int_0^1 \langle \nabla f(y), x-y\rangle dt} \quad &= \int_0^1 \langle \nabla f(y + t(x-y)) - \nabla f(y), (x-y)\rangle dt \\
{\scriptstyle y + t(x-y) - y = t(x-y)} \quad &= \int_0^1 t^{-1}\langle \nabla f(y + t(x-y)) - \nabla f(y), t(x-y)\rangle dt \\
&\geq \int_0^1 t^{-1}\mu\|t(x-y)\|^2 dt = \mu\|x-y\|^2 \int_0^1 t dt = \frac{\mu}{2}\|x-y\|_2^2
\end{aligned}
$$

Таким образом, мы полчаем выполнение критерия сильной выпуклости

$$
f(x) \geq f(y) + \langle \nabla f(y), x-y\rangle + \frac{\mu}{2}\|x-y\|_2^2
$$

Инструменты сходимости 💎💎💎💎

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu \|x-y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x-y)), x - y\rangle dt$:

$$
\begin{aligned}
f(x) - f(y) - \langle \nabla f(y), x - y\rangle &= \int_0^1 \langle \nabla f(y + t(x-y)), x - y\rangle dt - \langle \nabla f(y), x - y\rangle \\
\overset{\langle \nabla f(y), x-y\rangle = \int_0^1 \langle \nabla f(y), x-y\rangle dt}{} &= \int_0^1 \langle \nabla f(y + t(x-y)) - \nabla f(y), (x - y)\rangle dt \\
\overset{y + t(x-y) - y = t(x-y)}{} &= \int_0^1 t^{-1} \langle \nabla f(y + t(x-y)) - \nabla f(y), t(x - y)\rangle dt \\
&\geq \int_0^1 t^{-1} \mu \|t(x-y)\|^2 dt = \mu \|x-y\|^2 \int_0^1 t dt = \frac{\mu}{2}\|x-y\|_2^2
\end{aligned}
$$

Таким образом, мы полчаем выполнение критерия сильной выпуклости

$$
f(x) \geq f(y) + \langle \nabla f(y), x - y\rangle + \frac{\mu}{2}\|x-y\|_2^2 \text{ или, эквивалентно:}
$$

Инструменты сходимости

2. Для достаточности мы предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu\|x-y\|^2$. Используя теорему Ньютона-Лейбница $f(x) = f(y) + \int_0^1 \langle \nabla f(y + t(x-y)), x - y\rangle dt$:

$$
\begin{aligned}
f(x) - f(y) - \langle \nabla f(y), x - y\rangle &= \int_0^1 \langle \nabla f(y + t(x-y)), x - y\rangle dt - \langle \nabla f(y), x - y\rangle \\
{\scriptstyle \langle \nabla f(y), x-y\rangle = \int_0^1 \langle \nabla f(y), x-y\rangle dt} \quad &= \int_0^1 \langle \nabla f(y + t(x-y)) - \nabla f(y), (x - y)\rangle dt \\
{\scriptstyle y + t(x-y) - y = t(x-y)} \quad &= \int_0^1 t^{-1}\langle \nabla f(y + t(x-y)) - \nabla f(y), t(x - y)\rangle dt \\
&\geq \int_0^1 t^{-1}\mu\|t(x-y)\|^2 dt = \mu\|x-y\|^2 \int_0^1 t dt = \frac{\mu}{2}\|x-y\|_2^2
\end{aligned}
$$

Таким образом, мы полчаем выполнение критерия сильной выпуклости

$$
f(x) \geq f(y) + \langle \nabla f(y), x - y\rangle + \frac{\mu}{2}\|x-y\|_2^2 \text{ или, эквивалентно:}
$$

поменять x и y

$$
-\langle \nabla f(x), x - y\rangle \leq -\left(f(x) - f(y) + \frac{\mu}{2}\|x-y\|_2^2\right)
$$

Скорость сходимости для гладкой и выпуклой функции

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ выпуклая и дифференцируемая. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть есть минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Алгоритм проекции градиента с шагом $\frac{1}{L}$ достигает следующей сходимости после итерации $k > 0$:

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

Скорость сходимости для гладкой и выпуклой функции

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ выпуклая и дифференцируемая. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть есть минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Алгоритм проекции градиента с шагом $\frac{1}{L}$ достигает следующей сходимости после итерации $k > 0$:

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

1. Докажем лемму о достаточном убывании, предполагая, что $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ и правило косинусов $2x^T y = \|x\|^2 + \|y\|^2 - \|x-y\|^2$:

Скорость сходимости для гладкой и выпуклой функции

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ выпуклая и дифференцируемая. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть есть минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Алгоритм проекции градиента с шагом $\frac{1}{L}$ достигает следующей сходимости после итерации $k > 0$:

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

1. Докажем лемму о достаточном убывании, предполагая, что $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ и правило косинусов $2x^T y = \|x\|^2 + \|y\|^2 - \|x - y\|^2$:

Гладкость: $$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$$

Скорость сходимости для гладкой и выпуклой функции

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ выпуклая и дифференцируемая. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть есть минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Алгоритм проекции градиента с шагом $\frac{1}{L}$ достигает следующей сходимости после итерации $k > 0$:

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

1. Докажем лемму о достаточном убывании, предполагая, что $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ и правило косинусов $2x^T y = \|x\|^2 + \|y\|^2 - \|x - y\|^2$:

$$
\begin{aligned}
\text{Гладкость:} \qquad & f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\
\text{Метод:} \qquad & = f(x_k) - L\langle y_k - x_k, x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2
\end{aligned}
$$

Скорость сходимости для гладкой и выпуклой функции

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ выпуклая и дифференцируемая. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть есть минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Алгоритм проекции градиента с шагом $\frac{1}{L}$ достигает следующей сходимости после итерации $k > 0$:

