ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

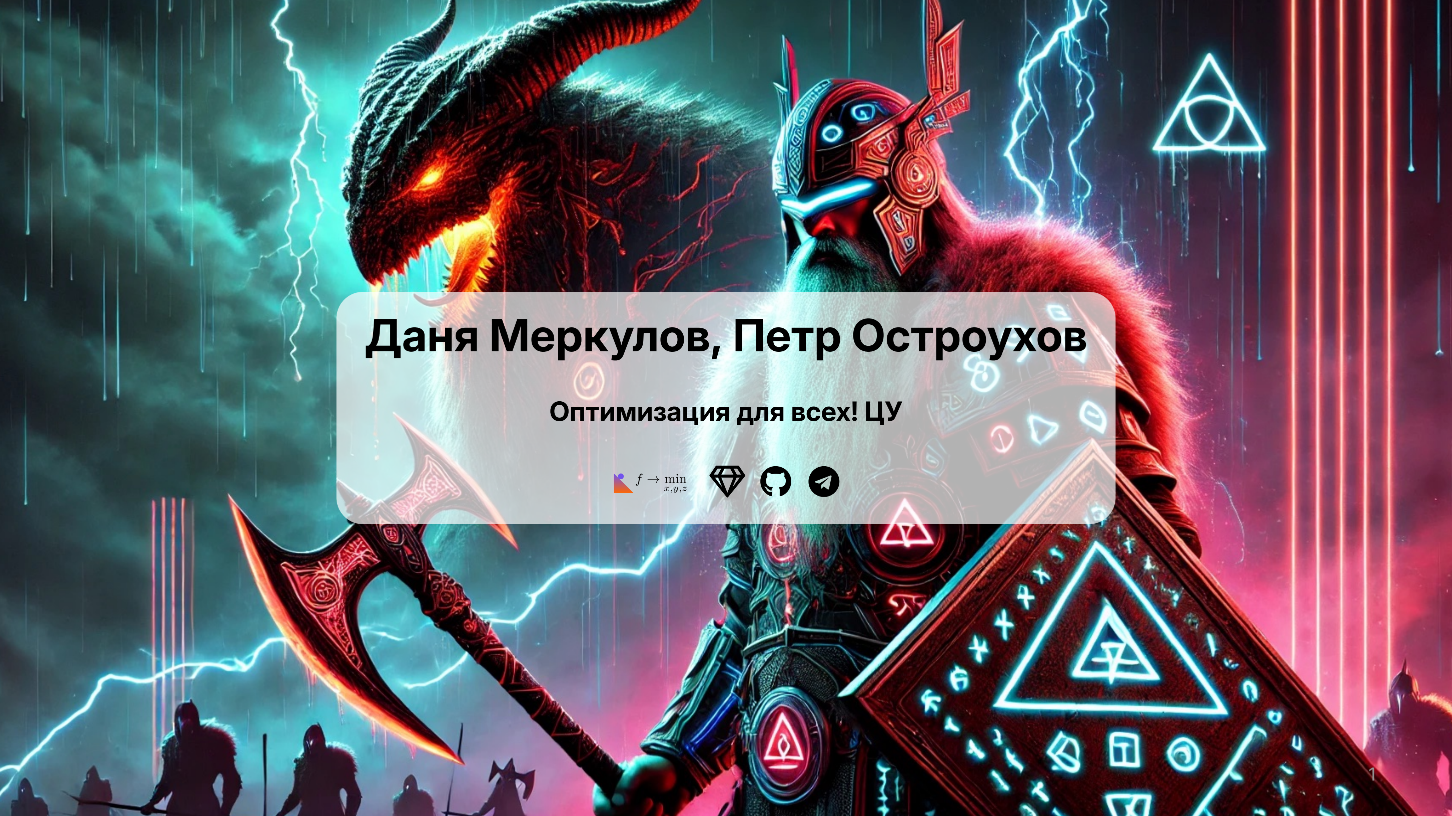
Градиентный спуск

МЕТОДЫ ВЫПУКЛОЙ ОПТИМИЗАЦИИ

НЕДЕЛЯ 7

Даня Меркулов

Пётр Остроухов



Даня Меркулов, Петр Остроухов

Оптимизация для всех! ЦУ

Повторение

Виды выпуклости

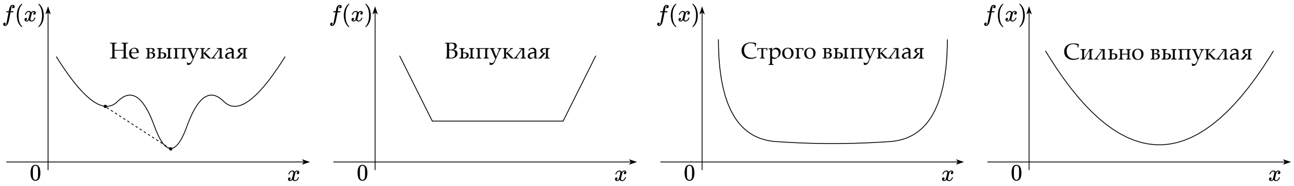


Рисунок 1. Примеры выпуклых функций

Гладкость

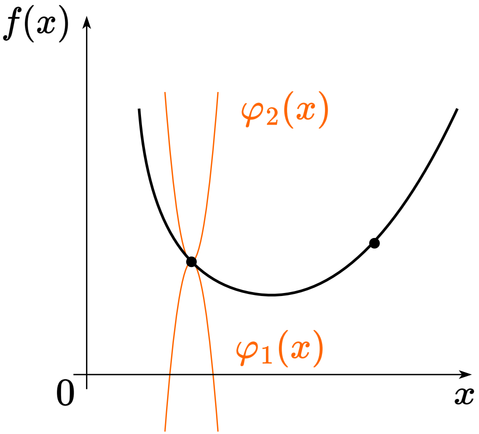


Рисунок 2. Иллюстрация Липшицевых парабол, между которыми зажата гладкая функция. Чаще нас интересует мажорирующая из них.

Определение: Будем говорить, что функция $f : \mathbb{R}^n \to \mathbb{R}$ является L -гладкой, если $\forall x, y \in \mathbb{R}^n$ выполнено:

$$\|\nabla f(y) - \nabla f(x)\| \leqslant L\|y - x\|$$

Гладкость

Определение: Будем говорить, что функция $f : \mathbb{R}^n \to \mathbb{R}$ является L -гладкой, если $\forall x, y \in \mathbb{R}^n$ выполнено:

$$\|\nabla f(y) - \nabla f(x)\| \leqslant L\|y - x\|$$

Обратим внимание, что значение константы гладкости (Липшицевости градиента) зависит от выбора нормы.
Если $f : \mathbb{R}^n \to \mathbb{R}$ - непрерывно дифференцируема и градиент Липшицев с константой L, то $\forall x, y \in \mathbb{R}^n$:

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

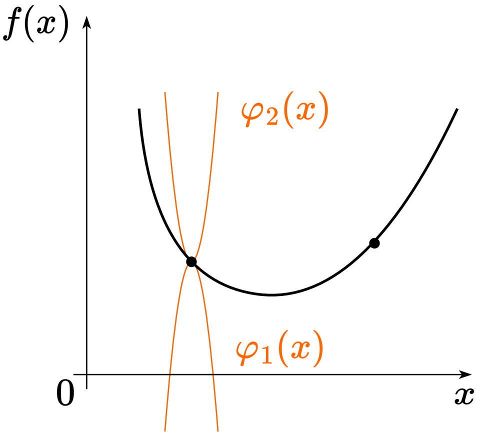


Рисунок 2. Иллюстрация Липшицевых парабол, между которыми зажата гладкая функция. Чаще нас интересует мажорирующая из них.

Гладкость

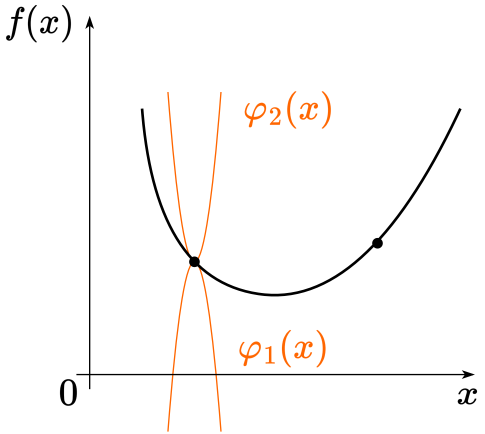


Рисунок 2. Иллюстрация Липшицевых парабол, между которыми зажата гладкая функция. Чаще нас интересует мажорирующая из них.

Определение: Будем говорить, что функция $f : \mathbb{R}^n \to \mathbb{R}$ является L -гладкой, если $\forall x, y \in \mathbb{R}^n$ выполнено:

$$\|\nabla f(y) - \nabla f(x)\| \leqslant L\|y - x\|$$

Обратим внимание, что значение константы гладкости (Липшицевости градиента) зависит от выбора нормы.
Если $f : \mathbb{R}^n \to \mathbb{R}$ - непрерывно дифференцируема и градиент Липшицев с константой L, то $\forall x, y \in \mathbb{R}^n$:

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

Если зафиксируем $x_0 \in \mathbb{R}^n$, то:

$$\varphi_1(x) = f(x_0) + \langle f(x_0), x - x_0 \rangle - \frac{L}{2}\|x - x_0\|^2$$

$$\varphi_2(x) = f(x_0) + \langle f(x_0), x - x_0 \rangle + \frac{L}{2}\|x - x_0\|^2$$

Гладкость

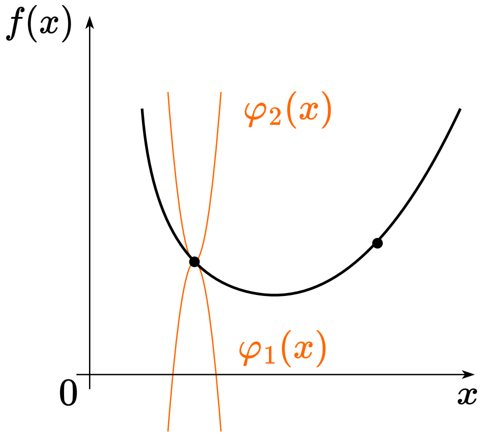


Рисунок 2. Иллюстрация Липшицевых парабол, между которыми зажата гладкая функция. Чаще нас интересует мажорирующая из них.

Определение: Будем говорить, что функция $f : \mathbb{R}^n \to \mathbb{R}$ является L -гладкой, если $\forall x, y \in \mathbb{R}^n$ выполнено:

$$\|\nabla f(y) - \nabla f(x)\| \leqslant L\|y - x\|$$

Обратим внимание, что значение константы гладкости (Липшицевости градиента) зависит от выбора нормы.
Если $f : \mathbb{R}^n \to \mathbb{R}$ - непрерывно дифференцируема и градиент Липшицев с константой L, то $\forall x, y \in \mathbb{R}^n$:

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

Если зафиксируем $x_0 \in \mathbb{R}^n$, то:

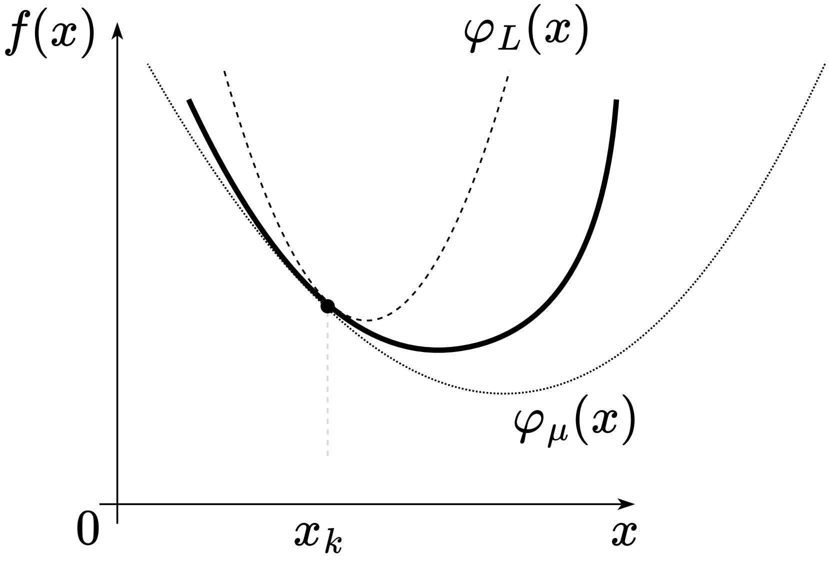
$$\varphi_1(x) = f(x_0) + \langle f(x_0), x - x_0\rangle - \frac{L}{2}\|x - x_0\|^2$$

$$\varphi_2(x) = f(x_0) + \langle f(x_0), x - x_0\rangle + \frac{L}{2}\|x - x_0\|^2$$

