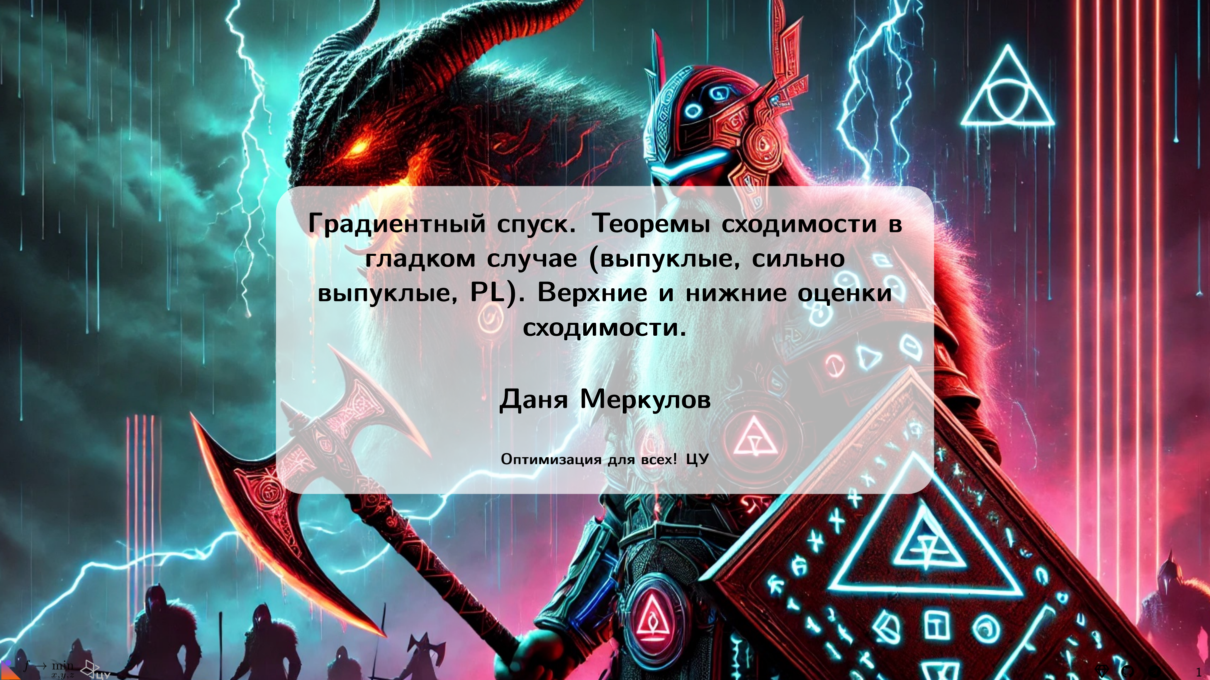


Градиентный спуск. Теоремы сходимости в гладком случае (выпуклые, сильно выпуклые, PL). Верхние и нижние оценки сходимости.

Даня Меркулов

Оптимизация для всех! ЦУ

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:

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

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

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:

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

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

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

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

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:

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

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

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

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

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:

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

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

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

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

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

$$h = -\frac{\nabla f(x)}{\|\nabla f(x)\|_2}$$

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:

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

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

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

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

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

$$h = -\frac{\nabla f(x)}{\|\nabla f(x)\|_2}$$

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

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

Дифференциальное уравнение градиентного потока

Рассмотрим дифференциальное уравнение градиентного потока:

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

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:

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

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 ♣

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 ♣

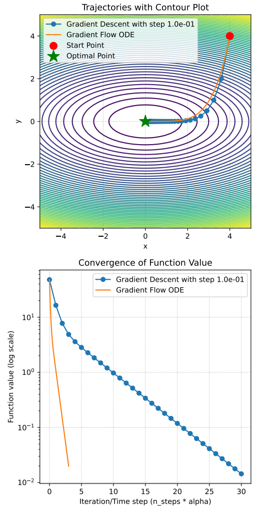
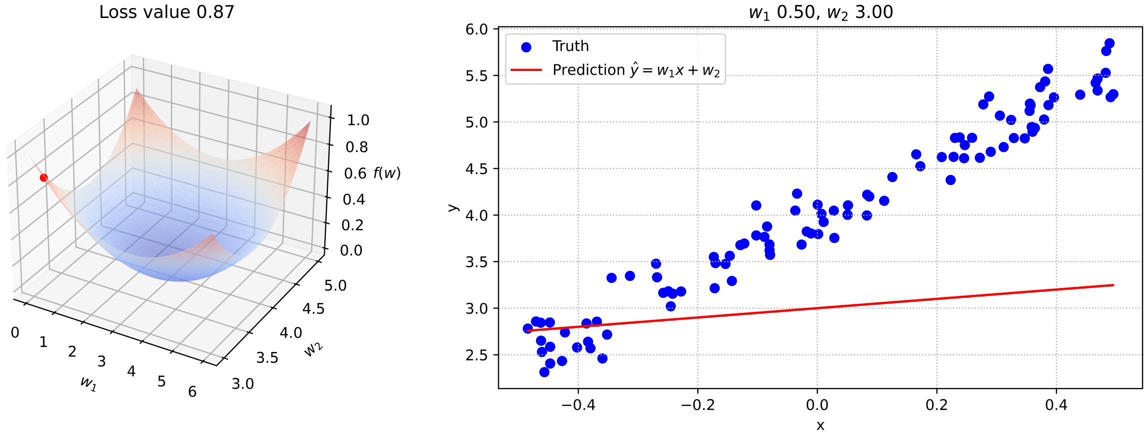


Рис. 1: Траектория градиентного потока

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)

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

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

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

Условия оптимальности:

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

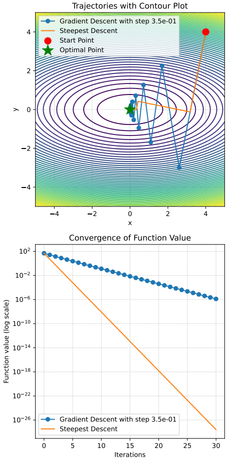


Рис. 2: Наискорейший спуск

Открыть в Colab ♣

Сильно выпуклые квадратичные функции

Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

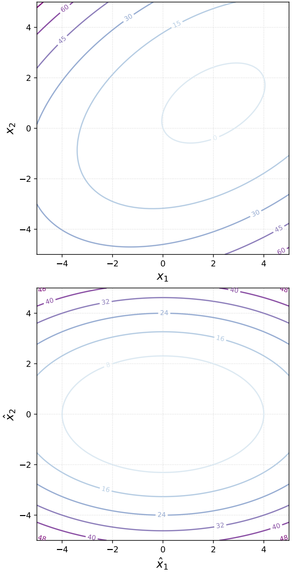
$$
\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{2} x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.
$$

Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

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

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

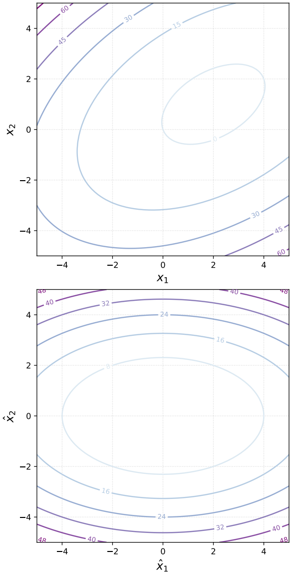


Сдвиг координат

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 $A = Q \Lambda Q^T$.

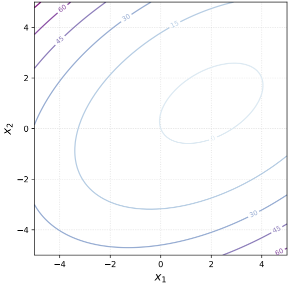


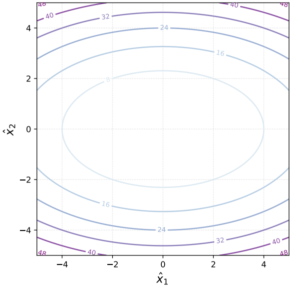
Сдвиг координат

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 $x = Q\hat{x} + x^*$.