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

1. Докажем лемму о достаточном убывании, предполагая, что $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ и правило косинусов $2x^Ty = \|x\|^2 + \|y\|^2 - \|x-y\|^2$:

Гладкость: $$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$$

Метод: $$= f(x_k) - L\langle y_k - x_k, x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2$$

Правило косинусов: $$= f(x_k) - \frac{L}{2}\left(\|y_k - x_k\|^2 + \|x_{k+1} - x_k\|^2 - \|y_k - x_{k+1}\|^2\right) + \frac{L}{2}\|x_{k+1} - x_k\|^2$$

Скорость сходимости для гладкой и выпуклой функции

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ выпуклая и дифференцируемая. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть есть минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Алгоритм проекции градиента с шагом $\frac{1}{L}$ достигает следующей сходимости после итерации $k > 0$:

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

1. Докажем лемму о достаточном убывании, предполагая, что $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ и правило косинусов $2x^Ty = \|x\|^2 + \|y\|^2 - \|x-y\|^2$:

$$
\begin{aligned}
&\text{Гладкость:} && f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\
&\text{Метод:} && = f(x_k) - L\langle y_k - x_k, x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\
&\text{Правило косинусов:} && = f(x_k) - \frac{L}{2}\left(\|y_k - x_k\|^2 + \|x_{k+1} - x_k\|^2 - \|y_k - x_{k+1}\|^2\right) + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\
& && = f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|y_k - x_{k+1}\|^2
\end{aligned}
$$

Скорость сходимости для гладкой и выпуклой функции

2. Теперь на каждом шаге прогресс не гарантирован. Снова используем формулу косинусов:

$$
\left\langle \frac{1}{L}\nabla f(x_k), x_k - x^* \right\rangle = \frac{1}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_k - x^* - \frac{1}{L}\nabla f(x_k)\|^2 \right)
$$

$$
\langle \nabla f(x_k), x_k - x^* \rangle = \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2 \right)
$$

Скорость сходимости для гладкой и выпуклой функции

2. Теперь на каждом шаге прогресс не гарантирован. Снова используем формулу косинусов:

$$\left\langle \frac{1}{L}\nabla f(x_k), x_k - x^* \right\rangle = \frac{1}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_k - x^* - \frac{1}{L}\nabla f(x_k)\|^2\right)$$
$$\langle \nabla f(x_k), x_k - x^* \rangle = \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2\right)$$

3. Теперь используем свойство проекции: $\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2$ с $x = x^*, y = y_k$:

$$\|x^* - \text{proj}_S(y_k)\|^2 + \|y_k - \text{proj}_S(y_k)\|^2 \leq \|x^* - y_k\|^2$$
$$\|y_k - x^*\|^2 \geq \|x^* - x_{k+1}\|^2 + \|y_k - x_{k+1}\|^2$$

Скорость сходимости для гладкой и выпуклой функции

2. Теперь на каждом шаге прогресс не гарантирован. Снова используем формулу косинусов:

$$
\left\langle \frac{1}{L}\nabla f(x_k), x_k - x^* \right\rangle = \frac{1}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_k - x^* - \frac{1}{L}\nabla f(x_k)\|^2\right)
$$

$$
\langle \nabla f(x_k), x_k - x^* \rangle = \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2\right)
$$

3. Теперь используем свойство проекции: $\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2$ с $x = x^*, y = y_k$:

$$
\|x^* - \text{proj}_S(y_k)\|^2 + \|y_k - \text{proj}_S(y_k)\|^2 \leq \|x^* - y_k\|^2
$$

$$
\|y_k - x^*\|^2 \geq \|x^* - x_{k+1}\|^2 + \|y_k - x_{k+1}\|^2
$$

4. Теперь, используя выпуклость и предыдущую часть:

Скорость сходимости для гладкой и выпуклой функции

2. Теперь на каждом шаге прогресс не гарантирован. Снова используем формулу косинусов:

$$\left\langle \frac{1}{L}\nabla f(x_k), x_k - x^* \right\rangle = \frac{1}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_k - x^* - \frac{1}{L}\nabla f(x_k)\|^2\right)$$

$$\langle \nabla f(x_k), x_k - x^* \rangle = \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2\right)$$

3. Теперь используем свойство проекции: $\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2$ с $x = x^*, y = y_k$:

$$\|x^* - \text{proj}_S(y_k)\|^2 + \|y_k - \text{proj}_S(y_k)\|^2 \leq \|x^* - y_k\|^2$$

$$\|y_k - x^*\|^2 \geq \|x^* - x_{k+1}\|^2 + \|y_k - x_{k+1}\|^2$$

4. Теперь, используя выпуклость и предыдущую часть:

Выпуклость: $$f(x_k) - f^* \leq \langle \nabla f(x_k), x_k - x^* \rangle$$

Скорость сходимости для гладкой и выпуклой функции

2. Теперь на каждом шаге прогресс не гарантирован. Снова используем формулу косинусов:

$$
\left\langle \frac{1}{L}\nabla f(x_k), x_k - x^* \right\rangle = \frac{1}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_k - x^* - \frac{1}{L}\nabla f(x_k)\|^2\right)
$$

$$
\langle \nabla f(x_k), x_k - x^* \rangle = \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2\right)
$$

3. Теперь используем свойство проекции: $\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2$ с $x = x^*, y = y_k$:

$$
\|x^* - \text{proj}_S(y_k)\|^2 + \|y_k - \text{proj}_S(y_k)\|^2 \leq \|x^* - y_k\|^2
$$

$$
\|y_k - x^*\|^2 \geq \|x^* - x_{k+1}\|^2 + \|y_k - x_{k+1}\|^2
$$

4. Теперь, используя выпуклость и предыдущую часть:

