



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  $\alpha \rightarrow 0$ :

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

## 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  $\alpha \rightarrow 0$ :

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

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  $\alpha \rightarrow 0$ :

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

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  $\alpha \rightarrow 0$ :

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

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:

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

# Дифференциальное уравнение градиентного потока

Рассмотрим дифференциальное уравнение градиентного потока:

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

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:

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

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:

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

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 ♣

(GF)



Рис. 1: Траектория градиентного потока

# Сходимость алгоритма градиентного спуска

Код для построения анимации ниже. Сходимость существенно зависит от выбора шага  $\alpha$ :

Loss value 0.87



$w_1$  0.50,  $w_2$  3.00



## 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

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

## 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  $\alpha_k$  даёт

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

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


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

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:

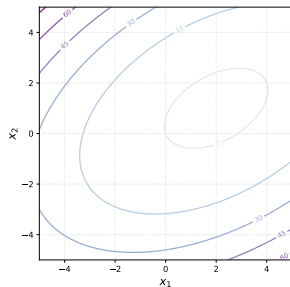
$$\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{2} x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}_{++}^d.$$

## Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

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

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



## Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

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

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



## Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

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

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



## Сдвиг координат

Рассмотрим следующую задачу квадратичной оптимизации:

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

- Во-первых, без ограничения общности мы можем установить  $c = 0$ , что не повлияет на процесс оптимизации.
- Во-вторых, у нас есть спектральное разложение матрицы  $A = Q\Lambda Q^\top$ .
- Покажем, что мы можем сделать сдвиг координат, чтобы сделать анализ немного проще. Пусть  $\hat{x} = Q^\top(x - x^*)$ , 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  $x = Q\hat{x} + x^*$ .

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





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

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



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

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



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

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



## 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  $\hat{x}$ )

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

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

## 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  $\hat{x}$ )

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$



## 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  $\hat{x}$ )

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

## 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  $\hat{x}$ )

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

## 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  $\hat{x}$ )

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

## 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  $\hat{x}$ )

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

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



## 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  $\hat{x}$ )

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

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

## 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  $\hat{x}$ )

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

• • •

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

$$\begin{array}{ll} |1 - \alpha\mu| < 1 & |1 - \alpha L| < 1 \\ -1 < 1 - \alpha\mu < 1 & -1 < 1 - \alpha L < 1 \end{array}$$

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

## Сильно выпуклые квадратичные функции

## 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  $\hat{x}$ )

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

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

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

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

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



## 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  $\hat{x}$ )

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

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

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

$$\alpha < \frac{2}{\mu}$$

$$\alpha \mu > 0$$

$$\alpha < \frac{2}{L}$$

$$\alpha L > 0$$





## 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  $\hat{x}$ )

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

Выберем  $\alpha$ , минимизирующий худший знаменатель прогрессии

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

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

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

...

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

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

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

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

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

$$\alpha < \frac{2}{\mu}$$

$$\alpha \mu > 0$$

$$\alpha < \frac{2}{L}$$

$$\alpha L > 0$$



## 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  $\hat{x}$ )

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

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

Выберем  $\alpha$ , минимизирующий худший знаменатель прогрессии

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

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

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

...

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

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

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

$$\alpha < \frac{2}{\mu}$$

$$\alpha \mu > 0$$

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

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

$$\alpha < \frac{2}{L}$$

$$\alpha L > 0$$



## 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  $\hat{x}$ )

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

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

Выберем  $\alpha$ , минимизирующий худший знаменатель прогрессии

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

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

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

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

...

