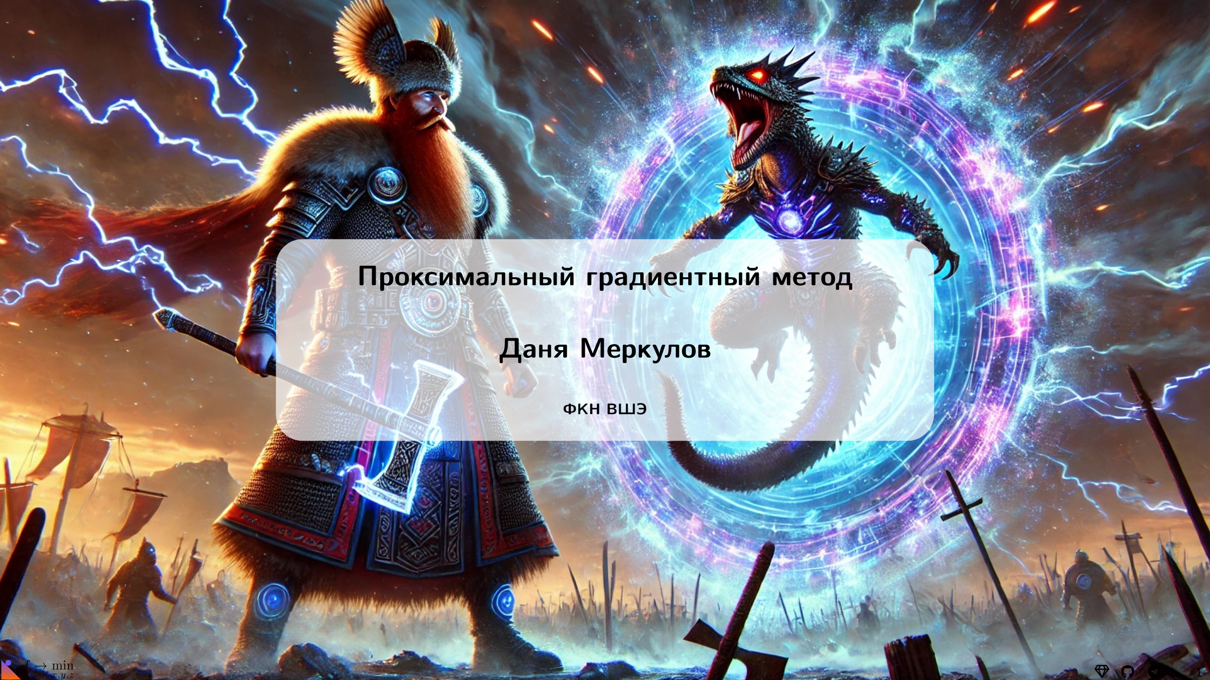


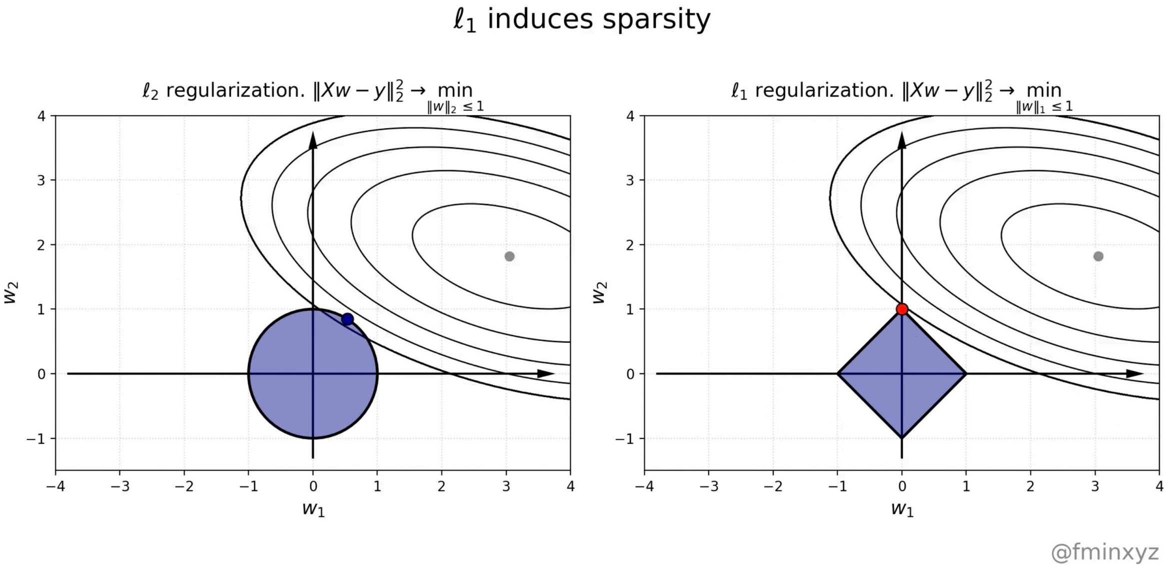
Проксимальный градиентный метод

Даня Меркулов

ФКН ВШЭ

Субградиентный метод

Негладкие задачи



Субградиентный метод

Субградиентный метод: $$\min_{x \in \mathbb{R}^n} f(x) \qquad x_{k+1} = x_k - \alpha_k g_k, \quad g_k \in \partial f(x_k)$$

Субградиентный метод

Субградиентный метод: $\min\limits_{x \in \mathbb{R}^n} f(x)$

$$\boxed{x_{k+1} = x_k - \alpha_k g_k, \quad g_k \in \partial f(x_k)}$$

выпуклый (негладкий)	сильно выпуклый (негладкий)
$f(x_k) - f^* \sim \mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$	$f(x_k) - f^* \sim \mathcal{O}\left(\frac{1}{k}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\varepsilon}\right)$

нижние оценки

$\frac{1}{\sqrt{k}}$ $\qquad$ $\frac{1}{k}$

Субградиентный метод

Субградиентный метод: $\min\limits_{x \in \mathbb{R}^n} f(x) \qquad x_{k+1} = x_k - \alpha_k g_k, \quad g_k \in \partial f(x_k)$

выпуклый (негладкий)	сильно выпуклый (негладкий)
$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{\sqrt{k}}\right)$	$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{k}\right)$
$k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon^2}\right)$	$k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon}\right)$

Theorem

Пусть f является G-липшицевой и выпуклой, тогда субградиентный метод сходится как:

$$f(\overline{x}) - f^* \leq \frac{GR}{\sqrt{k}},$$

где

- $\alpha = \frac{R}{G\sqrt{k}}$

Субградиентный метод

Субградиентный метод: $\min\limits_{x \in \mathbb{R}^n} f(x) \qquad x_{k+1} = x_k - \alpha_k g_k, \quad g_k \in \partial f(x_k)$

выпуклый (негладкий)	сильно выпуклый (негладкий)
$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{\sqrt{k}}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon^2}\right)$	$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{k}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon}\right)$

Theorem

Пусть f является G-липшицевой и выпуклой, тогда субградиентный метод сходится как:

$$f(\overline{x}) - f^* \leq \frac{GR}{\sqrt{k}},$$

где

- $\alpha = \frac{R}{G\sqrt{k}}$
- $R = \|x_0 - x^*\|$

Субградиентный метод

Субградиентный метод: $\min_{x \in \mathbb{R}^n} f(x) \qquad x_{k+1} = x_k - \alpha_k g_k, \quad g_k \in \partial f(x_k)$

выпуклый (негладкий)	сильно выпуклый (негладкий)
$f(x_k) - f^* \sim \mathcal{O}\left(\frac{1}{\sqrt{k}}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$	$f(x_k) - f^* \sim \mathcal{O}\left(\frac{1}{k}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\varepsilon}\right)$

Theorem

Пусть f является G-липшицевой и выпуклой, тогда субградиентный метод сходится как:

$$
f(\overline{x}) - f^* \leq \frac{GR}{\sqrt{k}},
$$

где

- $\alpha = \frac{R}{G\sqrt{k}}$
- $R = \|x_0 - x^*\|$
- $\overline{x} = \frac{1}{k}\sum_{i=0}^{k-1} x_i$

Нижние оценки для негладкой выпуклой оптимизации

выпуклый (негладкий)	сильно выпуклый (негладкий)
$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{\sqrt{k}}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon^2}\right)$	$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{k}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon}\right)$

$$x_{k+1} = x_k + \mathrm{Lin}(g_0 \ldots g_k)$$

Нижние оценки для негладкой выпуклой оптимизации

выпуклый (негладкий)	сильно выпуклый (негладкий)
$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{\sqrt{k}}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon^2}\right)$	$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{k}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon}\right)$

- Субградиентный метод является оптимальным для приведённых выше задач.

Нижние оценки для негладкой выпуклой оптимизации

выпуклый (негладкий)	сильно выпуклый (негладкий)
$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{\sqrt{k}}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon^2}\right)$	$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{k}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon}\right)$

- Субградиентный метод является оптимальным для приведённых выше задач.
- Можно использовать зеркальный спуск (обобщение субградиентного метода на случай неевклидового расстояния) с той же скоростью сходимости, чтобы лучше учитывать геометрию задачи.

Нижние оценки для негладкой выпуклой оптимизации

лин. скорость

выпуклый (негладкий)	сильно выпуклый (негладкий)
$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{\sqrt{k}}\right)$	$f(x_k) - f^* \sim \mathcal{O}\left(\dfrac{1}{k}\right)$
$k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon^2}\right)$	$k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon}\right)$

- Субградиентный метод является оптимальным для приведённых выше задач.
- Можно использовать зеркальный спуск (обобщение субградиентного метода на случай неевклидового расстояния) с той же скоростью сходимости, чтобы лучше учитывать геометрию задачи.
- Однако мы можем достичь стандартной скорости градиентного спуска $\mathcal{O}\left(\frac{1}{k}\right)$ (и даже ускоренной $\mathcal{O}\left(\frac{1}{k^2}\right)$), если будем использовать структуру задачи.

Проксимальный оператор

Интуиция проксимального отображения через схему Эйлера

Рассмотрим ODE градиентного потока:

$$\boxed{\frac{dx}{dt} = -\nabla f(x)}$$

Явная схема Эйлера:

Интуиция проксимального отображения через схему Эйлера

Рассмотрим ODE градиентного потока:

$$\frac{dx}{dt} = -\nabla f(x)$$

Явная схема Эйлера:

$$\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_k)$$

Приводит к обычному методу градиентного спуска

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$

Интуиция проксимального отображения через схему Эйлера

Рассмотрим ODE градиентного потока:

$$\frac{dx}{dt} = -\nabla f(x)$$

Явная схема Эйлера:

$$\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_k)$$

Приводит к обычному методу градиентного спуска

Неявная схема Эйлера:

$$\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_{k+1})$$

$$x_{k+1} = x_k - \alpha \nabla f(x_{k+1})$$

Интуиция проксимального отображения через схему Эйлера

Рассмотрим ODE градиентного потока:

$$
\frac{dx}{dt} = -\nabla f(x)
$$

Явная схема Эйлера:

$$
\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_k)
$$

Приводит к обычному методу градиентного спуска

Неявная схема Эйлера:

$$
\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_{k+1})
$$

$$
\frac{x_{k+1} - x_k}{\alpha} + \nabla f(x_{k+1}) = 0
$$

$$
\left. \underbrace{\frac{x - x_k}{\alpha} + \nabla f(x)}_{\nabla \Psi(x) = 0 \,\big|_{x = x_{k+1}}} = 0 \right|_{x = x_k}
$$

Интуиция проксимального отображения через схему Эйлера

Рассмотрим ODE градиентного потока:

$$\frac{dx}{dt} = -\nabla f(x)$$

Явная схема Эйлера:

$$\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_k)$$

Приводит к обычному методу градиентного спуска

Неявная схема Эйлера:

$$\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_{k+1})$$

$$\frac{x_{k+1} - x_k}{\alpha} + \nabla f(x_{k+1}) = 0$$

$$\frac{x - x_k}{\alpha} + \nabla f(x)\Big|_{x = x_{k+1}} = 0$$

Интуиция проксимального отображения через схему Эйлера

Рассмотрим ODE градиентного потока:

$$\frac{dx}{dt} = -\nabla f(x)$$

Явная схема Эйлера:

$$\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_k)$$

Приводит к обычному методу градиентного спуска

Неявная схема Эйлера:

$$\begin{aligned}
\frac{x_{k+1} - x_k}{\alpha} &= -\nabla f(x_{k+1}) \\
\frac{x_{k+1} - x_k}{\alpha} + \nabla f(x_{k+1}) &= 0 \\
\frac{x - x_k}{\alpha} + \nabla f(x)\Big|_{x = x_{k+1}} &= 0 \\
\nabla \left[\frac{1}{2\alpha} \|x - x_k\|_2^2 + f(x) \right]\Bigg|_{x = x_{k+1}} &= 0
\end{aligned}$$

Интуиция проксимального отображения через схему Эйлера

Рассмотрим ODE градиентного потока:

$$\frac{dx}{dt} = -\nabla f(x)$$

Явная схема Эйлера:

$$\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_k)$$

Приводит к обычному методу градиентного спуска

Неявная схема Эйлера:

$$\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_{k+1})$$

$$\frac{x_{k+1} - x_k}{\alpha} + \nabla f(x_{k+1}) = 0$$

$$\left.\frac{x - x_k}{\alpha} + \nabla f(x)\right|_{x = x_{k+1}} = 0$$

$$\left.\nabla \left[\frac{1}{2\alpha} \|x - x_k\|_2^2 + f(x) \right]\right|_{x = x_{k+1}} = 0$$

$$\boxed{x_{k+1} = \text{arg}\min_{x \in \mathbb{R}^n} \left[f(x) + \frac{1}{2\alpha} \|x - x_k\|_2^2 \right]}$$

Интуиция проксимального отображения через схему Эйлера

Рассмотрим ODE градиентного потока:

$$
\frac{dx}{dt} = -\nabla f(x)
$$

Явная схема Эйлера:

$$
\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_k)
$$

Приводит к обычному методу градиентного спуска

Неявная схема Эйлера:

$$
\begin{aligned}
\frac{x_{k+1} - x_k}{\alpha} &= -\nabla f(x_{k+1}) \\
\frac{x_{k+1} - x_k}{\alpha} + \nabla f(x_{k+1}) &= 0 \\
\left. \frac{x - x_k}{\alpha} + \nabla f(x) \right|_{x = x_{k+1}} &= 0 \\
\left. \nabla \left[\frac{1}{2\alpha} \|x - x_k\|_2^2 + f(x) \right] \right|_{x = x_{k+1}} &= 0 \\
x_{k+1} = \text{arg}\min_{x \in \mathbb{R}^n} \left[f(x) + \frac{1}{2\alpha} \|x - x_k\|_2^2 \right] &
\end{aligned}
$$

Интуиция проксимального отображения через схему Эйлера

Рассмотрим ODE градиентного потока:

$$\frac{dx}{dt} = -\nabla f(x)$$

Явная схема Эйлера:

$$\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_k)$$

Приводит к обычному методу градиентного спуска

$$\operatorname{PROX}_f(x_k) = \operatorname*{argmin}_{x \in \mathbb{R}^n} \left(f(x) + \frac{1}{2}\|x - x_k\|_2^2 \right)$$

Неявная схема Эйлера:

$$\frac{x_{k+1} - x_k}{\alpha} = -\nabla f(x_{k+1})$$

$$\frac{x_{k+1} - x_k}{\alpha} + \nabla f(x_{k+1}) = 0$$

$$\left.\frac{x - x_k}{\alpha} + \nabla f(x)\right|_{x = x_{k+1}} = 0$$

$$\left.\nabla \left[\frac{1}{2\alpha}\|x - x_k\|_2^2 + f(x)\right]\right|_{x = x_{k+1}} = 0$$

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^n} \left[f(x) + \frac{1}{2\alpha}\|x - x_k\|_2^2 \right]$$

❗ Проксимальный оператор

$$\text{prox}_{f,\alpha}(x_k) = \arg\min_{x \in \mathbb{R}^n} \left[f(x) + \frac{1}{2\alpha}\|x - x_k\|_2^2 \right] = \operatorname*{argmin}_{x \in \mathbb{R}^n} \left[\underbrace{\alpha f(x)}_{\tilde{f}_\alpha(x)} + \frac{1}{2}\|x - x_k\|_2^2 \right]$$

Визуализация проксимального оператора

$$
\mathrm{Prox}_f(x) = \operatorname*{argmin}_{x'} \frac{1}{2}\|x - x'\|^2 + f(x')
$$