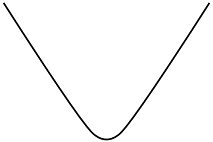
Это две параболы, и для них верно, что

$$\varphi_1(x) \leqslant f(x) \leqslant \varphi_2(x) \; \forall x$$

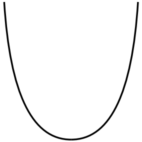
Гладкость и сильная выпуклость



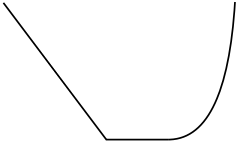
Гладкость и сильная выпуклость



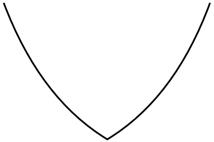
Гладкая
Выпуклая



Гладкая
μ - сильно выпуклая



Негладкая
Выпуклая



Негладкая
μ - сильно выпуклая

Градиентный спуск

Направление локального наискорейшего спуска

Рассмотрим линейное приближение
дифференцируемой функции f вдоль направления
h, где $\|h\|_2 = 1$:

Направление локального наискорейшего спуска

Рассмотрим линейное приближение дифференцируемой функции f вдоль направления h, где $\|h\|_2 = 1$:

$$f(x + \alpha h) = f(x) + \alpha \langle \nabla f(x), h \rangle + o(\alpha)$$

Направление локального наискорейшего спуска

Рассмотрим линейное приближение
дифференцируемой функции f вдоль направления
h, где $\|h\|_2 = 1$:

$$f(x + \alpha h) = f(x) + \alpha\langle \nabla f(x), h\rangle + o(\alpha)$$

Хотим, чтобы h было направлением убывания:

$$f(x + \alpha h) - f(x) < 0$$

$$\alpha\langle \nabla f(x), h\rangle + o(\alpha) < 0$$

Направление локального наискорейшего спуска

Рассмотрим линейное приближение дифференцируемой функции f вдоль направления h, где $\|h\|_2 = 1$:

$$f(x + \alpha h) = f(x) + \alpha \langle \nabla f(x), h \rangle + o(\alpha)$$

Хотим, чтобы h было направлением убывания:

$$f(x + \alpha h) - f(x) < 0$$

$$\alpha \langle \nabla f(x), h \rangle + o(\alpha) < 0$$

Переходя к пределу при $\alpha \to 0$:

$$\langle \nabla f(x), h \rangle < 0$$

Направление локального наискорейшего спуска

Рассмотрим линейное приближение дифференцируемой функции f вдоль направления h, где $\|h\|_2 = 1$:

$$f(x + \alpha h) = f(x) + \alpha \langle \nabla f(x), h \rangle + o(\alpha)$$

Хотим, чтобы h было направлением убывания:

$$f(x + \alpha h) - f(x) < 0$$

$$\alpha \langle \nabla f(x), h \rangle + o(\alpha) < 0$$

Переходя к пределу при $\alpha \to 0$:

$$\langle \nabla f(x), h \rangle < 0$$

Более того, мы хотим, чтобы разница $f(x) - f(x + \alpha h)$ была максимальна:

$$h = \arg\max_h \left(-\langle \nabla f(x), h \rangle\right) = \arg\min_h \langle \nabla f(x), h \rangle.$$

Также из неравенства Коши–Буняковского получаем:

$$|\langle \nabla f(x), h \rangle| \leq \|\nabla f(x)\|_2 \|h\|_2$$
$$\langle \nabla f(x), h \rangle \geq -\|\nabla f(x)\|_2 \|h\|_2 = -\|\nabla f(x)\|_2$$

Направление локального наискорейшего спуска

Рассмотрим линейное приближение дифференцируемой функции f вдоль направления h, где $\|h\|_2 = 1$:

$$f(x + \alpha h) = f(x) + \alpha\langle \nabla f(x), h\rangle + o(\alpha)$$

Хотим, чтобы h было направлением убывания:

$$f(x + \alpha h) - f(x) < 0$$

$$\alpha\langle \nabla f(x), h\rangle + o(\alpha) < 0$$

Переходя к пределу при $\alpha \to 0$:

$$\langle \nabla f(x), h\rangle < 0$$

Более того, мы хотим, чтобы разница $f(x) - f(x + \alpha h)$ была максимальна:

$$h = \arg\max_h \left(-\langle \nabla f(x), h\rangle\right) = \arg\min_h \langle \nabla f(x), h\rangle.$$

Также из неравенства Коши–Буняковского получаем:

$$|\langle \nabla f(x), h\rangle| \le \|\nabla f(x)\|_2 \|h\|_2$$
$$\langle \nabla f(x), h\rangle \ge -\|\nabla f(x)\|_2 \|h\|_2 = -\|\nabla f(x)\|_2$$

Таким образом, направление антиградиента

$$h = \arg\min_h \langle \nabla f(x), h\rangle = -\frac{\nabla f(x)}{\|\nabla f(x)\|_2}$$

представляет собой направление **наискорейшего локального** убывания функции f.

Направление локального наискорейшего спуска

Рассмотрим линейное приближение дифференцируемой функции f вдоль направления h, где $\|h\|_2 = 1$:

$$f(x + \alpha h) = f(x) + \alpha \langle \nabla f(x), h \rangle + o(\alpha)$$

Хотим, чтобы h было направлением убывания:

$$f(x + \alpha h) - f(x) < 0$$

$$\alpha \langle \nabla f(x), h \rangle + o(\alpha) < 0$$

Переходя к пределу при $\alpha \to 0$:

$$\langle \nabla f(x), h \rangle < 0$$

Более того, мы хотим, чтобы разница $f(x) - f(x + \alpha h)$ была максимальна:

$$h = \arg\max_h \left(-\langle \nabla f(x), h \rangle \right) = \arg\min_h \langle \nabla f(x), h \rangle.$$

Также из неравенства Коши–Буняковского получаем:

$$|\langle \nabla f(x), h \rangle| \leq \|\nabla f(x)\|_2 \|h\|_2$$
$$\langle \nabla f(x), h \rangle \geq -\|\nabla f(x)\|_2 \|h\|_2 = -\|\nabla f(x)\|_2$$

Таким образом, направление антиградиента

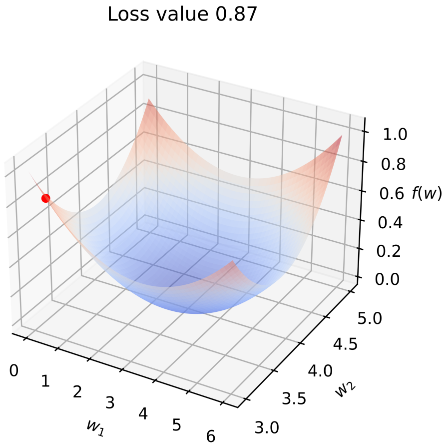
$$h = \arg\min_h \langle \nabla f(x), h \rangle = -\frac{\nabla f(x)}{\|\nabla f(x)\|_2}$$

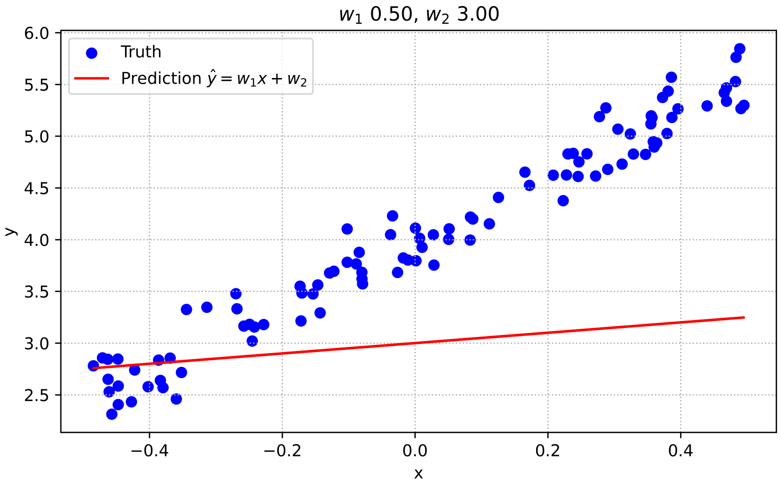
представляет собой направление **наискорейшего локального** убывания функции f.
Итерация метода имеет вид:

$$x^{k+1} = x^k - \alpha \nabla f(x^k)$$

Сходимость алгоритма градиентного спуска

Код для построения анимации ниже. Сходимость существенно зависит от выбора шага α:





Точный линейный поиск (метод наискорейшего спуска)

$$\alpha_k = \arg\min_{\alpha \in \mathbb{R}^+} f(x^k - \alpha \nabla f(x^k))$$

Подход скорее теоретический, чем практический: он удобен для анализа сходимости, но точный линейный поиск часто затруднён, если вычисление функции занимает слишком много времени или стоит слишком дорого.
Интересное теоретическое свойство этого метода заключается в том, что градиенты на соседних итерациях ортогональны. Условие оптимальности по α_k даёт

$$\frac{d}{d\alpha} f(x^k - \alpha \nabla f(x^k))\Big|_{\alpha = \alpha_k} = 0.$$

Точный линейный поиск (метод наискорейшего спуска)

$$
\alpha_k = \arg\min_{\alpha \in \mathbb{R}^+} f(x^k - \alpha \nabla f(x^k))
$$

Подход скорее теоретический, чем практический: он удобен для анализа сходимости, но точный линейный поиск часто затруднён, если вычисление функции занимает слишком много времени или стоит слишком дорого.
Интересное теоретическое свойство этого метода заключается в том, что градиенты на соседних итерациях ортогональны. Условие оптимальности по α_k даёт

$$
\frac{d}{d\alpha} f(x^k - \alpha \nabla f(x^k))\Big|_{\alpha=\alpha_k} = 0.
$$

Условия оптимальности:

Точный линейный поиск (метод наискорейшего спуска)

$$
\alpha_k = \underset{\alpha \in \mathbb{R}^+}{\text{arg min}}\, f(x^k - \alpha \nabla f(x^k))
$$

Подход скорее теоретический, чем практический: он удобен для анализа сходимости, но точный линейный поиск часто затруднён, если вычисление функции занимает слишком много времени или стоит слишком дорого.

Интересное теоретическое свойство этого метода заключается в том, что градиенты на соседних итерациях ортогональны. Условие оптимальности по α_k даёт

$$
\frac{d}{d\alpha} f(x^k - \alpha \nabla f(x^k))\Big|_{\alpha = \alpha_k} = 0.
$$

Условия оптимальности:

$$
\nabla f(x^{k+1})^\top \nabla f(x^k) = 0
$$

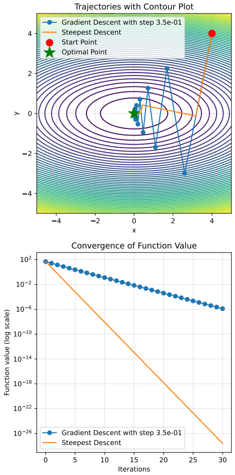


Рисунок 3. Наискорейший спуск

Открыть в Colab ♣

Сильно выпуклые квадратичные функции

Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

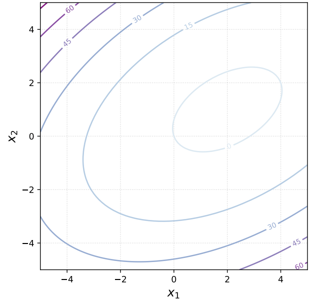
$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{2} x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.$$

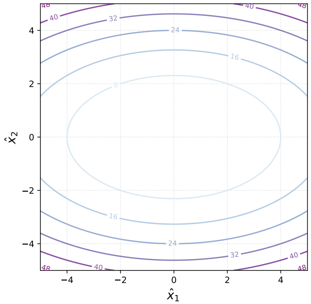
Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{2} x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.$$

- Во-первых, без ограничения общности мы можем установить $c = 0$, что не повлияет на процесс оптимизации.



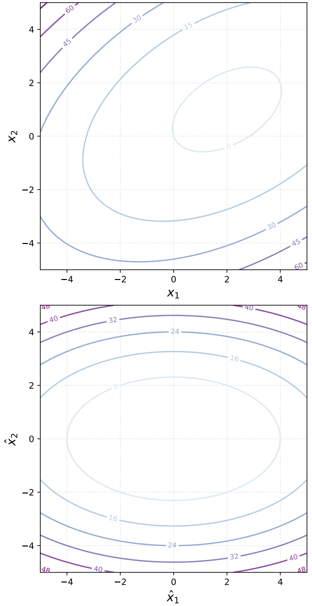


Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{2} x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.$$

- Во-первых, без ограничения общности мы можем установить $c = 0$, что не повлияет на процесс оптимизации.
- Во-вторых, у нас есть спектральное разложение матрицы $A = Q \Lambda Q^T$.