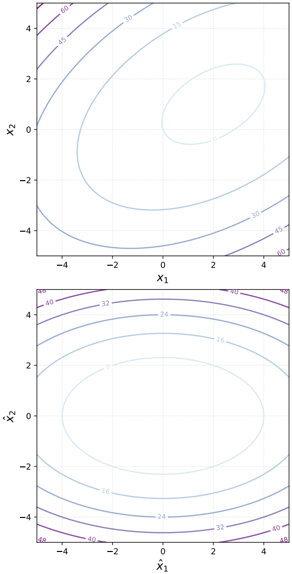
Сдвиг координат

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 $x = Q\hat{x} + x^*$.

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



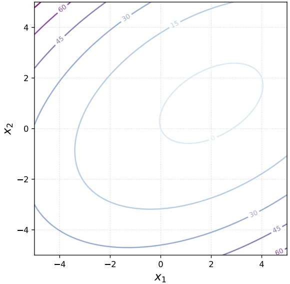
Сдвиг координат

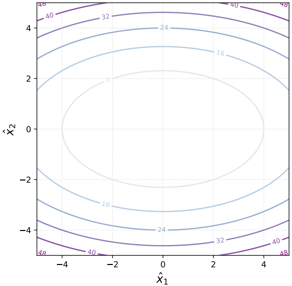
Рассмотрим следующую задачу квадратичной оптимизации:

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

- 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 $x = Q\hat{x} + x^*$.

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





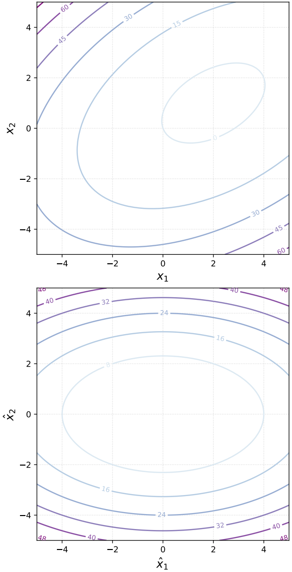
Сдвиг координат

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 $x = Q\hat{x} + x^*$.

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



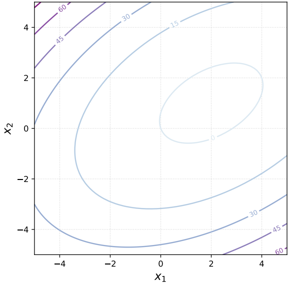
Сдвиг координат

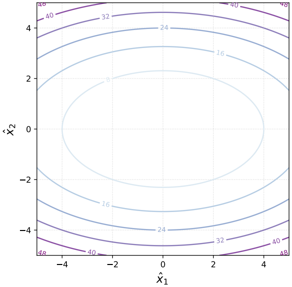
Рассмотрим следующую задачу квадратичной оптимизации:

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

- 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 $x = Q\hat{x} + x^*$.

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





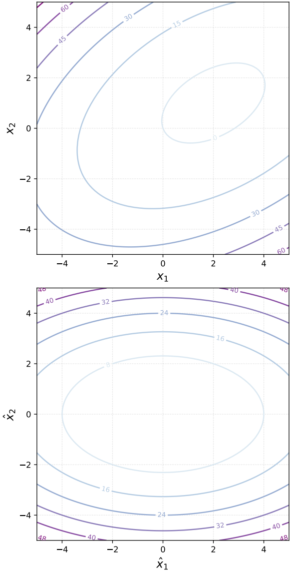
Сдвиг координат

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 $x = Q\hat{x} + x^*$.

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



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 $\hat{x}$)

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

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 $\hat{x}$)

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

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

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

. . .

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

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 $\hat{x}$)

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

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

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

. . .

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

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

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:

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

. . .

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:

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

. . .

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

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

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 $\hat{x}$)

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

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

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

. . .

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

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

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 $\hat{x}$)

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

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

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

. . .

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

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

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 $\hat{x}$)

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

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

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

. . .

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

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

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

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

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 $\hat{x}$)

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

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

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

. . .

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

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

Выберем α, минимизирующий худший знаменатель прогрессии

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

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 $\hat{x}$)

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

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

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

. . .

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

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

Выберем α, минимизирующий худший знаменатель прогрессии

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

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 $\hat{x}$)

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

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

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

. . .

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

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

Выберем α, минимизирующий худший знаменатель прогрессии

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

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:

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

. . .

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

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

Выберем α, минимизирующий худший знаменатель прогрессии

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

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:

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

. . .

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

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

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:

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

. . .