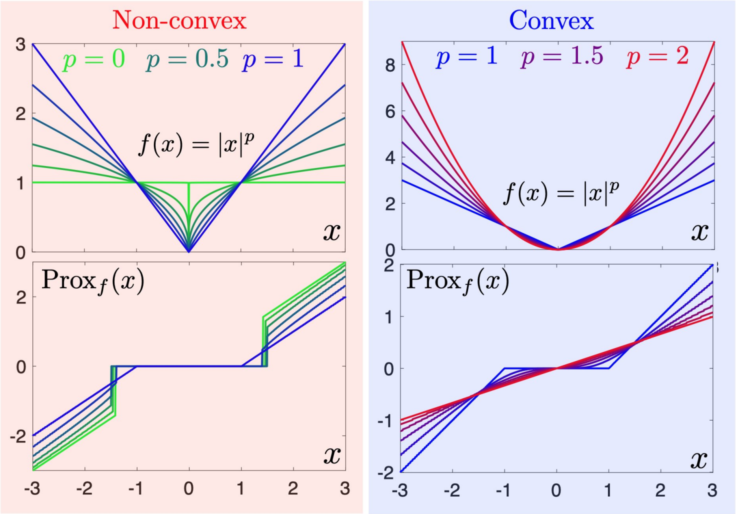


Figure 1: Источник

Связь проксимального оператора с ГС и методом Ньютона

- **Градиентный спуск из проксимального метода.** Вернёмся к дискретизации:

Связь проксимального оператора с ГС и методом Ньютона

- **Градиентный спуск из проксимального метода.** Вернёмся к дискретизации:

$$\boxed{x_{k+1} + \alpha \nabla f(x_{k+1}) = x_k}$$

Связь проксимального оператора с ГС и методом Ньютона

- **Градиентный спуск из проксимального метода.** Вернёмся к дискретизации:

$$
\begin{aligned}
x_{k+1} + \alpha \nabla f(x_{k+1}) &= x_k \\
(I + \alpha \nabla f)(x_{k+1}) &= x_k
\end{aligned}
$$

Связь проксимального оператора с ГС и методом Ньютона

- **Градиентный спуск из проксимального метода.** Вернёмся к дискретизации:

$$
\begin{aligned}
x_{k+1} + \alpha \nabla f(x_{k+1}) &= x_k \\
(I + \alpha \nabla f)(x_{k+1}) &= x_k \\
\boxed{x_{k+1} = (I + \alpha \nabla f)^{-1} x_k} &\overset{\alpha \to 0}{\approx} (I - \alpha \nabla f) x_k
\end{aligned}
$$

Связь проксимального оператора с ГС и методом Ньютона

- **Градиентный спуск из проксимального метода.** Вернёмся к дискретизации:

$$
\begin{aligned}
x_{k+1} + \alpha \nabla f(x_{k+1}) &= x_k \\
(I + \alpha \nabla f)(x_{k+1}) &= x_k
\end{aligned}
$$

$$
x_{k+1} = (I + \alpha \nabla f)^{-1} x_k \overset{\alpha \to 0}{\approx} (I - \alpha \nabla f) x_k
$$

$$= x_k - \alpha \cdot \nabla f(x_k)$$

$$x_{k+1} = \mathrm{PROX}_f(x_k)$$

Связь проксимального оператора с ГС и методом Ньютона

- **Градиентный спуск из проксимального метода.** Вернёмся к дискретизации:

$$
\begin{aligned}
x_{k+1} + \alpha \nabla f(x_{k+1}) &= x_k \\
(I + \alpha \nabla f)(x_{k+1}) &= x_k \\
x_{k+1} = (I + \alpha \nabla f)^{-1} x_k &\overset{\alpha \to 0}{\approx} (I - \alpha \nabla f) x_k
\end{aligned}
$$

  Таким образом, при $\alpha \to 0$ мы получаем обычный градиентный спуск: $x_{k+1} = x_k - \alpha \nabla f(x_k)$

- **Метод Ньютона из проксимального метода.** Рассмотрим проксимальное отображение от квадратичной аппроксимации Тейлора функции $f^{II}_{x_k}(x)$:

Связь проксимального оператора с ГС и методом Ньютона

- **Градиентный спуск из проксимального метода.** Вернёмся к дискретизации:

$$
\begin{aligned}
x_{k+1} + \alpha \nabla f(x_{k+1}) &= x_k \\
(I + \alpha \nabla f)(x_{k+1}) &= x_k \\
x_{k+1} = (I + \alpha \nabla f)^{-1} x_k &\overset{\alpha \to 0}{\approx} (I - \alpha \nabla f) x_k
\end{aligned}
$$

  Таким образом, при $\alpha \to 0$ мы получаем обычный градиентный спуск: $x_{k+1} = x_k - \alpha \nabla f(x_k)$

- **Метод Ньютона из проксимального метода.** Рассмотрим проксимальное отображение от квадратичной аппроксимации Тейлора функции $f^{II}_{x_k}(x)$:

$$
\boxed{x_{k+1} = \text{prox}_{f^{II}_{x_k}, \alpha}(x_k)} = \arg\min_{x \in \mathbb{R}^n} \Big[\underbrace{f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle \nabla^2 f(x_k)(x - x_k), x - x_k \rangle}_{f^{II}_{x_k}(x)} + \frac{1}{2\alpha} \|x - x_k\|_2^2 \Big]
$$

$X_{k+1} = \text{PROX}_{\cancel{f}\; f^{II}_{x_k}}(x_k)$

Связь проксимального оператора с ГС и методом Ньютона

- **Градиентный спуск из проксимального метода.** Вернёмся к дискретизации:

$$
\begin{aligned}
x_{k+1} + \alpha \nabla f(x_{k+1}) &= x_k \\
(I + \alpha \nabla f)(x_{k+1}) &= x_k \\
x_{k+1} = (I + \alpha \nabla f)^{-1} x_k &\overset{\alpha \to 0}{\approx} (I - \alpha \nabla f) x_k
\end{aligned}
$$

Таким образом, при $\alpha \to 0$ мы получаем обычный градиентный спуск: $x_{k+1} = x_k - \alpha \nabla f(x_k)$

- **Метод Ньютона из проксимального метода.** Рассмотрим проксимальное отображение от квадратичной аппроксимации Тейлора функции $f^{II}_{x_k}(x)$:

$$
x_{k+1} = \text{prox}_{f^{II}_{x_k}, \alpha}(x_k) = \text{arg}\min_{x \in \mathbb{R}^n} \left[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle \nabla^2 f(x_k)(x - x_k), x - x_k \rangle + \frac{1}{2\alpha} \|x - x_k\|_2^2 \right]
$$

$$
\left. \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) + \frac{1}{\alpha}(x - x_k) \right|_{x = x_{k+1}} = 0
$$

Связь проксимального оператора с ГС и методом Ньютона

- **Градиентный спуск из проксимального метода.** Вернёмся к дискретизации:

$$
\begin{aligned}
x_{k+1} + \alpha \nabla f(x_{k+1}) &= x_k \\
(I + \alpha \nabla f)(x_{k+1}) &= x_k \\
x_{k+1} = (I + \alpha \nabla f)^{-1} x_k &\overset{\alpha \to 0}{\approx} (I - \alpha \nabla f) x_k
\end{aligned}
$$

  Таким образом, при $\alpha \to 0$ мы получаем обычный градиентный спуск: $x_{k+1} = x_k - \alpha \nabla f(x_k)$

- **Метод Ньютона из проксимального метода.** Рассмотрим проксимальное отображение от квадратичной аппроксимации Тейлора функции $f^{II}_{x_k}(x)$:

$$
x_{k+1} = \text{prox}_{f^{II}_{x_k}, \alpha}(x_k) = \text{arg}\min_{x \in \mathbb{R}^n} \left[f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle \nabla^2 f(x_k)(x - x_k), x - x_k \rangle + \frac{1}{2\alpha} \|x - x_k\|_2^2 \right]
$$

$$
\left. \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) + \frac{1}{\alpha}(x - x_k) \right|_{x = x_{k+1}} = 0
$$

$$
\boxed{x_{k+1} = x_k - \left[\nabla^2 f(x_k) + \frac{1}{\alpha} I \right]^{-1} \nabla f(x_k)}
$$

$\alpha \to \infty$

$x_{k+1} = \text{Newton}(x_k)$

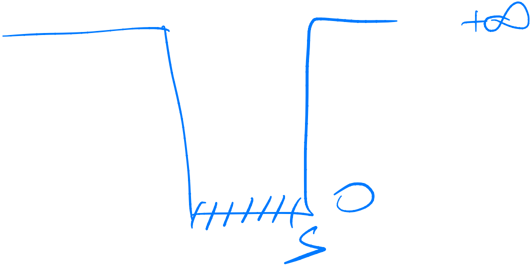
От проекций к проксимальности

Пусть $\mathbb{I}_S$ — индикаторная функция замкнутого выпуклого множества S. Напомним определение ортогональной проекции $\pi_S(y)$

От проекций к проксимальности

Пусть $\mathbb{I}_S$ — индикаторная функция замкнутого выпуклого множества S. Напомним определение ортогональной проекции $\pi_S(y)$

$$\pi_S(y) := \arg\min_{x \in S} \frac{1}{2}\|x - y\|_2^2.$$



$\mathbb{I}_S(x)$

От проекций к проксимальности

Пусть $\mathbb{I}_S$ — индикаторная функция замкнутого выпуклого множества S. Напомним определение ортогональной проекции $\pi_S(y)$

$$\pi_S(y) := \arg\min_{x \in S} \frac{1}{2}\|x - y\|_2^2.$$

С помощью индикаторной функции

$$\boxed{\mathbb{I}_S(x) = \begin{cases} 0, & x \in S, \\ \infty, & x \notin S, \end{cases}}$$

$$\underbrace{\mathrm{PROX}_{I_S}(x_k) = ?}\ \underbrace{\arg\min_{x \in \mathbb{R}^n}\left(I_S(x) + \frac{1}{2}\|x - x_k\|_2^2\right) = \arg\min_{x \in S'}\left(\frac{1}{2}\|x - x_k\|_2^2\right) = \mathrm{PROJ}_S(x_k)}$$

От проекций к проксимальности

Пусть $\mathbb{I}_S$ — индикаторная функция замкнутого выпуклого множества S. Напомним определение ортогональной проекции $\pi_S(y)$

$$\pi_S(y) := \arg\min_{x \in S} \frac{1}{2}\|x - y\|_2^2.$$

С помощью индикаторной функции

$$\mathbb{I}_S(x) = \begin{cases} 0, & x \in S, \\ \infty, & x \notin S, \end{cases}$$

Перепишем ортогональную проекцию $\pi_S(y)$ как

$$\pi_S(y) := \arg\min_{x \in \mathbb{R}^n} \frac{1}{2}\|x - y\|^2 + \mathbb{I}_S(x).$$

От проекций к проксимальности

Пусть $\mathbb{I}_S$ — индикаторная функция замкнутого выпуклого множества S. Напомним определение ортогональной проекции $\pi_S(y)$

$$\pi_S(y) := \arg\min_{x \in S} \frac{1}{2}\|x - y\|_2^2.$$

С помощью индикаторной функции

$$\mathbb{I}_S(x) = \begin{cases} 0, & x \in S, \\ \infty, & x \notin S, \end{cases}$$

Перепишем ортогональную проекцию $\pi_S(y)$ как

$$\pi_S(y) := \arg\min_{x \in \mathbb{R}^n} \left[\frac{1}{2}\|x - y\|^2 + \mathbb{I}_S(x)\right].$$

Проксимальность: заменим $\mathbb{I}_S$ на произвольную выпуклую функцию!

$$\text{prox}_r(y) = \text{prox}_{r,1}(y) := \arg\min_x \left[\frac{1}{2}\|x - y\|^2 + r(x)\right]$$

Композитная оптимизация

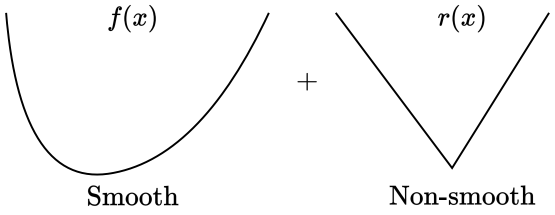
Регуляризованные / Композитные целевые функции

Многие негладкие задачи имеют вид

$$
\boxed{\min_{x \in \mathbb{R}^n} \varphi(x) = f(x) + r(x)}
$$

- **Lasso, L1-МНК, compressed sensing**

$$
f(x) = \frac{1}{2}\|Ax - b\|_2^2, r(x) = \lambda \|x\|_1
$$



Регуляризованные / Композитные целевые функции

Многие негладкие задачи имеют вид

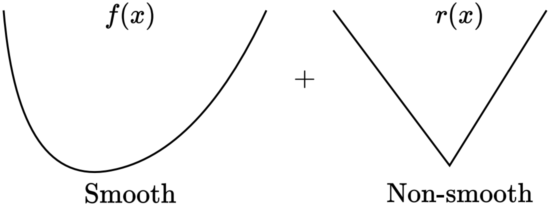
$$\min_{x \in \mathbb{R}^n} \varphi(x) = f(x) + r(x)$$

- **Lasso, L1-МНК, compressed sensing**

$$f(x) = \frac{1}{2}\|Ax - b\|_2^2, r(x) = \lambda\|x\|_1$$

- **L1-логистическая регрессия, разреженная ЛР**

$$f(x) = -y \log h(x) - (1-y)\log(1-h(x)), r(x) = \lambda\|x\|_1$$



Проксимальный градиентный метод из условий оптимальности

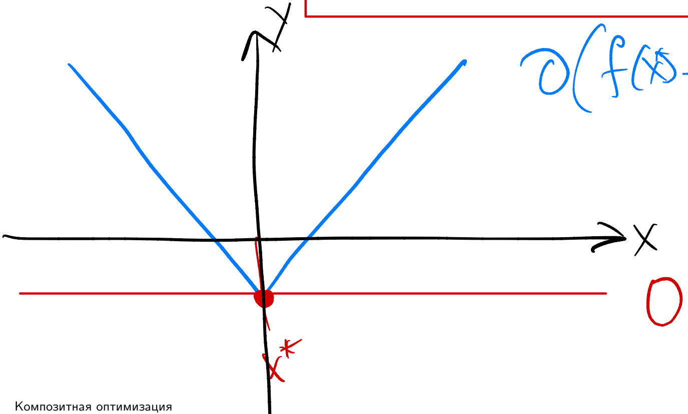
Условия оптимальности:

$$
0 \in \nabla f(x^*) + \partial r(x^*)
$$

$$
\boxed{0 \in \partial \varphi(x^*)}
$$

Принцип Ферма

$$
\partial(f(x^*) + r(x^*))
$$

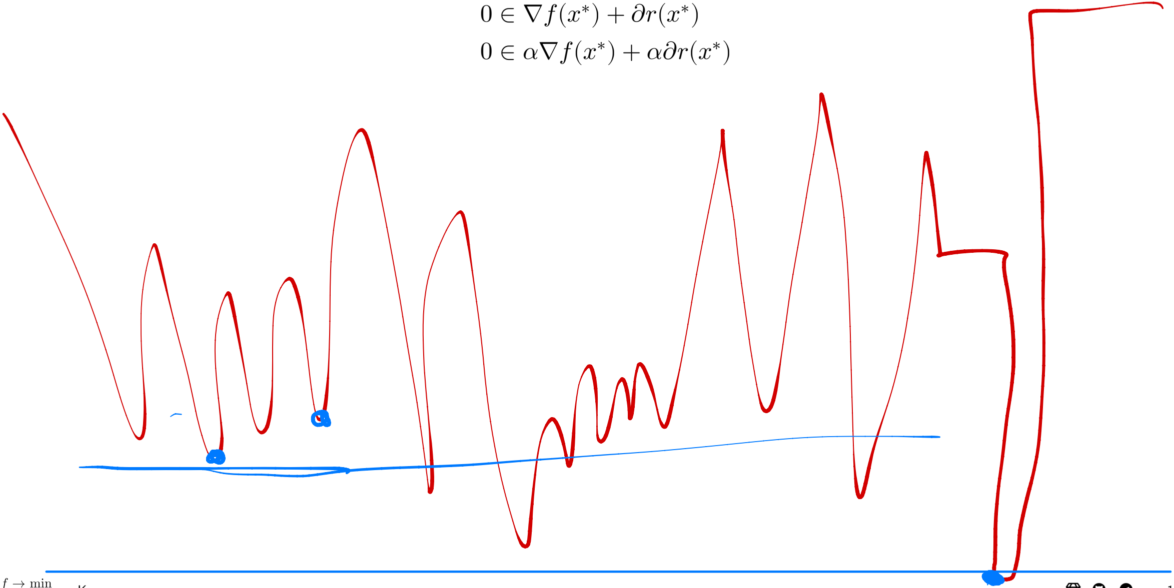


Проксимальный градиентный метод из условий оптимальности

Условия оптимальности:

$$0 \in \nabla f(x^*) + \partial r(x^*)$$

$$0 \in \alpha \nabla f(x^*) + \alpha \partial r(x^*)$$



Проксимальный градиентный метод из условий оптимальности

Условия оптимальности:

$$
\begin{aligned}
0 &\in \underline{\nabla f(x^*)} + \partial r(x^*) \\
0 &\in \alpha \nabla f(x^*) + \alpha \partial r(x^*) \\
x^* &\in \alpha \nabla f(x^*) + (I + \alpha \partial r)(x^*)
\end{aligned}
$$