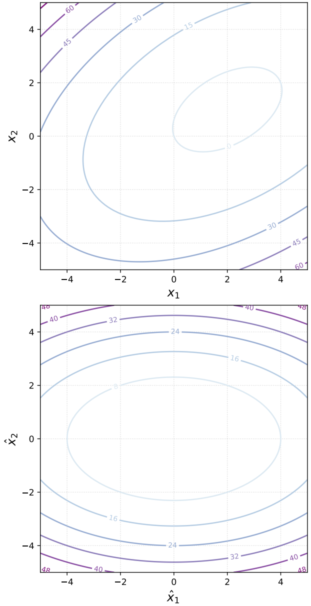


Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

$$\min_{x\in\mathbb{R}^d} f(x) = \min_{x\in\mathbb{R}^d} \frac{1}{2}x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.$$

- Во-первых, без ограничения общности мы можем установить $c = 0$, что не повлияет на процесс оптимизации.
- Во-вторых, у нас есть спектральное разложение матрицы $A = Q\Lambda Q^T$.
- Покажем, что мы можем сделать сдвиг координат, чтобы сделать анализ немного проще. Пусть $\hat{x} = Q^T(x - x^*)$, где x^* — точка минимума исходной функции, определяемая как $Ax^* = b$. При этом $x = Q\hat{x} + x^*$.



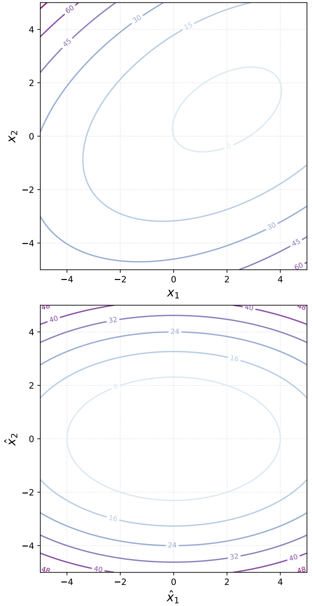
Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{2} x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.$$

- Во-первых, без ограничения общности мы можем установить $c = 0$, что не повлияет на процесс оптимизации.
- Во-вторых, у нас есть спектральное разложение матрицы $A = Q \Lambda Q^T$.
- Покажем, что мы можем сделать сдвиг координат, чтобы сделать анализ немного проще. Пусть $\hat{x} = Q^T(x - x^*)$, где x^* — точка минимума исходной функции, определяемая как $Ax^* = b$. При этом $x = Q\hat{x} + x^*$.

$$f(\hat{x}) = \frac{1}{2}(Q\hat{x} + x^*)^\top A (Q\hat{x} + x^*) - b^\top (Q\hat{x} + x^*)$$



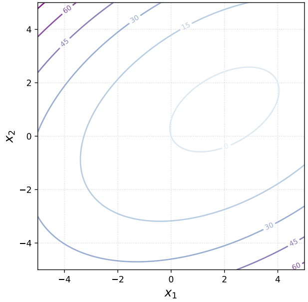
Сдвиг координат

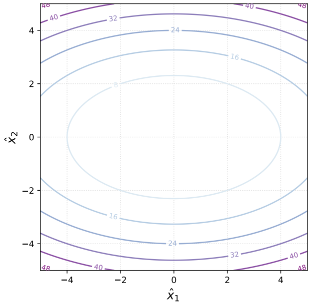
Рассмотрим следующую задачу квадратичной оптимизации:

$$\min_{x\in\mathbb{R}^d} f(x) = \min_{x\in\mathbb{R}^d} \frac{1}{2}x^\top Ax - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.$$

- Во-первых, без ограничения общности мы можем установить $c = 0$, что не повлияет на процесс оптимизации.
- Во-вторых, у нас есть спектральное разложение матрицы $A = Q\Lambda Q^T$.
- Покажем, что мы можем сделать сдвиг координат, чтобы сделать анализ немного проще. Пусть $\hat{x} = Q^T(x - x^*)$, где x^* — точка минимума исходной функции, определяемая как $Ax^* = b$. При этом $x = Q\hat{x} + x^*$.

$$\begin{aligned} f(\hat{x}) &= \frac{1}{2}(Q\hat{x} + x^*)^\top A(Q\hat{x} + x^*) - b^\top(Q\hat{x} + x^*) \\ &= \frac{1}{2}\hat{x}^T Q^T AQ\hat{x} + \frac{1}{2}(x^*)^T A(x^*) + (x^*)^T AQ\hat{x} - b^T Q\hat{x} - b^T x^* \end{aligned}$$





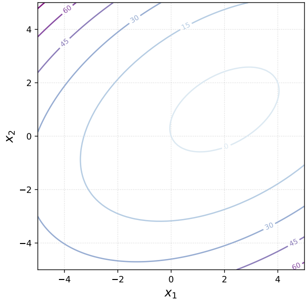
Сдвиг координат

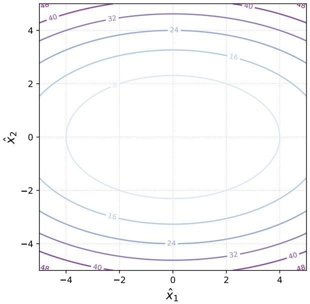
Рассмотрим следующую задачу квадратичной оптимизации:

$$\min_{x\in\mathbb{R}^d} f(x) = \min_{x\in\mathbb{R}^d} \frac{1}{2}x^\top Ax - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.$$

- Во-первых, без ограничения общности мы можем установить $c = 0$, что не повлияет на процесс оптимизации.
- Во-вторых, у нас есть спектральное разложение матрицы $A = Q\Lambda Q^T$.
- Покажем, что мы можем сделать сдвиг координат, чтобы сделать анализ немного проще. Пусть $\hat{x} = Q^T(x - x^*)$, где x^* — точка минимума исходной функции, определяемая как $Ax^* = b$. При этом $x = Q\hat{x} + x^*$.

$$\begin{aligned}
f(\hat{x}) &= \frac{1}{2}(Q\hat{x} + x^*)^\top A(Q\hat{x} + x^*) - b^\top(Q\hat{x} + x^*) \\
&= \frac{1}{2}\hat{x}^T Q^T AQ\hat{x} + \frac{1}{2}(x^*)^T A(x^*) + (x^*)^T AQ\hat{x} - b^T Q\hat{x} - b^T x^* \\
&= \frac{1}{2}\hat{x}^T \Lambda\hat{x} + \frac{1}{2}(x^*)^T A(x^*) + (x^*)^T AQ\hat{x} - (x^*)^T A^T Q\hat{x} - (x^*)^T Ax^*
\end{aligned}$$





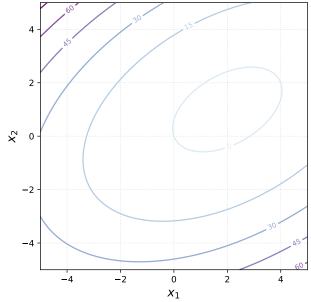
Сдвиг координат

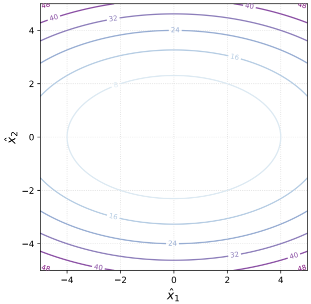
Рассмотрим следующую задачу квадратичной оптимизации:

$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{2} x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.$$

- Во-первых, без ограничения общности мы можем установить $c = 0$, что не повлияет на процесс оптимизации.
- Во-вторых, у нас есть спектральное разложение матрицы $A = Q \Lambda Q^T$.
- Покажем, что мы можем сделать сдвиг координат, чтобы сделать анализ немного проще. Пусть $\hat{x} = Q^T(x - x^*)$, где x^* — точка минимума исходной функции, определяемая как $Ax^* = b$. При этом $x = Q\hat{x} + x^*$.

$$\begin{aligned}
f(\hat{x}) &= \frac{1}{2}(Q\hat{x} + x^*)^\top A (Q\hat{x} + x^*) - b^\top (Q\hat{x} + x^*) \\
&= \frac{1}{2}\hat{x}^T Q^T A Q \hat{x} + \frac{1}{2}(x^*)^T A (x^*) + (x^*)^T A Q \hat{x} - b^T Q \hat{x} - b^T x^* \\
&= \frac{1}{2}\hat{x}^T \Lambda \hat{x} + \frac{1}{2}(x^*)^T A (x^*) + (x^*)^T A Q \hat{x} - (x^*)^T A^T Q \hat{x} - (x^*)^T A x^* \\
&= \frac{1}{2}\hat{x}^T \Lambda \hat{x} - \frac{1}{2}(x^*)^T A x^*
\end{aligned}$$





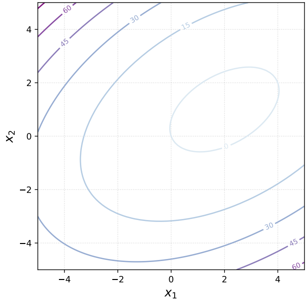
Сдвиг координат

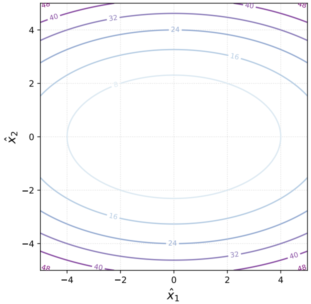
Рассмотрим следующую задачу квадратичной оптимизации:

$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{2} x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.$$

- Во-первых, без ограничения общности мы можем установить $c = 0$, что не повлияет на процесс оптимизации.
- Во-вторых, у нас есть спектральное разложение матрицы $A = Q \Lambda Q^T$.
- Покажем, что мы можем сделать сдвиг координат, чтобы сделать анализ немного проще. Пусть $\hat{x} = Q^T(x - x^*)$, где x^* — точка минимума исходной функции, определяемая как $Ax^* = b$. При этом $x = Q\hat{x} + x^*$.

$$\begin{aligned}
f(\hat{x}) &= \frac{1}{2}(Q\hat{x} + x^*)^\top A (Q\hat{x} + x^*) - b^\top (Q\hat{x} + x^*) \\
&= \frac{1}{2}\hat{x}^T Q^T A Q \hat{x} + \frac{1}{2}(x^*)^T A (x^*) + (x^*)^T A Q \hat{x} - b^T Q \hat{x} - b^T x^* \\
&= \frac{1}{2}\hat{x}^T \Lambda \hat{x} + \frac{1}{2}(x^*)^T A (x^*) + (x^*)^T A Q \hat{x} - (x^*)^T A^T Q \hat{x} - (x^*)^T A x^* \\
&= \frac{1}{2}\hat{x}^T \Lambda \hat{x} - \frac{1}{2}(x^*)^T A x^* \simeq \frac{1}{2}\hat{x}^T \Lambda \hat{x}
\end{aligned}$$





Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T \Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
x^{k+1} = x^k - \alpha^k \nabla f(x^k)
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
x^{k+1} = x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= \left(1 - \alpha^k \lambda_{(i)}\right) x^k_{(i)} \quad \text{для } i\text{-й координаты}
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T \Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= \left(1 - \alpha^k \lambda_{(i)}\right) x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha \lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha \lambda_{(i)}| < 1
$$

. . .

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
|1 - \alpha\mu| < 1
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T \Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= \left(1 - \alpha^k \lambda_{(i)}\right) x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha \lambda_{(i)}| < 1
$$

...

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 \\
&-1 < 1 - \alpha\mu < 1
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T \Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha \lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 \\
&-1 < 1 - \alpha\mu < 1 \\
&\alpha < \frac{2}{\mu} \qquad \alpha\mu > 0
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha \lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && \\
&\alpha < \frac{2}{\mu} \qquad \alpha\mu > 0 &&
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^{k}_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha\lambda_{(i)}| < 1
$$

...

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \frac{2}{\mu} \qquad \alpha\mu > 0 &&
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha\lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \tfrac{2}{\mu} \qquad \alpha\mu > 0 && \alpha < \tfrac{2}{L} \qquad \alpha L > 0
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha\lambda_{(i)}| < 1
$$

$\dots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \frac{2}{\mu} \quad\quad \alpha\mu > 0 && \alpha < \frac{2}{L} \quad\quad \alpha L > 0
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T \Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha \lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \frac{2}{\mu} \qquad \alpha\mu > 0 && \alpha < \frac{2}{L} \qquad \alpha L > 0
\end{aligned}
$$

Выберем α, минимизирующий худший знаменатель прогрессии

$$
\rho^* = \min_\alpha \rho(\alpha)
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha\lambda_{(i)}| < 1
$$

...