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

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

Выберем α, минимизирующий худший знаменатель прогрессии

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

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

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

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

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

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

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:

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

. . .

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

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

Выберем α, минимизирующий худший знаменатель прогрессии

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

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

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

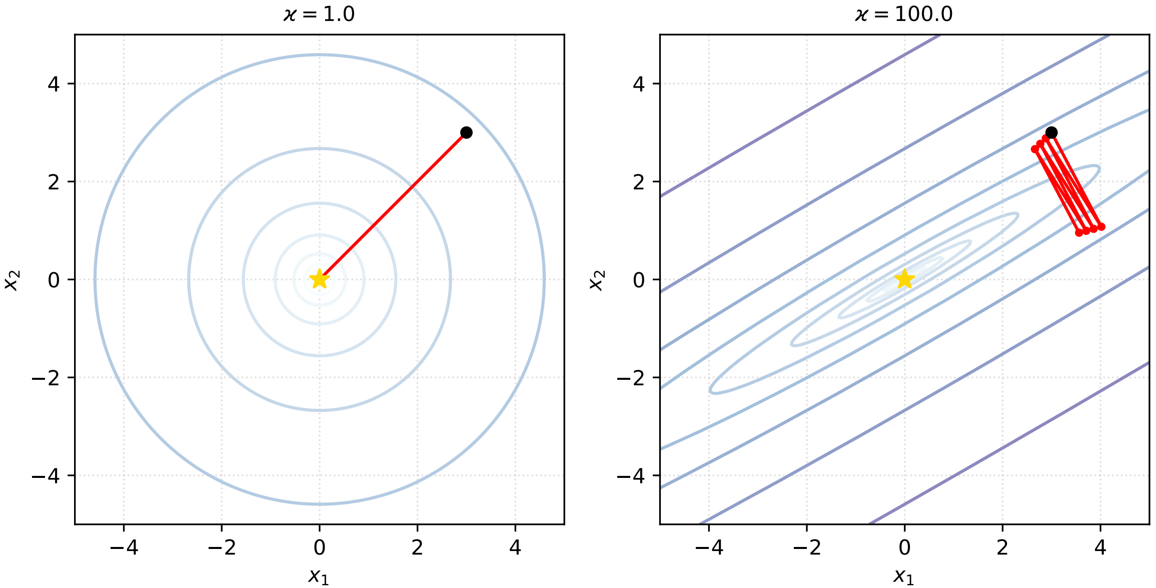
$$
\|x^k\|_2 \leq \left(\frac{L - \mu}{L + \mu}\right)^k \|x^0\|_2 \quad f(x^k) \leq \left(\frac{L - \mu}{L + \mu}\right)^{2k} f(x^0)
$$

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

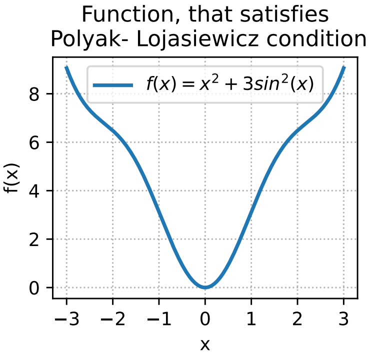
Говорят, что f удовлетворяет условию Поляка-Лоясиевича (PL), если для некоторого $\mu > 0$ выполняется

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

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

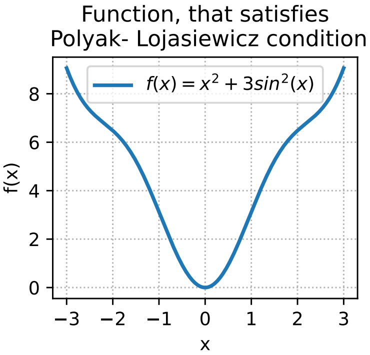
$$\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*) \quad \forall x$$

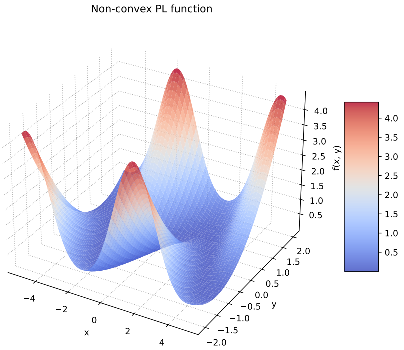
Интересно, что градиентный спуск может сходиться линейно даже без выпуклости.