Выпуклость:

$$
\begin{aligned}
f(x_k) - f^* &\leq \langle \nabla f(x_k), x_k - x^* \rangle \\
&\leq \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 - \|y_k - x
\end{aligned}
$$

Скорость сходимости для гладкой и выпуклой функции

2. Теперь на каждом шаге прогресс не гарантирован. Снова используем формулу косинусов:

$$
\begin{aligned}
\left\langle \frac{1}{L}\nabla f(x_k), x_k - x^* \right\rangle &= \frac{1}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_k - x^* - \frac{1}{L}\nabla f(x_k)\|^2 \right) \\
\langle \nabla f(x_k), x_k - x^* \rangle &= \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|y_k - x^*\|^2 \right)
\end{aligned}
$$

3. Теперь используем свойство проекции: $\|x - \text{proj}_S(y)\|^2 + \|y - \text{proj}_S(y)\|^2 \leq \|x - y\|^2$ с $x = x^*, y = y_k$:

$$
\begin{aligned}
\|x^* - \text{proj}_S(y_k)\|^2 + \|y_k - \text{proj}_S(y_k)\|^2 &\leq \|x^* - y_k\|^2 \\
\|y_k - x^*\|^2 &\geq \|x^* - x_{k+1}\|^2 + \|y_k - x_{k+1}\|^2
\end{aligned}
$$

4. Теперь, используя выпуклость и предыдущую часть:

Выпуклость:
$$
\begin{aligned}
f(x_k) - f^* &\leq \langle \nabla f(x_k), x_k - x^* \rangle \\
&\leq \frac{L}{2}\left(\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 - \|y_k -
\end{aligned}
$$

Просуммируем для $i = 0, k-1$
$$
\sum_{i=0}^{k-1} [f(x_i) - f^*] \leq \sum_{i=0}^{k-1} \frac{1}{2L}\|\nabla f(x_i)\|^2 + \frac{L}{2}\|x_0 - x^*\|^2 - \frac{L}{2}\sum_{i=0}^{i-1}\|y_i - x_{i+1}\|^2
$$

Скорость сходимости для гладкой и выпуклой функции

5. Оценим градиенты с помощью неравенства о достаточном убывании Уравнение 7:

Скорость сходимости для гладкой и выпуклой функции

5. Оценим градиенты с помощью неравенства о достаточном убывании Уравнение 7:

$$\sum_{i=0}^{k-1} [f(x_i) - f^*] \leq \sum_{i=0}^{k-1} \left[f(x_i) - f(x_{i+1}) + \frac{L}{2} \|y_i - x_{i+1}\|^2 \right] + \frac{L}{2} \|x_0 - x^*\|^2 - \frac{L}{2} \sum_{i=0}^{i-1} \|y_i - x_{i+1}\|^2$$

Скорость сходимости для гладкой и выпуклой функции

5. Оценим градиенты с помощью неравенства о достаточном убывании Уравнение 7:

$$
\begin{aligned}
\sum_{i=0}^{k-1} [f(x_i) - f^*] &\leq \sum_{i=0}^{k-1} \left[f(x_i) - f(x_{i+1}) + \frac{L}{2} \|y_i - x_{i+1}\|^2 \right] + \frac{L}{2} \|x_0 - x^*\|^2 - \frac{L}{2} \sum_{i=0}^{i-1} \|y_i - x_{i+1}\|^2 \\
&\leq f(x_0) - f(x_k) + \frac{L}{2} \sum_{i=0}^{i-1} \|y_i - x_{i+1}\|^2 + \frac{L}{2} \|x_0 - x^*\|^2 - \frac{L}{2} \sum_{i=0}^{i-1} \|y_i - x_{i+1}\|^2
\end{aligned}
$$

Скорость сходимости для гладкой и выпуклой функции

5. Оценим градиенты с помощью неравенства о достаточном убывании Уравнение 7:

$$
\begin{aligned}
\sum_{i=0}^{k-1}[f(x_i) - f^*] &\leq \sum_{i=0}^{k-1}\left[f(x_i) - f(x_{i+1}) + \frac{L}{2}\|y_i - x_{i+1}\|^2\right] + \frac{L}{2}\|x_0 - x^*\|^2 - \frac{L}{2}\sum_{i=0}^{i-1}\|y_i - x_{i+1}\|^2 \\
&\leq f(x_0) - f(x_k) + \frac{L}{2}\sum_{i=0}^{i-1}\|y_i - x_{i+1}\|^2 + \frac{L}{2}\|x_0 - x^*\|^2 - \frac{L}{2}\sum_{i=0}^{i-1}\|y_i - x_{i+1}\|^2 \\
&\leq f(x_0) - f(x_k) + \frac{L}{2}\|x_0 - x^*\|^2
\end{aligned}
$$

Скорость сходимости для гладкой и выпуклой функции

5. Оценим градиенты с помощью неравенства о достаточном убывании Уравнение 7:

$$
\begin{aligned}
\sum_{i=0}^{k-1}[f(x_i) - f^*] &\leq \sum_{i=0}^{k-1}\left[f(x_i) - f(x_{i+1}) + \frac{L}{2}\|y_i - x_{i+1}\|^2\right] + \frac{L}{2}\|x_0 - x^*\|^2 - \frac{L}{2}\sum_{i=0}^{i-1}\|y_i - x_{i+1}\|^2 \\
&\leq f(x_0) - f(x_k) + \frac{L}{2}\sum_{i=0}^{i-1}\|y_i - x_{i+1}\|^2 + \frac{L}{2}\|x_0 - x^*\|^2 - \frac{L}{2}\sum_{i=0}^{i-1}\|y_i - x_{i+1}\|^2 \\
&\leq f(x_0) - f(x_k) + \frac{L}{2}\|x_0 - x^*\|^2
\end{aligned}
$$

$$
\sum_{i=0}^{k-1} f(x_i) - kf^* \leq f(x_0) - f(x_k) + \frac{L}{2}\|x_0 - x^*\|^2
$$