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{array}{ll}
|1 - \alpha\mu| < 1 & |1 - \alpha L| < 1 \\
-1 < 1 - \alpha\mu < 1 & -1 < 1 - \alpha L < 1 \\
\alpha < \frac{2}{\mu} \quad \alpha\mu > 0 & \alpha < \frac{2}{L} \quad \alpha L > 0
\end{array}
$$

Выберем α, минимизирующий худший знаменатель прогрессии

$$
\rho^* = \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha\lambda_{(i)}|
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha \lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \frac{2}{\mu} \qquad \alpha\mu > 0 && \alpha < \frac{2}{L} \qquad \alpha L > 0
\end{aligned}
$$

Выберем α, минимизирующий худший знаменатель прогрессии

$$
\begin{aligned}
\rho^* &= \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha \lambda_{(i)}| \\
&= \min_\alpha \max \left\{ |1 - \alpha\mu|, |1 - \alpha L| \right\}
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha\lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \frac{2}{\mu} \qquad \alpha\mu > 0 && \alpha < \frac{2}{L} \qquad \alpha L > 0
\end{aligned}
$$

Выберем α, минимизирующий худший знаменатель прогрессии

$$
\begin{aligned}
\rho^* &= \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha\lambda_{(i)}| \\
&= \min_\alpha \max \{|1 - \alpha\mu|, |1 - \alpha L|\}
\end{aligned}
$$

$$
\alpha^* : \quad 1 - \alpha^*\mu = \alpha^* L - 1
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^{k}_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha \lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \tfrac{2}{\mu} \qquad \alpha\mu > 0 && \alpha < \tfrac{2}{L} \qquad \alpha L > 0
\end{aligned}
$$

Выберем α, минимизирующий худший знаменатель прогрессии

$$
\begin{aligned}
\rho^* &= \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha \lambda_{(i)}| \\
&= \min_\alpha \max \left\{ |1 - \alpha\mu|, |1 - \alpha L| \right\} \\
\alpha^* &: \quad 1 - \alpha^*\mu = \alpha^* L - 1 \\
\alpha^* &= \frac{2}{\mu + L}
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha\lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \tfrac{2}{\mu} \qquad \alpha\mu > 0 && \alpha < \tfrac{2}{L} \qquad \alpha L > 0
\end{aligned}
$$

Выберем α, минимизирующий худший знаменатель прогрессии

$$
\begin{aligned}
\rho^* &= \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha\lambda_{(i)}| \\
&= \min_\alpha \max\left\{|1 - \alpha\mu|, |1 - \alpha L|\right\} \\
\alpha^* &: \quad 1 - \alpha^*\mu = \alpha^* L - 1 \\
\alpha^* &= \frac{2}{\mu + L} \quad \rho^* = \frac{L - \mu}{L + \mu}
\end{aligned}
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^{k}_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha \lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \tfrac{2}{\mu} \qquad \alpha\mu > 0 && \alpha < \tfrac{2}{L} \qquad \alpha L > 0
\end{aligned}
$$

Выберем α, минимизирующий худший знаменатель прогрессии

$$
\begin{aligned}
\rho^* &= \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha\lambda_{(i)}| \\
&= \min_\alpha \max \left\{ |1 - \alpha\mu|, |1 - \alpha L| \right\}
\end{aligned}
$$

$$
\alpha^* : \quad 1 - \alpha^*\mu = \alpha^* L - 1
$$

$$
\alpha^* = \frac{2}{\mu + L} \quad \rho^* = \frac{L - \mu}{L + \mu}
$$

$$
|x^k_{(i)}| \leq \left(\frac{L - \mu}{L + \mu} \right)^k |x^0_{(i)}|
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha\lambda_{(i)}| < 1
$$

...

Помним, что $\lambda_{\text{min}} = \mu > 0$, $\lambda_{\text{max}} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \frac{2}{\mu} \qquad \alpha\mu > 0 && \alpha < \frac{2}{L} \qquad \alpha L > 0
\end{aligned}
$$

Выберем α, минимизирующий худший знаменатель прогрессии

$$
\begin{aligned}
\rho^* &= \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha\lambda_{(i)}| \\
&= \min_\alpha \max \left\{ |1 - \alpha\mu|, |1 - \alpha L| \right\}
\end{aligned}
$$

$$
\alpha^* : \quad 1 - \alpha^*\mu = \alpha^* L - 1
$$

$$
\alpha^* = \frac{2}{\mu + L} \qquad \rho^* = \frac{L - \mu}{L + \mu}
$$

$$
|x^k_{(i)}| \leq \left(\frac{L - \mu}{L + \mu} \right)^k |x^0_{(i)}|
$$

$$
\|x^k\|_2 \leq \left(\frac{L - \mu}{L + \mu} \right)^k \|x^0\|_2
$$

Анализ сходимости

Теперь мы можем работать с функцией $f(x) = \frac{1}{2}x^T\Lambda x$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$$
\begin{aligned}
x^{k+1} &= x^k - \alpha^k \nabla f(x^k) = x^k - \alpha^k \Lambda x^k \\
&= (I - \alpha^k \Lambda) x^k \\
x^{k+1}_{(i)} &= (1 - \alpha^k \lambda_{(i)})\, x^k_{(i)} \quad \text{для } i\text{-й координаты} \\
x^k_{(i)} &= (1 - \alpha\, \lambda_{(i)})^k\, x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha
\end{aligned}
$$

Используем постоянный шаг $\alpha^k = \alpha$. Условие сходимости:

$$
\rho(\alpha) = \max_i |1 - \alpha \lambda_{(i)}| < 1
$$

$\ldots$

Помним, что $\lambda_{\min} = \mu > 0$, $\lambda_{\max} = L \geq \mu$.

$$
\begin{aligned}
&|1 - \alpha\mu| < 1 && |1 - \alpha L| < 1 \\
&-1 < 1 - \alpha\mu < 1 && -1 < 1 - \alpha L < 1 \\
&\alpha < \tfrac{2}{\mu} \qquad \alpha\mu > 0 && \alpha < \tfrac{2}{L} \qquad \alpha L > 0
\end{aligned}
$$

Выберем α, минимизирующий худший знаменатель прогрессии

$$
\begin{aligned}
\rho^* &= \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha \lambda_{(i)}| \\
&= \min_\alpha \max \left\{ |1 - \alpha\mu|, |1 - \alpha L| \right\}
\end{aligned}
$$

$$
\alpha^* : \quad 1 - \alpha^* \mu = \alpha^* L - 1
$$

$$
\alpha^* = \frac{2}{\mu + L} \quad \rho^* = \frac{L - \mu}{L + \mu}
$$

$$
|x^k_{(i)}| \leq \left(\frac{L - \mu}{L + \mu} \right)^k |x^0_{(i)}|
$$

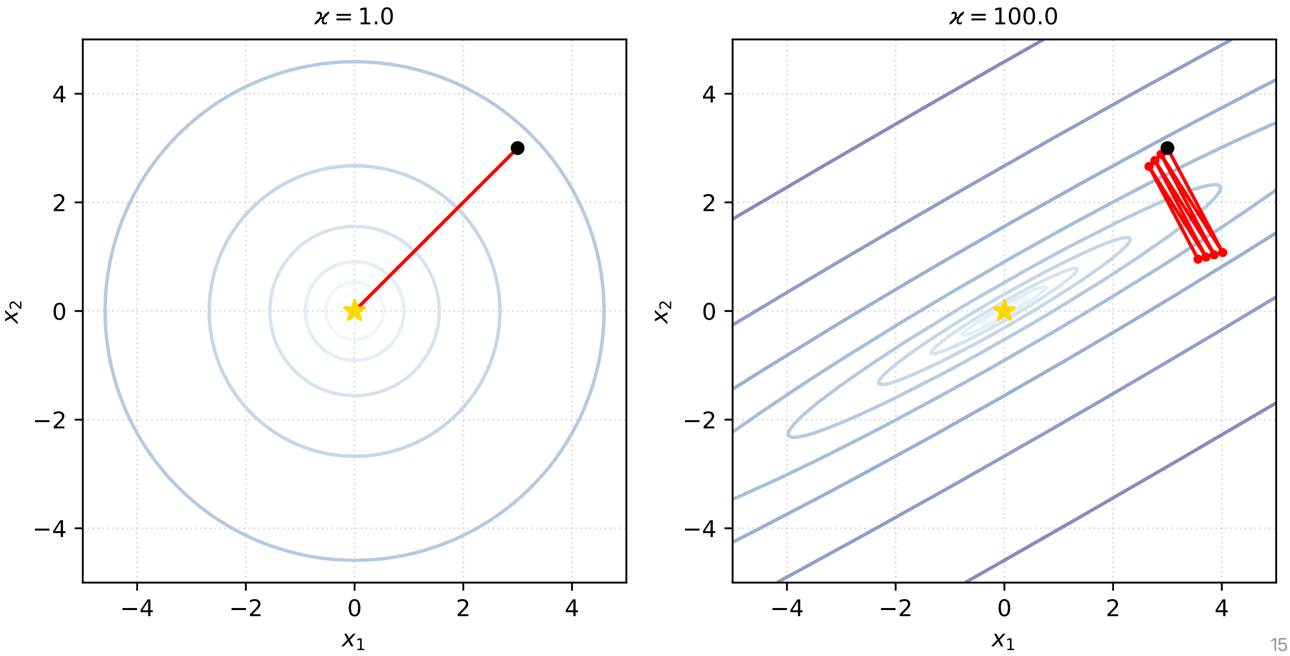
$$
\|x^k\|_2 \leq \left(\frac{L - \mu}{L + \mu} \right)^k \|x^0\|_2 \quad f(x^k) \leq \left(\frac{L - \mu}{L + \mu} \right)^{2k} f(x^0)
$$

Анализ сходимости

Таким образом, имеем линейную сходимость по аргументу со скоростью $\frac{\varkappa-1}{\varkappa+1} = 1 - \frac{2}{\varkappa+1}$, где $\varkappa = \frac{L}{\mu}$ — *число обусловленности* квадратичной задачи.

$\varkappa$	ρ	Итераций до уменьшения ошибки по аргументу в 10 раз	Итераций до уменьшения ошибки по функции в 10 раз
1.1	0.05	1	1
2	0.33	3	2
5	0.67	6	3
10	0.82	12	6
50	0.96	58	29
100	0.98	116	58
500	0.996	576	288
1000	0.998	1152	576

Число обусловленности $\varkappa$



Случай PL-функций

PL-функции. Линейная сходимость градиентного спуска без выпуклости

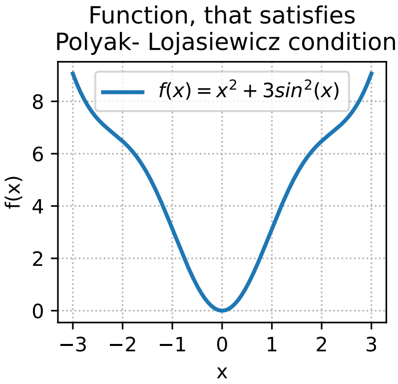
Говорят, что f удовлетворяет условию Поляка-Лоясиевича (PL), если для некоторого $\mu > 0$ выполняется

$$\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*) \quad \forall x$$

Интересно, что градиентный спуск может сходиться линейно даже без выпуклости.

Следующие функции удовлетворяют условию PL, но не являются выпуклыми. Код

$$f(x) = x^2 + 3\sin^2(x)$$



PL-функции. Линейная сходимость градиентного спуска без выпуклости

Говорят, что f удовлетворяет условию Поляка-Лоясиевича (PL), если для некоторого $\mu > 0$ выполняется

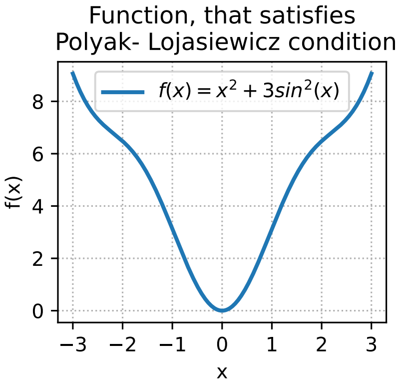
$$\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*) \quad \forall x$$

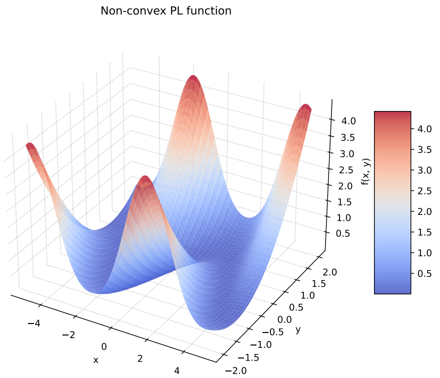
Интересно, что градиентный спуск может сходиться линейно даже без выпуклости.