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

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

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

$$\alpha < \frac{2}{\mu}$$

$$\alpha \mu > 0$$

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

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

$$\alpha < \frac{2}{L}$$

$$\alpha L > 0$$



## 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  $\hat{x}$ )

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

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

$$\alpha < \frac{2}{\mu}$$

$$\alpha \mu > 0$$

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

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

$$\alpha < \frac{2}{L}$$

$$\alpha L > 0$$

Выберем  $\alpha$ , минимизирующий худший знаменатель прогрессии

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

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

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

## 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  $\hat{x}$ )

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

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

$$\alpha < \frac{2}{\mu}$$

$$\alpha \mu > 0$$

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

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

$$\alpha < \frac{2}{L}$$

$$\alpha L > 0$$

Выберем  $\alpha$ , минимизирующий худший знаменатель прогрессии

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

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

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

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

## 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  $\hat{x}$ )

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

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

$$\alpha < \frac{2}{\mu}$$

$$\alpha \mu > 0$$

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

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

$$\alpha < \frac{2}{L}$$

$$\alpha L > 0$$

Выберем  $\alpha$ , минимизирующий худший знаменатель прогрессии

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

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

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

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

## 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  $\hat{x}$ )

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

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

$$\alpha < \frac{2}{\mu}$$

$$\alpha \mu > 0$$

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

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

$$\alpha < \frac{2}{L}$$

$$\alpha L > 0$$

Выберем  $\alpha$ , минимизирующий худший знаменатель прогрессии

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

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

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

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

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

## 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  $\hat{x}$ )

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

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

$$\alpha < \frac{2}{\mu}$$

$$\alpha \mu > 0$$

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

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

$$\alpha < \frac{2}{L}$$

$$\alpha L > 0$$

Выберем  $\alpha$ , минимизирующий худший знаменатель прогрессии

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

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

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

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

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

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



## 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  $\hat{x}$ )

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

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

$$x_{(i)}^{k+1} = (1 - \alpha^k \lambda_{(i)}) x_{(i)}^k \quad \text{для } i\text{-й координаты}$$

$$x_{(i)}^k = (1 - \alpha \lambda_{(i)})^k x_{(i)}^0 \quad \text{при постоянном шаге } \alpha^k = \alpha$$

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

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

...