Скорость сходимости для гладкой и выпуклой функции

5. Оценим градиенты с помощью неравенства о достаточном убывании Уравнение 7:

$$
\begin{aligned}
\sum_{i=0}^{k-1} [f(x_i) - f^*] &\leq \sum_{i=0}^{k-1} \left[f(x_i) - f(x_{i+1}) + \frac{L}{2} \|y_i - x_{i+1}\|^2 \right] + \frac{L}{2} \|x_0 - x^*\|^2 - \frac{L}{2} \sum_{i=0}^{i-1} \|y_i - x_{i+1}\|^2 \\
&\leq f(x_0) - f(x_k) + \frac{L}{2} \sum_{i=0}^{i-1} \|y_i - x_{i+1}\|^2 + \frac{L}{2} \|x_0 - x^*\|^2 - \frac{L}{2} \sum_{i=0}^{i-1} \|y_i - x_{i+1}\|^2 \\
&\leq f(x_0) - f(x_k) + \frac{L}{2} \|x_0 - x^*\|^2 \\
\sum_{i=0}^{k-1} f(x_i) - kf^* &\leq f(x_0) - f(x_k) + \frac{L}{2} \|x_0 - x^*\|^2 \\
\sum_{i=1}^{k} [f(x_i) - f^*] &\leq \frac{L}{2} \|x_0 - x^*\|^2
\end{aligned}
$$

Скорость сходимости для гладкой и выпуклой функции

6. Из неравенства о достаточном убывании

$$
f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|y_k - x_{k+1}\|^2,
$$

Скорость сходимости для гладкой и выпуклой функции

6. Из неравенства о достаточном убывании

$$
f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|y_k - x_{k+1}\|^2,
$$

Скорость сходимости для гладкой и выпуклой функции

6. Из неравенства о достаточном убывании

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|y_k - x_{k+1}\|^2,$$

используем факт, что $x_{k+1} = \text{proj}_S(y_k)$. По определению проекции,

$$\|y_k - x_{k+1}\| \leq \|y_k - x_k\|,$$

Скорость сходимости для гладкой и выпуклой функции

6. Из неравенства о достаточном убывании

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|y_k - x_{k+1}\|^2,$$

используем факт, что $x_{k+1} = \text{proj}_S(y_k)$. По определению проекции,

$$\|y_k - x_{k+1}\| \leq \|y_k - x_k\|,$$

и вспомним, что $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ указывает на $\|y_k - x_k\| = \frac{1}{L}\|\nabla f(x_k)\|$. Следовательно,

$$\frac{L}{2}\|y_k - x_{k+1}\|^2 \leq \frac{L}{2}\|y_k - x_k\|^2 = \frac{L}{2}\frac{1}{L^2}\|\nabla f(x_k)\|^2 = \frac{1}{2L}\|\nabla f(x_k)\|^2.$$

Скорость сходимости для гладкой и выпуклой функции

6. Из неравенства о достаточном убывании

$$
f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|y_k - x_{k+1}\|^2,
$$

используем факт, что $x_{k+1} = \text{proj}_S(y_k)$. По определению проекции,

$$
\|y_k - x_{k+1}\| \leq \|y_k - x_k\|,
$$

и вспомним, что $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ указывает на $\|y_k - x_k\| = \frac{1}{L}\|\nabla f(x_k)\|$. Следовательно,

$$
\frac{L}{2}\|y_k - x_{k+1}\|^2 \leq \frac{L}{2}\|y_k - x_k\|^2 = \frac{L}{2}\frac{1}{L^2}\|\nabla f(x_k)\|^2 = \frac{1}{2L}\|\nabla f(x_k)\|^2.
$$

Заменим обратно в $(*)$:

$$
f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{1}{2L}\|\nabla f(x_k)\|^2 = f(x_k).
$$

Скорость сходимости для гладкой и выпуклой функции

6. Из неравенства о достаточном убывании

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{L}{2}\|y_k - x_{k+1}\|^2,$$

используем факт, что $x_{k+1} = \text{proj}_S(y_k)$. По определению проекции,

$$\|y_k - x_{k+1}\| \leq \|y_k - x_k\|,$$

и вспомним, что $y_k = x_k - \frac{1}{L}\nabla f(x_k)$ указывает на $\|y_k - x_k\| = \frac{1}{L}\|\nabla f(x_k)\|$. Следовательно,

$$\frac{L}{2}\|y_k - x_{k+1}\|^2 \leq \frac{L}{2}\|y_k - x_k\|^2 = \frac{L}{2}\frac{1}{L^2}\|\nabla f(x_k)\|^2 = \frac{1}{2L}\|\nabla f(x_k)\|^2.$$

Заменим обратно в $(*)$:

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 + \frac{1}{2L}\|\nabla f(x_k)\|^2 = f(x_k).$$

Следовательно,

$$f(x_{k+1}) \leq f(x_k) \quad \text{для каждого } k,$$

$\{f(x_k)\}$ является монотонно неубывающей последовательностью.

