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

Даниил Меркулов

Методы оптимизации. МФТИ

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

шаг

## 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:

$$\boxed{f(x + \alpha h) < f(x)}$$

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

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:

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

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

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:

$$\boxed{|\langle \nabla f(x), h\rangle| \leq \|\nabla f(x)\|_2 \|h\|_2}$$

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

$$-\|\cdot\|\,\|\cdot\| \leq \langle \quad \rangle \leq \|\cdot\|\,\|\cdot\|$$

# 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

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

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:

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

# Дифференциальное уравнение градиентного потока

Рассмотрим дифференциальное уравнение градиентного потока:

$x \in \mathbb{R}^n$

$x(t)$

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

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

$\alpha$ $\quad$ $0$

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

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

# 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 $\alpha$:

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

# 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 $\alpha$:

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

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:

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

$\nabla f(x_k) = a x_k$

$\frac{dx}{dt} = -ax$

$x(t) = e^{-at} \cdot x(0)$

Дискретизируем его на равномерной сетке с шагом $\alpha$:

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

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 ♣

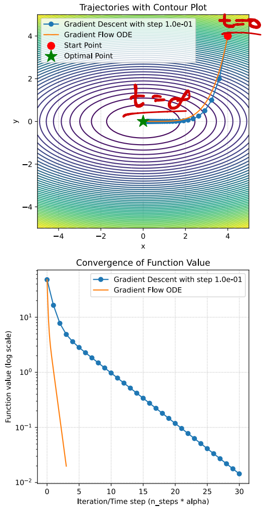
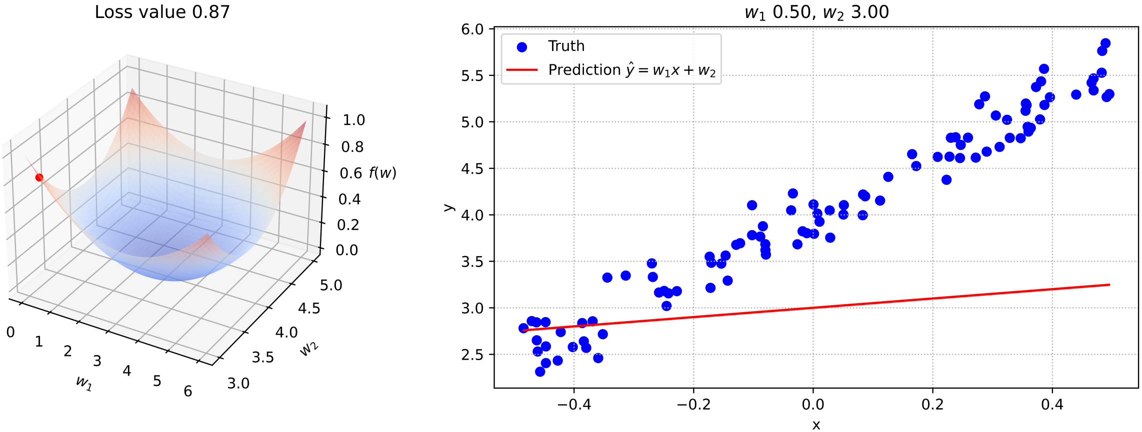



Рис. 1: Траектория градиентного потока

# Сходимость градиентного спуска

Существенно зависит от выбора шага $\alpha$:



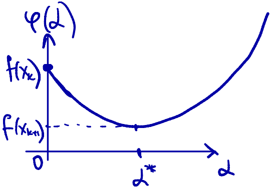
# Точный линейный поиск (метод наискорейшего спуска)

$$
\alpha_k = \arg\min_{\alpha \in \mathbb{R}^+} f(\underbrace{x^k - \alpha \nabla f(x^k)}_{x^{k+1}(\alpha)})
$$

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

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.

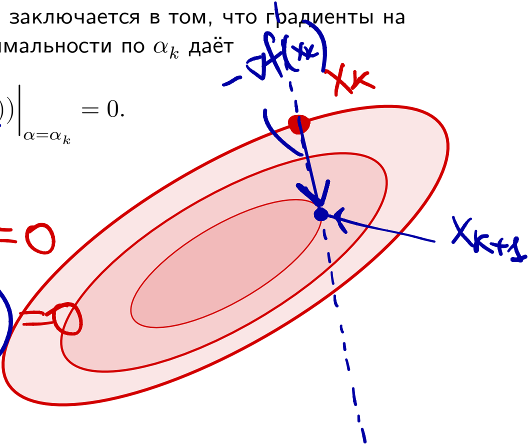
Интересное теоретическое свойство этого метода заключается в том, что градиенты на соседних итерациях ортогональны. Условие оптимальности по $\alpha_k$ даёт




$$
\frac{d}{d\alpha} f(\underbrace{x^k - \alpha \nabla f(x^k)}_{x^{k+1}(\alpha)})\Big|_{\alpha = \alpha_k} = 0.
$$

$$\frac{\partial f}{\partial x_{k+1}} \cdot \frac{\partial x_{k+1}}{\partial \alpha} = 0$$

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



# 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 $\alpha_k$ 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 $\alpha_k$ даёт

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

Условия оптимальности:

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

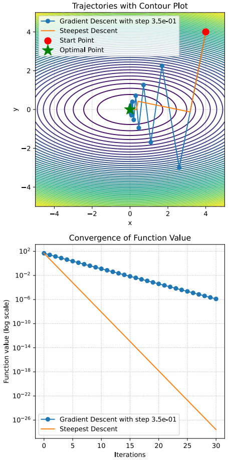



Рис. 2: Наискорейший спуск

Открыть в Colab ♣

# Сильно выпуклые квадратичные функции

# Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

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

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

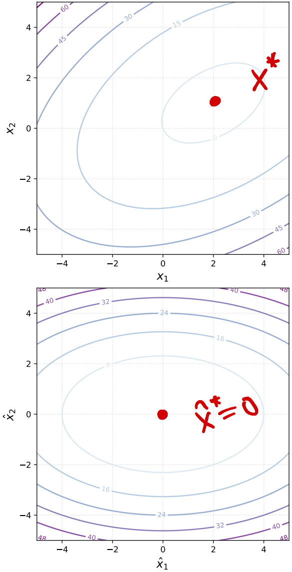
nx1 nx1 1x1 nx1

# 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.

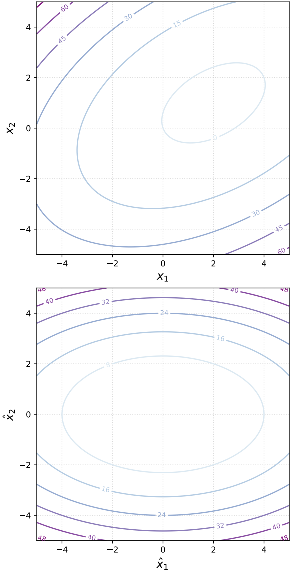


# Сдвиг координат

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 $\boxed{A = Q \Lambda Q^T}$.