Следующие функции удовлетворяют условию PL, но не являются выпуклыми. Код

$$f(x) = x^2 + 3\sin^2(x)$$

$$f(x, y) = \frac{(y - \sin x)^2}{2}$$





Анализ сходимости

Theorem

Рассмотрим задачу

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

и предположим, что f является PL-функцией с константой μ и L-гладкой, для некоторых $L \geq \mu > 0$.
Рассмотрим последовательность $(x^k)_{k \in \mathbb{N}}$, сгенерированную методом градиентного спуска из точки x^0 с постоянным шагом α, удовлетворяющим $0 < \alpha \leq \frac{1}{L}$. Пусть $f^* = \min_{x \in \mathbb{R}^d} f(x)$. Тогда:

$$f(x^k) - f^* \leq (1 - \alpha\mu)^k (f(x^0) - f^*).$$

Анализ сходимости

Используем L-гладкость вместе с правилом обновления, чтобы записать:

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

Анализ сходимости

Используем L-гладкость вместе с правилом обновления, чтобы записать:

$$
\begin{aligned}
f(x^{k+1}) &\leq f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 \\
&= f(x^k) - \alpha \|\nabla f(x^k)\|^2 + \frac{L\alpha^2}{2}\|\nabla f(x^k)\|^2
\end{aligned}
$$

Анализ сходимости

Используем L-гладкость вместе с правилом обновления, чтобы записать:

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

Анализ сходимости

Используем L-гладкость вместе с правилом обновления, чтобы записать:

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

Анализ сходимости

Используем L-гладкость вместе с правилом обновления, чтобы записать:

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

Анализ сходимости

Используем L-гладкость вместе с правилом обновления, чтобы записать:

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

где в последнем неравенстве использована гипотеза о шаге $\alpha L \leq 1$.

Анализ сходимости

Используем L-гладкость вместе с правилом обновления, чтобы записать:

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

где в последнем неравенстве использована гипотеза о шаге $\alpha L \le 1$.

Теперь используем свойство PL-функции и получаем:

$$
f(x^{k+1}) \le f(x^k) - \alpha\mu(f(x^k) - f^*).
$$

Вычтя f^* из обеих частей этого неравенства и применив рекурсию, мы получим искомый результат.

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

Theorem

Если функция $f(x)$ дифференцируема и μ-сильно выпукла, то она является PL-функцией.

Доказательство

По критерию сильной выпуклости первого порядка:

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

Положим $y = x^*$:

$$f(x^*) \geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2}\|x^* - x\|_2^2$$

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

Theorem

Если функция $f(x)$ дифференцируема и μ-сильно выпукла, то она является PL-функцией.

Доказательство

По критерию сильной выпуклости первого порядка:

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

Положим $y = x^*$:

$$f(x^*) \geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2}\|x^* - x\|_2^2$$
$$f(x) - f(x^*) \leq \nabla f(x)^T (x - x^*) - \frac{\mu}{2}\|x^* - x\|_2^2 =$$

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

Theorem

Если функция $f(x)$ дифференцируема и μ-сильно выпукла, то она является PL-функцией.

Доказательство

По критерию сильной выпуклости первого порядка:

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

Положим $y = x^*$:

$$\begin{aligned} f(x^*) &\geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2}\|x^* - x\|_2^2 \\ f(x) - f(x^*) &\leq \nabla f(x)^T (x - x^*) - \frac{\mu}{2}\|x^* - x\|_2^2 = \\ &= \left(\nabla f(x) - \frac{\mu}{2}(x^* - x)\right)^T (x - x^*) = \end{aligned}$$

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

Theorem

Если функция $f(x)$ дифференцируема и μ-сильно выпукла, то она является PL-функцией.

Доказательство

По критерию сильной выпуклости первого порядка:

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

Положим $y = x^*$:

$$\begin{aligned}
f(x^*) &\geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2} \|x^* - x\|_2^2 \\
f(x) - f(x^*) &\leq \nabla f(x)^T (x - x^*) - \frac{\mu}{2} \|x^* - x\|_2^2 = \\
&= \left(\nabla f(x) - \frac{\mu}{2}(x^* - x)\right)^T (x - x^*) = \\
&= \frac{1}{2} \left(\frac{2}{\sqrt{\mu}} \nabla f(x) - \sqrt{\mu}(x^* - x)\right)^T \sqrt{\mu}(x - x^*)
\end{aligned}$$

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

Theorem

Если функция $f(x)$ дифференцируема и μ-сильно выпукла, то она является PL-функцией.

Доказательство

По критерию сильной выпуклости первого порядка:

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

Положим $y = x^*$:

$$
\begin{aligned}
f(x^*) &\geq f(x) + \nabla f(x)^T(x^*-x) + \tfrac{\mu}{2}\|x^*-x\|_2^2 \\
f(x) - f(x^*) &\leq \nabla f(x)^T(x-x^*) - \tfrac{\mu}{2}\|x^*-x\|_2^2 = \\
&= \left(\nabla f(x) - \tfrac{\mu}{2}(x^*-x)\right)^T(x-x^*) = \\
&= \frac{1}{2}\left(\tfrac{2}{\sqrt{\mu}}\nabla f(x) - \sqrt{\mu}(x^*-x)\right)^T\sqrt{\mu}(x-x^*)
\end{aligned}
$$

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

Theorem

Если функция $f(x)$ дифференцируема и μ-сильно выпукла, то она является PL-функцией.

Доказательство

По критерию сильной выпуклости первого порядка:

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

Положим $y = x^*$:

$$\begin{aligned}
f(x^*) &\geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2}\|x^* - x\|_2^2 \\
f(x) - f(x^*) &\leq \nabla f(x)^T (x - x^*) - \frac{\mu}{2}\|x^* - x\|_2^2 = \\
&= \left(\nabla f(x) - \frac{\mu}{2}(x^* - x)\right)^T (x - x^*) = \\
&= \frac{1}{2}\left(\frac{2}{\sqrt{\mu}}\nabla f(x) - \sqrt{\mu}(x^* - x)\right)^T \sqrt{\mu}(x - x^*)
\end{aligned}$$

Пусть $a = \frac{1}{\sqrt{\mu}}\nabla f(x)$ и
$b = \sqrt{\mu}(x - x^*) - \frac{1}{\sqrt{\mu}}\nabla f(x)$

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

Theorem

Если функция $f(x)$ дифференцируема и μ-сильно выпукла, то она является PL-функцией.

Доказательство

По критерию сильной выпуклости первого порядка:

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

Положим $y = x^*$:

$$\begin{aligned} f(x^*) &\geq f(x) + \nabla f(x)^T(x^*-x) + \frac{\mu}{2}\|x^*-x\|_2^2 \\ f(x) - f(x^*) &\leq \nabla f(x)^T(x-x^*) - \frac{\mu}{2}\|x^*-x\|_2^2 = \\ &= \left(\nabla f(x) - \frac{\mu}{2}(x^*-x)\right)^T(x-x^*) = \\ &= \frac{1}{2}\left(\frac{2}{\sqrt{\mu}}\nabla f(x) - \sqrt{\mu}(x^*-x)\right)^T\sqrt{\mu}(x-x^*) \end{aligned}$$

Пусть $a = \frac{1}{\sqrt{\mu}}\nabla f(x)$ и

$b = \sqrt{\mu}(x-x^*) - \frac{1}{\sqrt{\mu}}\nabla f(x)$

Тогда $a + b = \sqrt{\mu}(x-x^*)$ и

$a - b = \frac{2}{\sqrt{\mu}}\nabla f(x) - \sqrt{\mu}(x-x^*)$

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

$$f(x) - f(x^*) \leq \frac{1}{2}\left(\frac{1}{\mu}\|\nabla f(x)\|_2^2 - \left\|\sqrt{\mu}(x - x^*) - \frac{1}{\sqrt{\mu}}\nabla f(x)\right\|_2^2\right)$$

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

$$f(x) - f(x^*) \leq \frac{1}{2}\left(\frac{1}{\mu}\|\nabla f(x)\|_2^2 - \left\|\sqrt{\mu}(x - x^*) - \frac{1}{\sqrt{\mu}}\nabla f(x)\right\|_2^2\right)$$

$$f(x) - f(x^*) \leq \frac{1}{2\mu}\|\nabla f(x)\|_2^2,$$

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

$$f(x) - f(x^*) \leq \frac{1}{2}\left(\frac{1}{\mu}\|\nabla f(x)\|_2^2 - \left\|\sqrt{\mu}(x - x^*) - \frac{1}{\sqrt{\mu}}\nabla f(x)\right\|_2^2\right)$$

$$f(x) - f(x^*) \leq \frac{1}{2\mu}\|\nabla f(x)\|_2^2,$$

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

$$
f(x) - f(x^*) \leq \frac{1}{2} \left(\frac{1}{\mu} \|\nabla f(x)\|_2^2 - \left\| \sqrt{\mu}(x - x^*) - \frac{1}{\sqrt{\mu}} \nabla f(x) \right\|_2^2 \right)
$$

$$
f(x) - f(x^*) \leq \frac{1}{2\mu} \|\nabla f(x)\|_2^2,
$$

которое является точным условием PL. Это означает, что мы уже имеем доказательство линейной сходимости для любой сильно выпуклой функции.

Выпуклый гладкий случай

Выпуклый гладкий случай

Theorem

Предположим, что $f : \mathbb{R}^d \to \mathbb{R}$ является выпуклой и L-гладкой функцией, для некоторого $L > 0$.
Пусть $(x^k)_{k\in\mathbb{N}}$ — последовательность итераций, сгенерированная методом градиентного спуска из точки x^0 с постоянным шагом α, удовлетворяющим $0 < \alpha \leq \frac{1}{L}$. Тогда для всех $k \in \mathbb{N}$ справедливо:

$$f(x^k) - f(x) \leq \frac{\|x^0 - x\|^2}{2\alpha k}.$$

Заметим, что мы здесь никак не упоминаем точку минимума. То есть, это сходимость $\forall x \in \mathbb{R}^d$ (в том числе и до точки минимума).

Анализ сходимости

Наш инструментарий:

1. Выпуклость:

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \,,\ \forall x, y. \tag{1}$$

Анализ сходимости

Наш инструментарий:

1. Выпуклость:

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \,, \; \forall x, y. \tag{1}$$

2. Гладкость:

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \left\| x - y \right\|^2 \,, \; \forall x, y. \tag{2}$$

Анализ сходимости

Наш инструментарий:

1. Выпуклость:

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \, , \; \forall x, y. \tag{1}$$

2. Гладкость:

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \|x - y\|^2 \, , \; \forall x, y. \tag{2}$$

3. 3-point identity (по сути, квадрат разности):

$$\|a - b\|^2 = \|a - c - (b - c)\|^2 = \|a - c\|^2 - 2\langle a - c, b - c \rangle + \|b - c\|^2$$

переносим справа все кроме $\|b - c\|^2$ налево и меняем местами все факторы внутри каждого из перенесенных членов:

$$\|b - c\|^2 = \|b - a\|^2 + 2\langle c - a, c - b \rangle - \|c - a\|^2. \tag{3}$$

Анализ сходимости

Наш инструментарий:

1. Выпуклость:
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle\ ,\ \forall x, y. \tag{1}$$

2. Гладкость:
$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \|x - y\|^2\ ,\ \forall x, y. \tag{2}$$

3. 3-point identity (по сути, квадрат разности):
$$\|a - b\|^2 = \|a - c - (b - c)\|^2 = \|a - c\|^2 - 2\langle a - c, b - c \rangle + \|b - c\|^2$$
переносим справа все кроме $\|b - c\|^2$ налево и меняем местами все факторы внутри каждого из перенесенных членов:
$$\|b - c\|^2 = \|b - a\|^2 + 2\langle c - a, c - b \rangle - \|c - a\|^2. \tag{3}$$

- Подставляем в (3) $b \equiv x$, $c \equiv x^{k+1}$, $a \equiv x_k$ и домножаем все на $\frac{1}{2}$:

(4)

Анализ сходимости

Наш инструментарий:

1. Выпуклость:

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \ , \ \forall x, y. \tag{1}$$

2. Гладкость:

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \left\| x - y \right\|^2 , \ \forall x, y. \tag{2}$$

3. 3-point identity (по сути, квадрат разности):

$$\|a - b\|^2 = \|a - c - (b - c)\|^2 = \|a - c\|^2 - 2\langle a - c, b - c \rangle + \|b - c\|^2$$

переносим справа все кроме $\|b - c\|^2$ налево и меняем местами все факторы внутри каждого из перенесенных членов:

$$\|b - c\|^2 = \|b - a\|^2 + 2\langle c - a, c - b \rangle - \|c - a\|^2. \tag{3}$$

- Подставляем в (3) $b \equiv x$, $c \equiv x^{k+1}$, $a \equiv x_k$ и домножаем все на $\frac{1}{2}$:

$$\frac{1}{2}\|x - x^{k+1}\|^2 = \frac{1}{2}\left\|x - x^k\right\|^2 + \left\langle x^{k+1} - x^k, x^{k+1} - x \right\rangle - \frac{1}{2}\left\|x^{k+1} - x^k\right\|^2 \tag{4}$$

Анализ сходимости

Наш инструментарий:

1. Выпуклость:
$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle\ ,\ \forall x, y. \tag{1}$$
2. Гладкость:
$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \left\| x - y \right\|^2\ ,\ \forall x, y. \tag{2}$$
3. 3-point identity (по сути, квадрат разности):
$$\|a - b\|^2 = \|a - c - (b - c)\|^2 = \|a - c\|^2 - 2\langle a - c, b - c\rangle + \|b - c\|^2$$
переносим справа все кроме $\|b - c\|^2$ налево и меняем местами все факторы внутри каждого из перенесенных членов:
$$\|b - c\|^2 = \|b - a\|^2 + 2\langle c - a, c - b\rangle - \|c - a\|^2. \tag{3}$$

- Подставляем в (3) $b \equiv x$, $c \equiv x^{k+1}$, $a \equiv x_k$ и домножаем все на $\frac{1}{2}$:

$$\begin{aligned} \frac{1}{2}\|x - x^{k+1}\|^2 &= \frac{1}{2}\left\|x - x^k\right\|^2 + \left\langle x^{k+1} - x^k, x^{k+1} - x \right\rangle - \frac{1}{2}\left\|x^{k+1} - x^k\right\|^2 \\ &= \frac{1}{2}\left\|x - x^k\right\|^2 - \alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle - \frac{1}{2}\left\|x^{k+1} - x^k\right\|^2. \end{aligned} \tag{4}$$

Анализ сходимости

- Посмотрим внимательнее на скалярное произведение $-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle$ и воспользуемся сначала выпуклостью (1), а потом – гладкостью (2):

Анализ сходимости

- Посмотрим внимательнее на скалярное произведение $-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle$ и воспользуемся сначала выпуклостью (1), а потом – гладкостью (2):

$$-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle = \alpha \left(\left\langle \nabla f(x^k), x - x^k \right\rangle + \left\langle \nabla f(x^k), x^k - x^{k+1} \right\rangle \right)$$

Анализ сходимости

- Посмотрим внимательнее на скалярное произведение $-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle$ и воспользуемся сначала выпуклостью (1), а потом – гладкостью (2):

$$
\begin{aligned}
-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle &= \alpha \left(\left\langle \nabla f(x^k), x - x^k \right\rangle + \left\langle \nabla f(x^k), x^k - x^{k+1} \right\rangle \right) \\
&\overset{(1)}{\leq} \alpha \left(f(x) - f(x^k) + \left\langle \nabla f(x^k), x^k - x^{k+1} \right\rangle \right)
\end{aligned}
$$

Анализ сходимости

- Посмотрим внимательнее на скалярное произведение $-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle$ и воспользуемся сначала выпуклостью (1), а потом – гладкостью (2):

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

Анализ сходимости

- Посмотрим внимательнее на скалярное произведение $-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle$ и воспользуемся сначала выпуклостью (1), а потом – гладкостью (2):

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

- Подставляем это все обратно в (4) и используем условие на размер шага $\alpha \leq \frac{1}{L}$:

Анализ сходимости

- Посмотрим внимательнее на скалярное произведение $-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle$ и воспользуемся сначала выпуклостью (1), а потом – гладкостью (2):

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

- Подставляем это все обратно в (4) и используем условие на размер шага $\alpha \leq \frac{1}{L}$:

$$
\frac{1}{2} \left\| x - x^{k+1} \right\|^2 \leq \frac{1}{2} \left\| x - x^k \right\|^2 + \alpha \left(f(x) - f(x^{k+1}) \right) + \left(\frac{\alpha L}{2} - \frac{1}{2} \right) \left\| x^{k+1} - x^k \right\|^2
$$

Анализ сходимости

- Посмотрим внимательнее на скалярное произведение $-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle$ и воспользуемся сначала выпуклостью (1), а потом – гладкостью (2):

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

- Подставляем это все обратно в (4) и используем условие на размер шага $\alpha \leq \frac{1}{L}$:

$$
\frac{1}{2} \left\| x - x^{k+1} \right\|^2 \leq \frac{1}{2} \left\| x - x^k \right\|^2 + \alpha \left(f(x) - f(x^{k+1}) \right) + \left(\frac{\alpha L}{2} - \frac{1}{2} \right) \left\| x^{k+1} - x^k \right\|^2
$$

$$
\frac{1}{2} \left\| x - x^{k+1} \right\|^2 - \frac{1}{2} \left\| x - x^k \right\|^2 \leq \alpha \left(f(x) - f(x^{k+1}) \right) + \left(\frac{\alpha L}{2} - \frac{1}{2} \right) \left\| x^{k+1} - x^k \right\|^2
$$

Анализ сходимости

- Посмотрим внимательнее на скалярное произведение $-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle$ и воспользуемся сначала выпуклостью (1), а потом – гладкостью (2):

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

- Подставляем это все обратно в (4) и используем условие на размер шага $\alpha \leq \frac{1}{L}$:

$$
\begin{aligned}
\frac{1}{2} \left\| x - x^{k+1} \right\|^2 &\leq \frac{1}{2} \left\| x - x^k \right\|^2 + \alpha \left(f(x) - f(x^{k+1}) \right) + \left(\frac{\alpha L}{2} - \frac{1}{2} \right) \left\| x^{k+1} - x^k \right\|^2 \\
\frac{1}{2} \left\| x - x^{k+1} \right\|^2 - \frac{1}{2} \left\| x - x^k \right\|^2 &\leq \alpha \left(f(x) - f(x^{k+1}) \right) + \left(\frac{\alpha L}{2} - \frac{1}{2} \right) \left\| x^{k+1} - x^k \right\|^2 \\
&\overset{(\alpha \leq 1/L)}{\leq} \frac{1}{L} \left(f(x) - f(x^{k+1}) \right).
\end{aligned}
$$

Анализ сходимости

- Посмотрим внимательнее на скалярное произведение $-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle$ и воспользуемся сначала выпуклостью (1), а потом – гладкостью (2):

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

- Подставляем это все обратно в (4) и используем условие на размер шага $\alpha \leq \frac{1}{L}$:

$$
\begin{aligned}
\frac{1}{2} \left\| x - x^{k+1} \right\|^2 &\leq \frac{1}{2} \left\| x - x^k \right\|^2 + \alpha \left(f(x) - f(x^{k+1}) \right) + \left(\frac{\alpha L}{2} - \frac{1}{2} \right) \left\| x^{k+1} - x^k \right\|^2 \\
\frac{1}{2} \left\| x - x^{k+1} \right\|^2 - \frac{1}{2} \left\| x - x^k \right\|^2 &\leq \alpha \left(f(x) - f(x^{k+1}) \right) + \left(\frac{\alpha L}{2} - \frac{1}{2} \right) \left\| x^{k+1} - x^k \right\|^2 \\
&\overset{(\alpha \leq 1/L)}{\leq} \frac{1}{L} \left(f(x) - f(x^{k+1}) \right).
\end{aligned}
$$

- Переносим правую часть влево, левую - вправо и домножаем на L:

.

Анализ сходимости

- Посмотрим внимательнее на скалярное произведение $-\alpha \left\langle \nabla f(x^k), x^{k+1} - x \right\rangle$ и воспользуемся сначала выпуклостью (1), а потом – гладкостью (2):

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

- Подставляем это все обратно в (4) и используем условие на размер шага $\alpha \leq \frac{1}{L}$:

$$
\begin{aligned}
\frac{1}{2} \left\| x - x^{k+1} \right\|^2 \leq \frac{1}{2} \left\| x - x^k \right\|^2 + \alpha \left(f(x) - f(x^{k+1}) \right) + \left(\frac{\alpha L}{2} - \frac{1}{2} \right) \left\| x^{k+1} - x^k \right\|^2 \\
\frac{1}{2} \left\| x - x^{k+1} \right\|^2 - \frac{1}{2} \left\| x - x^k \right\|^2 \leq \alpha \left(f(x) - f(x^{k+1}) \right) + \left(\frac{\alpha L}{2} - \frac{1}{2} \right) \left\| x^{k+1} - x^k \right\|^2 \\
\overset{(\alpha \leq 1/L)}{\leq} \frac{1}{L} \left(f(x) - f(x^{k+1}) \right).
\end{aligned}
$$

- Переносим правую часть влево, левую - вправо и домножаем на L:

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

Анализ сходимости

- Берем среднее от левой и правой частей от по всем k от 0 до $N-1$:

(5)

Анализ сходимости

- Берем среднее от левой и правой частей от по всем k от 0 до $N-1$:

$$\frac{1}{N}\sum_{k=0}^{N-1}\left(f(x^{k+1}) - f(x)\right) \leq \frac{L}{2N}\sum_{k=0}^{N-1}\left(\left\|x - x^k\right\|^2 - \left\|x - x^{k+1}\right\|\right) \tag{5}$$

Анализ сходимости

- Берем среднее от левой и правой частей от по всем k от 0 до $N-1$:

$$
\begin{aligned}
\frac{1}{N}\sum_{k=0}^{N-1}\left(f(x^{k+1})-f(x)\right) &\leq \frac{L}{2N}\sum_{k=0}^{N-1}\left(\left\|x-x^k\right\|^2-\left\|x-x^{k+1}\right\|\right)\\
&= \frac{L}{2N}\left(\left\|x-x^0\right\|^2-\left\|x-x^{k+1}\right\|^2\right)
\end{aligned} \tag{5}
$$

Анализ сходимости

- Берем среднее от левой и правой частей от по всем k от 0 до $N-1$:

$$
\begin{aligned}
\frac{1}{N}\sum_{k=0}^{N-1}\left(f(x^{k+1})-f(x)\right) &\leq \frac{L}{2N}\sum_{k=0}^{N-1}\left(\left\|x-x^k\right\|^2-\left\|x-x^{k+1}\right\|\right)\\
&= \frac{L}{2N}\left(\left\|x-x^0\right\|^2-\left\|x-x^{k+1}\right\|^2\right)\\
&\leq \frac{L}{2N}\left\|x-x^0\right\|^2.
\end{aligned}
\tag{5}
$$

Анализ сходимости

- Берем среднее от левой и правой частей от по всем k от 0 до $N-1$:

$$
\begin{aligned}
\frac{1}{N}\sum_{k=0}^{N-1}\left(f(x^{k+1})-f(x)\right) &\le \frac{L}{2N}\sum_{k=0}^{N-1}\left(\left\|x-x^{k}\right\|^2-\left\|x-x^{k+1}\right\|\right)\\
&= \frac{L}{2N}\left(\left\|x-x^{0}\right\|^2-\left\|x-x^{k+1}\right\|^2\right)\\
&\le \frac{L}{2N}\left\|x-x^{0}\right\|^2.
\end{aligned} \tag{5}
$$

- Так как для выпуклых функций (1) градиентный спуск монотонен:

$$
\begin{aligned}
f(x^k) &\ge f(x^{k+1}) + \langle \nabla f(x^{k+1}), x^k - x^{k+1}\rangle\\
&= f(x^{k+1}) + \alpha\left\|\nabla f(x^{k+1})\right\|^2\\
&\ge f(x^{k+1}),
\end{aligned}
$$

Анализ сходимости

- Берем среднее от левой и правой частей от по всем k от 0 до $N-1$:

$$\begin{aligned}
\frac{1}{N}\sum_{k=0}^{N-1}\left(f(x^{k+1})-f(x)\right) &\leq \frac{L}{2N}\sum_{k=0}^{N-1}\left(\left\|x-x^k\right\|^2-\left\|x-x^{k+1}\right\|\right)\\
&= \frac{L}{2N}\left(\left\|x-x^0\right\|^2-\left\|x-x^{k+1}\right\|^2\right)\\
&\leq \frac{L}{2N}\left\|x-x^0\right\|^2.
\end{aligned} \tag{5}$$

- Так как для выпуклых функций (1) градиентный спуск монотонен:

$$\begin{aligned}
f(x^k) &\geq f(x^{k+1}) + \langle \nabla f(x^{k+1}), x^k - x^{k+1}\rangle\\
&= f(x^{k+1}) + \alpha\left\|\nabla f(x^{k+1})\right\|^2\\
&\geq f(x^{k+1}),
\end{aligned}$$

то $\frac{1}{N}\sum_{i=0}^{N-1}\left(f(x^{k+1})-f(x)\right) \geq \min_{i=0,\dots,N-1} f(x^{i+1}) - f(x) = f(x^N) - f(x)$.