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

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

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 $f^* = \min\limits_{x \in \mathbb{R}^d} f(x)$. Тогда:

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

Любая μ-сильно выпуклая дифференцируемая функция является PL-функцией

ℹ Theorem

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

Выпуклый гладкий случай

Выпуклый гладкий случай

Theorem

Рассмотрим задачу

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

и предположим, что f 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 $0 < \alpha \leq \frac{1}{L}$. Пусть $f^* = \min_{x \in \mathbb{R}^d} f(x)$. Тогда для всех $x^* \in \operatorname{argmin} f$ и всех $k \in \mathbb{N}$ справедливо:

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

Итог

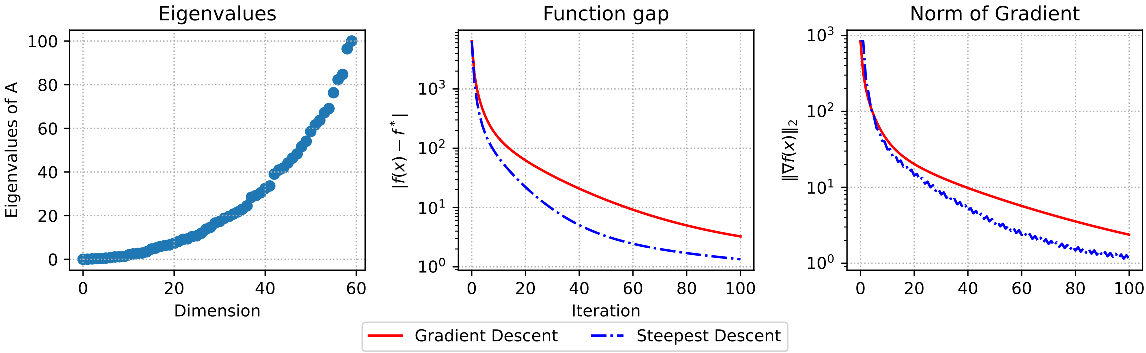
Градиентный спуск: $\min\limits_{x \in \mathbb{R}^n} f(x)$ $x^{k+1} = x^k - \alpha^k \nabla f(x^k)$

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

Численные эксперименты

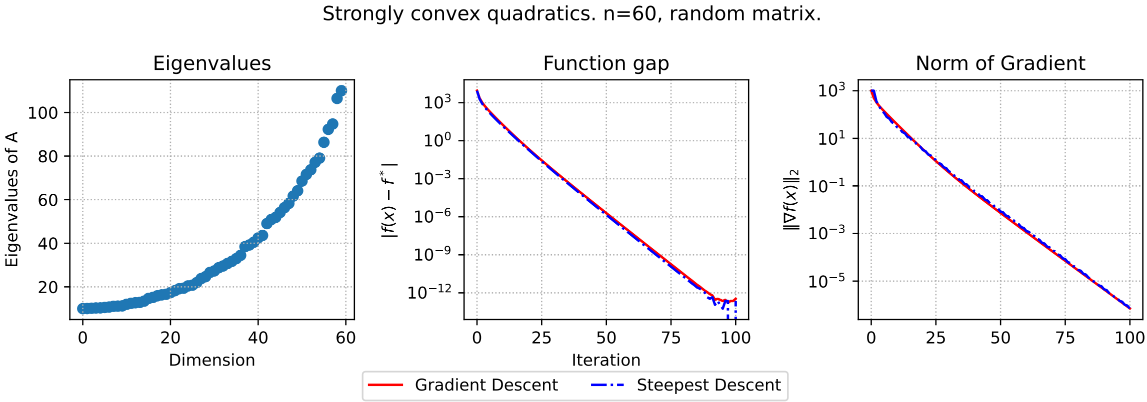
$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$

Convex quadratics. n=60, random matrix.