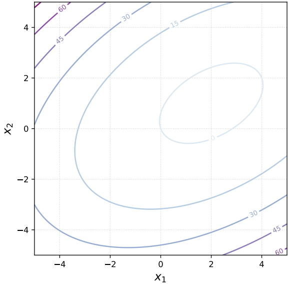


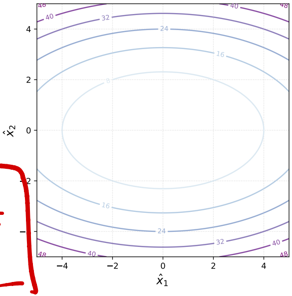
# Сдвиг координат

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 $\boxed{\hat{x} = Q^T(x - x^*)}$, где $x^*$ — точка минимума исходной функции, определяемая как $Ax^* = b$. При этом $x = Q\hat{x} + x^*$.




$$\hat{x} = Q^T(x - x^*) \quad \Big| \; Q \cdot$$

$$Q\hat{x} = \underbrace{QQ^T}(x - x^*)$$

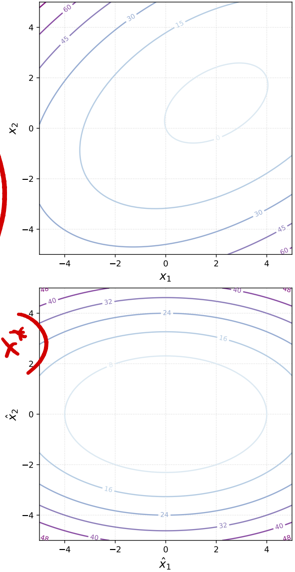
$$Q\hat{x} \overset{I}{=} x - x^* \Rightarrow \boxed{x = Q\hat{x} + x^*}$$

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




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

$=$

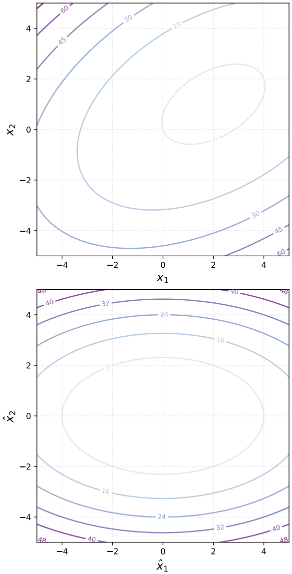
# Сдвиг координат

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

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

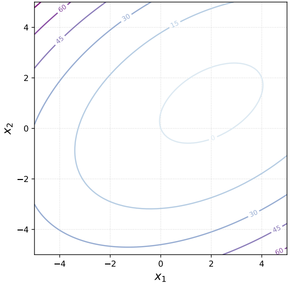


# Сдвиг координат

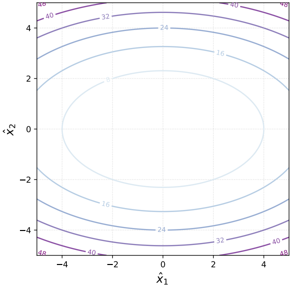
Рассмотрим следующую задачу квадратичной оптимизации:

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

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




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




$$\hat{x}^T \underbrace{Q^T Q}_{I} \Lambda \underbrace{Q^T Q}_{I} \hat{x}$$

$$\frac{1}{2} \hat{x}^T \Lambda \hat{x}$$

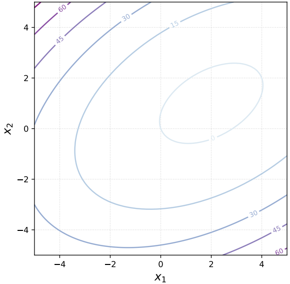
# Сдвиг координат

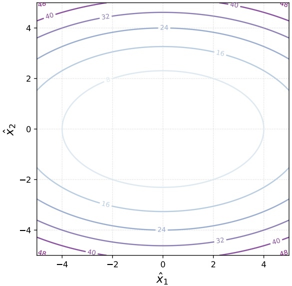
Рассмотрим следующую задачу квадратичной оптимизации:

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

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

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



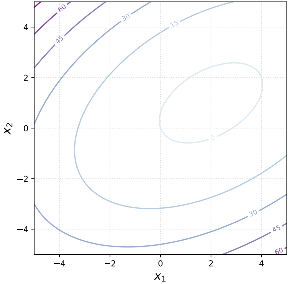
## Сдвиг координат

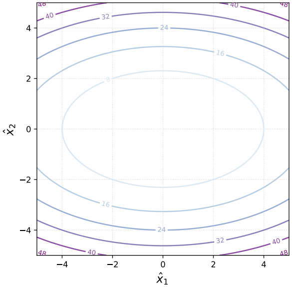
Рассмотрим следующую задачу квадратичной оптимизации:

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

- 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

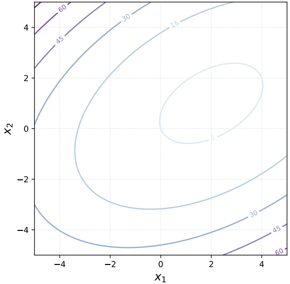
Рассмотрим следующую задачу квадратичной оптимизации:

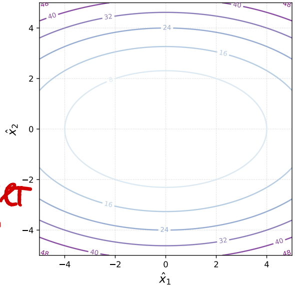
$$
\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{2} x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}^d_{++}.
$$

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

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

НАС НЕ интересует КОНСТАНТА



# Анализ сходимости

$\alpha^k = \alpha = const$

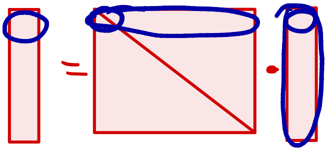
Теперь мы можем работать с функцией $\boxed{f(x) = \frac{1}{2}x^T\Lambda x}$ с $x^* = 0$ без ограничения общности (убрав крышку из $\hat{x}$)

