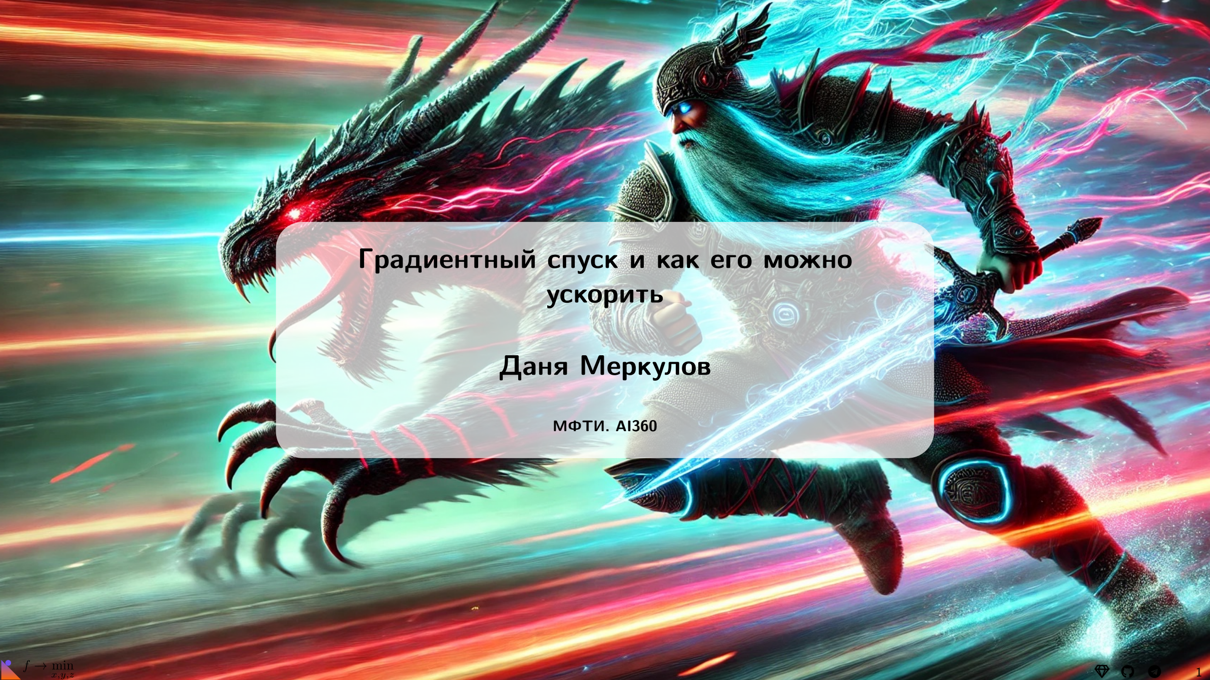


Градиентный спуск и как его можно ускорить

Даня Меркулов

МФТИ. AI360

Градиентный спуск

Направление локального наискорейшего спуска

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

Направление локального наискорейшего спуска

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

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

Направление локального наискорейшего спуска

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

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

Мы хотим, чтобы направление h было направлением убывания функции:

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

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

Направление локального наискорейшего спуска

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

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

Мы хотим, чтобы направление h было направлением убывания функции:

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

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

перейдя к пределу при $\alpha \to 0$:

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

Направление локального наискорейшего спуска

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

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

Мы хотим, чтобы направление h было направлением убывания функции:

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

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

перейдя к пределу при $\alpha \to 0$:

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

Также из неравенства Коши-Буняковского-Шварца:

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

Направление локального наискорейшего спуска

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

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

Мы хотим, чтобы направление h было направлением убывания функции:

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

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

перейдя к пределу при $\alpha \to 0$:

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

Также из неравенства Коши-Буняковского-Шварца:

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

Таким образом, направление антиградиента

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

даёт направление **наискорейшего локального** убывания функции f.

Направление локального наискорейшего спуска

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

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

Мы хотим, чтобы направление h было направлением убывания функции:

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

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

перейдя к пределу при $\alpha \to 0$:

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

Также из неравенства Коши-Буняковского-Шварца:

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

Таким образом, направление антиградиента

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

даёт направление **наискорейшего локального** убывания функции f.