Скорость сходимости для гладкой и выпуклой функции

7. Финальная оценка сходимости Из шага 5, мы уже установили

$$
\sum_{i=0}^{k-1}[f(x_i) - f^*] \leq \frac{L}{2}\|x_0 - x^*\|_2^2.
$$

Скорость сходимости для гладкой и выпуклой функции

7. Финальная оценка сходимости Из шага 5, мы уже установили

$$
\sum_{i=0}^{k-1} [f(x_i) - f^*] \leq \frac{L}{2} \|x_0 - x^*\|_2^2.
$$

Скорость сходимости для гладкой и выпуклой функции

7. Финальная оценка сходимости Из шага 5, мы уже установили

$$\sum_{i=0}^{k-1} [f(x_i) - f^*] \leq \frac{L}{2} \|x_0 - x^*\|_2^2.$$

Поскольку $f(x_i)$ убывает в i, в частности $f(x_k) \leq f(x_i)$ для всех $i \leq k$. Следовательно,

$$k\left[f(x_k) - f^*\right] \leq \sum_{i=0}^{k-1} [f(x_i) - f^*] \leq \frac{L}{2} \|x_0 - x^*\|_2^2,$$

Скорость сходимости для гладкой и выпуклой функции

7. Финальная оценка сходимости Из шага 5, мы уже установили

$$
\sum_{i=0}^{k-1} [f(x_i) - f^*] \le \frac{L}{2} \|x_0 - x^*\|_2^2.
$$

Поскольку $f(x_i)$ убывает в i, в частности $f(x_k) \le f(x_i)$ для всех $i \le k$. Следовательно,

$$
k\,[f(x_k) - f^*] \le \sum_{i=0}^{k-1} [f(x_i) - f^*] \le \frac{L}{2} \|x_0 - x^*\|_2^2,
$$

которое сразу дает

$$
f(x_k) - f^* \le \frac{L\|x_0 - x^*\|_2^2}{2k}.
$$

Это завершает доказательство скорости сходимости $\mathcal{O}(\frac{1}{k})$ для выпуклой и L-гладкой f с ограничениями на проекцию.

Скорость сходимости для гладкой сильно выпуклой функции

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ является μ-сильно выпуклой. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть есть минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Алгоритм проекции градиента с шагом $\alpha \leq \frac{1}{L}$ достигает следующей сходимости после $k > 0$:

$$\|x_k - x^*\|_2^2 \leq (1 - \alpha\mu)^k \|x_0 - x^*\|_2^2$$

Доказательство

1. Сначала докажем свойство стационарной точки: $\text{proj}_S(x^* - \alpha\nabla f(x^*)) = x^*$.

   Это следует из критерия проекции и условия оптимальности первого порядка для x^*. Пусть $y = x^* - \alpha\nabla f(x^*)$. Мы должны показать, что $\langle y - x^*, x - x^*\rangle \leq 0$ для всех $x \in S$.

   $$\langle (x^* - \alpha\nabla f(x^*)) - x^*, x - x^*\rangle = -\alpha\langle\nabla f(x^*), x - x^*\rangle \leq 0$$

   Неравенство выполняется, потому что $\alpha > 0$ и $\langle\nabla f(x^*), x - x^*\rangle \geq 0$ является условием оптимальности для x^*.

Скорость сходимости для гладкой сильно выпуклой функции

1. Учитывая расстояние до решения и используя свойство стационарной точки:

Скорость сходимости для гладкой сильно выпуклой функции

1. Учитывая расстояние до решения и используя свойство стационарной точки:

$$
\|x_{k+1} - x^*\|_2^2 = \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2
$$

Скорость сходимости для гладкой сильно выпуклой функции

1. Учитывая расстояние до решения и используя свойство стационарной точки:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
\text{Свойство стационарной точки} &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - \text{proj}_S(x^* - \alpha \nabla f(x^*))\|_2^2
\end{aligned}
$$

Скорость сходимости для гладкой сильно выпуклой функции

1. Учитывая расстояние до решения и используя свойство стационарной точки:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
\text{Свойство стационарной точки} &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - \text{proj}_S(x^* - \alpha \nabla f(x^*))\|_2^2 \\
\text{нерастяжимость} &\leq \|x_k - \alpha \nabla f(x_k) - (x^* - \alpha \nabla f(x^*))\|_2^2
\end{aligned}
$$

Скорость сходимости для гладкой сильно выпуклой функции

1. Учитывая расстояние до решения и используя свойство стационарной точки:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
{\text{Свойство стационарной точки}} &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - \text{proj}_S(x^* - \alpha \nabla f(x^*))\|_2^2 \\
{\text{нерастяжимость}} &\leq \|x_k - \alpha \nabla f(x_k) - (x^* - \alpha \nabla f(x^*))\|_2^2 \\
&= \|x_k - x^*\|^2 - 2\alpha \langle \nabla f(x_k) - \nabla f(x^*), x_k - x^* \rangle + \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

Скорость сходимости для гладкой сильно выпуклой функции

1. Учитывая расстояние до решения и используя свойство стационарной точки:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
\text{Свойство стационарной точки} &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - \text{proj}_S(x^* - \alpha \nabla f(x^*))\|_2^2 \\
\text{нерастяжимость} &\leq \|x_k - \alpha \nabla f(x_k) - (x^* - \alpha \nabla f(x^*))\|_2^2 \\
&= \|x_k - x^*\|^2 - 2\alpha \langle \nabla f(x_k) - \nabla f(x^*), x_k - x^* \rangle + \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

2. Теперь используем гладкость из инструментов сходимости и сильную выпуклость:

Скорость сходимости для гладкой сильно выпуклой функции

1. Учитывая расстояние до решения и используя свойство стационарной точки:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
\text{Свойство стационарной точки} &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - \text{proj}_S(x^* - \alpha \nabla f(x^*))\|_2^2 \\
\text{нерастяжимость} &\leq \|x_k - \alpha \nabla f(x_k) - (x^* - \alpha \nabla f(x^*))\|_2^2 \\
&= \|x_k - x^*\|^2 - 2\alpha \langle \nabla f(x_k) - \nabla f(x^*), x_k - x^* \rangle + \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

2. Теперь используем гладкость из инструментов сходимости и сильную выпуклость:

$$
\text{гладкость} \quad \|\nabla f(x_k) - \nabla f(x^*)\|_2^2 \leq 2L \left(f(x_k) - f(x^*) - \langle \nabla f(x^*), x_k - x^* \rangle \right)
$$