Анализ сходимости

- Берем среднее от левой и правой частей от по всем k от 0 до $N-1$:

$$
\begin{aligned}
\frac{1}{N}\sum_{k=0}^{N-1}\left(f(x^{k+1})-f(x)\right) &\leq \frac{L}{2N}\sum_{k=0}^{N-1}\left(\left\|x-x^k\right\|^2-\left\|x-x^{k+1}\right\|\right)\\
&= \frac{L}{2N}\left(\left\|x-x^0\right\|^2-\left\|x-x^{k+1}\right\|^2\right)\\
&\leq \frac{L}{2N}\left\|x-x^0\right\|^2.
\end{aligned} \tag{5}
$$

- Так как для выпуклых функций (1) градиентный спуск монотонен:

$$
\begin{aligned}
f(x^k) &\geq f(x^{k+1}) + \langle \nabla f(x^{k+1}), x^k - x^{k+1}\rangle\\
&= f(x^{k+1}) + \alpha\left\|\nabla f(x^{k+1})\right\|^2\\
&\geq f(x^{k+1}),
\end{aligned}
$$

то $\frac{1}{N}\sum_{i=0}^{N-1}\left(f(x^{k+1})-f(x)\right) \geq \min_{i=0,\dots,N-1} f(x^{i+1}) - f(x) = f(x^N) - f(x)$. Подставляя это в (5), получаем искомый результат.

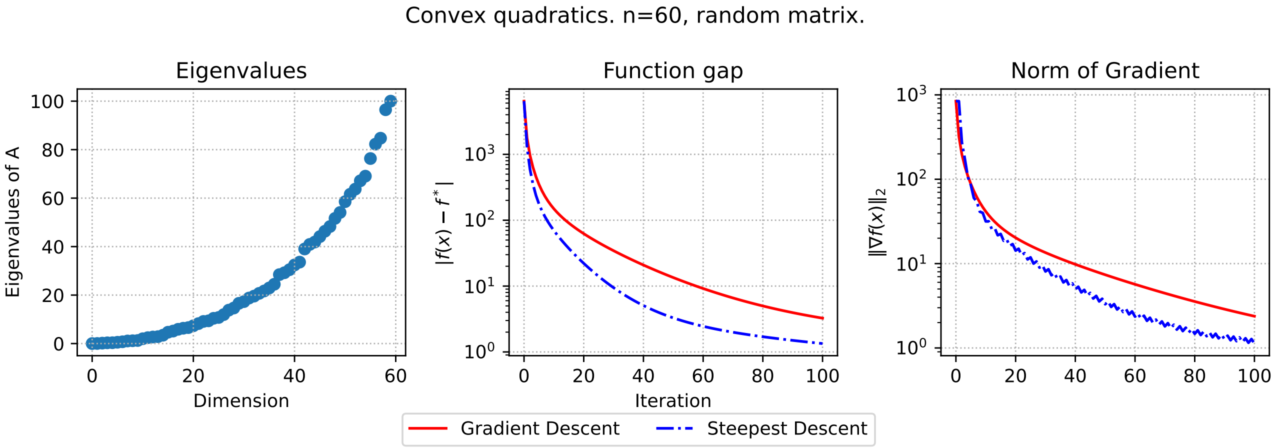
Итог

Градиентный спуск: $\min\limits_{x \in \mathbb{R}^n} f(x)$ $\qquad x^{k+1} = x^k - \alpha^k \nabla f(x^k)$

гладкий (не выпуклый)	гладкий и выпуклый	гладкий и сильно выпуклый (или PL)
$\|\nabla f(x^k)\|^2 \sim \mathcal{O}\left(\dfrac{1}{k}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon}\right)$	$f(x^k) - f^* \sim \mathcal{O}\left(\dfrac{1}{k}\right)$ $k_\varepsilon \sim \mathcal{O}\left(\dfrac{1}{\varepsilon}\right)$	$\|x^k - x^*\|^2 \sim \mathcal{O}\left(\left(1 - \dfrac{\mu}{L}\right)^k\right)$ $k_\varepsilon \sim \mathcal{O}\left(\varkappa \log \dfrac{1}{\varepsilon}\right)$

Численные эксперименты

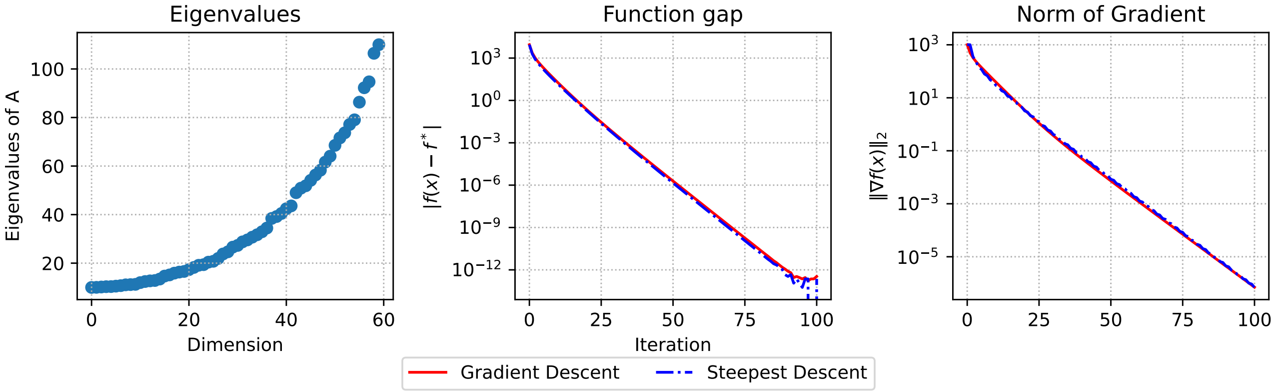
$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{1}{2} x^T A x - b^T x \right\},\ \mu = 0,\ L = 100.$$



Численные эксперименты

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{1}{2} x^T A x - b^T x \right\}, \ \mu = 10, \ L = 110.$$

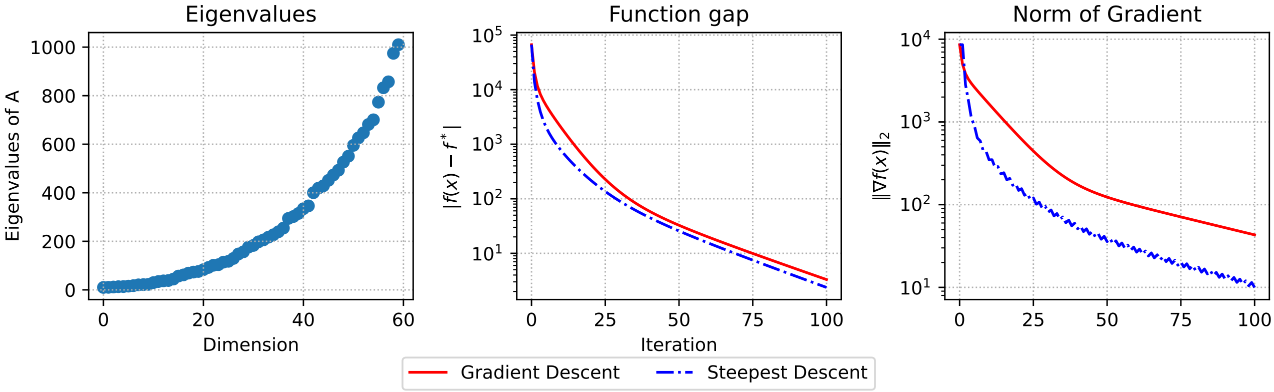
Strongly convex quadratics. n=60, random matrix.



Численные эксперименты

$$
\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{1}{2} x^T A x - b^T x \right\},\ \mu = 10,\ L = 1000.
$$

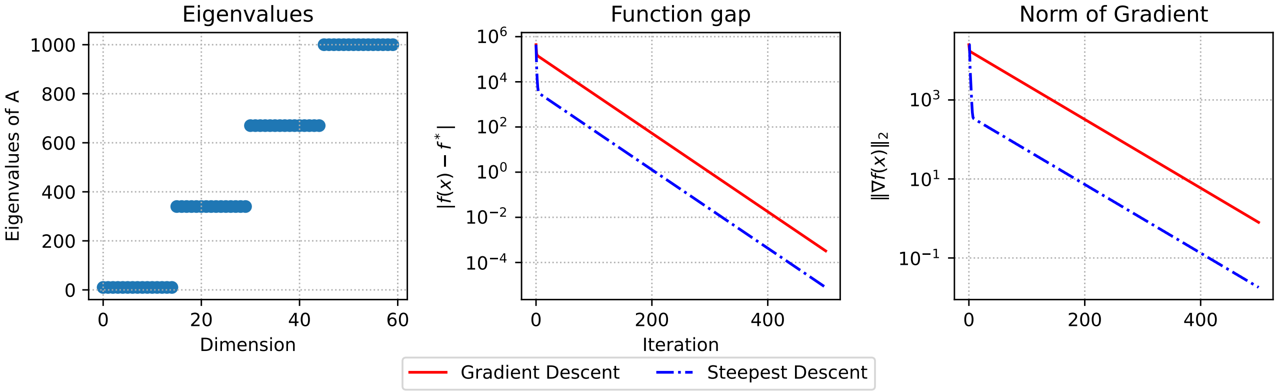
Strongly convex quadratics. n=60, random matrix.



Численные эксперименты

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{1}{2} x^T A x - b^T x \right\}, \; \mu = 10, \; L = 1000.$$

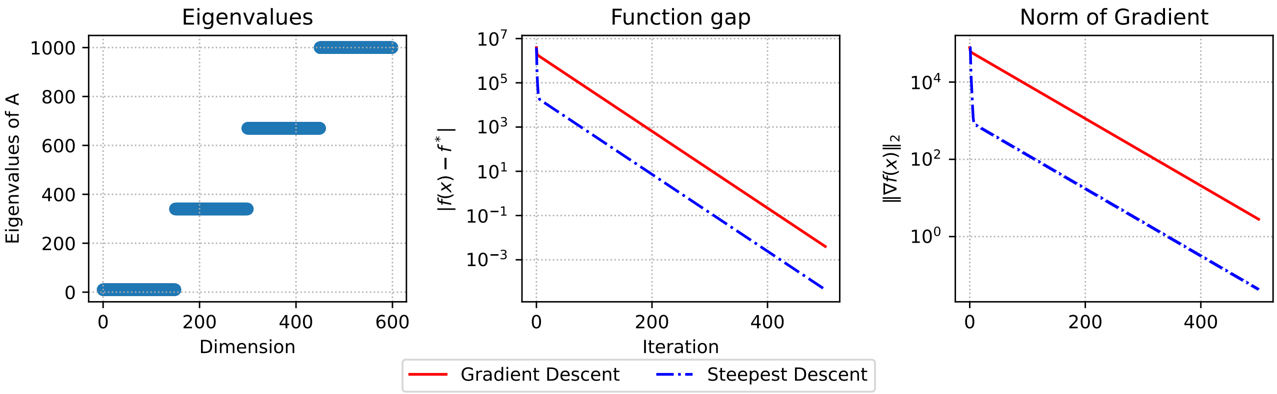
Strongly convex quadratics. n=60, clustered matrix.



Численные эксперименты

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{1}{2} x^T A x - b^T x \right\}, \ \mu = 10, \ L = 1000.$$

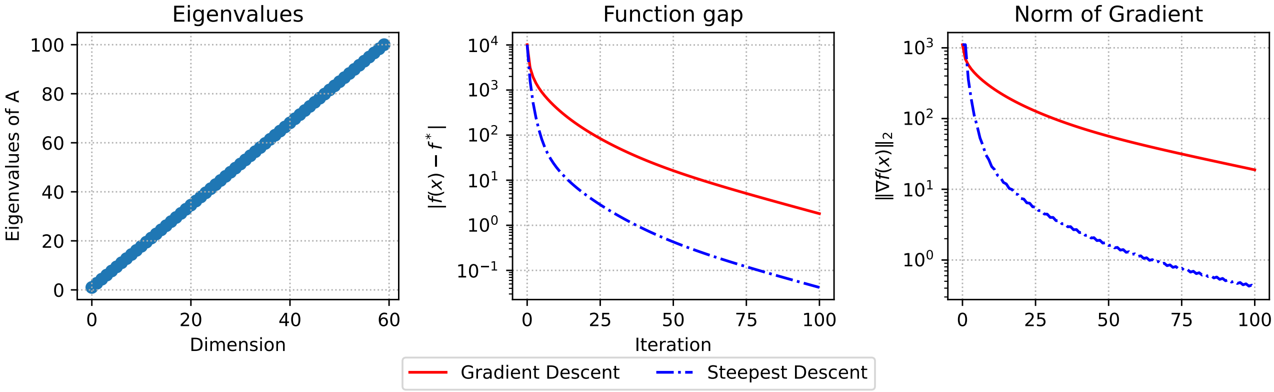
Strongly convex quadratics. n=600, clustered matrix.