Численные эксперименты

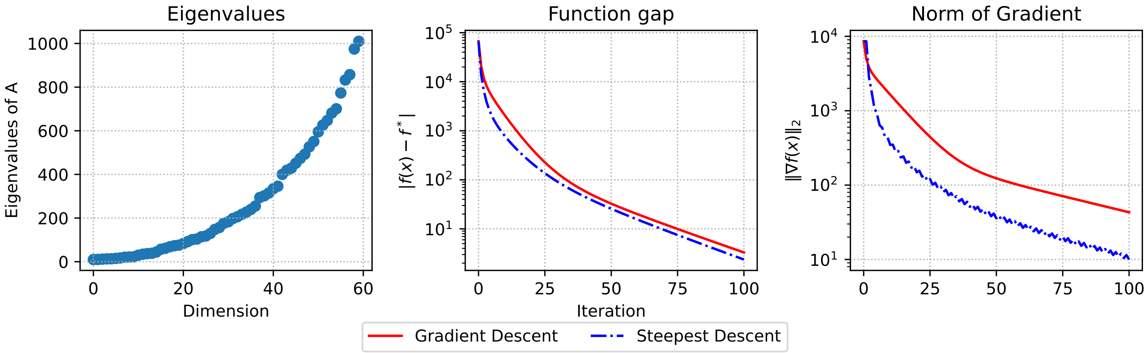
$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$



Численные эксперименты

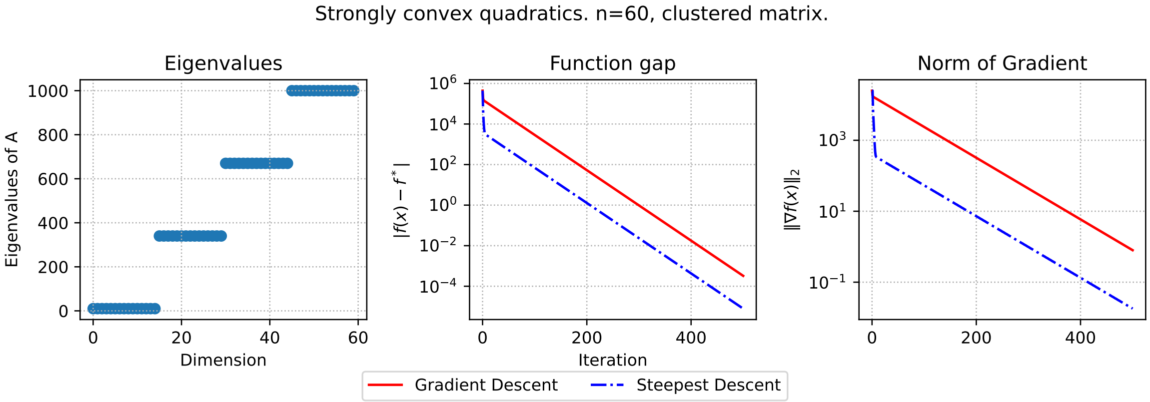
$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$

Strongly convex quadratics. n=60, random matrix.



Численные эксперименты

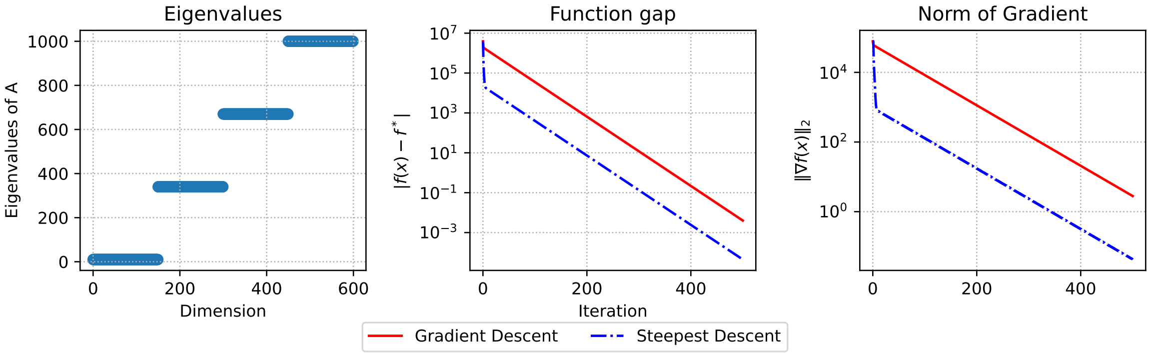
$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$