$\nabla f = \Lambda x, \quad \nabla^2 f = \Lambda$

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

$$= x_k - \alpha \Lambda x_k =$$

$$x_{k+1} = \underline{(I - \alpha \Lambda)} x_k$$



# 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 $\hat{x}$)

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

# 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 $\hat{x}$)

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

$$
\boxed{x^{k+1}_{(i)} = \left(1 - \alpha^k \lambda_{(i)}\right) x^k_{(i)} \quad \text{для } i\text{-й координаты}}
$$

$$|1 - \alpha\lambda| < 1$$

$$
\begin{aligned}
x_{k+1} &= (1 - \alpha\lambda) x_k = \\
&= (1 - \alpha\lambda)^2 x_{k-1} = \\
&= (1 - \alpha\lambda)^3 x_{k-2}
\end{aligned}
$$

# 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 $\hat{x}$)

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

# 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 $\hat{x}$)

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

$\lambda_{min}(A) =$

$= \lambda_{min}(\Lambda) = \mu$

$\lambda_{max}(\Lambda) = L$

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

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

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

$|1 - \alpha\lambda_1|, |1 - \alpha\lambda_2|, |1 - \alpha\lambda_3| \ldots$

$n$ штук

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

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

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

$0 < 2 - \alpha\mu < 2$

$\alpha < \frac{2}{\mu} \qquad \alpha > 0$

$|1 - \alpha L| < 1$

$\alpha < \frac{2}{L}, \quad \alpha > 0$

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

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

$$
|1 - \alpha L| < 1
$$

## 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:

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

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:

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

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

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

# 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:

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

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 $\hat{x}$)

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

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

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

Выберем $\alpha$, 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 $\hat{x}$)

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

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

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

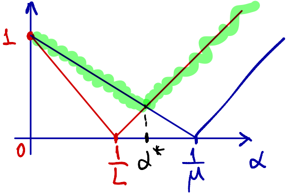
Помним, что $\lambda_{\min} = \mu > 0$, $\lambda_{\max} = L \geq \mu$.

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

Выберем $\alpha$, минимизирующий худший знаменатель прогрессии

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

max




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

$$\boxed{\alpha^* = \frac{2}{\mu + L}}$$

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

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

Выберем $\alpha$, минимизирующий худший знаменатель прогрессии

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

# 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 $\alpha$, минимизирующий худший знаменатель прогрессии

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

# 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 $\hat{x}$)

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

$$= (I - \alpha^k \Lambda) x^k$$

$$x^{k+1}_{(i)} = \left(1 - \alpha^k \lambda_{(i)}\right) x^k_{(i)} \quad \text{для } i\text{-й координаты}$$

$$x^k_{(i)} = \left(1 - \alpha\, \lambda_{(i)}\right)^k x^0_{(i)} \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

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

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

$$-1 < 1 - \alpha\mu < 1 \qquad\qquad -1 < 1 - \alpha L < 1$$

$$\alpha < \frac{2}{\mu} \quad \alpha\mu > 0 \qquad\qquad \alpha < \frac{2}{L} \quad \alpha L > 0$$

Выберем $\alpha$, минимизирующий худший знаменатель прогрессии

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

$$= \min_\alpha \left\{|1 - \alpha\mu|, |1 - \alpha L|\right\}$$

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

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

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

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

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

Выберем $\alpha$, минимизирующий худший знаменатель прогрессии

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

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

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

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

$\|x^k - x^*\|$

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

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

Выберем $\alpha$, минимизирующий худший знаменатель прогрессии

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

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

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

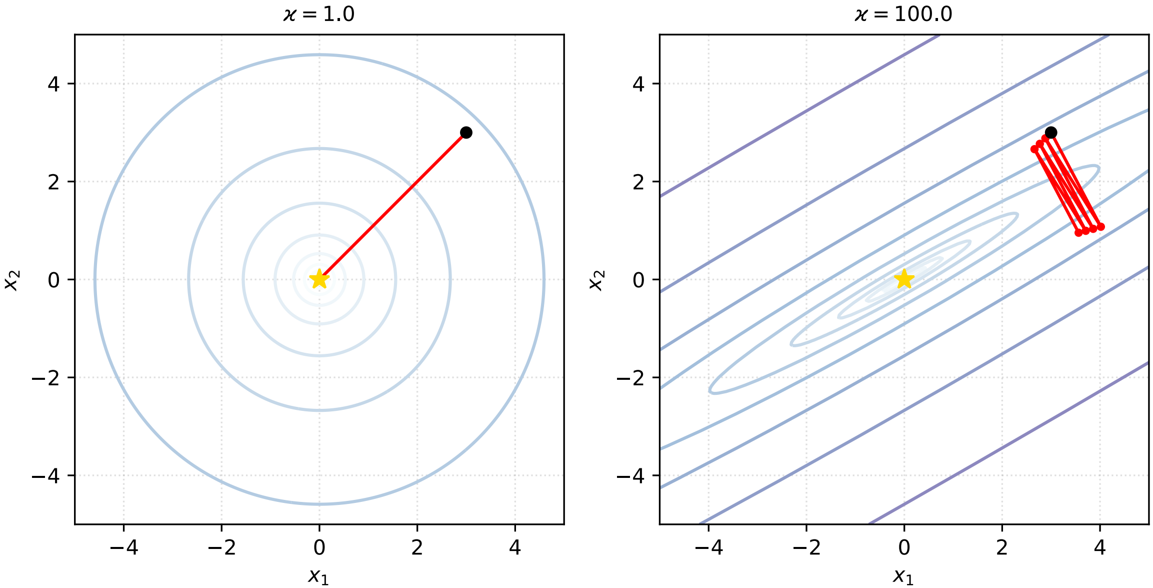
$$
\|x^k\|_2 \leq \left(\frac{L - \mu}{L + \mu}\right)^k \|x^0\|_2 \qquad f(x^k) \leq \left(\frac{L - \mu}{L + \mu}\right)^{2k} f(x^0)
$$

# 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.

| $\varkappa$ | $\rho$ | 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

Поляка - Лоясиевича

# PL-функции. Линейная сходимость градиентного спуска без выпуклости

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

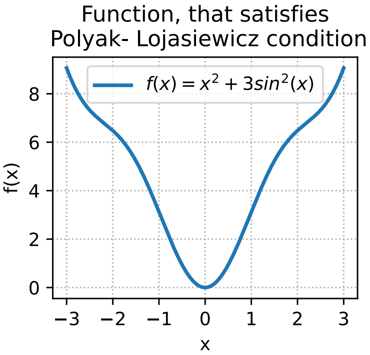
$$\boxed{\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*) \quad \forall x}$$

$f^*$

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

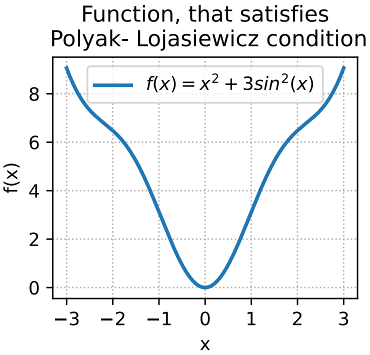
$$\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*) \quad \forall x$$

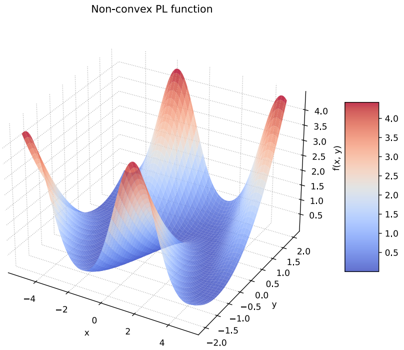
Интересно, что градиентный спуск может сходиться линейно даже без выпуклости.