Численные эксперименты

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{1}{2} x^T A x - b^T x \right\},\ \mu = 10,\ L = 1000.$$

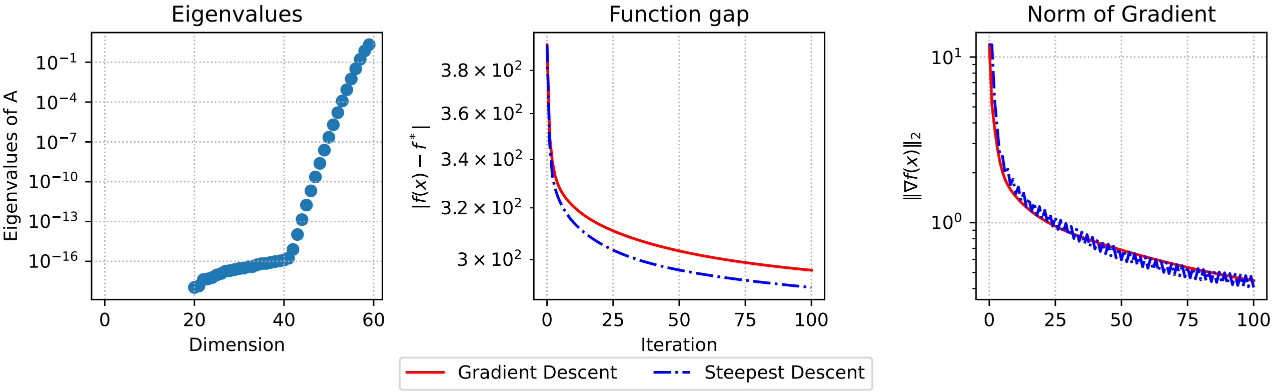
Strongly convex quadratics. n=60, uniform spectrum matrix.



Численные эксперименты

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{1}{2} x^T A x - b^T x \right\}$$

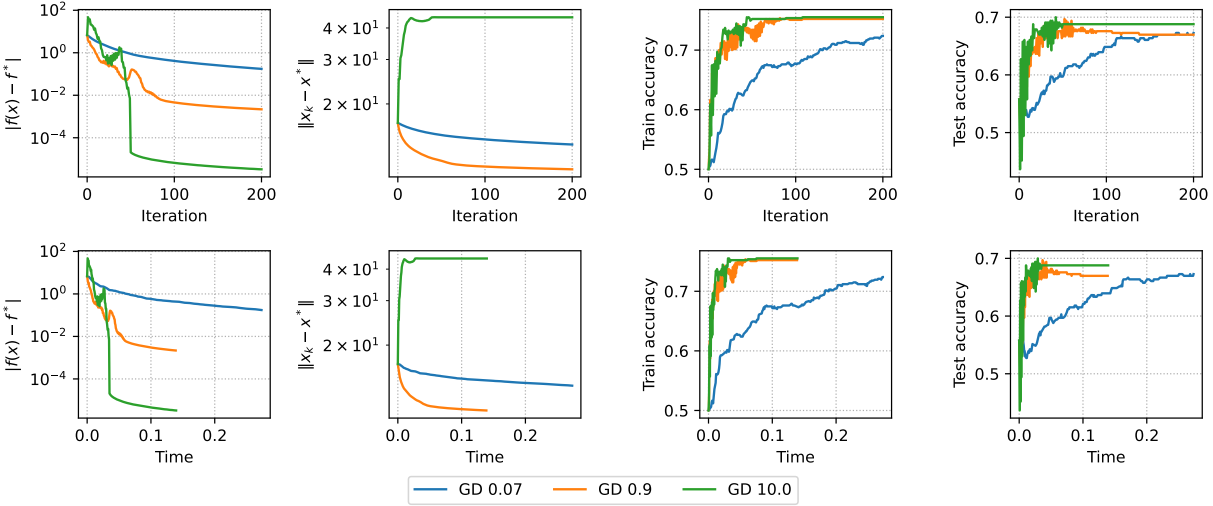
Strongly convex quadratics. n=60, Hilbert matrix.



Численные эксперименты

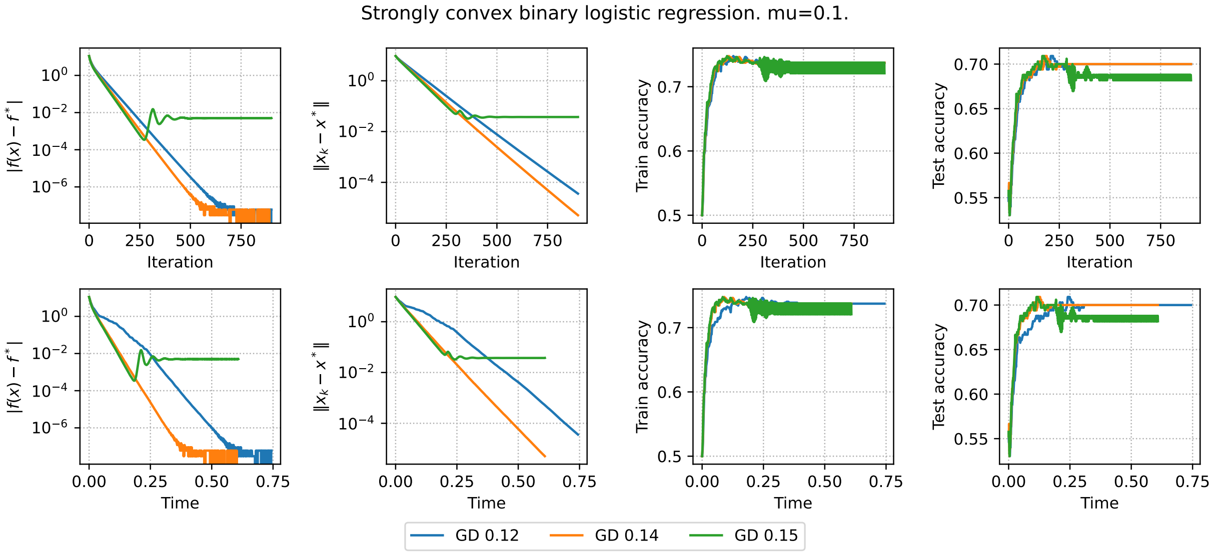
$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{\mu}{2}\|x\|_2^2 + \frac{1}{m}\sum_{i=1}^{m} \log(1 + \exp(-y_i\langle a_i, x\rangle)) \right\}$$

Convex binary logistic regression. mu=0.



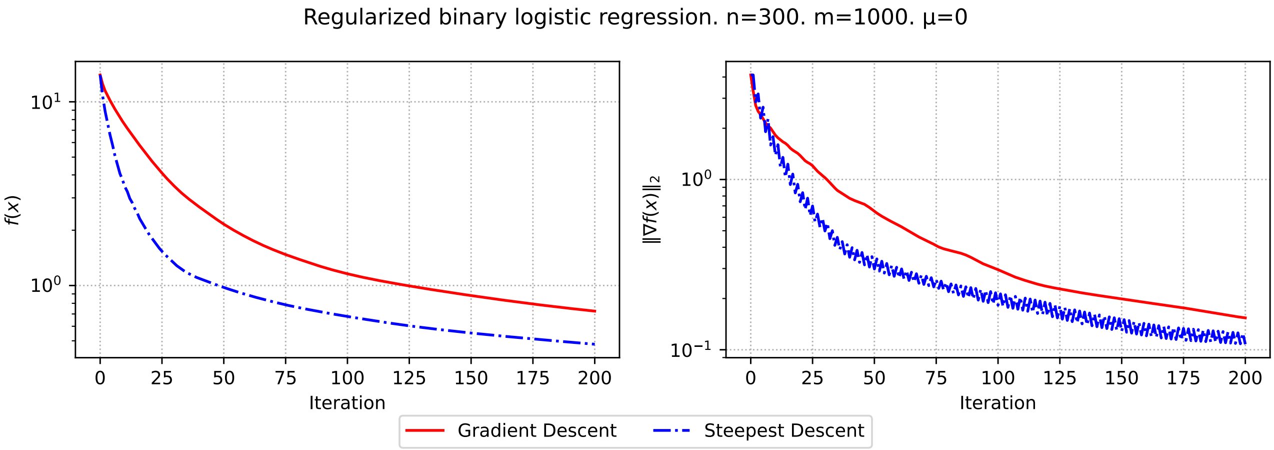
Численные эксперименты

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{\mu}{2}\|x\|_2^2 + \frac{1}{m}\sum_{i=1}^{m} \log(1 + \exp(-y_i \langle a_i, x \rangle)) \right\}$$



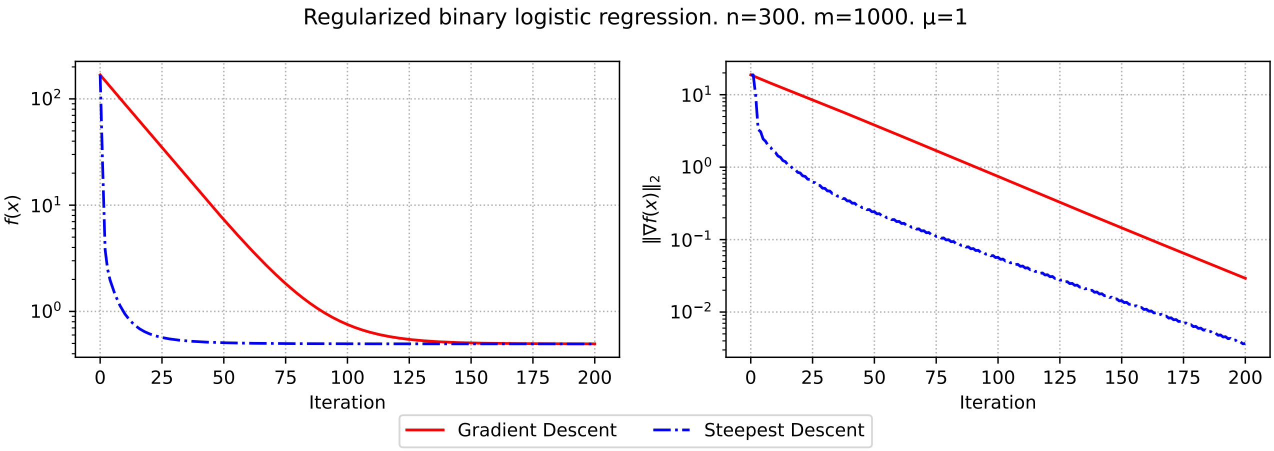
Численные эксперименты

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{\mu}{2}\|x\|_2^2 + \frac{1}{m}\sum_{i=1}^{m} \log(1 + \exp(-y_i\langle a_i, x\rangle)) \right\}$$



Численные эксперименты

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) = \frac{\mu}{2}\|x\|_2^2 + \frac{1}{m}\sum_{i=1}^{m} \log(1 + \exp(-y_i\langle a_i, x\rangle)) \right\}$$



Экстра: Дифференциальное уравнение градиентного потока

Рассмотрим дифференциальное уравнение градиентного потока:

$$\frac{dx}{dt} = -\nabla f(x(t)). \tag{GF}$$

Экстра: Дифференциальное уравнение градиентного потока

Рассмотрим дифференциальное уравнение градиентного потока:

$$\frac{dx}{dt} = -\nabla f(x(t)). \tag{GF}$$

Дискретизируем его на равномерной сетке с шагом α:

$$\frac{x^{k+1} - x^k}{\alpha} = -\nabla f(x^k),$$

Экстра: Дифференциальное уравнение градиентного потока

Рассмотрим дифференциальное уравнение градиентного потока:

$$\frac{dx}{dt} = -\nabla f(x(t)). \tag{GF}$$

Дискретизируем его на равномерной сетке с шагом α:

$$\frac{x^{k+1} - x^k}{\alpha} = -\nabla f(x^k),$$

где $x^k \equiv x(t_k)$ и $\alpha = t_{k+1} - t_k$ — шаг сетки.
Отсюда получаем выражение для x^{k+1}:

$$x^{k+1} = x^k - \alpha \nabla f(x^k),$$

являющееся точной формулой обновления градиентного спуска.
Открыть в Colab ♣

Экстра: Дифференциальное уравнение градиентного потока

Рассмотрим дифференциальное уравнение градиентного потока:

$$\frac{dx}{dt} = -\nabla f(x(t)). \tag{GF}$$

Дискретизируем его на равномерной сетке с шагом α:

$$\frac{x^{k+1} - x^k}{\alpha} = -\nabla f(x^k),$$

где $x^k \equiv x(t_k)$ и $\alpha = t_{k+1} - t_k$ — шаг сетки.
Отсюда получаем выражение для x^{k+1}:

$$x^{k+1} = x^k - \alpha \nabla f(x^k),$$

являющееся точной формулой обновления градиентного спуска.
Открыть в Colab ♣

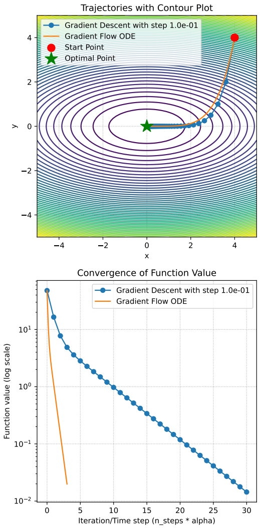


Рисунок 6. Траектория градиентного потока