Скорость сходимости для гладкой сильно выпуклой функции

1. Учитывая расстояние до решения и используя свойство стационарной точки:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
\text{Свойство стационарной точки} &= \|\text{proj}_S(x_k - \alpha \nabla f(x_k)) - \text{proj}_S(x^* - \alpha \nabla f(x^*))\|_2^2 \\
\text{нерастяжимость} &\leq \|x_k - \alpha \nabla f(x_k) - (x^* - \alpha \nabla f(x^*))\|_2^2 \\
&= \|x_k - x^*\|^2 - 2\alpha\langle \nabla f(x_k) - \nabla f(x^*), x_k - x^*\rangle + \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

2. Теперь используем гладкость из инструментов сходимости и сильную выпуклость:

$$
\begin{aligned}
\text{гладкость} \quad & \|\nabla f(x_k) - \nabla f(x^*)\|_2^2 \leq 2L\left(f(x_k) - f(x^*) - \langle \nabla f(x^*), x_k - x^*\rangle\right) \\
\text{сильная выпуклость} \quad & -\langle \nabla f(x_k) - \nabla f(x^*), x_k - x^*\rangle \leq -\left(f(x_k) - f(x^*) + \frac{\mu}{2}\|x_k - x^*\|_2^2\right) - \langle \nabla f(x^*), x_k - x
\end{aligned}
$$

Скорость сходимости для гладкой сильно выпуклой функции

3. Заменим:

Скорость сходимости для гладкой сильно выпуклой функции

3. Заменим:

$$\|x_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|^2 - 2\alpha \left(f(x_k) - f(x^*) + \frac{\mu}{2}\|x_k - x^*\|_2^2 \right) - 2\alpha\langle \nabla f(x^*), x_k - x^*\rangle +$$
$$+ \alpha^2 2L \left(f(x_k) - f(x^*) - \langle \nabla f(x^*), x_k - x^*\rangle \right)$$

Скорость сходимости для гладкой сильно выпуклой функции

3. Заменим:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &\leq \|x_k - x^*\|^2 - 2\alpha \left(f(x_k) - f(x^*) + \frac{\mu}{2}\|x_k - x^*\|_2^2 \right) - 2\alpha \langle \nabla f(x^*), x_k - x^* \rangle + \\
&+ \alpha^2 2L \left(f(x_k) - f(x^*) - \langle \nabla f(x^*), x_k - x^* \rangle \right) \\
&\leq (1 - \alpha\mu)\|x_k - x^*\|^2 + 2\alpha(\alpha L - 1) \left(f(x_k) - f(x^*) - \langle \nabla f(x^*), x_k - x^* \rangle \right)
\end{aligned}
$$

Скорость сходимости для гладкой сильно выпуклой функции 💎💎💎

3. Заменим:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|^2 - 2\alpha \left(f(x_k) - f(x^*) + \frac{\mu}{2}\|x_k - x^*\|_2^2\right) - 2\alpha\langle \nabla f(x^*), x_k - x^*\rangle + \\
+ \alpha^2 2L \left(f(x_k) - f(x^*) - \langle \nabla f(x^*), x_k - x^*\rangle\right) \\
\leq (1 - \alpha\mu)\|x_k - x^*\|^2 + 2\alpha(\alpha L - 1)\left(f(x_k) - f(x^*) - \langle \nabla f(x^*), x_k - x^*\rangle\right)
\end{aligned}
$$

4. В силу выпуклости f: $f(x_k) - f(x^*) - \langle \nabla f(x^*), x_k - x^*\rangle \geq 0$. Следовательно, при выборе $\alpha \leq \frac{1}{L}$:

$$
\|x_{k+1} - x^*\|_2^2 \leq (1 - \alpha\mu)\|x_k - x^*\|^2,
$$

что и дает линейную сходимость метода со скоростью не хуже $1 - \frac{\mu}{L}$.

Скорость сходимости для гладкой и выпуклой функции

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ является выпуклой и дифференцируемой. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть существует минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Метод Франк-Вульфа с шагом $\gamma_k = \frac{k-1}{k+1}$ достигает следующей сходимости после $k > 0$ итераций:

$$
f(x_k) - f^* \leq \frac{2LR^2}{k+1}
$$

где $R = \max\limits_{x,y \in S} \|x - y\|$ является диаметром множества S.

Скорость сходимости для гладкой и выпуклой функции

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R}$ является выпуклой и дифференцируемой. Пусть $S \subseteq \mathbb{R}^n$ замкнутое выпуклое множество, и пусть существует минимизатор x^* функции f на S; кроме того, пусть f гладкая на S с параметром L. Метод Франк-Вульфа с шагом $\gamma_k = \frac{k-1}{k+1}$ достигает следующей сходимости после $k > 0$ итераций:

$$
f(x_k) - f^* \leq \frac{2LR^2}{k+1}
$$

где $R = \max\limits_{x,y \in S} \|x - y\|$ является диаметром множества S.