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

шаг $\alpha = \frac{1}{L}$

константа лин. сх-ти: $1 - \frac{\mu}{L}$

**Theorem**

Рассмотрим задачу

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

и предположим, что $f$ является PL-функцией с константой $\mu$ 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 $\alpha$, удовлетворяющим $\boxed{0 < \alpha \leq \frac{1}{L}}$. Пусть $f^* = \min_{x \in \mathbb{R}^d} f(x)$. Тогда:

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

$$\left(1 - \frac{\mu}{L}\right)^k \left(f(x_0) - f^*\right)$$

# Анализ сходимости

Липшицева парабола

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

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

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

# Анализ сходимости

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

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

$$
= f(x_k) + \left(-\alpha + \frac{L\alpha^2}{2}\right)\|\nabla f(x_k)\|_2^2
$$

$$
f(x_k) - f(x_{k+1}) \geq \left(\alpha - \frac{L\alpha^2}{2}\right)\|\nabla f(x_k)\|_2^2
$$

# Анализ сходимости

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

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

# Анализ сходимости

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

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

$\alpha L \leq 1$

$2 - L\alpha \geq 1$

$-(2 - L\alpha) \leq -1$

$\|\nabla f(x_k)\|_2^2 \geq 2\mu\left(f(x_k) - f^*\right)$

# 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:

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

$\geq 2\mu(f - f^*)$

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

Теперь используем свойство PL-функции и получаем:

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

$-f^*$ $\quad$ $-f^*$

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

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

# Любая $\mu$-сильно выпуклая дифференцируемая функция является PL-функцией

**Theorem**

Если функция $f(x)$ дифференцируема и $\mu$-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 $\mu$-сильно выпуклая дифференцируемая функция является PL-функцией (с константой $\mu$)

**Theorem**

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

**Доказательство**

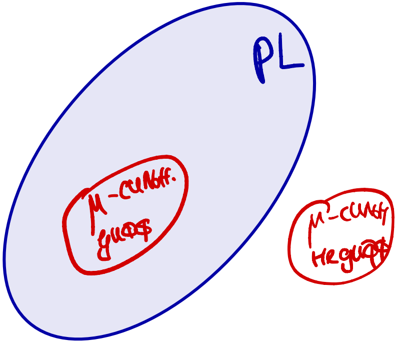
По критерию сильной выпуклости первого порядка:

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

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 $\mu$-сильно выпуклая дифференцируемая функция является PL-функцией

**Theorem**

Если функция $f(x)$ дифференцируема и $\mu$-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 $y = x^*$:

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

# Любая $\mu$-сильно выпуклая дифференцируемая функция является PL-функцией

**Theorem**

Если функция $f(x)$ дифференцируема и $\mu$-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 $y = x^*$:

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

# Любая $\mu$-сильно выпуклая дифференцируемая функция является PL-функцией

**Theorem**

Если функция $f(x)$ дифференцируема и $\mu$-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 $y = x^*$:

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

# Любая $\mu$-сильно выпуклая дифференцируемая функция является PL-функцией

> **Theorem**
>
> Если функция $f(x)$ дифференцируема и $\mu$-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 $y = x^*$:

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

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

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

# Любая $\mu$-сильно выпуклая дифференцируемая функция является PL-функцией

**Theorem**

Если функция $f(x)$ дифференцируема и $\mu$-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 $y = x^*$:

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

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 $\mu$-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 $\mu$-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 $\mu$-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 $\mu$-сильно выпуклая дифференцируемая функция является PL-функцией

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

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

$R = \|x_0 - x^*\|$

**Theorem**

Рассмотрим задачу

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

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 $\alpha$, 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:

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

$f(x_k) - f^* \leq \frac{L \cdot R^2}{2k}$

# Анализ сходимости

* Как и раньше, сначала используем гладкость:

(1)

# Анализ сходимости

- Как и раньше, сначала используем гладкость:

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

Липшицева
парабола

# Анализ сходимости

- Как и раньше, сначала используем гладкость:

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

$$
f(x_k) + \left(\frac{L\alpha^2}{2} - \alpha\right)\|\nabla f(x_k)\|_2^2
$$

# Анализ сходимости

- Как и раньше, сначала используем гладкость:

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

# Анализ сходимости

- Как и раньше, сначала используем гладкость:

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

## Анализ сходимости

- Как и раньше, сначала используем гладкость:

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

# Анализ сходимости

- Как и раньше, сначала используем гладкость:

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

$$
\boxed{f(x^k) - f(x^{k+1}) \geq \frac{1}{2L} \|\nabla f(x^k)\|^2 \quad \text{если } \alpha = \frac{1}{L}}
$$

## Анализ сходимости

- Как и раньше, сначала используем гладкость:

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

  Обычно для сходящегося градиентного спуска чем больше допустимый шаг, тем быстрее сходимость, поэтому часто берут $\alpha = \frac{1}{L}$.

- После этого используем выпуклость:

$$
\tag{2}
$$

# Анализ сходимости

- Как и раньше, сначала используем гладкость:

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

$$
f(x^k) - f(x^{k+1}) \geq \frac{1}{2L}\|\nabla f(x^k)\|^2 \quad \text{если } \alpha = \frac{1}{L}
$$

Обычно для сходящегося градиентного спуска чем больше допустимый шаг, тем быстрее сходимость, поэтому часто берут $\alpha = \frac{1}{L}$.

- После этого используем выпуклость:

$$
\underline{f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle} \tag{2}
$$

$x = x_k$

$y =$

# Анализ сходимости

- Как и раньше, сначала используем гладкость:

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

  Обычно для сходящегося градиентного спуска чем больше допустимый шаг, тем быстрее сходимость, поэтому часто берут $\alpha = \frac{1}{L}$.