Численные эксперименты

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

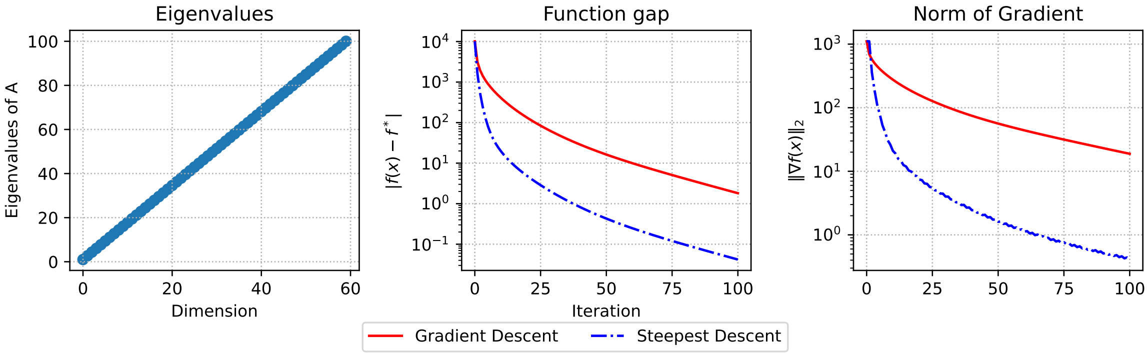
Strongly convex quadratics. n=600, clustered matrix.



Численные эксперименты

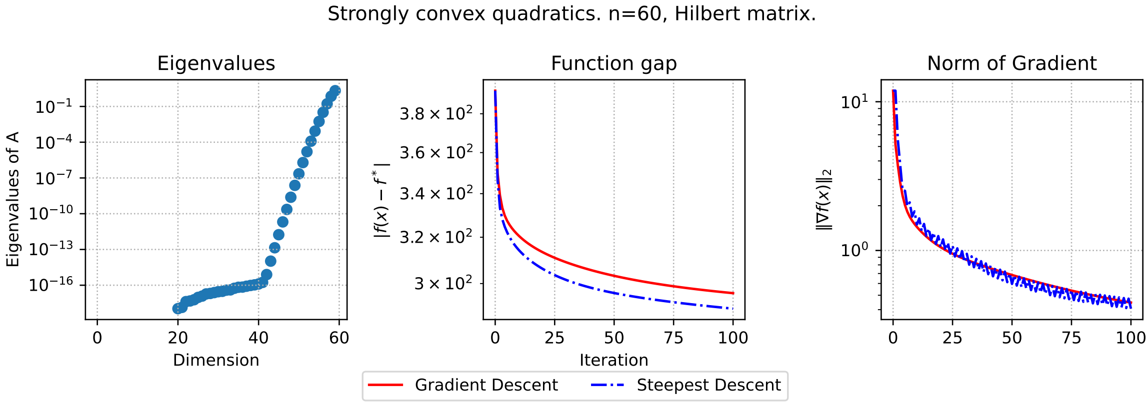
$$f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}$$

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



Численные эксперименты

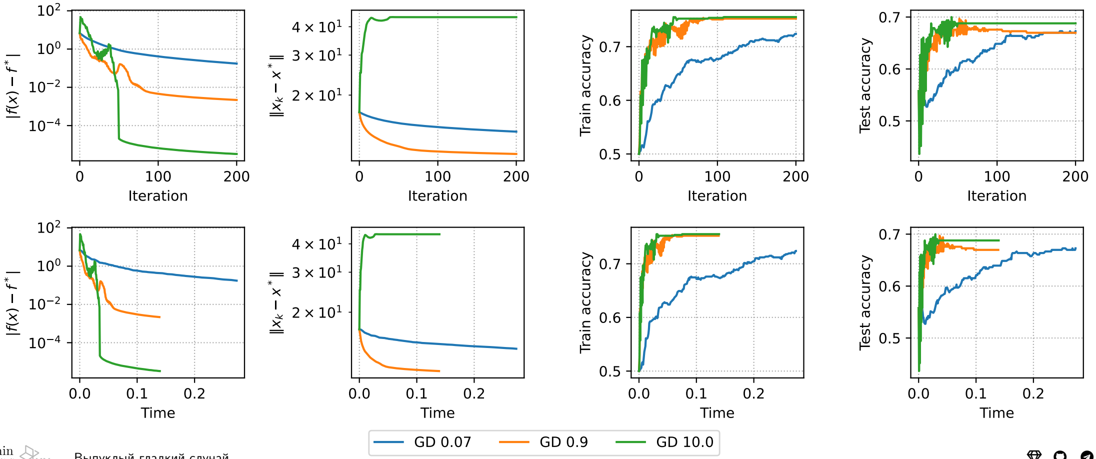
$$f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}$$