1. Благодаря L-гладкости функции f, имеем:

$$
\begin{aligned}
f\left(x_{k+1}\right) - f\left(x_k\right) &\leq \left\langle \nabla f\left(x_k\right), x_{k+1} - x_k \right\rangle + \frac{L}{2}\left\|x_{k+1} - x_k\right\|^2 \\
&= (1-\gamma_k)\left\langle \nabla f\left(x_k\right), y_k - x_k \right\rangle + \frac{L(1-\gamma_k)^2}{2}\left\|y_k - x_k\right\|^2
\end{aligned}
$$

Скорость сходимости для гладкой и выпуклой функции

2. Благодаря выпуклости функции f, для любой точки $x \in S$, включая x^*:

$$\langle \nabla f(x_k), x - x_k \rangle \leq f(x) - f(x_k)$$

В частности, для $x = x^*$:

$$\langle \nabla f(x_k), x^* - x_k \rangle \leq f(x^*) - f(x_k)$$

Скорость сходимости для гладкой и выпуклой функции

2. Благодаря выпуклости функции f, для любой точки $x \in S$, включая x^*:

$$\langle \nabla f(x_k), x - x_k \rangle \leq f(x) - f(x_k)$$

   В частности, для $x = x^*$:

$$\langle \nabla f(x_k), x^* - x_k \rangle \leq f(x^*) - f(x_k)$$

3. По определению y_k, имеем $\langle \nabla f(x_k), y_k \rangle \leq \langle \nabla f(x_k), x^* \rangle$, следовательно:

$$\langle \nabla f(x_k), y_k - x_k \rangle \leq \langle \nabla f(x_k), x^* - x_k \rangle \leq f(x^*) - f(x_k)$$

Скорость сходимости для гладкой и выпуклой функции

2. Благодаря выпуклости функции f, для любой точки $x \in S$, включая x^*:
$$\langle \nabla f(x_k), x - x_k \rangle \leq f(x) - f(x_k)$$
В частности, для $x = x^*$:
$$\langle \nabla f(x_k), x^* - x_k \rangle \leq f(x^*) - f(x_k)$$
3. По определению y_k, имеем $\langle \nabla f(x_k), y_k \rangle \leq \langle \nabla f(x_k), x^* \rangle$, следовательно:
$$\langle \nabla f(x_k), y_k - x_k \rangle \leq \langle \nabla f(x_k), x^* - x_k \rangle \leq f(x^*) - f(x_k)$$
4. Объединяя неравенства:
$$\begin{aligned} f\left(x_{k+1}\right) - f\left(x_k\right) &\leq (1-\gamma_k)\left\langle \nabla f\left(x_k\right), y_k - x_k \right\rangle + \frac{L(1-\gamma_k)^2}{2}\left\|y_k - x_k\right\|^2 \\ &\leq (1-\gamma_k)\left(f(x^*) - f(x_k)\right) + \frac{L(1-\gamma_k)^2}{2} R^2 \end{aligned}$$

Скорость сходимости для гладкой и выпуклой функции

2. Благодаря выпуклости функции f, для любой точки $x \in S$, включая x^*:

$$\langle \nabla f(x_k), x - x_k \rangle \leq f(x) - f(x_k)$$

   В частности, для $x = x^*$:

$$\langle \nabla f(x_k), x^* - x_k \rangle \leq f(x^*) - f(x_k)$$

3. По определению y_k, имеем $\langle \nabla f(x_k), y_k \rangle \leq \langle \nabla f(x_k), x^* \rangle$, следовательно:

$$\langle \nabla f(x_k), y_k - x_k \rangle \leq \langle \nabla f(x_k), x^* - x_k \rangle \leq f(x^*) - f(x_k)$$

4. Объединяя неравенства:

$$\begin{aligned} f\left(x_{k+1}\right) - f\left(x_k\right) &\leq (1 - \gamma_k) \left\langle \nabla f\left(x_k\right), y_k - x_k \right\rangle + \frac{L(1-\gamma_k)^2}{2} \|y_k - x_k\|^2 \\ &\leq (1 - \gamma_k) \left(f(x^*) - f(x_k)\right) + \frac{L(1-\gamma_k)^2}{2} R^2 \end{aligned}$$

5. Перегруппируем слагаемые:

$$f\left(x_{k+1}\right) - f(x^*) \leq \gamma_k \left(f(x_k) - f(x^*)\right) + (1-\gamma_k)^2 \frac{LR^2}{2}$$

Скорость сходимости для гладкой и выпуклой функции

6. Обозначим $\delta_k = \frac{f(x_k) - f(x^*)}{LR^2}$, получим:

$$
\delta_{k+1} \leq \gamma_k \delta_k + \frac{(1-\gamma_k)^2}{2} = \frac{k-1}{k+1}\delta_k + \frac{2}{(k+1)^2}
$$

Скорость сходимости для гладкой и выпуклой функции

6. Обозначим $\delta_k = \frac{f(x_k) - f(x^*)}{LR^2}$, получим:

$$\delta_{k+1} \leq \gamma_k \delta_k + \frac{(1-\gamma_k)^2}{2} = \frac{k-1}{k+1}\delta_k + \frac{2}{(k+1)^2}$$

7. Докажем, что $\delta_k \leq \frac{2}{k+1}$ индукцией.

что дает нам желаемый результат:

$$f(x_k) - f^* \leq \frac{2LR^2}{k+1}$$

Скорость сходимости для гладкой и выпуклой функции

6. Обозначим $\delta_k = \frac{f(x_k)-f(x^*)}{LR^2}$, получим:

$$\delta_{k+1} \leq \gamma_k \delta_k + \frac{(1-\gamma_k)^2}{2} = \frac{k-1}{k+1}\delta_k + \frac{2}{(k+1)^2}$$

7. Докажем, что $\delta_k \leq \frac{2}{k+1}$ индукцией.
   - База: $\delta_2 \leq \frac{1}{2} < \frac{2}{3}$

что дает нам желаемый результат:

$$f(x_k) - f^* \leq \frac{2LR^2}{k+1}$$

Скорость сходимости для гладкой и выпуклой функции

6. Обозначим $\delta_k = \frac{f(x_k) - f(x^*)}{LR^2}$, получим:

$$\delta_{k+1} \leq \gamma_k \delta_k + \frac{(1-\gamma_k)^2}{2} = \frac{k-1}{k+1}\delta_k + \frac{2}{(k+1)^2}$$