- После этого используем выпуклость:

$$
f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle \text{ где } y = x^*, x = x^k \tag{2}
$$

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

## Анализ сходимости

- Как и раньше, сначала используем гладкость:

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

$$
f(x^k) - f(x^{k+1}) \geq \frac{1}{2L}\|\nabla f(x^k)\|^2 \quad \text{если } \alpha = \frac{1}{L}
$$

Обычно для сходящегося градиентного спуска чем больше допустимый шаг, тем быстрее сходимость, поэтому часто берут $\alpha = \frac{1}{L}$.

- После этого используем выпуклость:

$$
f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \text{ где } y = x^*, x = x^k
$$

$$
\boxed{f(x^k) - f^* \leq \langle \nabla f(x^k), x^k - x^* \rangle} \leq f^* + \langle \nabla f(x^k), x^k - x^* \rangle
\tag{2}
$$

# Анализ сходимости

- Теперь подставляем (2) в (1):

## Анализ сходимости

- Теперь подставляем (2) в (1):

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

# Анализ сходимости

- Теперь подставляем (2) в (1):

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

# Анализ сходимости

- Теперь подставляем (2) в (1):

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

# Анализ сходимости

- Теперь подставляем (2) в (1):

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

Пусть $\underline{a = x^k - x^*}$ и $\underline{b = x^k - x^* - \alpha \nabla f(x^k)}$.

# Анализ сходимости

- Теперь подставляем (2) в (1):

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

Пусть $a = x^k - x^*$ и $b = x^k - x^* - \alpha\nabla f(x^k)$. Тогда $\underline{a - b = \alpha\nabla f(x^k)}$ и $\underline{a + b = 2\left(x^k - x^* - \frac{\alpha}{2}\nabla f(x^k)\right)}$.

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

# Анализ сходимости

- Теперь подставляем (2) в (1):

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

Пусть $a = x^k - x^*$ и $b = x^k - x^* - \alpha\nabla f(x^k)$. Тогда $a + b = \alpha\nabla f(x^k)$ и $a - b = 2\left(x^k - x^* - \frac{\alpha}{2}\nabla f(x^k)\right)$.

$$
f(x^{k+1}) \leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^k - x^* - \alpha\nabla f(x^k)\|_2^2\right]
$$

# Анализ сходимости

- Теперь подставляем (2) в (1):

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

Пусть $a = x^k - x^*$ и $b = x^k - x^* - \alpha \nabla f(x^k)$. Тогда $a + b = \alpha \nabla f(x^k)$ и $a - b = 2\left(x^k - x^* - \frac{\alpha}{2}\nabla f(x^k)\right)$.

$x^{k+1}$

$$
\begin{aligned}
f(x^{k+1}) &\leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^k - x^* - \alpha \nabla f(x^k)\|_2^2\right] \\
&\leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2\right]
\end{aligned}
$$

# Анализ сходимости

- Теперь подставляем (2) в (1):

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

Пусть $a = x^k - x^*$ и $b = x^k - x^* - \alpha\nabla f(x^k)$. Тогда $a + b = \alpha\nabla f(x^k)$ и $a - b = 2\left(x^k - x^* - \frac{\alpha}{2}\nabla f(x^k)\right)$.

$$
\begin{aligned}
f(x^{k+1}) &\leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^k - x^* - \alpha\nabla f(x^k)\|_2^2\right] \\
&\leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2\right] \\
2\alpha\left(f(x^{k+1}) - f^*\right) &\leq \|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2
\end{aligned}
$$

ТЕЛЕСКОПЕВ
ЧУ

# Анализ сходимости

- Теперь подставляем (2) в (1):

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

Пусть $a = x^k - x^*$ и $b = x^k - x^* - \alpha \nabla f(x^k)$. Тогда $a + b = \alpha \nabla f(x^k)$ и $a - b = 2\left(x^k - x^* - \frac{\alpha}{2}\nabla f(x^k)\right)$.

$$
\begin{aligned}
f(x^{k+1}) &\leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^k - x^* - \alpha \nabla f(x^k)\|_2^2\right] \\
&\leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2\right] \\
2\alpha\left(f(x^{k+1}) - f^*\right) &\leq \|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2
\end{aligned}
$$

- Просуммируем по $i = 0, \ldots, k-1$. Большинство слагаемых обнуляется из-за телескопической суммы:

$$(3)$$

## Анализ сходимости

- Теперь подставляем (2) в (1):

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

Пусть $a = x^k - x^*$ и $b = x^k - x^* - \alpha\nabla f(x^k)$. Тогда $a + b = \alpha \nabla f(x^k)$ и $a - b = 2\left(x^k - x^* - \frac{\alpha}{2}\nabla f(x^k)\right)$.

$$
\begin{aligned}
f(x^{k+1}) &\leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^k - x^* - \alpha\nabla f(x^k)\|_2^2\right] \\
&\leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2\right] \\
2\alpha\left(f(x^{k+1}) - f^*\right) &\leq \|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2
\end{aligned}
$$

- Просуммируем по $i = 0, \ldots, k-1$. Большинство слагаемых обнуляется из-за телескопической суммы:

$$
2\alpha \sum_{i=0}^{k-1}\left(f(x^{i+1}) - f^*\right) \leq \|x^0 - x^*\|_2^2 - \cancel{\|x^k - x^*\|_2^2} \tag{3}
$$

# Анализ сходимости

- Теперь подставляем (2) в (1):

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

Пусть $a = x^k - x^*$ и $b = x^k - x^* - \alpha \nabla f(x^k)$. Тогда $a + b = \alpha \nabla f(x^k)$ и $a - b = 2\left(x^k - x^* - \frac{\alpha}{2}\nabla f(x^k)\right)$.

$$
\begin{aligned}
f(x^{k+1}) &\leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^k - x^* - \alpha \nabla f(x^k)\|_2^2\right] \\
&\leq f^* + \frac{1}{2\alpha}\left[\|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2\right] \\
2\alpha\left(f(x^{k+1}) - f^*\right) &\leq \|x^k - x^*\|_2^2 - \|x^{k+1} - x^*\|_2^2
\end{aligned}
$$

- Просуммируем по $i = 0, \ldots, k-1$. Большинство слагаемых обнуляется из-за телескопической суммы:

$$
\boxed{2\alpha \sum_{i=0}^{k-1}\left(f(x^{i+1}) - f^*\right) \leq \|x^0 - x^*\|_2^2 - \|x^k - x^*\|_2^2 \leq \|x^0 - x^*\|_2^2} \tag{3}
$$