$0 \in \partial \varphi(x^*)$

$0 \in \partial \left(f(x^*) + r(x^*) \right)$

$0 \in \underline{\partial f(x^*)} + \partial r(x^*)$

Проксимальный градиентный метод из условий оптимальности

Условия оптимальности:

$$
\begin{aligned}
0 &\in \nabla f(x^*) + \partial r(x^*) \\
0 &\in \alpha \nabla f(x^*) + \alpha \partial r(x^*) \\
x^* &\in \alpha \nabla f(x^*) + (I + \alpha \partial r)(x^*) \\
x^* - \alpha \nabla f(x^*) &\in (I + \alpha \partial r)(x^*)
\end{aligned}
$$

Проксимальный градиентный метод из условий оптимальности

Условия оптимальности:

$$
\begin{aligned}
0 &\in \nabla f(x^*) + \partial r(x^*) \\
0 &\in \alpha \nabla f(x^*) + \alpha \partial r(x^*) \\
x^* &\in \alpha \nabla f(x^*) + (I + \alpha \partial r)(x^*) \\
x^* - \alpha \nabla f(x^*) &\in (I + \alpha \partial r)(x^*) \\
x^* &= (I + \alpha \partial r)^{-1}(x^* - \alpha \nabla f(x^*))
\end{aligned}
$$

Проксимальный градиентный метод из условий оптимальности

Условия оптимальности:

$$
\begin{aligned}
0 &\in \nabla f(x^*) + \partial r(x^*) \\
0 &\in \alpha \nabla f(x^*) + \alpha \partial r(x^*) \\
x^* &\in \alpha \nabla f(x^*) + (I + \alpha \partial r)(x^*) \\
x^* - \alpha \nabla f(x^*) &\in (I + \alpha \partial r)(x^*) \\
x^* &= (I + \alpha \partial r)^{-1}(x^* - \alpha \nabla f(x^*)) \\
x^* &= \text{prox}_{r,\alpha}(x^* - \alpha \nabla f(x^*))
\end{aligned}
$$

Проксимальный градиентный метод из условий оптимальности

Условия оптимальности:

$$
\begin{aligned}
0 &\in \nabla f(x^*) + \partial r(x^*) \\
0 &\in \alpha \nabla f(x^*) + \alpha \partial r(x^*) \\
x^* &\in \alpha \nabla f(x^*) + (I + \alpha \partial r)(x^*) \\
x^* - \alpha \nabla f(x^*) &\in (I + \alpha \partial r)(x^*) \\
x^* &= (I + \alpha \partial r)^{-1}(x^* - \alpha \nabla f(x^*)) \\
x^* &= \text{prox}_{r,\alpha}(x^* - \alpha \nabla f(x^*))
\end{aligned}
$$

Проксимальный градиентный метод из условий оптимальности

Условия оптимальности:

$$
\begin{aligned}
0 &\in \nabla f(x^*) + \partial r(x^*) \\
0 &\in \alpha \nabla f(x^*) + \alpha \partial r(x^*) \\
x^* &\in \alpha \nabla f(x^*) + (I + \alpha \partial r)(x^*) \\
x^* - \alpha \nabla f(x^*) &\in (I + \alpha \partial r)(x^*) \\
x^* &= (I + \alpha \partial r)^{-1}(x^* - \alpha \nabla f(x^*)) \\
x^* &= \text{prox}_{r,\alpha}(x^* - \alpha \nabla f(x^*))
\end{aligned}
$$

Что приводит к проксимальному градиентному методу:

$$
\boxed{x_{k+1} = \text{prox}_{r,\alpha}(x_k - \alpha \nabla f(x_k))}
$$

И этот метод сходится со скоростью $\mathcal{O}(\frac{1}{k})$!

Проксимальный градиентный метод из условий оптимальности

Условия оптимальности:

$$
\begin{aligned}
0 &\in \nabla f(x^*) + \partial r(x^*) \\
0 &\in \alpha \nabla f(x^*) + \alpha \partial r(x^*) \\
x^* &\in \alpha \nabla f(x^*) + (I + \alpha \partial r)(x^*) \\
x^* - \alpha \nabla f(x^*) &\in (I + \alpha \partial r)(x^*) \\
x^* &= \underline{(I + \alpha \partial r)}^{-1}(x^* - \alpha \nabla f(x^*)) \\
x^* &= \underline{\text{prox}_{r,\alpha}}(x^* - \alpha \nabla f(x^*))
\end{aligned}
$$

Что приводит к проксимальному градиентному методу:

$$
\underline{x_{k+1} = \text{prox}_{r,\alpha}(x_k - \alpha \nabla f(x_k))}
$$

И этот метод сходится со скоростью $\mathcal{O}(\frac{1}{k})$!

Альтернативная форма проксимального оператора

$$
\text{prox}_{f,\alpha}(x_k) = \text{prox}_{\alpha f}(x_k) = \text{arg}\min_{x \in \mathbb{R}^n} \left[\alpha f(x) + \frac{1}{2} \|x - x_k\|_2^2 \right] \qquad \text{prox}_f(x_k) = \text{arg}\min_{x \in \mathbb{R}^n} \left[f(x) + \frac{1}{2} \|x - x_k\|_2^2 \right]
$$

Вывод проксимального оператора для ℓ_1-нормы

Пусть $\boxed{r(x) = \lambda\|x\|_1}$, $\lambda > 0$. По определению:

$$\boxed{\text{prox}_r(y) = \arg\min_{x\in\mathbb{R}^n} \left[\lambda\|x\|_1 + \frac{1}{2}\|x - y\|_2^2\right]}$$

$$= \arg\min_{x\in\mathbb{R}^n} \left[\lambda \cdot \sum_{i=1}^{n} |x_i| + \frac{1}{2}\sum_{i=1}^{n}(x_i - y_i)^2\right] =$$

$$= \arg\min_{x} \sum_{i=1}^{n} \left[\lambda|x_i| + \frac{1}{2}(x_i - y_i)^2\right]$$

$$\arg\min_{x} \lambda|x| + \frac{1}{2}(x-y)^2$$

Вывод проксимального оператора для ℓ_1-нормы

Пусть $r(x) = \lambda\|x\|_1$, $\lambda > 0$. По определению:

$$\text{prox}_r(y) = \arg\min_{x\in\mathbb{R}^n} \left[\lambda\|x\|_1 + \frac{1}{2}\|x-y\|_2^2\right]$$

Так как $\|x\|_1 = \sum_i |x_i|$ и $\|x-y\|_2^2 = \sum_i (x_i - y_i)^2$, задача **разделяется** по координатам:

$$[\text{prox}_r(y)]_i = \arg\min_{x_i\in\mathbb{R}} \underbrace{\left[\lambda|x_i| + \frac{1}{2}(x_i - y_i)^2\right]}_{h(x_i)}$$

Вывод проксимального оператора для ℓ_1-нормы

Пусть $r(x) = \lambda\|x\|_1$, $\lambda > 0$. По определению:

$$\text{prox}_r(y) = \arg\min_{x\in\mathbb{R}^n} \left[\lambda\|x\|_1 + \frac{1}{2}\|x - y\|_2^2\right]$$

Так как $\|x\|_1 = \sum_i |x_i|$ и $\|x - y\|_2^2 = \sum_i (x_i - y_i)^2$, задача **разделяется** по координатам:

$$[\text{prox}_r(y)]_i = \arg\min_{x_i\in\mathbb{R}} \underbrace{\left[\lambda|x_i| + \frac{1}{2}(x_i - y_i)^2\right]}_{h(x_i)}$$

Рассмотрим три случая для каждой скалярной задачи:

Вывод проксимального оператора для ℓ_1-нормы

Пусть $r(x) = \lambda\|x\|_1$, $\lambda > 0$. По определению:

$$\text{prox}_r(y) = \arg\min_{x\in\mathbb{R}^n} \left[\lambda\|x\|_1 + \frac{1}{2}\|x - y\|_2^2\right]$$

Так как $\|x\|_1 = \sum_i |x_i|$ и $\|x - y\|_2^2 = \sum_i (x_i - y_i)^2$, задача **разделяется** по координатам:

$$[\text{prox}_r(y)]_i = \arg\min_{x_i\in\mathbb{R}} \underbrace{\left[\lambda|x_i| + \frac{1}{2}(x_i - y_i)^2\right]}_{h(x_i)}$$

Рассмотрим три случая для каждой скалярной задачи:

1. $x_i > 0$: $h(x_i) = \lambda x_i + \frac{1}{2}(x_i - y_i)^2$. Берём производную: $h'(x_i) = \boxed{\lambda + x_i - y_i = 0}$, откуда $\boxed{x_i^* = y_i - \lambda.}$ Решение допустимо $(x_i^* > 0)$ при $y_i > \lambda$.

Вывод проксимального оператора для ℓ_1-нормы

Пусть $r(x) = \lambda\|x\|_1$, $\lambda > 0$. По определению:

$$\text{prox}_r(y) = \arg\min_{x \in \mathbb{R}^n} \left[\lambda\|x\|_1 + \frac{1}{2}\|x - y\|_2^2\right]$$

Так как $\|x\|_1 = \sum_i |x_i|$ и $\|x - y\|_2^2 = \sum_i (x_i - y_i)^2$, задача **разделяется** по координатам:

$$[\text{prox}_r(y)]_i = \arg\min_{x_i \in \mathbb{R}} \underbrace{\left[\lambda|x_i| + \frac{1}{2}(x_i - y_i)^2\right]}_{h(x_i)}$$

Рассмотрим три случая для каждой скалярной задачи:

1. $x_i > 0$**:** $h(x_i) = \lambda x_i + \frac{1}{2}(x_i - y_i)^2$. Берём производную: $h'(x_i) = \lambda + x_i - y_i = 0$, откуда $x_i^* = y_i - \lambda$. Решение допустимо ($x_i^* > 0$) при $y_i > \lambda$.

Вывод проксимального оператора для ℓ_1-нормы

Пусть $r(x) = \lambda\|x\|_1$, $\lambda > 0$. По определению:

$$\text{prox}_r(y) = \arg\min_{x\in\mathbb{R}^n} \left[\lambda\|x\|_1 + \frac{1}{2}\|x - y\|_2^2\right]$$

Так как $\|x\|_1 = \sum_i |x_i|$ и $\|x - y\|_2^2 = \sum_i (x_i - y_i)^2$, задача **разделяется** по координатам:

$$[\text{prox}_r(y)]_i = \arg\min_{x_i\in\mathbb{R}} \underbrace{\left[\lambda|x_i| + \frac{1}{2}(x_i - y_i)^2\right]}_{h(x_i)}$$

Рассмотрим три случая для каждой скалярной задачи:

1. $x_i > 0$: $h(x_i) = \lambda x_i + \frac{1}{2}(x_i - y_i)^2$. Берём производную: $h'(x_i) = \lambda + x_i - y_i = 0$, откуда $\boxed{x_i^* = y_i - \lambda.}$ Решение допустимо $(x_i^* > 0)$ при $y_i > \lambda$.
2. $x_i < 0$: $h(x_i) = -\lambda x_i + \frac{1}{2}(x_i - y_i)^2$. Берём производную: $h'(x_i) = -\lambda + x_i - y_i = 0$, откуда $\boxed{x_i^* = y_i + \lambda.}$ Решение допустимо $(x_i^* < 0)$ при $y_i < -\lambda$.

Вывод проксимального оператора для ℓ_1-нормы

Пусть $r(x) = \lambda\|x\|_1$, $\lambda > 0$. По определению:

$$\text{prox}_r(y) = \arg\min_{x\in\mathbb{R}^n} \left[\lambda\|x\|_1 + \frac{1}{2}\|x-y\|_2^2\right]$$

Так как $\|x\|_1 = \sum_i |x_i|$ и $\|x-y\|_2^2 = \sum_i (x_i - y_i)^2$, задача **разделяется** по координатам:

$$[\text{prox}_r(y)]_i = \arg\min_{x_i\in\mathbb{R}} \underbrace{\left[\lambda|x_i| + \frac{1}{2}(x_i - y_i)^2\right]}_{h(x_i)}$$

Рассмотрим три случая для каждой скалярной задачи:

1. $x_i > 0$: $h(x_i) = \lambda x_i + \frac{1}{2}(x_i - y_i)^2$. Берём производную: $h'(x_i) = \lambda + x_i - y_i = 0$, откуда $x_i^* = y_i - \lambda$. Решение допустимо ($x_i^* > 0$) при $y_i > \lambda$.
2. $x_i < 0$: $h(x_i) = -\lambda x_i + \frac{1}{2}(x_i - y_i)^2$. Берём производную: $h'(x_i) = -\lambda + x_i - y_i = 0$, откуда $x_i^* = y_i + \lambda$. Решение допустимо ($x_i^* < 0$) при $y_i < -\lambda$.

Вывод проксимального оператора для ℓ_1-нормы

Пусть $r(x) = \lambda\|x\|_1$, $\lambda > 0$. По определению:

$$\text{prox}_r(y) = \arg\min_{x\in\mathbb{R}^n} \left[\lambda\|x\|_1 + \frac{1}{2}\|x - y\|_2^2\right]$$

Так как $\|x\|_1 = \sum_i |x_i|$ и $\|x - y\|_2^2 = \sum_i (x_i - y_i)^2$, задача **разделяется** по координатам:

$$[\text{prox}_r(y)]_i = \arg\min_{x_i\in\mathbb{R}} \underbrace{\left[\lambda|x_i| + \frac{1}{2}(x_i - y_i)^2\right]}_{h(x_i)}$$

Рассмотрим три случая для каждой скалярной задачи:

1. $x_i > 0$: $h(x_i) = \lambda x_i + \frac{1}{2}(x_i - y_i)^2$. Берём производную: $h'(x_i) = \lambda + x_i - y_i = 0$, откуда $x_i^* = y_i - \lambda$. Решение допустимо ($x_i^* > 0$) при $y_i > \lambda$.
2. $x_i < 0$: $h(x_i) = -\lambda x_i + \frac{1}{2}(x_i - y_i)^2$. Берём производную: $h'(x_i) = -\lambda + x_i - y_i = 0$, откуда $x_i^* = y_i + \lambda$. Решение допустимо ($x_i^* < 0$) при $y_i < -\lambda$.

$\partial(|x|)(0)$

3. $x_i = 0$: Из условия оптимальности $\boxed{0 \in \partial h(0):}$ $0 \in \lambda \cdot [-1, 1] - y_i$, т.е. $\boxed{|y_i| \le \lambda.}$

Вывод проксимального оператора для ℓ_1-нормы

Объединяя три случая:

$$
[\text{prox}_r(y)]_i = \begin{cases} y_i - \lambda, & y_i > \lambda, \\ 0, & |y_i| \leq \lambda, \\ y_i + \lambda, & y_i < -\lambda. \end{cases}
$$

Вывод проксимального оператора для ℓ_1-нормы

Объединяя три случая:

$$
[\text{prox}_r(y)]_i = \begin{cases} y_i - \lambda, & y_i > \lambda, \\ 0, & |y_i| \leq \lambda, \\ y_i + \lambda, & y_i < -\lambda. \end{cases}
$$

Оператор мягкого порога (soft-thresholding):

$$
\boxed{[\text{prox}_r(y)]_i = \text{sign}(y_i) \cdot [|y_i| - \lambda]_+}
$$

Вывод проксимального оператора для ℓ_1-нормы

Объединяя три случая:

$$[\mathrm{prox}_r(y)]_i = \begin{cases} y_i - \lambda, & y_i > \lambda, \\ 0, & |y_i| \le \lambda, \\ y_i + \lambda, & y_i < -\lambda. \end{cases}$$

Оператор мягкого порога (soft-thresholding):

$$\boxed{[\mathrm{prox}_r(y)]_i = \mathrm{sign}(y_i) \cdot [|y_i| - \lambda]_+}$$

«Сжимает» каждую координату к нулю на λ.
Компоненты с $|y_i| \le \lambda$ обнуляются — ℓ_1-регуляризация порождает **разреженность**.