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

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

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

$$\alpha < \frac{2}{\mu}$$

$$\alpha \mu > 0$$

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

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

$$\alpha < \frac{2}{L}$$

$$\alpha L > 0$$

Выберем  $\alpha$ , минимизирующий худший знаменатель прогрессии

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

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

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

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

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

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



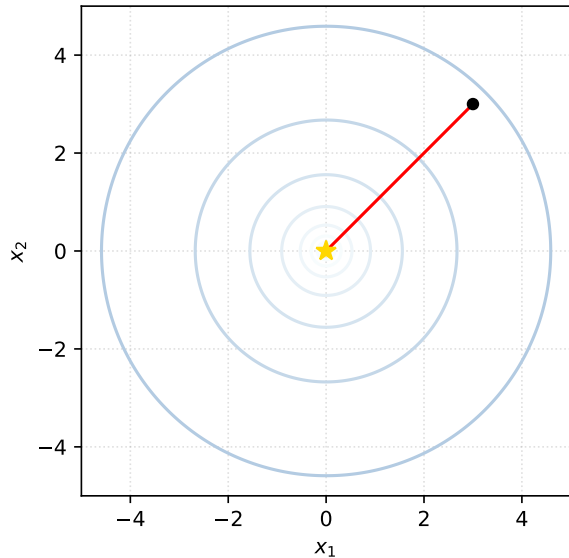
## Анализ сходимости

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

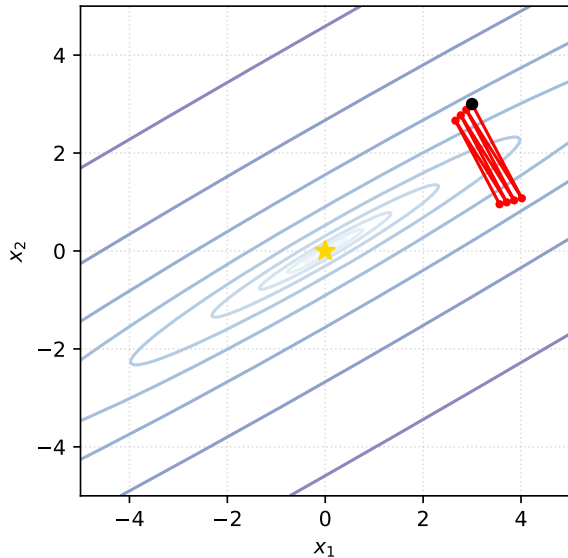
$\kappa$	$\rho$	Итераций до уменьшения ошибки по аргументу в 10 раз	Итераций до уменьшения ошибки по функции в 10 раз
1.1	0.05	1	1
2	0.33	3	2
5	0.67	6	3
10	0.82	12	6
50	0.96	58	29
100	0.98	116	58
500	0.996	576	288
1000	0.998	1152	576

# Число обусловленности $\kappa$

$\kappa = 1.0$



$\kappa = 100.0$



## 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

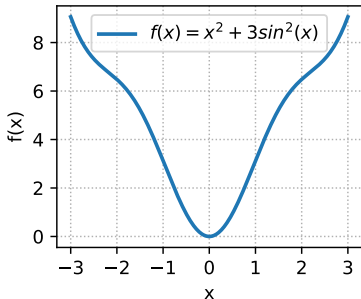
$$\|\nabla f(x)\|^2 \geq 2\mu(f(x) - f^*) \quad \forall x$$

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

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

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

Function, that satisfies  
Polyak-Lojasiewicz condition



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

Non-convex PL function



## i Theorem

Рассмотрим задачу

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

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  $0 < \alpha \leq \frac{1}{L}$ . Пусть  $f^* = \min_{x \in \mathbb{R}^d} f(x)$ . Тогда:

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

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

## Theorem

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



## Выпуклый гладкий случай

# Выпуклый гладкий случай

## i Theorem

Рассмотрим задачу

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

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:

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

Градиентный спуск:

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

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

гладкий (не выпуклый)

$$\|\nabla f(x^k)\|^2 \sim \mathcal{O}\left(\frac{1}{k}\right)$$

$$k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\varepsilon}\right)$$

гладкий и выпуклый

$$f(x^k) - f^* \sim \mathcal{O}\left(\frac{1}{k}\right)$$

$$k_\varepsilon \sim \mathcal{O}\left(\frac{1}{\varepsilon}\right)$$

гладкий и сильно выпуклый (или PL)

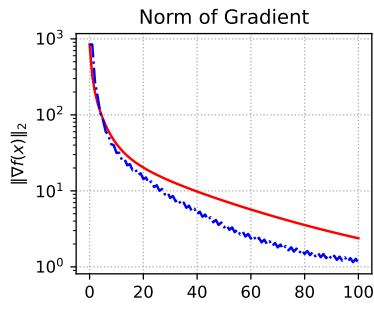
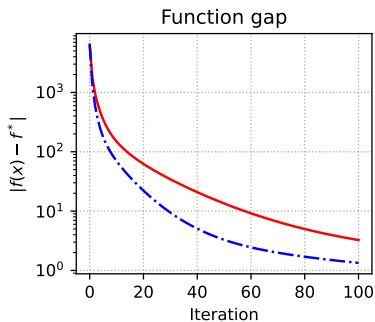
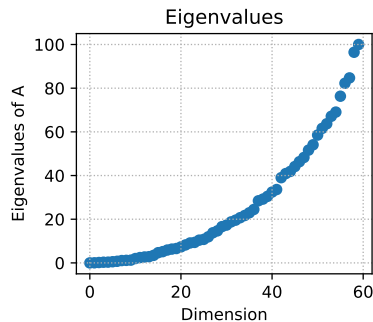
$$\|x^k - x^*\|^2 \sim \mathcal{O}\left(\left(1 - \frac{\mu}{L}\right)^k\right)$$

$$k_\varepsilon \sim \mathcal{O}\left(\kappa \log \frac{1}{\varepsilon}\right)$$

# Численные эксперименты

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

Convex quadratics.  $n=60$ , random matrix.