# Анализ сходимости

- Поскольку на каждой итерации $\boxed{f(x^{i+1}) \leq f(x^i)}$, то

$$
kf(x^k) \leq \sum_{i=0}^{k-1} f(x^{i+1})
$$

$$
f(x^k) \leq \frac{1}{k} \sum_{i=0}^{k-1} f(x^{i+1})
$$

## Анализ сходимости

- Поскольку на каждой итерации $f(x^{i+1}) \leq f(x^i)$, то

$$kf(x^k) \leq \sum_{i=0}^{k-1} f(x^{i+1})$$

- Теперь подставим это в (3):

# Анализ сходимости

- Поскольку на каждой итерации $f(x^{i+1}) \leq f(x^i)$, то

$$
kf(x^k) \leq \sum_{i=0}^{k-1} f(x^{i+1})
$$

- Теперь подставим это в (3):

$$
2\alpha k f(x^k) - 2\alpha k f^* \leq 2\alpha \sum_{i=0}^{k-1} \left(f(x^{i+1}) - f^*\right) \leq \|x^0 - x^*\|_2^2
$$

# Анализ сходимости

- Поскольку на каждой итерации $f(x^{i+1}) \leq f(x^i)$, то

$$kf(x^k) \leq \sum_{i=0}^{k-1} f(x^{i+1})$$

- Теперь подставим это в (3):

$$2\alpha k f(x^k) - 2\alpha k f^* \leq 2\alpha \sum_{i=0}^{k-1} \left( f(x^{i+1}) - f^* \right) \leq \|x^0 - x^*\|_2^2$$

$$\boxed{f(x^k) - f^* \leq \frac{\|x^0 - x^*\|_2^2}{2\alpha k}}$$

# Анализ сходимости

- Поскольку на каждой итерации $f(x^{i+1}) \leq f(x^i)$, то

$$kf(x^k) \leq \sum_{i=0}^{k-1} f(x^{i+1})$$

- Теперь подставим это в (3):

$$2\alpha k f(x^k) - 2\alpha k f^* \leq 2\alpha \sum_{i=0}^{k-1} \left(f(x^{i+1}) - f^*\right) \leq \|x^0 - x^*\|_2^2$$

$$f(x^k) - f^* \leq \frac{\|x^0 - x^*\|_2^2}{2\alpha k} \leq \frac{L\|x^0 - x^*\|_2^2}{2k}$$

$\alpha = \frac{1}{L}$

# Итог

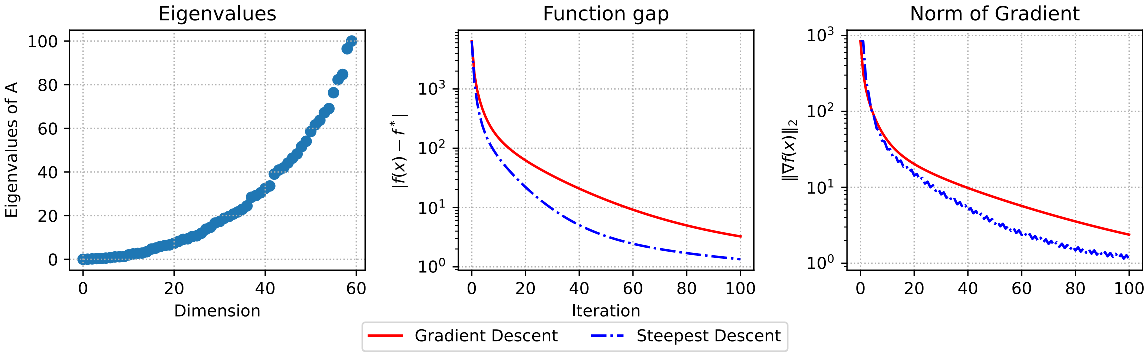
Градиентный спуск: $$\min_{x \in \mathbb{R}^n} f(x) \qquad x^{k+1} = x^k - \alpha^k \nabla f(x^k)$$

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

# Численные эксперименты

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

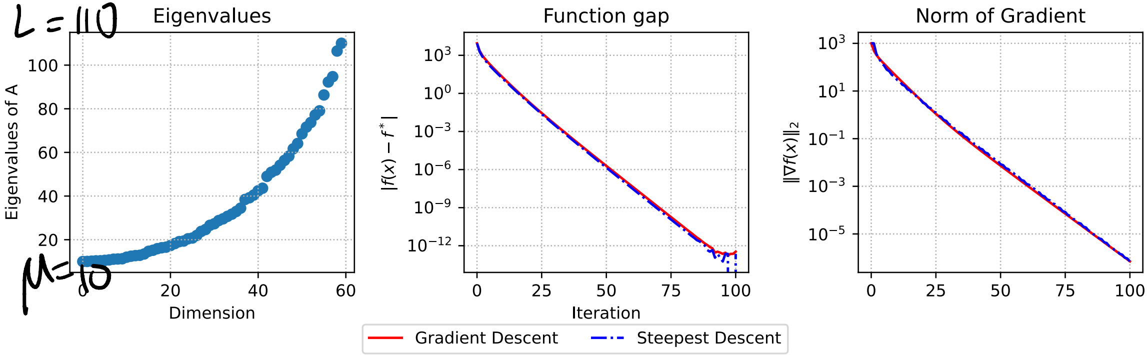
Convex quadratics. n=60, random matrix.



# Численные эксперименты

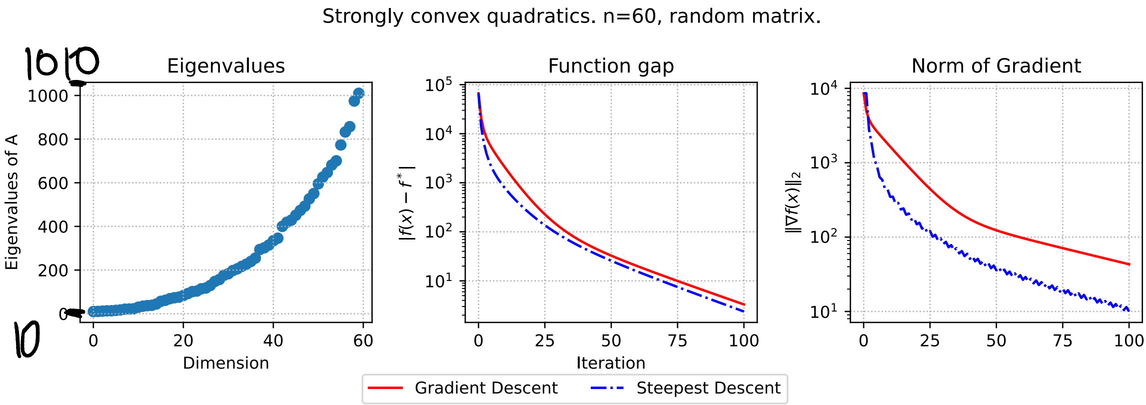
$$f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}$$

Strongly convex quadratics. n=60, random matrix.