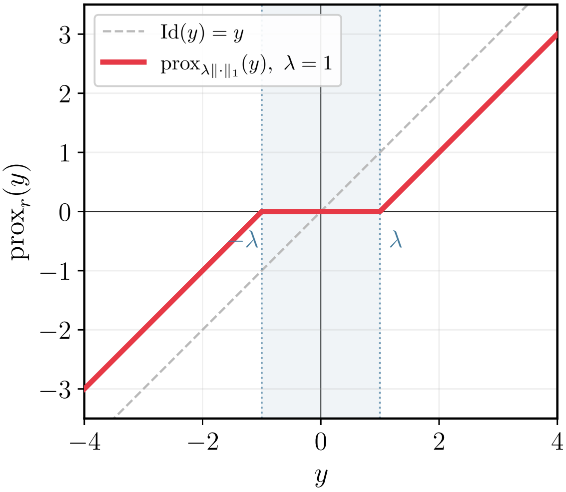


Figure 2: Оператор мягкого порога: «мёртвая зона» $[-\lambda, \lambda]$

Вывод проксимального оператора для $\frac{\lambda}{2}\|x\|_2^2$

Пусть $r(x) = \frac{\lambda}{2}\|x\|_2^2$, $\lambda > 0$. По определению:

$$\boxed{\mathrm{prox}_r(y) = \arg\min_{x\in\mathbb{R}^n}\left[\frac{\lambda}{2}\|x\|_2^2 + \frac{1}{2}\|x-y\|_2^2\right]}$$

Вывод проксимального оператора для $\frac{\lambda}{2}\|x\|_2^2$

Пусть $r(x) = \frac{\lambda}{2}\|x\|_2^2$, $\lambda > 0$. По определению:

$$\text{prox}_r(y) = \arg\min_{x \in \mathbb{R}^n} \left[\frac{\lambda}{2}\|x\|_2^2 + \frac{1}{2}\|x - y\|_2^2 \right]$$

Функция гладкая и сильно выпуклая. Градиент $= 0$:

$$\lambda x^* + (x^* - y) = 0$$

$$x^* = \boxed{\text{prox}_r(y) = \frac{y}{1 + \lambda}}$$

Вывод проксимального оператора для $\frac{\lambda}{2}\|x\|_2^2$

Пусть $r(x) = \frac{\lambda}{2}\|x\|_2^2$, $\lambda > 0$. По определению:

$$\text{prox}_r(y) = \arg\min_{x\in\mathbb{R}^n} \left[\frac{\lambda}{2}\|x\|_2^2 + \frac{1}{2}\|x-y\|_2^2\right]$$

Функция гладкая и сильно выпуклая. Градиент = 0:

$$\lambda x + (x - y) = 0$$

$$\boxed{\text{prox}_r(y) = \frac{y}{1+\lambda}}$$

Равномерное сжатие к нулю с коэффициентом $\frac{1}{1+\lambda}$. В отличие от ℓ_1, ни одна компонента не обнуляется — ℓ_2 не порождает разреженность.

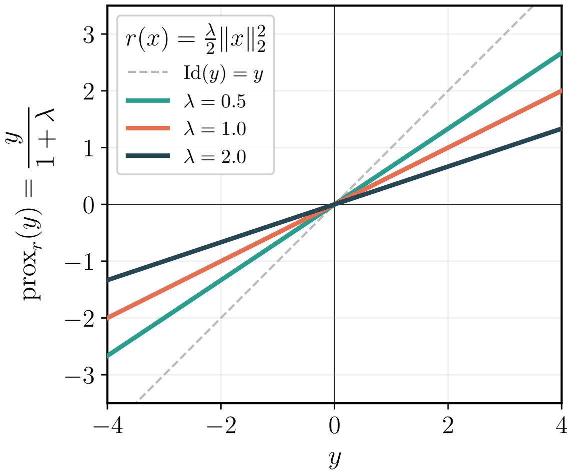


Figure 3: Линейное сжатие для разных λ

Вывод проксимального оператора для $\mathbb{I}_S(x)$ (проекция)

Пусть $r(x) = \mathbb{I}_S(x)$ — индикаторная функция замкнутого выпуклого S. По определению:

$$\text{prox}_r(y) = \arg\min_{x \in \mathbb{R}^n} \left[\mathbb{I}_S(x) + \frac{1}{2}\|x - y\|_2^2 \right]$$

Вывод проксимального оператора для $\mathbb{I}_S(x)$ (проекция)

Пусть $r(x) = \mathbb{I}_S(x)$ — индикаторная функция замкнутого выпуклого S. По определению:

$$\text{prox}_r(y) = \arg\min_{x \in \mathbb{R}^n} \left[\mathbb{I}_S(x) + \frac{1}{2}\|x - y\|_2^2 \right]$$

Так как $\mathbb{I}_S(x) = 0$ при $x \in S$ и $+\infty$ иначе, минимизация допустима только по $x \in S$:

$$\text{prox}_r(y) = \arg\min_{x \in S} \frac{1}{2}\|x - y\|_2^2 = \text{proj}_S(y)$$

Вывод проксимального оператора для $\mathbb{I}_S(x)$ (проекция)

Пусть $r(x) = \mathbb{I}_S(x)$ — индикаторная функция замкнутого выпуклого S. По определению:

$$
\text{prox}_r(y) = \arg\min_{x \in \mathbb{R}^n} \left[\mathbb{I}_S(x) + \frac{1}{2}\|x - y\|_2^2 \right]
$$

Так как $\mathbb{I}_S(x) = 0$ при $x \in S$ и $+\infty$ иначе, минимизация допустима только по $x \in S$:

$$
\text{prox}_r(y) = \arg\min_{x \in S} \frac{1}{2}\|x - y\|_2^2 = \text{proj}_S(y)
$$

$$
\boxed{\text{prox}_{\mathbb{I}_S}(y) = \text{proj}_S(y)}
$$

Вывод проксимального оператора для $\mathbb{I}_S(x)$ (проекция)

Пусть $r(x) = \mathbb{I}_S(x)$ — индикаторная функция замкнутого выпуклого S. По определению:

$$\text{prox}_r(y) = \arg\min_{x\in\mathbb{R}^n} \left[\mathbb{I}_S(x) + \frac{1}{2}\|x-y\|_2^2\right]$$

Так как $\mathbb{I}_S(x) = 0$ при $x \in S$ и $+\infty$ иначе, минимизация допустима только по $x \in S$:

$$\text{prox}_r(y) = \arg\min_{x\in S} \frac{1}{2}\|x-y\|_2^2 = \text{proj}_S(y)$$

$$\boxed{\text{prox}_{\mathbb{I}_S}(y) = \text{proj}_S(y)}$$

Проксимальный оператор индикаторной функции — это **ортогональная проекция** на множество S.

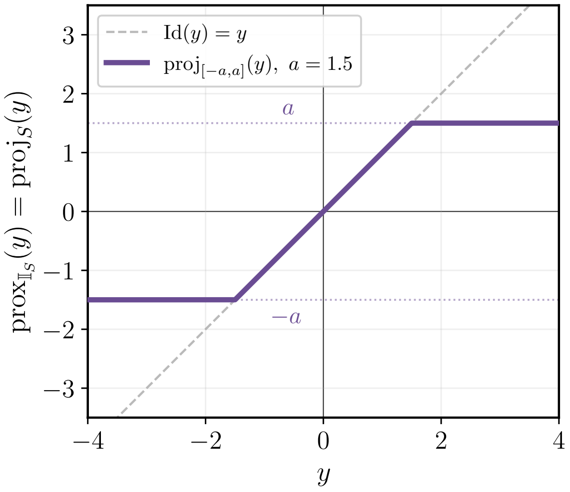


Figure 4: Проекция на отрезок $[-a, a]$ — скалярный пример

Свойства проксимального оператора

Theorem

Пусть $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклая функция, для которой определён prox_r. Если существует такая точка $\hat{x} \in \mathbb{R}^n$, что $r(\hat{x}) < +\infty$, то проксимальный оператор определён однозначно (т.е. всегда возвращает единственное значение).

Доказательство:

Свойства проксимального оператора

Theorem

Пусть $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклая функция, для которой определён prox_r. Если существует такая точка $\hat{x} \in \mathbb{R}^n$, что $r(\hat{x}) < +\infty$, то проксимальный оператор определён однозначно (т.е. всегда возвращает единственное значение).

Доказательство:

Проксимальный оператор возвращает минимум некоторой задачи оптимизации.

Свойства проксимального оператора

Theorem

Пусть $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклая функция, для которой определён prox_r. Если существует такая точка $\hat{x} \in \mathbb{R}^n$, что $r(\hat{x}) < +\infty$, то проксимальный оператор определён однозначно (т.е. всегда возвращает единственное значение).

Доказательство:

Проксимальный оператор возвращает минимум некоторой задачи оптимизации.

Вопрос: что можно сказать об этой задаче?

Свойства проксимального оператора

Theorem

Пусть $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклая функция, для которой определён prox_r. Если существует такая точка $\hat{x} \in \mathbb{R}^n$, что $r(\hat{x}) < +\infty$, то проксимальный оператор определён однозначно (т.е. всегда возвращает единственное значение).

Доказательство:

Проксимальный оператор возвращает минимум некоторой задачи оптимизации.

Вопрос: что можно сказать об этой задаче?

Она является сильно выпуклой, а значит имеет ровно один минимум (существование $\hat{x}$ необходимо для того, чтобы $r(\tilde{x}) + \frac{1}{2}\|x - \tilde{x}\|_2^2$ принимало конечное значение хотя бы в одной точке).

Свойства проксимального оператора

Theorem

Пусть $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклая функция, для которой определён prox_r. Тогда для любых $x, y \in \mathbb{R}^n$ следующие три условия эквивалентны:

- $\text{prox}_r(x) = y$,

Доказательство

Свойства проксимального оператора

Theorem

Пусть $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклая функция, для которой определён prox_r. Тогда для любых $x, y \in \mathbb{R}^n$ следующие три условия эквивалентны:

- $\text{prox}_r(x) = y$,
- $x - y \in \partial r(y)$,

Доказательство

Свойства проксимального оператора

Theorem

Пусть $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклая функция, для которой определён prox_r. Тогда для любых $x, y \in \mathbb{R}^n$ следующие три условия эквивалентны:

- $\text{prox}_r(x) = y$,
- $x - y \in \partial r(y)$,
- $\langle x - y, z - y \rangle \leq r(z) - r(y)$ для любого $z \in \mathbb{R}^n$.

Доказательство

Свойства проксимального оператора

Theorem

Пусть $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклая функция, для которой определён prox_r. Тогда для любых $x, y \in \mathbb{R}^n$ следующие три условия эквивалентны:

- $\text{prox}_r(x) = y$,
- $x - y \in \partial r(y)$,
- $\langle x - y, z - y \rangle \leq r(z) - r(y)$ для любого $z \in \mathbb{R}^n$.

Доказательство

1. Установим эквивалентность первого и второго условий. Первое условие можно переписать как

$$
y = \arg\min_{\tilde{x} \in \mathbb{R}^n} \left(r(\tilde{x}) + \frac{1}{2}\|x - \tilde{x}\|^2 \right).
$$

Из условия оптимальности для выпуклой функции r это эквивалентно:

$$
0 \in \partial \left(r(\tilde{x}) + \frac{1}{2}\|x - \tilde{x}\|^2 \right) \bigg|_{\tilde{x} = y} = \partial r(y) + y - x.
$$

Свойства проксимального оператора

Theorem

Пусть $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклая функция, для которой определён prox_r. Тогда для любых $x, y \in \mathbb{R}^n$ следующие три условия эквивалентны:

- $\text{prox}_r(x) = y$,
- $x - y \in \partial r(y)$,
- $\langle x - y, z - y \rangle \leq r(z) - r(y)$ для любого $z \in \mathbb{R}^n$.

Доказательство

1. Установим эквивалентность первого и второго условий. Первое условие можно переписать как

$$y = \arg\min_{\tilde{x} \in \mathbb{R}^n} \left(r(\tilde{x}) + \frac{1}{2}\|x - \tilde{x}\|^2 \right).$$

Из условия оптимальности для выпуклой функции r это эквивалентно:

$$0 \in \partial \left(r(\tilde{x}) + \frac{1}{2}\|x - \tilde{x}\|^2 \right)\Bigg|_{\tilde{x}=y} = \partial r(y) + y - x.$$

2. Из определения субдифференциала для любого субградиента $g \in \partial r(y)$ и для любого $z \in \mathbb{R}^n$:

$$\langle g, z - y \rangle \leq r(z) - r(y).$$

В частности, это выполняется для $g = x - y$. Обратно, это также очевидно: для $g = x - y$ приведённое неравенство выполняется, что означает $g \in \partial r(y)$.

Свойства проксимального оператора

Theorem

Оператор $\text{prox}_r(x)$ является строго нерастягивающим (FNE)

$$\|\text{prox}_r(x) - \text{prox}_r(y)\|_2^2 \leq \langle \text{prox}_r(x) - \text{prox}_r(y), x - y \rangle$$

и нерастягивающим:

$$\|\text{prox}_r(x) - \text{prox}_r(y)\|_2 \leq \|x - y\|_2$$

Доказательство

1. Пусть $u = \text{prox}_r(x)$, $v = \text{prox}_r(y)$. Тогда из предыдущего свойства:

$$\langle x - u, z_1 - u \rangle \leq r(z_1) - r(u)$$
$$\langle y - v, z_2 - v \rangle \leq r(z_2) - r(v).$$

Свойства проксимального оператора

Theorem

Оператор $\text{prox}_r(x)$ является строго нерастягивающим (FNE)

$$\|\text{prox}_r(x) - \text{prox}_r(y)\|_2^2 \leq \langle \text{prox}_r(x) - \text{prox}_r(y), x - y \rangle$$

и нерастягивающим:

$$\|\text{prox}_r(x) - \text{prox}_r(y)\|_2 \leq \|x - y\|_2$$

Доказательство

1. Пусть $u = \text{prox}_r(x)$, $v = \text{prox}_r(y)$. Тогда из предыдущего свойства:

$$\begin{aligned}
\langle x - u, z_1 - u \rangle &\leq r(z_1) - r(u) \\
\langle y - v, z_2 - v \rangle &\leq r(z_2) - r(v).
\end{aligned}$$

2. Подставим $z_1 = v$ и $z_2 = u$. Складывая, получаем:

$$\begin{aligned}
\langle x - u, v - u \rangle + \langle y - v, u - v \rangle &\leq 0, \\
\langle x - y, v - u \rangle + \|v - u\|_2^2 &\leq 0.
\end{aligned}$$

Свойства проксимального оператора

Theorem

Оператор $\text{prox}_r(x)$ является строго нерастягивающим (FNE)

$$\|\text{prox}_r(x) - \text{prox}_r(y)\|_2^2 \leq \langle \text{prox}_r(x) - \text{prox}_r(y), x - y \rangle$$

и нерастягивающим:

$$\|\text{prox}_r(x) - \text{prox}_r(y)\|_2 \leq \|x - y\|_2$$

Доказательство

1. Пусть $u = \text{prox}_r(x)$, $v = \text{prox}_r(y)$. Тогда из предыдущего свойства:

$$\begin{aligned} \langle x - u, z_1 - u \rangle &\leq r(z_1) - r(u) \\ \langle y - v, z_2 - v \rangle &\leq r(z_2) - r(v). \end{aligned}$$

2. Подставим $z_1 = v$ и $z_2 = u$. Складывая, получаем:

$$\begin{aligned} \langle x - u, v - u \rangle + \langle y - v, u - v \rangle &\leq 0, \\ \langle x - y, v - u \rangle + \|v - u\|_2^2 &\leq 0. \end{aligned}$$

3. Что в точности является тем, что нужно доказать после подстановки u, v:

$$\|u - v\|_2^2 \leq \langle x - y, u - v \rangle$$

Свойства проксимального оператора

Theorem

Оператор $\text{prox}_r(x)$ является строго нерастягивающим (FNE)

$$\|\text{prox}_r(x) - \text{prox}_r(y)\|_2^2 \leq \langle \text{prox}_r(x) - \text{prox}_r(y), x - y \rangle$$

и нерастягивающим:

$$\|\text{prox}_r(x) - \text{prox}_r(y)\|_2 \leq \|x - y\|_2$$

Доказательство

1. Пусть $u = \text{prox}_r(x)$, $v = \text{prox}_r(y)$. Тогда из предыдущего свойства:

$$\begin{aligned} \langle x - u, z_1 - u \rangle &\leq r(z_1) - r(u) \\ \langle y - v, z_2 - v \rangle &\leq r(z_2) - r(v). \end{aligned}$$

2. Подставим $z_1 = v$ и $z_2 = u$. Складывая, получаем:

$$\begin{aligned} \langle x - u, v - u \rangle + \langle y - v, u - v \rangle &\leq 0, \\ \langle x - y, v - u \rangle + \|v - u\|_2^2 &\leq 0. \end{aligned}$$

3. Что в точности является тем, что нужно доказать после подстановки u, v:

$$\|u - v\|_2^2 \leq \langle x - y, u - v \rangle$$

4. Нерастяжимость следует из неравенства Коши — Буняковского — Шварца.