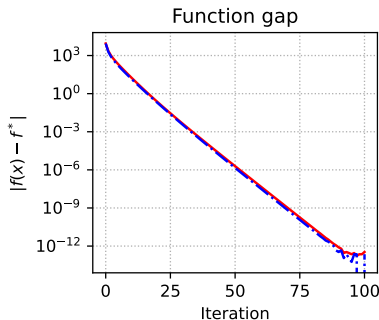


— Gradient Descent    - - - Steepest Descent

# Численные эксперименты

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

Strongly convex quadratics.  $n=60$ , random matrix.

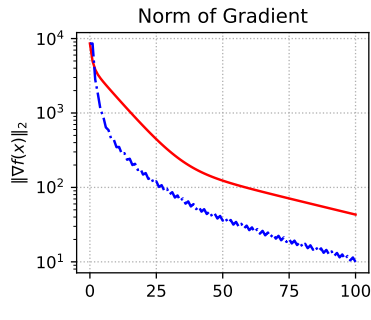
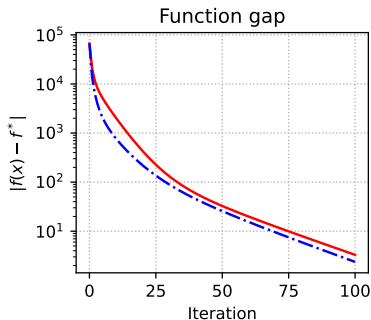
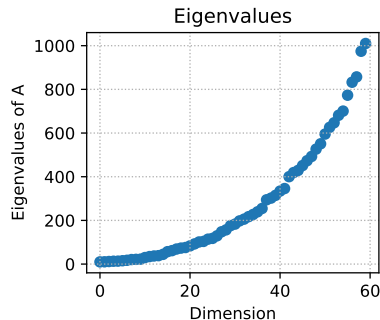


— Gradient Descent    - - - Steepest Descent

# Численные эксперименты

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

Strongly convex quadratics.  $n=60$ , random matrix.

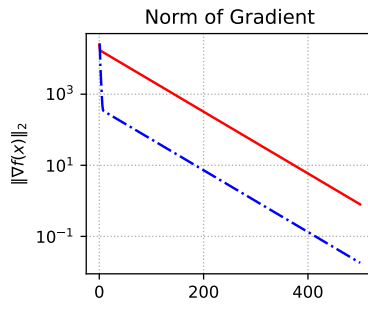


— Gradient Descent    - - - Steepest Descent

# Численные эксперименты

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

Strongly convex quadratics.  $n=60$ , clustered matrix.

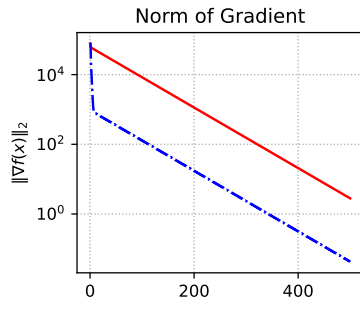
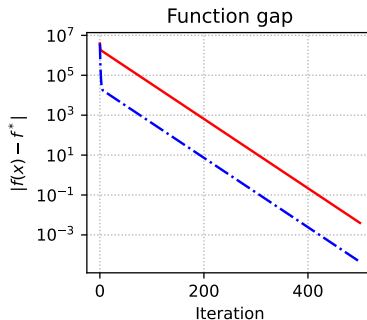
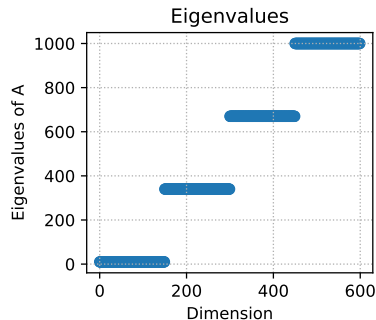


— Gradient Descent    -.- Steepest Descent

# Численные эксперименты

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

Strongly convex quadratics.  $n=600$ , clustered matrix.