## Численные эксперименты

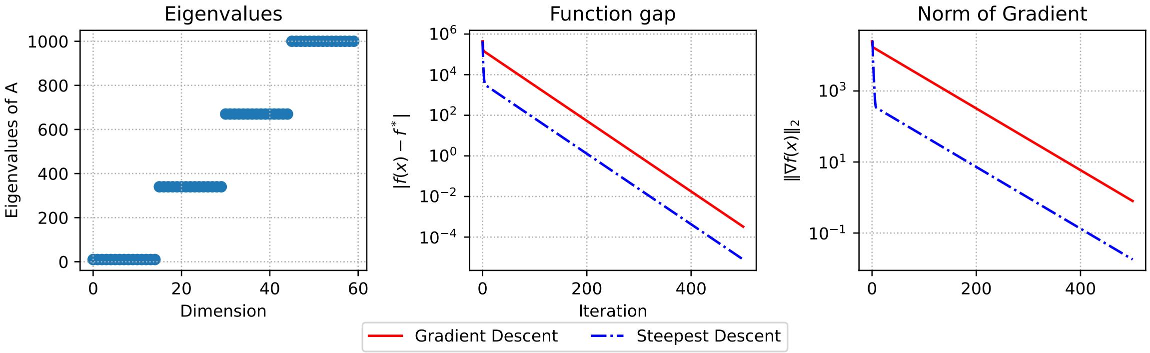
$$f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}$$



# Численные эксперименты

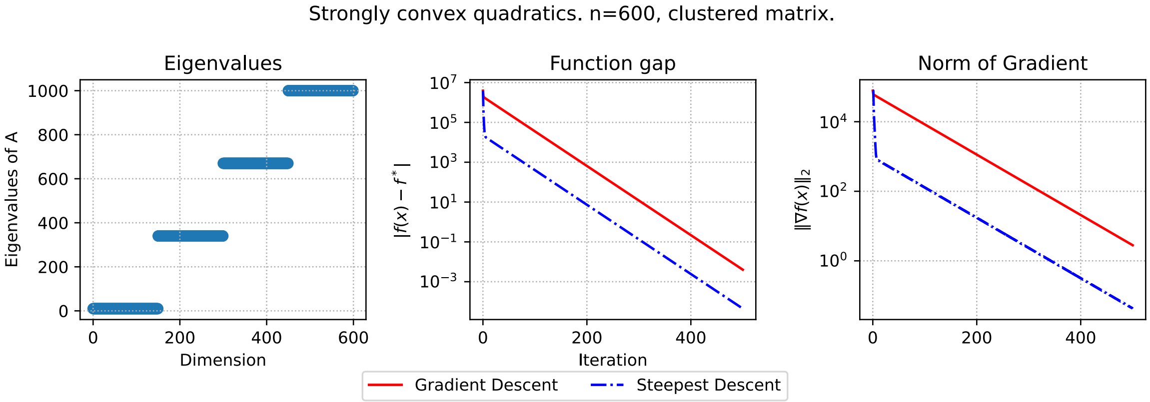
$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$

Strongly convex quadratics. n=60, clustered matrix.



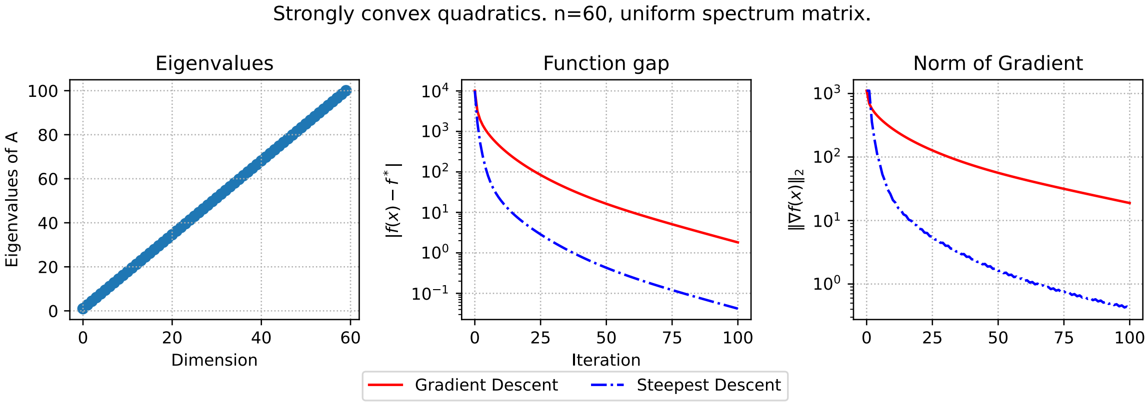
# Численные эксперименты

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



# Численные эксперименты

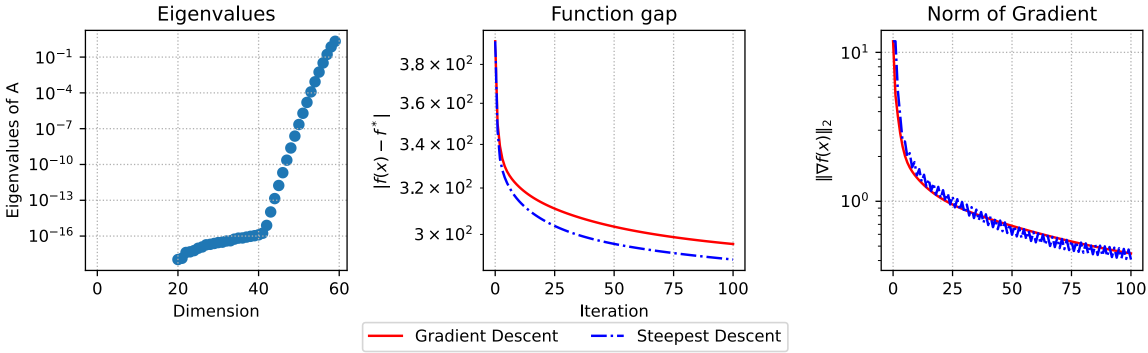
$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$