Свойства проксимального оператора

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ и $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклые функции. Дополнительно предположим, что f непрерывно дифференцируема и L-гладкая, а для r определён prox_r. Тогда x^* является решением композитной задачи оптимизации тогда и только тогда, когда для любого $\alpha > 0$ выполняется:

$$
x^* = \text{prox}_{r,\alpha}(x^* - \alpha \nabla f(x^*))
$$

Доказательство

1. Условия оптимальности:

Свойства проксимального оператора

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ и $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклые функции. Дополнительно предположим, что f непрерывно дифференцируема и L-гладкая, а для r определён prox_r. Тогда x^* является решением композитной задачи оптимизации тогда и только тогда, когда для любого $\alpha > 0$ выполняется:

$$x^* = \text{prox}_{r,\alpha}(x^* - \alpha \nabla f(x^*))$$

Доказательство

1. Условия оптимальности:

$$0 \in \nabla f(x^*) + \partial r(x^*)$$

Свойства проксимального оператора

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ и $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклые функции. Дополнительно предположим, что f непрерывно дифференцируема и L-гладкая, а для r определён prox_r. Тогда x^* является решением композитной задачи оптимизации тогда и только тогда, когда для любого $\alpha > 0$ выполняется:

$$
x^* = \text{prox}_{r,\alpha}(x^* - \alpha \nabla f(x^*))
$$

Доказательство

1. Условия оптимальности:

$$
\begin{aligned}
0 \in & \nabla f(x^*) + \partial r(x^*) \\
-\alpha \nabla f(x^*) \in & \alpha \partial r(x^*)
\end{aligned}
$$

Свойства проксимального оператора

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ и $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклые функции. Дополнительно предположим, что f непрерывно дифференцируема и L-гладкая, а для r определён prox_r. Тогда x^* является решением композитной задачи оптимизации тогда и только тогда, когда для любого $\alpha > 0$ выполняется:

$$x^* = \text{prox}_{r,\alpha}(x^* - \alpha \nabla f(x^*))$$

Доказательство

1. Условия оптимальности:

$$\begin{aligned} 0 \in& \nabla f(x^*) + \partial r(x^*) \\ -\alpha \nabla f(x^*) \in& \alpha \partial r(x^*) \\ x^* - \alpha \nabla f(x^*) - x^* \in& \alpha \partial r(x^*) \end{aligned}$$

Свойства проксимального оператора

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ и $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклые функции. Дополнительно предположим, что f непрерывно дифференцируема и L-гладкая, а для r определён prox_r. Тогда x^* является решением композитной задачи оптимизации тогда и только тогда, когда для любого $\alpha > 0$ выполняется:

$$x^* = \text{prox}_{r,\alpha}(x^* - \alpha \nabla f(x^*))$$

Доказательство

1. Условия оптимальности:

$$\begin{aligned} 0 \in & \nabla f(x^*) + \partial r(x^*) \\ -\alpha \nabla f(x^*) \in & \alpha \partial r(x^*) \\ x^* - \alpha \nabla f(x^*) - x^* \in & \alpha \partial r(x^*) \end{aligned}$$

2. Из предыдущей леммы (применённой к αr):

$$\text{prox}_{\alpha r}(x) = y \Leftrightarrow x - y \in \alpha \partial r(y)$$

Свойства проксимального оператора

Theorem

Пусть $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ и $r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ — выпуклые функции. Дополнительно предположим, что f непрерывно дифференцируема и L-гладкая, а для r определён prox_r. Тогда x^* является решением композитной задачи оптимизации тогда и только тогда, когда для любого $\alpha > 0$ выполняется:

$$x^* = \text{prox}_{r,\alpha}(x^* - \alpha \nabla f(x^*))$$

Доказательство

1. Условия оптимальности:
$$\begin{aligned} 0 \in & \nabla f(x^*) + \partial r(x^*) \\ -\alpha \nabla f(x^*) \in & \alpha \partial r(x^*) \\ x^* - \alpha \nabla f(x^*) - x^* \in & \alpha \partial r(x^*) \end{aligned}$$
2. Из предыдущей леммы (применённой к αr):
$$\text{prox}_{\alpha r}(x) = y \Leftrightarrow x - y \in \alpha \partial r(y)$$
3. Следовательно,
$$x^* = \text{prox}_{\alpha r}(x^* - \alpha \nabla f(x^*)) = \text{prox}_{r,\alpha}(x^* - \alpha \nabla f(x^*))$$

Теоретические инструменты для анализа сходимости

Инструменты для анализа сходимости

Theorem

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

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

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

Доказательство

1. Рассмотрим вспомогательную функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$. Очевидно, что это выпуклая функция (как сумма выпуклых). Легко проверить, что она L-гладкая по определению, поскольку $\nabla \varphi(y) = \nabla f(y) - \nabla f(x)$ и $\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leq L\|y_1 - y_2\|$.

Инструменты для анализа сходимости

Theorem

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

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

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

Доказательство

1. Рассмотрим вспомогательную функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$. Очевидно, что это выпуклая функция (как сумма выпуклых). Легко проверить, что она L-гладкая по определению, поскольку $\nabla \varphi(y) = \nabla f(y) - \nabla f(x)$ и $\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leq L\|y_1 - y_2\|$.
2. Запишем верхнюю параболическую оценку для функции $\varphi(y)$:

Инструменты для анализа сходимости

Theorem

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

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

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

Доказательство

1. Рассмотрим вспомогательную функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$. Очевидно, что это выпуклая функция (как сумма выпуклых). Легко проверить, что она L-гладкая по определению, поскольку $\nabla \varphi(y) = \nabla f(y) - \nabla f(x)$ и $\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leq L\|y_1 - y_2\|$.
2. Запишем верхнюю параболическую оценку для функции $\varphi(y)$:

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

Инструменты для анализа сходимости

Theorem

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

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

Доказательство

1. Рассмотрим вспомогательную функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$. Очевидно, что это выпуклая функция (как сумма выпуклых). Легко проверить, что она L-гладкая по определению, поскольку $\nabla \varphi(y) = \nabla f(y) - \nabla f(x)$ и $\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leq L\|y_1 - y_2\|$.
2. Запишем верхнюю параболическую оценку для функции $\varphi(y)$:

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

Инструменты для анализа сходимости

Theorem

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

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

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

Доказательство

1. Рассмотрим вспомогательную функцию $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$. Очевидно, что это выпуклая функция (как сумма выпуклых). Легко проверить, что она L-гладкая по определению, поскольку $\nabla \varphi(y) = \nabla f(y) - \nabla f(x)$ и $\|\nabla \varphi(y_1) - \nabla \varphi(y_2)\| = \|\nabla f(y_1) - \nabla f(y_2)\| \leq L\|y_1 - y_2\|$.
2. Запишем верхнюю параболическую оценку для функции $\varphi(y)$:

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

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

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

Инструменты для анализа сходимости

3. Минимизатор φ: $\nabla \varphi(y) = \nabla f(y) - \nabla f(x) = 0$ при $y = x$, следовательно $\varphi(x) \leq \varphi(z)$ для любого z. В частности:
$$
\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla \varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla \varphi(y)\|_2^2
$$

Инструменты для анализа сходимости

3. Минимизатор φ: $\nabla \varphi(y) = \nabla f(y) - \nabla f(x) = 0$ при $y = x$, следовательно $\varphi(x) \leq \varphi(z)$ для любого z. В частности:
$$\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$
4. Подставим $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$:

Инструменты для анализа сходимости

3. Минимизатор φ: $\nabla \varphi(y) = \nabla f(y) - \nabla f(x) = 0$ при $y = x$, следовательно $\varphi(x) \leq \varphi(z)$ для любого z. В частности:

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

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

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

Инструменты для анализа сходимости

3. Минимизатор φ: $\nabla\varphi(y) = \nabla f(y) - \nabla f(x) = 0$ при $y = x$, следовательно $\varphi(x) \leq \varphi(z)$ для любого z. В частности:

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

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

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

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

Инструменты для анализа сходимости

3. Минимизатор φ: $\nabla \varphi(y) = \nabla f(y) - \nabla f(x) = 0$ при $y = x$, следовательно $\varphi(x) \leq \varphi(z)$ для любого z. В частности:
$$\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla\varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla\varphi(y)\|_2^2$$
4. Подставим $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$:
$$f(x) - \langle \nabla f(x), x \rangle \leq f(y) - \langle \nabla f(x), y \rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2$$
$$f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y)$$
$$\|\nabla f(y) - \nabla f(x)\|_2^2 \leq 2L\left(f(y) - f(x) - \langle \nabla f(x), y - x \rangle\right)$$

Инструменты для анализа сходимости

3. Минимизатор φ: $\nabla \varphi(y) = \nabla f(y) - \nabla f(x) = 0$ при $y = x$, следовательно $\varphi(x) \leq \varphi(z)$ для любого z. В частности:
$$
\varphi(x) \leq \varphi\left(y - \frac{1}{L}\nabla \varphi(y)\right) \leq \varphi(y) - \frac{1}{2L}\|\nabla \varphi(y)\|_2^2
$$
4. Подставим $\varphi(y) = f(y) - \langle \nabla f(x), y \rangle$:
$$
\begin{aligned}
& f(x) - \langle \nabla f(x), x \rangle \leq f(y) - \langle \nabla f(x), y \rangle - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2 \\
& f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 \leq f(y) \\
& \|\nabla f(y) - \nabla f(x)\|_2^2 \leq 2L\left(f(y) - f(x) - \langle \nabla f(x), y - x \rangle\right) \\
\text{меняем x и y} \quad & \|\nabla f(x) - \nabla f(y)\|_2^2 \leq 2L\left(f(x) - f(y) - \langle \nabla f(y), x - y \rangle\right)
\end{aligned}
$$

Инструменты для анализа сходимости

3. Минимизатор φ: $\nabla \varphi(y) = \nabla f(y) - \nabla f(x) = 0$ при $y = x$, следовательно $\varphi(x) \leq \varphi(z)$ для любого z. В частности:

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

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

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

Инструменты для анализа сходимости

3. Минимизатор φ: $\nabla \varphi(y) = \nabla f(y) - \nabla f(x) = 0$ при $y = x$, следовательно $\varphi(x) \leq \varphi(z)$ для любого z. В частности:

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

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

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

Лемма доказана. На первый взгляд она не имеет большого геометрического смысла, но мы будем использовать её как удобный инструмент для оценки разности градиентов.

Инструменты для анализа 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 $x, y \in \mathbb{R}^n$ выполняется:

$$
\begin{aligned}
\text{Сильно выпуклый случай } \mu > 0 \quad & \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2 \\
\text{Выпуклый случай } \mu = 0 \quad & \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0
\end{aligned}
$$

Доказательство

1. Докажем только сильно выпуклый случай, выпуклый следует из него при $\mu = 0$. Начнём 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 для анализа сходимости

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

Инструменты для анализа сходимости

2. Для достаточности предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2$. Используя формулу 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 для анализа сходимости

2. Для достаточности предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2$. Используя формулу 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 для анализа сходимости

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

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

Инструменты для анализа сходимости

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

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

Инструменты для анализа сходимости

2. Для достаточности предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \mu \|x-y\|^2$. Используя формулу 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 для анализа сходимости

2. Для достаточности предполагаем, что $\langle \nabla f(x) - \nabla f(y), x-y\rangle \geq \mu\|x-y\|^2$. Используя формулу 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 для анализа сходимости

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

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

Инструменты для анализа сходимости

2. Для достаточности предполагаем, что $\langle \nabla f(x) - \nabla f(y), x-y\rangle \geq \mu\|x-y\|^2$. Используя формулу 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, критерий сильной выпуклости выполнен

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

Инструменты для анализа сходимости

2. Для достаточности предполагаем, что $\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu \|x - y\|^2$. Используя формулу 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, критерий сильной выпуклости выполнен

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

Инструменты для анализа сходимости

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

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

Таким образом, критерий сильной выпуклости выполнен

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

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

Проксимальный градиентный метод. Выпуклый случай

Сходимость

Theorem

Рассмотрим проксимальный градиентный метод

$$\boxed{x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right)}$$

Для критерия $\boxed{\varphi(x) = f(x) + r(x)}$ предполагаем:

- f выпукла, дифференцируема, $\text{dom}(f) = \mathbb{R}^n$, и ∇f липшицев с константой $L > 0$.
- r выпукла, и $\text{prox}_{\alpha r}(x_k) = \text{arg}\min\limits_{x \in \mathbb{R}^n} \left[\alpha r(x) + \frac{1}{2}\|x - x_k\|_2^2\right]$ может быть вычислен.

Проксимальный градиентный спуск с фиксированным шагом $\alpha = 1/L$ удовлетворяет

$$\boxed{\varphi(x_k) - \varphi^* \leq \frac{L\|x_0 - x^*\|^2}{2k},}$$

Проксимальный градиентный спуск имеет скорость сходимости $O(1/k)$ или $O(1/\varepsilon)$. Это совпадает со скоростью обычного градиентного спуска! (Но помните о стоимости проксимальной операции)

Сходимость (доказательство)

1. Введём **градиентное отображение**, обозначаемое $G_\alpha(x)$, которое играет роль «градиентоподобного объекта»:

$$
\begin{aligned}
x_{k+1} &= \text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) \\
x_{k+1} &= x_k - \alpha G_\alpha(x_k).
\end{aligned}
$$

где $G_\alpha(x)$ определяется как:

$$
G_\alpha(x) = \frac{1}{\alpha}\left(x - \text{prox}_{\alpha r}\left(x - \alpha \nabla f\left(x\right)\right)\right)
$$

Заметим, что $G_\alpha(x) = 0$ тогда и только тогда, когда x оптимально. Таким образом, G_α аналогичен ∇f. Если x локально оптимально, то $G_\alpha(x) = 0$ даже для невыпуклой f. Это показывает, что проксимальный градиентный метод эффективно комбинирует градиентный спуск по f с проксимальным оператором от r, позволяя работать с негладкими компонентами.

Сходимость (доказательство)

1. Введём **градиентное отображение**, обозначаемое $G_\alpha(x)$, которое играет роль «градиентоподобного объекта»:

$$
\begin{aligned}
x_{k+1} &= \text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) \\
x_{k+1} &= x_k - \alpha G_\alpha(x_k).
\end{aligned}
$$

где $G_\alpha(x)$ определяется как:

$$
G_\alpha(x) = \frac{1}{\alpha}\left(x - \text{prox}_{\alpha r}\left(x - \alpha \nabla f\left(x\right)\right)\right)
$$

Заметим, что $G_\alpha(x) = 0$ тогда и только тогда, когда x оптимально. Таким образом, G_α аналогичен ∇f. Если x локально оптимально, то $G_\alpha(x) = 0$ даже для невыпуклой f. Это показывает, что проксимальный градиентный метод эффективно комбинирует градиентный спуск по f с проксимальным оператором от r, позволяя работать с негладкими компонентами.

2. Используем гладкость и выпуклость f для произвольной точки x:

Сходимость (доказательство)

1. Введём **градиентное отображение**, обозначаемое $G_\alpha(x)$, которое играет роль «градиентоподобного объекта»:

$$
\begin{aligned}
x_{k+1} &= \text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) \\
x_{k+1} &= x_k - \alpha G_\alpha(x_k).
\end{aligned}
$$

где $G_\alpha(x)$ определяется как:

$$
G_\alpha(x) = \frac{1}{\alpha}\left(x - \text{prox}_{\alpha r}\left(x - \alpha \nabla f\left(x\right)\right)\right)
$$

Заметим, что $G_\alpha(x) = 0$ тогда и только тогда, когда x оптимально. Таким образом, G_α аналогичен ∇f. Если x локально оптимально, то $G_\alpha(x) = 0$ даже для невыпуклой f. Это показывает, что проксимальный градиентный метод эффективно комбинирует градиентный спуск по f с проксимальным оператором от r, позволяя работать с негладкими компонентами.

2. Используем гладкость и выпуклость f для произвольной точки x:

$$
\text{гладкость} \quad f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_2^2
$$

Сходимость (доказательство)

1. Введём **градиентное отображение**, обозначаемое $G_\alpha(x)$, которое играет роль «градиентоподобного объекта»:

$$
\begin{aligned}
x_{k+1} &= \text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) \\
x_{k+1} &= x_k - \alpha G_\alpha(x_k).
\end{aligned}
$$

где $G_\alpha(x)$ определяется как:

$$
G_\alpha(x) = \frac{1}{\alpha}\left(x - \text{prox}_{\alpha r}\left(x - \alpha \nabla f\left(x\right)\right)\right)
$$

Заметим, что $G_\alpha(x) = 0$ тогда и только тогда, когда x оптимально. Таким образом, G_α аналогичен ∇f. Если x локально оптимально, то $G_\alpha(x) = 0$ даже для невыпуклой f. Это показывает, что проксимальный градиентный метод эффективно комбинирует градиентный спуск по f с проксимальным оператором от r, позволяя работать с негладкими компонентами.