7. Докажем, что $\delta_k \leq \frac{2}{k+1}$ индукцией.
   - База: $\delta_2 \leq \frac{1}{2} < \frac{2}{3}$
   - Предположение: $\delta_k \leq \frac{2}{k+1}$

   что дает нам желаемый результат:

$$f(x_k) - f^* \leq \frac{2LR^2}{k+1}$$

Скорость сходимости для гладкой и выпуклой функции 💎💎💎

6. Обозначим $\delta_k = \frac{f(x_k)-f(x^*)}{LR^2}$, получим:

$$\delta_{k+1} \leq \gamma_k \delta_k + \frac{(1-\gamma_k)^2}{2} = \frac{k-1}{k+1}\delta_k + \frac{2}{(k+1)^2}$$

7. Докажем, что $\delta_k \leq \frac{2}{k+1}$ индукцией.
   - База: $\delta_2 \leq \frac{1}{2} < \frac{2}{3}$
   - Предположение: $\delta_k \leq \frac{2}{k+1}$
   - Тогда $\delta_{k+1} \leq \frac{k-1}{k+1} \cdot \frac{2}{k+1} + \frac{2}{(k+1)^2} = \frac{2k}{k^2+2k+1} < \frac{2}{k+2}$ 🎓

   что дает нам желаемый результат:

$$f(x_k) - f^* \leq \frac{2LR^2}{k+1}$$

Нижняя граница для метода Франк-Вульфа [4]

Theorem

Рассмотрим любой алгоритм, который использует только линейный минимизатор (LMO). Пусть диаметр множества S равен R. Существует L-гладкая сильно выпуклая функция $f : \mathbb{R}^n \to \mathbb{R}$ такая, что этот алгоритм требует по крайней мере

$$
\min\left(\frac{n}{2}, \frac{LR^2}{16\varepsilon}\right)
$$

итераций (т.е. вызовов LMO) для построения точки $\hat{x} \in S$ с $f(\hat{x}) - \min\limits_{x \in S} f(x) \leq \varepsilon$. Нижняя граница применима как для выпуклых, так и для сильно выпуклых функций.

[4] The Complexity of Large-scale Convex Programming under a Linear Optimization Oracle

Нижняя граница для метода Франк-Вульфа [4]

Theorem

Рассмотрим любой алгоритм, который использует только линейный минимизатор (LMO). Пусть диаметр множества S равен R. Существует L-гладкая сильно выпуклая функция $f : \mathbb{R}^n \to \mathbb{R}$ такая, что этот алгоритм требует по крайней мере

$$\min\left(\frac{n}{2}, \frac{LR^2}{16\varepsilon}\right)$$

итераций (т.е. вызовов LMO) для построения точки $\hat{x} \in S$ с $f(\hat{x}) - \min\limits_{x \in S} f(x) \leq \varepsilon$. Нижняя граница применима как для выпуклых, так и для сильно выпуклых функций.

Схема доказательства. Рассмотрим следующую задачу оптимизации:

$$\min_{x \in S} f(x) = \min_{x \in S} \frac{1}{2}\|x\|_2^2$$

$$S = \left\{ x \in \mathbb{R}^n \mid x \geq 0, \sum_{i=1}^n x_i = 1 \right\}$$

Обратим внимание, что:

- f является 1-гладкой;
- диаметр S равен $R = 2$;
- f сильно выпукла.

[4] The Complexity of Large-scale Convex Programming under a Linear Optimization Oracle

Нижняя граница для метода Франк-Вульфа [5]

1. Оптимальное решение равно

$$
x^* = \frac{1}{n}\mathbf{1} = \frac{1}{n}\sum_{i=1}^{n} e_i, \quad \text{и} \quad f(x^*) = \frac{1}{2n},
$$

где $e_i = (0, \dots, 0, \underbrace{1}_{i\text{-ая позиция}}, 0, \dots, 0)^\top$ является i-м стандартным базисным вектором.

Нижняя граница для метода Франк-Вульфа [5]

1. Оптимальное решение равно

$$
x^* = \frac{1}{n}\mathbf{1} = \frac{1}{n}\sum_{i=1}^{n} e_i, \quad \text{и} \quad f(x^*) = \frac{1}{2n},
$$

где $e_i = (0, \ldots, 0, \underset{i\text{-ая позиция}}{1}, 0, \ldots, 0)^\top$ является i-м стандартным базисным вектором.

2. Линейный минимизатор (LMO) на S возвращает вершину e_i. После k итераций, метод обнаружит не более k различных базисных векторов $e_{i_1}, \ldots, e_{i_k}$. Лучшая выпуклая комбинация, которую можно сформировать, равна

$$
\hat{x} = \frac{1}{k}\sum_{j=1}^{k} e_{i_j}.
$$

[5] The Complexity of Large-scale Convex Programming under a Linear Optimization Oracle

Нижняя граница для метода Франк-Вульфа [6]

1. Оценивая функцию в $\hat{x}$, получаем:

$$f(\hat{x}) - f(x^*) \geq \frac{1}{2}\left(\frac{1}{\min\{k, n\}} - \frac{1}{n}\right).$$

Нижняя граница для метода Франк-Вульфа [6]

1. Оценивая функцию в $\hat{x}$, получаем:

$$
f(\hat{x}) - f(x^*) \geq \frac{1}{2}\left(\frac{1}{\min\{k, n\}} - \frac{1}{n}\right).
$$

2. Чтобы гарантировать, что $f(\hat{x}) - f(x^*) \leq \varepsilon$, необходимо, чтобы (полное доказательство приведено в статье):

$$
k \geq \min\left\{\frac{n}{2}, \frac{1}{4\varepsilon}\right\} = \min\left\{\frac{n}{2}, \frac{LR^2}{16\varepsilon}\right\}.
$$

[6] The Complexity of Large-scale Convex Programming under a Linear Optimization Oracle