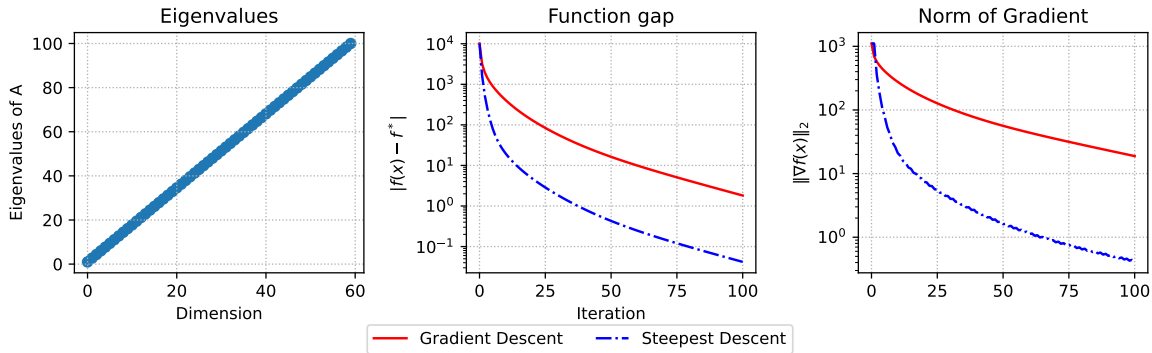
— Gradient Descent    -.- Steepest Descent



# Численные эксперименты

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

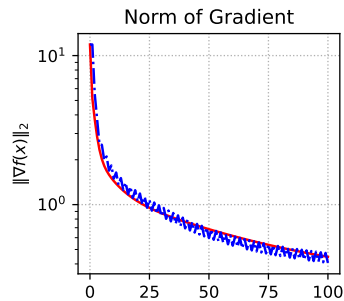
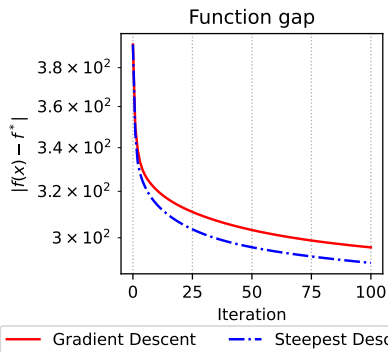
Strongly convex quadratics.  $n=60$ , uniform spectrum matrix.



# Численные эксперименты

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

Strongly convex quadratics.  $n=60$ , Hilbert matrix.



# Численные эксперименты

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

Convex binary logistic regression.  $\mu=0$ .



— GD 0.07 — GD 0.9 — GD 10.0

# Численные эксперименты

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

Strongly convex binary logistic regression.  $\mu=0.1$ .



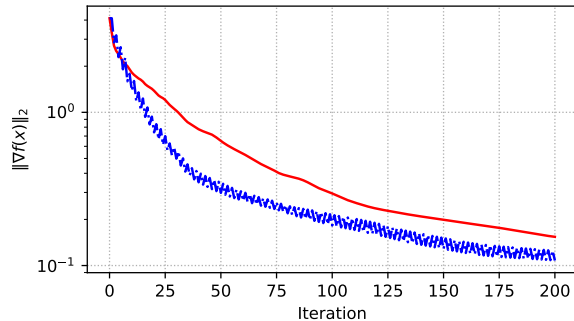
## Численные эксперименты

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

Regularized binary logistic regression.  $n=300$ .  $m=1000$ .  $\mu=0$



— Gradient Descent    -.- Steepest Descent



## Численные эксперименты

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

Regularized binary logistic regression.  $n=300$ .  $m=1000$ .  $\mu=1$



— Gradient Descent    -.- Steepest Descent