2. Используем гладкость и выпуклость f для произвольной точки x:

$$
\text{гладкость} \quad f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_2^2
$$

$$
\text{выпуклость} \quad f(x) \geq f(x_k) + \langle \nabla f(x_k), x - x_k \rangle
$$

Сходимость (доказательство)

1. Введём **градиентное отображение**, обозначаемое $G_\alpha(x)$, которое играет роль «градиентоподобного объекта»:

$$
\begin{aligned}
x_{k+1} &= \text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) \\
x_{k+1} &= x_k - \alpha G_\alpha(x_k).
\end{aligned}
$$

где $G_\alpha(x)$ определяется как:

$$
G_\alpha(x) = \frac{1}{\alpha}\left(x - \text{prox}_{\alpha r}\left(x - \alpha \nabla f\left(x\right)\right)\right)
$$

Заметим, что $G_\alpha(x) = 0$ тогда и только тогда, когда x оптимально. Таким образом, G_α аналогичен ∇f. Если x локально оптимально, то $G_\alpha(x) = 0$ даже для невыпуклой f. Это показывает, что проксимальный градиентный метод эффективно комбинирует градиентный спуск по f с проксимальным оператором от r, позволяя работать с негладкими компонентами.

2. Используем гладкость и выпуклость f для произвольной точки x:

$$
\begin{aligned}
\text{гладкость} \quad f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_2^2 \\
\text{выпуклость } f(x) \geq f(x_k) + \langle \nabla f(x_k), x - x_k \rangle \quad &\leq f(x) - \langle \nabla f(x_k), x - x_k \rangle + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{\alpha^2 L}{2}\|G_\alpha(x_k)\|_2^2
\end{aligned}
$$

Сходимость (доказательство)

1. Введём **градиентное отображение**, обозначаемое $G_\alpha(x)$, которое играет роль «градиентоподобного объекта»:

$$
\begin{aligned}
x_{k+1} &= \text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) \\
x_{k+1} &= x_k - \alpha G_\alpha(x_k).
\end{aligned}
$$

где $G_\alpha(x)$ определяется как:

$$
G_\alpha(x) = \frac{1}{\alpha}\left(x - \text{prox}_{\alpha r}\left(x - \alpha \nabla f\left(x\right)\right)\right)
$$

Заметим, что $G_\alpha(x) = 0$ тогда и только тогда, когда x оптимально. Таким образом, G_α аналогичен ∇f. Если x локально оптимально, то $G_\alpha(x) = 0$ даже для невыпуклой f. Это показывает, что проксимальный градиентный метод эффективно комбинирует градиентный спуск по f с проксимальным оператором от r, позволяя работать с негладкими компонентами.

2. Используем гладкость и выпуклость f для произвольной точки x:

$$
\begin{aligned}
\text{гладкость} \quad f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{L}{2}\|x_{k+1} - x_k\|_2^2 \\
\text{выпуклость } f(x) \geq f(x_k) + \langle \nabla f(x_k), x - x_k \rangle \quad &\leq f(x) - \langle \nabla f(x_k), x - x_k \rangle + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{\alpha^2 L}{2}\|G_\alpha(x_k)\|_2^2 \\
&\leq f(x) + \langle \nabla f(x_k), x_{k+1} - x \rangle + \frac{\alpha^2 L}{2}\|G_\alpha(x_k)\|_2^2
\end{aligned}
$$

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$
x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad \Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r(x_{k+1})
$$

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$
x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad \Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r(x_{k+1})
$$

$$
\text{Поскольку } x_{k+1} - x_k = -\alpha G_\alpha(x_k) \qquad \Rightarrow \qquad \alpha G_\alpha(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r(x_{k+1})
$$

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$
\begin{aligned}
x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad &\Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r(x_{k+1}) \\
\text{Поскольку } x_{k+1} - x_k = -\alpha G_\alpha(x_k) \qquad &\Rightarrow \qquad \alpha G_\alpha(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r(x_{k+1}) \\
&\qquad\qquad G_\alpha(x_k) - \nabla f(x_k) \in \partial r(x_{k+1})
\end{aligned}
$$

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$
\begin{aligned}
x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad &\Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r(x_{k+1}) \\
\text{Поскольку } x_{k+1} - x_k = -\alpha G_\alpha(x_k) \qquad &\Rightarrow \qquad \alpha G_\alpha(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r(x_{k+1}) \\
& \qquad\qquad G_\alpha(x_k) - \nabla f(x_k) \in \partial r(x_{k+1})
\end{aligned}
$$

4. По определению субградиента выпуклой функции r для любой точки x:

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$
\begin{aligned}
x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad &\Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r(x_{k+1}) \\
\text{Поскольку } x_{k+1} - x_k = -\alpha G_\alpha(x_k) \qquad &\Rightarrow \qquad \alpha G_\alpha(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r(x_{k+1}) \\
&\qquad\qquad G_\alpha(x_k) - \nabla f(x_k) \in \partial r(x_{k+1})
\end{aligned}
$$

4. По определению субградиента выпуклой функции r для любой точки x:

$$
r(x) \geq r(x_{k+1}) + \langle g, x - x_{k+1} \rangle, \quad g \in \partial r(x_{k+1})
$$

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$
\begin{aligned}
x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad &\Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r (x_{k+1}) \\
\text{Поскольку } x_{k+1} - x_k = - \alpha G_{\alpha}(x_k) \qquad &\Rightarrow \qquad \alpha G_{\alpha}(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r (x_{k+1}) \\
&\qquad\qquad G_{\alpha}(x_k) - \nabla f(x_k) \in \partial r (x_{k+1})
\end{aligned}
$$

4. По определению субградиента выпуклой функции r для любой точки x:

$$
\begin{aligned}
& r(x) \geq r(x_{k+1}) + \langle g, x - x_{k+1} \rangle, \quad g \in \partial r (x_{k+1}) \\
\text{подставляем конкретный субградиент} \qquad & r(x) \geq r(x_{k+1}) + \langle G_{\alpha}(x_k) - \nabla f(x_k), x - x_{k+1} \rangle
\end{aligned}
$$

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$
\begin{aligned}
x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad &\Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r (x_{k+1}) \\
\text{Поскольку } x_{k+1} - x_k = -\alpha G_\alpha(x_k) \qquad &\Rightarrow \qquad \alpha G_\alpha(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r (x_{k+1}) \\
&\qquad\qquad G_\alpha(x_k) - \nabla f(x_k) \in \partial r (x_{k+1})
\end{aligned}
$$

4. По определению субградиента выпуклой функции r для любой точки x:

$$
\begin{aligned}
& r(x) \geq r(x_{k+1}) + \langle g, x - x_{k+1} \rangle, \quad g \in \partial r (x_{k+1}) \\
\text{подставляем конкретный субградиент} \qquad & r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k) - \nabla f(x_k), x - x_{k+1} \rangle \\
& r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k), x - x_{k+1} \rangle - \langle \nabla f(x_k), x - x_{k+1} \rangle
\end{aligned}
$$

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad \Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r(x_{k+1})$$

$$\text{Поскольку } x_{k+1} - x_k = -\alpha G_\alpha(x_k) \qquad \Rightarrow \qquad \alpha G_\alpha(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r(x_{k+1})$$

$$G_\alpha(x_k) - \nabla f(x_k) \in \partial r(x_{k+1})$$

4. По определению субградиента выпуклой функции r для любой точки x:

$$r(x) \geq r(x_{k+1}) + \langle g, x - x_{k+1} \rangle, \quad g \in \partial r(x_{k+1})$$

подставляем конкретный субградиент

$$r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k) - \nabla f(x_k), x - x_{k+1} \rangle$$

$$r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k), x - x_{k+1} \rangle - \langle \nabla f(x_k), x - x_{k+1} \rangle$$

$$\langle \nabla f(x_k), x_{k+1} - x \rangle \leq r(x) - r(x_{k+1}) - \langle G_\alpha(x_k), x - x_{k+1} \rangle$$

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$
\begin{aligned}
x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad &\Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r(x_{k+1}) \\
\text{Поскольку } x_{k+1} - x_k = -\alpha G_\alpha(x_k) \qquad &\Rightarrow \qquad \alpha G_\alpha(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r(x_{k+1}) \\
&\qquad\qquad G_\alpha(x_k) - \nabla f(x_k) \in \partial r(x_{k+1})
\end{aligned}
$$

4. По определению субградиента выпуклой функции r для любой точки x:

$$
\begin{aligned}
& r(x) \geq r(x_{k+1}) + \langle g, x - x_{k+1} \rangle, \quad g \in \partial r(x_{k+1}) \\
\text{подставляем конкретный субградиент} \qquad & r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k) - \nabla f(x_k), x - x_{k+1} \rangle \\
& r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k), x - x_{k+1} \rangle - \langle \nabla f(x_k), x - x_{k+1} \rangle \\
& \langle \nabla f(x_k), x_{k+1} - x \rangle \leq r(x) - r(x_{k+1}) - \langle G_\alpha(x_k), x - x_{k+1} \rangle
\end{aligned}
$$

5. Учитывая полученную оценку, вернёмся к гладкости и выпуклости:

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad \Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r(x_{k+1})$$

$$\text{Поскольку } x_{k+1} - x_k = -\alpha G_\alpha(x_k) \qquad \Rightarrow \qquad \alpha G_\alpha(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r(x_{k+1})$$

$$G_\alpha(x_k) - \nabla f(x_k) \in \partial r(x_{k+1})$$

4. По определению субградиента выпуклой функции r для любой точки x:

$$r(x) \geq r(x_{k+1}) + \langle g, x - x_{k+1} \rangle, \quad g \in \partial r(x_{k+1})$$

$$\text{подставляем конкретный субградиент} \qquad r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k) - \nabla f(x_k), x - x_{k+1} \rangle$$

$$r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k), x - x_{k+1} \rangle - \langle \nabla f(x_k), x - x_{k+1} \rangle$$

$$\langle \nabla f(x_k), x_{k+1} - x \rangle \leq r(x) - r(x_{k+1}) - \langle G_\alpha(x_k), x - x_{k+1} \rangle$$

5. Учитывая полученную оценку, вернёмся к гладкости и выпуклости:

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

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$
\begin{aligned}
x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad &\Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r(x_{k+1}) \\
\text{Поскольку } x_{k+1} - x_k = -\alpha G_\alpha(x_k) \qquad &\Rightarrow \qquad \alpha G_\alpha(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r(x_{k+1}) \\
& \qquad G_\alpha(x_k) - \nabla f(x_k) \in \partial r(x_{k+1})
\end{aligned}
$$

4. По определению субградиента выпуклой функции r для любой точки x:

$$
\begin{aligned}
& r(x) \geq r(x_{k+1}) + \langle g, x - x_{k+1} \rangle, \quad g \in \partial r(x_{k+1}) \\
\text{подставляем конкретный субградиент} \qquad & r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k) - \nabla f(x_k), x - x_{k+1} \rangle \\
& r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k), x - x_{k+1} \rangle - \langle \nabla f(x_k), x - x_{k+1} \rangle \\
& \langle \nabla f(x_k), x_{k+1} - x \rangle \leq r(x) - r(x_{k+1}) - \langle G_\alpha(x_k), x - x_{k+1} \rangle
\end{aligned}
$$

5. Учитывая полученную оценку, вернёмся к гладкости и выпуклости:

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

$$
f(x_{k+1}) \leq f(x) + r(x) - r(x_{k+1}) - \langle G_\alpha(x_k), x - x_{k+1} \rangle + \frac{\alpha^2 L}{2} \|G_\alpha(x_k)\|_2^2
$$

Сходимость

3. Теперь используем свойство проксимального отображения, доказанное ранее:

$$x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right) \qquad \Leftrightarrow \qquad x_k - \alpha \nabla f(x_k) - x_{k+1} \in \partial \alpha r(x_{k+1})$$

$$\text{Поскольку } x_{k+1} - x_k = -\alpha G_\alpha(x_k) \qquad \Rightarrow \qquad \alpha G_\alpha(x_k) - \alpha \nabla f(x_k) \in \partial \alpha r(x_{k+1})$$

$$G_\alpha(x_k) - \nabla f(x_k) \in \partial r(x_{k+1})$$

4. По определению субградиента выпуклой функции r для любой точки x:

$$r(x) \geq r(x_{k+1}) + \langle g, x - x_{k+1} \rangle, \quad g \in \partial r(x_{k+1})$$

$$\text{подставляем конкретный субградиент} \qquad r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k) - \nabla f(x_k), x - x_{k+1} \rangle$$

$$r(x) \geq r(x_{k+1}) + \langle G_\alpha(x_k), x - x_{k+1} \rangle - \langle \nabla f(x_k), x - x_{k+1} \rangle$$

$$\langle \nabla f(x_k), x_{k+1} - x \rangle \leq r(x) - r(x_{k+1}) - \langle G_\alpha(x_k), x - x_{k+1} \rangle$$

5. Учитывая полученную оценку, вернёмся к гладкости и выпуклости:

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

$$f(x_{k+1}) \leq f(x) + r(x) - r(x_{k+1}) - \langle G_\alpha(x_k), x - x_{k+1} \rangle + \frac{\alpha^2 L}{2} \|G_\alpha(x_k)\|_2^2$$

$$f(x_{k+1}) + r(x_{k+1}) \leq f(x) + r(x) - \langle G_\alpha(x_k), x - x_k + \alpha G_\alpha(x_k) \rangle + \frac{\alpha^2 L}{2} \|G_\alpha(x_k)\|_2^2$$

Сходимость

6. Используя $\varphi(x) = f(x) + r(x)$, докажем чрезвычайно полезное неравенство, которое позволит продемонстрировать монотонное убывание итерации:

Сходимость

6. Используя $\varphi(x) = f(x) + r(x)$, докажем чрезвычайно полезное неравенство, которое позволит продемонстрировать монотонное убывание итерации:

$$\varphi(x_{k+1}) \leq \varphi(x) - \langle G_\alpha(x_k), x - x_k \rangle - \langle G_\alpha(x_k), \alpha G_\alpha(x_k) \rangle + \frac{\alpha^2 L}{2} \|G_\alpha(x_k)\|_2^2$$

Сходимость

6. Используя $\varphi(x) = f(x) + r(x)$, докажем чрезвычайно полезное неравенство, которое позволит продемонстрировать монотонное убывание итерации:

$$
\varphi(x_{k+1}) \le \varphi(x) - \langle G_\alpha(x_k), x - x_k \rangle - \langle G_\alpha(x_k), \alpha G_\alpha(x_k) \rangle + \frac{\alpha^2 L}{2} \|G_\alpha(x_k)\|_2^2
$$

$$
\varphi(x_{k+1}) \le \varphi(x) + \langle G_\alpha(x_k), x_k - x \rangle + \frac{\alpha}{2} \left(\alpha L - 2\right) \|G_\alpha(x_k)\|_2^2
$$

Сходимость

6. Используя $\varphi(x) = f(x) + r(x)$, докажем чрезвычайно полезное неравенство, которое позволит продемонстрировать монотонное убывание итерации:

$$\varphi(x_{k+1}) \leq \varphi(x) - \langle G_\alpha(x_k), x - x_k \rangle - \langle G_\alpha(x_k), \alpha G_\alpha(x_k) \rangle + \frac{\alpha^2 L}{2} \|G_\alpha(x_k)\|_2^2$$

$$\varphi(x_{k+1}) \leq \varphi(x) + \langle G_\alpha(x_k), x_k - x \rangle + \frac{\alpha}{2} (\alpha L - 2) \|G_\alpha(x_k)\|_2^2$$

$\alpha \leq \frac{1}{L} \Rightarrow \frac{\alpha}{2}(\alpha L - 2) \leq -\frac{\alpha}{2}$