Численные эксперименты

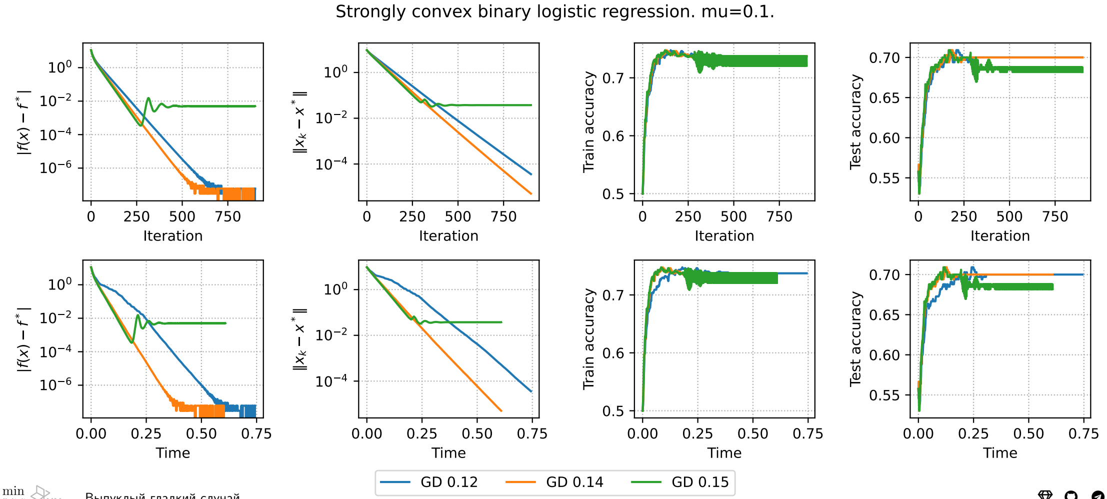
$$f(x) = \frac{\mu}{2}\|x\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\log(1+\exp(-y_i\langle a_i, x\rangle)) \to \min_{x\in\mathbb{R}^n}$$

Convex binary logistic regression. mu=0.



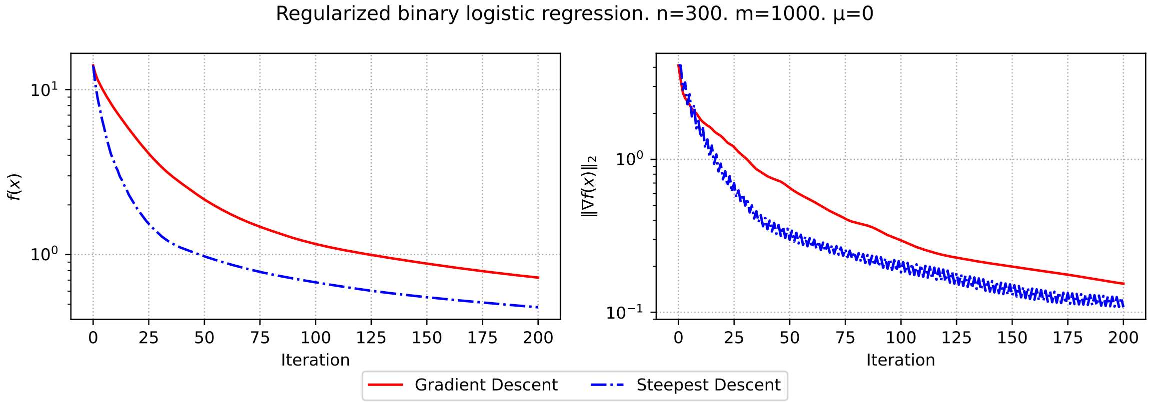
Численные эксперименты

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



Численные эксперименты

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



Численные эксперименты

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

Regularized binary logistic regression. n=300. m=1000. μ=1