# Численные эксперименты

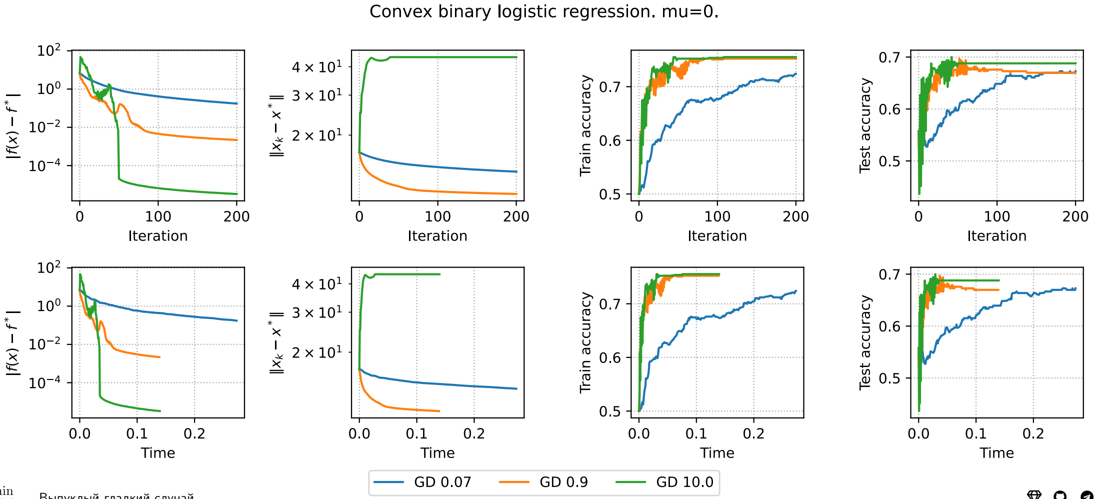
$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$

Strongly convex quadratics. n=60, Hilbert matrix.



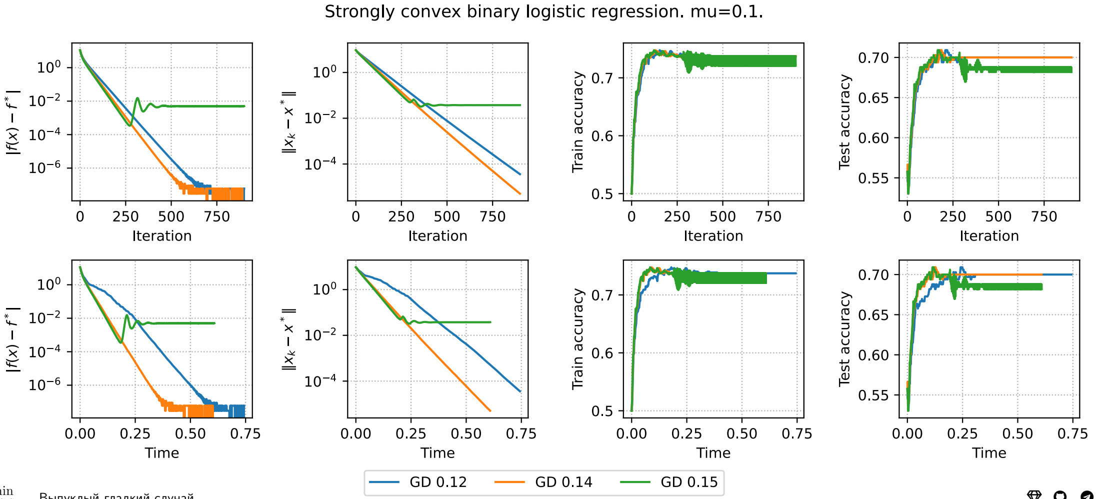
# Численные эксперименты

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



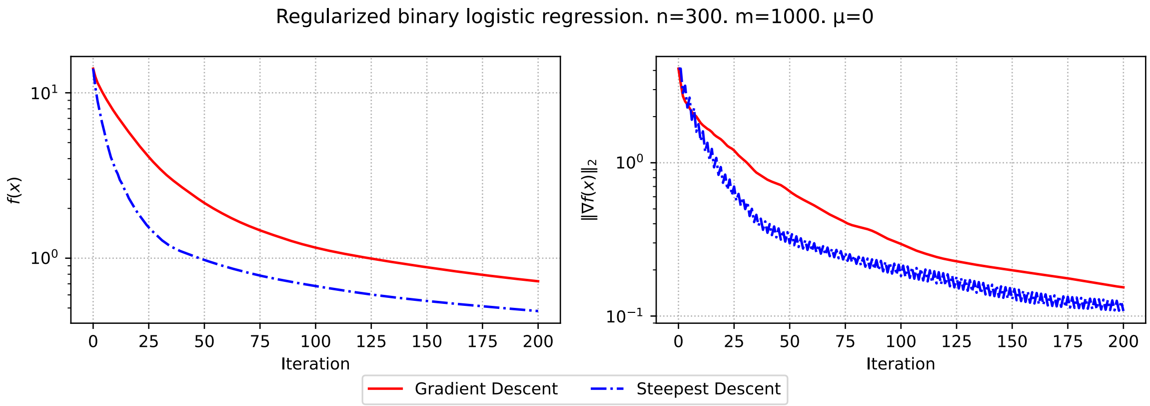
# Численные эксперименты

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



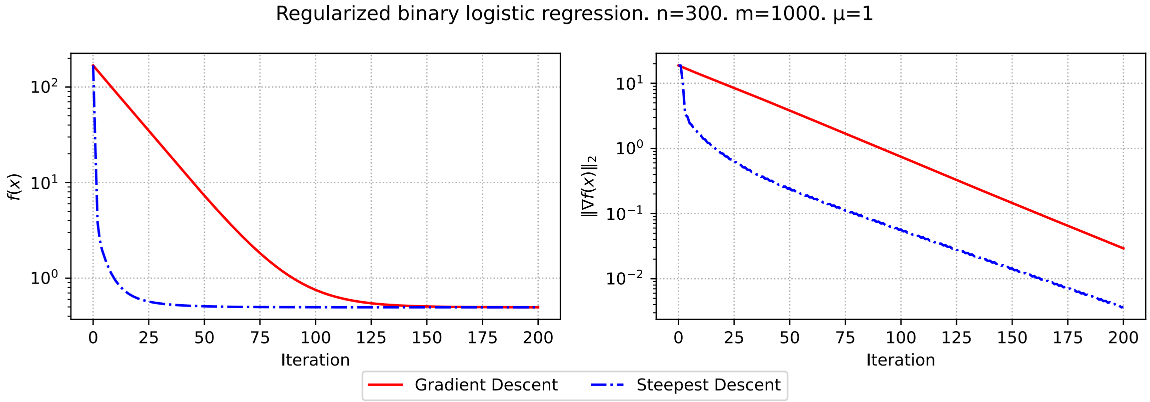
# Численные эксперименты

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



# 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

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