Сходимость

6. Используя $\varphi(x) = f(x) + r(x)$, докажем чрезвычайно полезное неравенство, которое позволит продемонстрировать монотонное убывание итерации:

$$
\begin{aligned}
&\varphi(x_{k+1}) \le \varphi(x) - \langle G_\alpha(x_k), x - x_k \rangle - \langle G_\alpha(x_k), \alpha G_\alpha(x_k) \rangle + \frac{\alpha^2 L}{2} \|G_\alpha(x_k)\|_2^2 \\
&\varphi(x_{k+1}) \le \varphi(x) + \langle G_\alpha(x_k), x_k - x \rangle + \frac{\alpha}{2} \left(\alpha L - 2\right) \|G_\alpha(x_k)\|_2^2 \\
\alpha \le \tfrac{1}{L} \Rightarrow \tfrac{\alpha}{2}(\alpha L - 2) \le -\tfrac{\alpha}{2} \quad &\varphi(x_{k+1}) \le \varphi(x) + \langle G_\alpha(x_k), x_k - x \rangle - \frac{\alpha}{2} \|G_\alpha(x_k)\|_2^2
\end{aligned}
$$

Сходимость

6. Используя $\varphi(x) = f(x) + r(x)$, докажем чрезвычайно полезное неравенство, которое позволит продемонстрировать монотонное убывание итерации:

$$\varphi(x_{k+1}) \le \varphi(x) - \langle G_\alpha(x_k), x - x_k \rangle - \langle G_\alpha(x_k), \alpha G_\alpha(x_k) \rangle + \frac{\alpha^2 L}{2} \|G_\alpha(x_k)\|_2^2$$

$$\varphi(x_{k+1}) \le \varphi(x) + \langle G_\alpha(x_k), x_k - x \rangle + \frac{\alpha}{2} \left(\alpha L - 2\right) \|G_\alpha(x_k)\|_2^2$$

$$\alpha \le \frac{1}{L} \Rightarrow \frac{\alpha}{2}(\alpha L - 2) \le -\frac{\alpha}{2} \qquad \varphi(x_{k+1}) \le \varphi(x) + \langle G_\alpha(x_k), x_k - x \rangle - \frac{\alpha}{2} \|G_\alpha(x_k)\|_2^2$$

7. Теперь легко проверить, что при $x = x_k$ мы получаем монотонное убывание для проксимального градиентного алгоритма:

$$\varphi(x_{k+1}) \le \varphi(x_k) - \frac{\alpha}{2} \|G_\alpha(x_k)\|_2^2$$

Сходимость

8. При $x = x^*$:

Сходимость

8. При $x = x^*$:

$$\varphi(x_{k+1}) \leq \varphi(x^*) + \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2$$

Сходимость

8. При $x = x^*$:

$$\varphi(x_{k+1}) \leq \varphi(x^*) + \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2$$

$$\varphi(x_{k+1}) - \varphi(x^*) \leq \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2$$

Сходимость

8. При $x = x^*$:

$$
\begin{aligned}
\varphi(x_{k+1}) &\leq \varphi(x^*) + \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2 \\
\varphi(x_{k+1}) - \varphi(x^*) &\leq \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2 \\
&\leq \frac{1}{2\alpha}\left[2\langle \alpha G_\alpha(x_k), x_k - x^* \rangle - \|\alpha G_\alpha(x_k)\|_2^2\right]
\end{aligned}
$$

Сходимость

8. При $x = x^*$:

$$
\begin{aligned}
\varphi(x_{k+1}) &\leq \varphi(x^*) + \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2 \\
\varphi(x_{k+1}) - \varphi(x^*) &\leq \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2 \\
&\leq \frac{1}{2\alpha}\left[2\langle \alpha G_\alpha(x_k), x_k - x^* \rangle - \|\alpha G_\alpha(x_k)\|_2^2\right] \\
&\leq \frac{1}{2\alpha}\left[2\langle \alpha G_\alpha(x_k), x_k - x^* \rangle - \|\alpha G_\alpha(x_k)\|_2^2 - \|x_k - x^*\|_2^2 + \|x_k - x^*\|_2^2\right]
\end{aligned}
$$

Сходимость

8. При $x = x^*$:

$$
\begin{aligned}
\varphi(x_{k+1}) &\leq \varphi(x^*) + \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2 \\
\varphi(x_{k+1}) - \varphi(x^*) &\leq \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2 \\
&\leq \frac{1}{2\alpha}\left[2\langle \alpha G_\alpha(x_k), x_k - x^* \rangle - \|\alpha G_\alpha(x_k)\|_2^2\right] \\
&\leq \frac{1}{2\alpha}\left[2\langle \alpha G_\alpha(x_k), x_k - x^* \rangle - \|\alpha G_\alpha(x_k)\|_2^2 - \|x_k - x^*\|_2^2 + \|x_k - x^*\|_2^2\right] \\
&\leq \frac{1}{2\alpha}\left[-\|x_k - x^* - \alpha G_\alpha(x_k)\|_2^2 + \|x_k - x^*\|_2^2\right]
\end{aligned}
$$

Сходимость

8. При $x = x^*$:

$$
\begin{aligned}
\varphi(x_{k+1}) &\le \varphi(x^*) + \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2 \\
\varphi(x_{k+1}) - \varphi(x^*) &\le \langle G_\alpha(x_k), x_k - x^* \rangle - \frac{\alpha}{2}\|G_\alpha(x_k)\|_2^2 \\
&\le \frac{1}{2\alpha}\left[2\langle \alpha G_\alpha(x_k), x_k - x^* \rangle - \|\alpha G_\alpha(x_k)\|_2^2\right] \\
&\le \frac{1}{2\alpha}\left[2\langle \alpha G_\alpha(x_k), x_k - x^* \rangle - \|\alpha G_\alpha(x_k)\|_2^2 - \|x_k - x^*\|_2^2 + \|x_k - x^*\|_2^2\right] \\
&\le \frac{1}{2\alpha}\left[-\|x_k - x^* - \alpha G_\alpha(x_k)\|_2^2 + \|x_k - x^*\|_2^2\right] \\
&\le \frac{1}{2\alpha}\left[\|x_k - x^*\|_2^2 - \|x_{k+1} - x^*\|_2^2\right]
\end{aligned}
$$

Сходимость

9. Запишем неравенство выше для всех итераций $i = 0, \ldots, k-1$ и просуммируем:

Что даёт стандартную оценку $\frac{L\|x_0 - x^*\|_2^2}{2k}$ при $\alpha = \frac{1}{L}$, т.е. ту же скорость $\mathcal{O}\left(\frac{1}{k}\right)$, что и у градиентного спуска для гладких выпуклых задач!

Сходимость

9. Запишем неравенство выше для всех итераций $i = 0, \ldots, k-1$ и просуммируем:

$$\sum_{i=0}^{k-1} [\varphi(x_{i+1}) - \varphi(x^*)] \leq \frac{1}{2\alpha} \left[\|x_0 - x^*\|_2^2 - \|x_k - x^*\|_2^2\right]$$

Что даёт стандартную оценку $\frac{L\|x_0 - x^*\|_2^2}{2k}$ при $\alpha = \frac{1}{L}$, т.е. ту же скорость $\mathcal{O}\left(\frac{1}{k}\right)$, что и у градиентного спуска для гладких выпуклых задач!

Сходимость

9. Запишем неравенство выше для всех итераций $i = 0, \dots, k-1$ и просуммируем:

$$\sum_{i=0}^{k-1} [\varphi(x_{i+1}) - \varphi(x^*)] \leq \frac{1}{2\alpha} \left[\|x_0 - x^*\|_2^2 - \|x_k - x^*\|_2^2\right]$$

$$\leq \frac{1}{2\alpha} \|x_0 - x^*\|_2^2$$

Что даёт стандартную оценку $\frac{L\|x_0 - x^*\|_2^2}{2k}$ при $\alpha = \frac{1}{L}$, т.е. ту же скорость $\mathcal{O}\left(\frac{1}{k}\right)$, что и у градиентного спуска для гладких выпуклых задач!

Сходимость

9. Запишем неравенство выше для всех итераций $i = 0, \ldots, k-1$ и просуммируем:

$$\sum_{i=0}^{k-1} [\varphi(x_{i+1}) - \varphi(x^*)] \leq \frac{1}{2\alpha} \left[\|x_0 - x^*\|_2^2 - \|x_k - x^*\|_2^2\right] \leq \frac{1}{2\alpha} \|x_0 - x^*\|_2^2$$

10. Поскольку $\varphi(x_k)$ является убывающей последовательностью, следует:

Что даёт стандартную оценку $\frac{L\|x_0 - x^*\|_2^2}{2k}$ при $\alpha = \frac{1}{L}$, т.е. ту же скорость $\mathcal{O}\left(\frac{1}{k}\right)$, что и у градиентного спуска для гладких выпуклых задач!

Сходимость

9. Запишем неравенство выше для всех итераций $i = 0, \dots, k-1$ и просуммируем:

$$
\begin{aligned}
\sum_{i=0}^{k-1} [\varphi(x_{i+1}) - \varphi(x^*)] &\leq \frac{1}{2\alpha} \left[\|x_0 - x^*\|_2^2 - \|x_k - x^*\|_2^2 \right] \\
&\leq \frac{1}{2\alpha} \|x_0 - x^*\|_2^2
\end{aligned}
$$

10. Поскольку $\varphi(x_k)$ является убывающей последовательностью, следует:

$$
\sum_{i=0}^{k-1} \varphi(x_k) = k\varphi(x_k) \leq \sum_{i=0}^{k-1} \varphi(x_{i+1})
$$

Что даёт стандартную оценку $\frac{L\|x_0 - x^*\|_2^2}{2k}$ при $\alpha = \frac{1}{L}$, т.е. ту же скорость $\mathcal{O}\left(\frac{1}{k}\right)$, что и у градиентного спуска для гладких выпуклых задач!

Сходимость

9. Запишем неравенство выше для всех итераций $i = 0, \dots, k-1$ и просуммируем:

$$\sum_{i=0}^{k-1} [\varphi(x_{i+1}) - \varphi(x^*)] \leq \frac{1}{2\alpha} \left[\|x_0 - x^*\|_2^2 - \|x_k - x^*\|_2^2\right]$$

$$\leq \frac{1}{2\alpha} \|x_0 - x^*\|_2^2$$

10. Поскольку $\varphi(x_k)$ является убывающей последовательностью, следует:

$$\sum_{i=0}^{k-1} \varphi(x_k) = k\varphi(x_k) \leq \sum_{i=0}^{k-1} \varphi(x_{i+1})$$

$$\varphi(x_k) \leq \frac{1}{k} \sum_{i=0}^{k-1} \varphi(x_{i+1})$$

Что даёт стандартную оценку $\frac{L\|x_0 - x^*\|_2^2}{2k}$ при $\alpha = \frac{1}{L}$, т.е. ту же скорость $\mathcal{O}\left(\frac{1}{k}\right)$, что и у градиентного спуска для гладких выпуклых задач!

Сходимость

9. Запишем неравенство выше для всех итераций $i = 0, \dots, k-1$ и просуммируем:

$$\sum_{i=0}^{k-1} [\varphi(x_{i+1}) - \varphi(x^*)] \leq \frac{1}{2\alpha} \left[\|x_0 - x^*\|_2^2 - \|x_k - x^*\|_2^2\right]$$

$$\leq \frac{1}{2\alpha} \|x_0 - x^*\|_2^2$$

10. Поскольку $\varphi(x_k)$ является убывающей последовательностью, следует:

$$\sum_{i=0}^{k-1} \varphi(x_k) = k\varphi(x_k) \leq \sum_{i=0}^{k-1} \varphi(x_{i+1})$$

$$\varphi(x_k) \leq \frac{1}{k} \sum_{i=0}^{k-1} \varphi(x_{i+1})$$

$$\varphi(x_k) - \varphi(x^*) \leq \frac{1}{k} \sum_{i=0}^{k-1} [\varphi(x_{i+1}) - \varphi(x^*)] \leq \frac{\|x_0 - x^*\|_2^2}{2\alpha k}$$

Что даёт стандартную оценку $\frac{L\|x_0 - x^*\|_2^2}{2k}$ при $\alpha = \frac{1}{L}$, т.е. ту же скорость $\mathcal{O}\left(\frac{1}{k}\right)$, что и у градиентного спуска для гладких выпуклых задач!

Проксимальный градиентный метод. Сильно выпуклый случай

Сходимость

ℹ Theorem

Рассмотрим проксимальный градиентный метод

$$
x_{k+1} = \text{prox}_{\alpha r}\left(x_k - \alpha \nabla f(x_k)\right)
$$

Для критерия $\varphi(x) = f(x) + r(x)$ предполагаем:

- f является μ-сильно выпуклой, дифференцируемой, $\text{dom}(f) = \mathbb{R}^n$, и ∇f липшицев с константой $L > 0$.
- r выпукла, и $\text{prox}_{\alpha r}(x_k) = \text{arg}\min\limits_{x \in \mathbb{R}^n} \left[\alpha r(x) + \frac{1}{2}\|x - x_k\|_2^2\right]$ может быть вычислен.

Проксимальный градиентный спуск с фиксированным шагом $\alpha \leq 1/L$ удовлетворяет

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

Это в точности скорость сходимости градиентного спуска. Заметим, что исходная задача даже негладкая!

Сходимость

Доказательство

1. Рассмотрим расстояние до решения и используем лемму о стационарной точке:

Сходимость

Доказательство

1. Рассмотрим расстояние до решения и используем лемму о стационарной точке:

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

Сходимость

Доказательство

1. Рассмотрим расстояние до решения и используем лемму о стационарной точке:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
\text{лемма о стац. точке} &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - \text{prox}_{\alpha r}(x^* - \alpha \nabla f(x^*))\|_2^2
\end{aligned}
$$

Сходимость

Доказательство

1. Рассмотрим расстояние до решения и используем лемму о стационарной точке:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
{\scriptstyle \text{лемма о стац. точке}} &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - \text{prox}_{\alpha r}(x^* - \alpha \nabla f(x^*))\|_2^2 \\
{\scriptstyle \text{нерастяжимость}} &\leq \|x_k - \alpha \nabla f(x_k) - x^* + \alpha \nabla f(x^*)\|_2^2
\end{aligned}
$$

Сходимость

Доказательство

1. Рассмотрим расстояние до решения и используем лемму о стационарной точке:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
\text{лемма о стац. точке} &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - \text{prox}_{\alpha r}(x^* - \alpha \nabla f(x^*))\|_2^2 \\
\text{нерастяжимость} &\leq \|x_k - \alpha \nabla f(x_k) - x^* + \alpha \nabla f(x^*)\|_2^2 \\
&= \|x_k - x^*\|^2 - 2\alpha \langle \nabla f(x_k) - \nabla f(x^*), x_k - x^* \rangle + \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

Сходимость

Доказательство

1. Рассмотрим расстояние до решения и используем лемму о стационарной точке:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
\text{лемма о стац. точке} &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - \text{prox}_{\alpha r}(x^* - \alpha \nabla f(x^*))\|_2^2 \\
\text{нерастяжимость} &\leq \|x_k - \alpha \nabla f(x_k) - x^* + \alpha \nabla f(x^*)\|_2^2 \\
&= \|x_k - x^*\|^2 - 2\alpha \langle \nabla f(x_k) - \nabla f(x^*), x_k - x^* \rangle + \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

2. Используем лемму о гладкости из инструментов сходимости и сильную выпуклость:

Сходимость

Доказательство

1. Рассмотрим расстояние до решения и используем лемму о стационарной точке:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
\text{лемма о стац. точке} &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - \text{prox}_{\alpha r}(x^* - \alpha \nabla f(x^*))\|_2^2 \\
\text{нерастяжимость} &\leq \|x_k - \alpha \nabla f(x_k) - x^* + \alpha \nabla f(x^*)\|_2^2 \\
&= \|x_k - x^*\|^2 - 2\alpha \langle \nabla f(x_k) - \nabla f(x^*), x_k - x^* \rangle + \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

2. Используем лемму о гладкости из инструментов сходимости и сильную выпуклость:

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

Сходимость

Доказательство

1. Рассмотрим расстояние до решения и используем лемму о стационарной точке:

$$
\begin{aligned}
\|x_{k+1} - x^*\|_2^2 &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - x^*\|_2^2 \\
{\scriptstyle \text{лемма о стац. точке}} &= \|\text{prox}_{\alpha r}(x_k - \alpha \nabla f(x_k)) - \text{prox}_{\alpha r}(x^* - \alpha \nabla f(x^*))\|_2^2 \\
{\scriptstyle \text{нерастяжимость}} &\leq \|x_k - \alpha \nabla f(x_k) - x^* + \alpha \nabla f(x^*)\|_2^2 \\
&= \|x_k - x^*\|^2 - 2\alpha \langle \nabla f(x_k) - \nabla f(x^*), x_k - x^* \rangle + \alpha^2 \|\nabla f(x_k) - \nabla f(x^*)\|_2^2
\end{aligned}
$$

2. Используем лемму о гладкости из инструментов сходимости и сильную выпуклость:

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

Сходимость

3. Подставим:

Сходимость

3. Подставим:

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

Сходимость

3. Подставим:

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

Сходимость

3. Подставим:

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

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

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

   что в точности является линейной сходимостью метода со скоростью до $1 - \frac{\mu}{L}$.

Ускоренный проксимальный метод

Ускоренный проксимальный метод

Пусть $x_0 = y_0 \in \mathrm{dom}(r)$. Для $k \geq 1$:

$$
\begin{aligned}
x_k &= \mathrm{prox}_{\alpha_k r}(y_{k-1} - \alpha_k \nabla f(y_{k-1})) \\
y_k &= x_k + \frac{k-1}{k+2}(x_k - x_{k-1})
\end{aligned}
$$

Достигает

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

Ускоренный проксимальный метод

FISTA

Ускоренный проксимальный метод

Пусть $x_0 = y_0 \in \text{dom}(r)$. Для $k \geq 1$:

$$x_k = \text{prox}_{\alpha_k r}(y_{k-1} - \alpha_k \nabla f(y_{k-1}))$$
$$y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1})$$

$\frac{1}{\sqrt{k}} \to \frac{1}{k^2}$

Достигает

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

Схема предложена: Нестеров (1983, 2004); также Бек, Тебулле (2009). Упрощённый анализ: Цэн (2008).

- Использует дополнительную «память» для интерполяции

Ускоренный проксимальный метод

Ускоренный проксимальный метод

Пусть $x_0 = y_0 \in \text{dom}(r)$. Для $k \geq 1$:

$$
x_k = \text{prox}_{\alpha_k r}(y_{k-1} - \alpha_k \nabla f(y_{k-1}))
$$
$$
y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1})
$$

Достигает

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

Схема предложена: Нестеров (1983, 2004); также Бек, Тебулле (2009). Упрощённый анализ: Цэн (2008).

- Использует дополнительную «память» для интерполяции
- Та же вычислительная стоимость, что и у обычного проксимального градиентного метода

Ускоренный проксимальный метод

Ускоренный проксимальный метод

Пусть $x_0 = y_0 \in \mathrm{dom}(r)$. Для $k \geq 1$:

$$x_k = \mathrm{prox}_{\alpha_k r}(y_{k-1} - \alpha_k \nabla f(y_{k-1}))$$
$$y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1})$$

Достигает

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

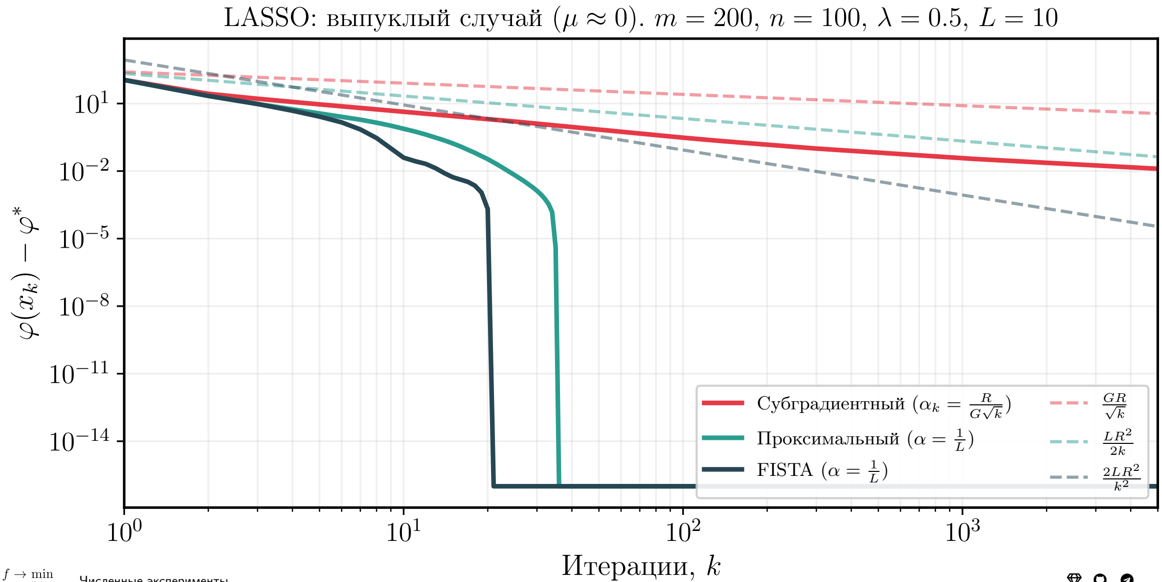
Схема предложена: Нестеров (1983, 2004); также Бек, Тебулле (2009). Упрощённый анализ: Цэн (2008).

- Использует дополнительную «память» для интерполяции
- Та же вычислительная стоимость, что и у обычного проксимального градиентного метода
- Скорость сходимости теоретически оптимальна

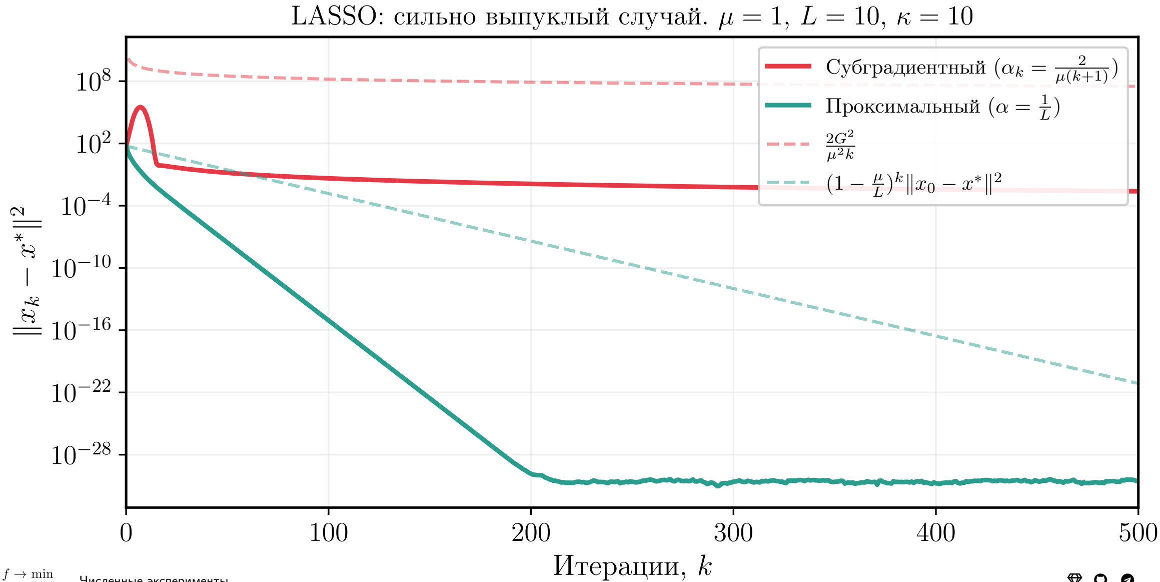
Численные эксперименты

LASSO: выпуклый случай — теоретические гарантии

$\frac{1}{2}\|Ax-B\|_2^2 + \lambda\|x\|_1$



LASSO: сильно выпуклый случай — теоретические гарантии



LASSO: разреженность решения

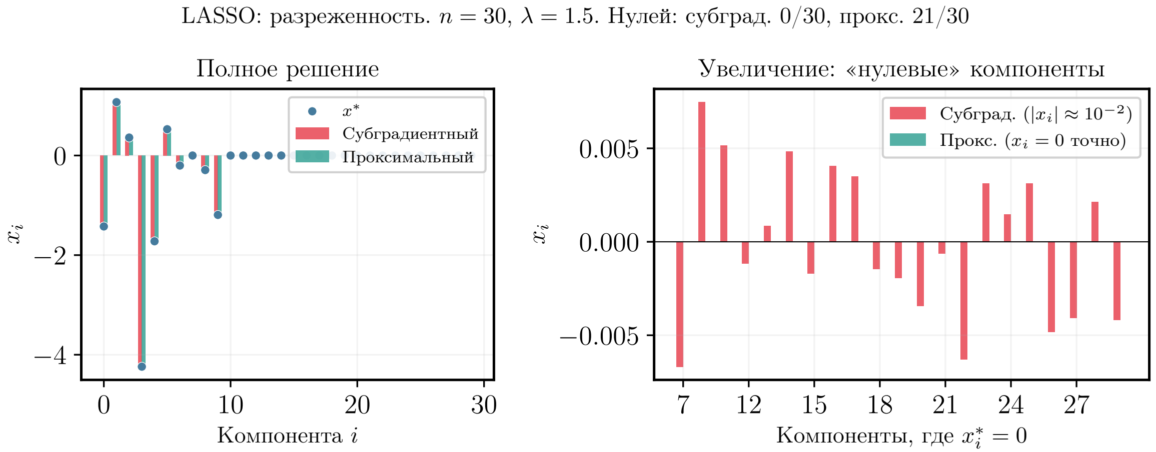
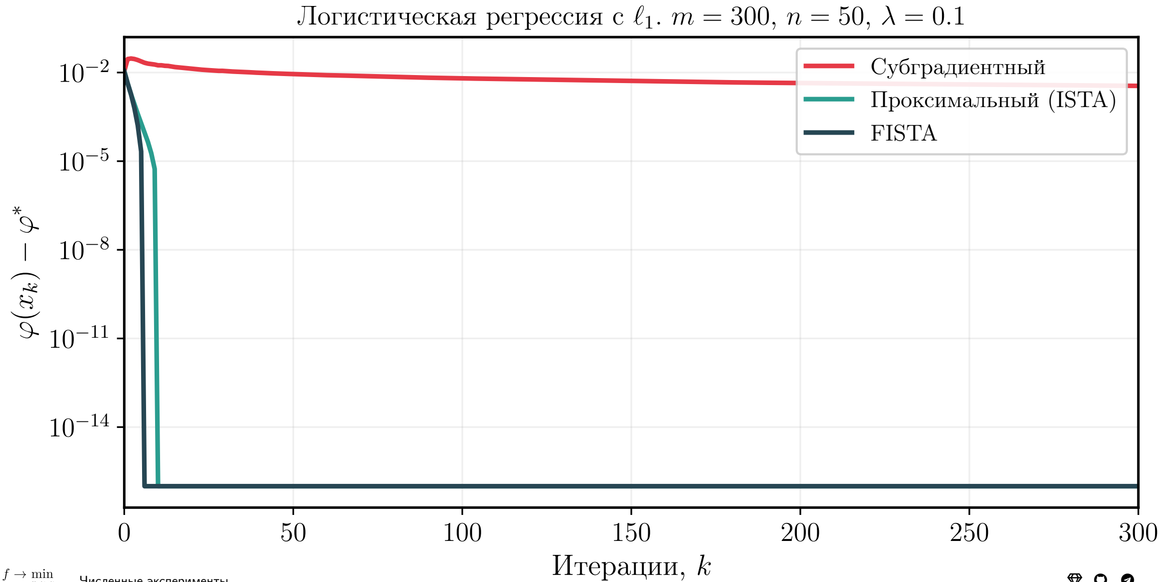


Figure 7: Проксимальный метод порождает точные нули (21/30 компонент), в точности воспроизводя структуру оптимального решения x^*. Субградиентный метод даёт малые, но ненулевые значения (0 точных нулей) — soft-thresholding принципиально отличается от вычитания субградиента.

Логистическая регрессия с ℓ_1-регуляризацией



Восстановление матриц: ядерная норма

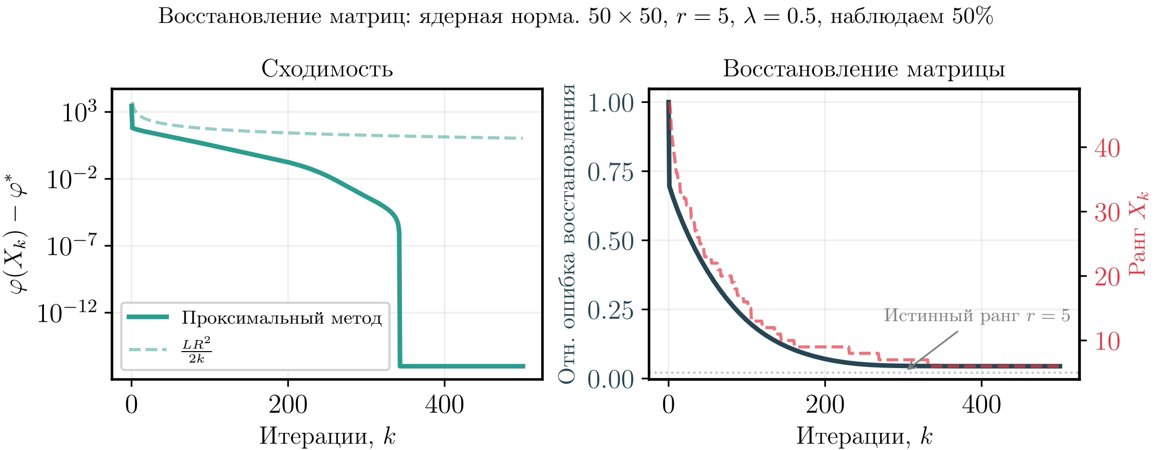


Figure 9: Проксимальный метод с SVD мягким порогом восстанавливает матрицу ранга $r = 5$ из 50% наблюдений. Ранг итерата быстро стабилизируется около истинного. Проксимальный оператор ядерной нормы — мягкое пороговое отсечение сингулярных значений.

Итоги

- Используя структуру задачи, можно преодолеть нижние оценки для неструктурированной задачи.

Итоги

- Используя структуру задачи, можно преодолеть нижние оценки для неструктурированной задачи.
- Проксимальный градиентный метод для композитной задачи с L-гладкой выпуклой функцией f и выпуклой проксимально-дружественной функцией r имеет ту же скорость сходимости, что и градиентный спуск для функции f. Свойства гладкости/негладкости r не влияют на сходимость.

Итоги

- Используя структуру задачи, можно преодолеть нижние оценки для неструктурированной задачи.
- Проксимальный градиентный метод для композитной задачи с L-гладкой выпуклой функцией f и выпуклой проксимально-дружественной функцией r имеет ту же скорость сходимости, что и градиентный спуск для функции f. Свойства гладкости/негладкости r не влияют на сходимость.
- Кажется, что положив $f = 0$, любую негладкую задачу можно решить с помощью такого метода. Вопрос: верно ли это?

Итоги

- Используя структуру задачи, можно преодолеть нижние оценки для неструктурированной задачи.
- Проксимальный градиентный метод для композитной задачи с L-гладкой выпуклой функцией f и выпуклой проксимально-дружественной функцией r имеет ту же скорость сходимости, что и градиентный спуск для функции f. Свойства гладкости/негладкости r не влияют на сходимость.
- Кажется, что положив $f = 0$, любую негладкую задачу можно решить с помощью такого метода. Вопрос: верно ли это?

Итоги

- Используя структуру задачи, можно преодолеть нижние оценки для неструктурированной задачи.
- Проксимальный градиентный метод для композитной задачи с L-гладкой выпуклой функцией f и выпуклой проксимально-дружественной функцией r имеет ту же скорость сходимости, что и градиентный спуск для функции f. Свойства гладкости/негладкости r не влияют на сходимость.
- Кажется, что положив $f = 0$, любую негладкую задачу можно решить с помощью такого метода. Вопрос: верно ли это?

  Если допустить неточное вычисление проксимального оператора (численно), то действительно любую негладкую задачу оптимизации можно решить. Но это не лучше с точки зрения теории, чем решение задачи субградиентным спуском, поскольку для решения проксимальной подзадачи используется некоторый вспомогательный метод (например, тот же субградиентный спуск).
- Проксимальный метод является общей современной основой для многих численных методов. Дальнейшее развитие включает ускоренные, стохастические, прямо-двойственные модификации и т.д.

Итоги

- Используя структуру задачи, можно преодолеть нижние оценки для неструктурированной задачи.
- Проксимальный градиентный метод для композитной задачи с L-гладкой выпуклой функцией f и выпуклой проксимально-дружественной функцией r имеет ту же скорость сходимости, что и градиентный спуск для функции f. Свойства гладкости/негладкости r не влияют на сходимость.
- Кажется, что положив $f = 0$, любую негладкую задачу можно решить с помощью такого метода. Вопрос: верно ли это?

  Если допустить неточное вычисление проксимального оператора (численно), то действительно любую негладкую задачу оптимизации можно решить. Но это не лучше с точки зрения теории, чем решение задачи субградиентным спуском, поскольку для решения проксимальной подзадачи используется некоторый вспомогательный метод (например, тот же субградиентный спуск).
- Проксимальный метод является общей современной основой для многих численных методов. Дальнейшее развитие включает ускоренные, стохастические, прямо-двойственные модификации и т.д.
- Дополнительная литература: Proximal operator splitting, Douglas-Rachford splitting, Best approximation problem, Three operator splitting.