Результатом этого является метод градиентного спуска:

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

Направление локального наискорейшего спуска

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

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

Мы хотим, чтобы направление h было направлением убывания функции:

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

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

перейдя к пределу при $\alpha \to 0$:

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

Также из неравенства Коши-Буняковского-Шварца:

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

Таким образом, направление антиградиента

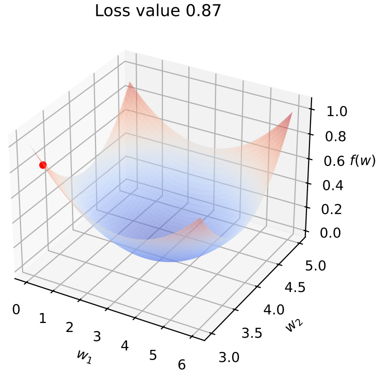
$$h = -\frac{f'(x)}{\|f'(x)\|_2}$$

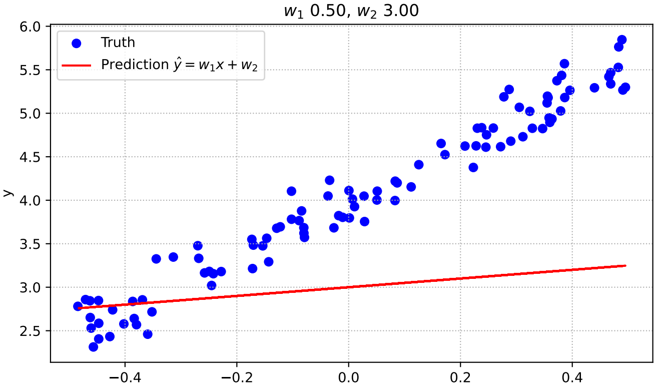
даёт направление **наискорейшего локального** убывания функции f.
Результатом этого является метод градиентного спуска:

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

Сходимость градиентного спуска

Сходимость градиентного спуска сильно зависит от выбора шага α:





Наискорейший спуск

$$
\alpha_k = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_{k+1}) = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_k - \alpha \nabla f(x_k))
$$

Более теоретический, чем практический подход. Он также позволяет анализировать сходимость, но часто точный поиск вдоль направления может быть сложным, если вычисление функции занимает слишком много времени или стоит дорого.
Интересное теоретическое свойство этого метода заключается в том, что каждая следующая итерация ортогональна предыдущей:

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

Наискорейший спуск

$$
\alpha_k = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_{k+1}) = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_k - \alpha \nabla f(x_k))
$$

Более теоретический, чем практический подход. Он также позволяет анализировать сходимость, но часто точный поиск вдоль направления может быть сложным, если вычисление функции занимает слишком много времени или стоит дорого.

Интересное теоретическое свойство этого метода заключается в том, что каждая следующая итерация ортогональна предыдущей:

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

Условие оптимальности:

Наискорейший спуск

$$\alpha_k = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_{k+1}) = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_k - \alpha \nabla f(x_k))$$

Более теоретический, чем практический подход. Он также позволяет анализировать сходимость, но часто точный поиск вдоль направления может быть сложным, если вычисление функции занимает слишком много времени или стоит дорого.

Интересное теоретическое свойство этого метода заключается в том, что каждая следующая итерация ортогональна предыдущей:

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

Условие оптимальности:

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

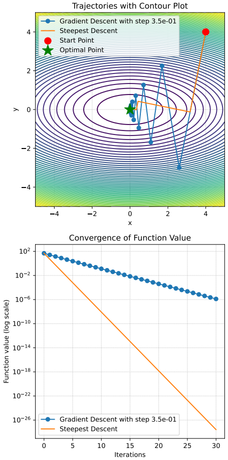


Рис. 1: Наискорейший спуск

Open In Colab ♣

Сходимость для сильно выпуклых квадратичных функций

Сдвиг координат

Рассмотрим следующую квадратичную оптимизационную задачу:

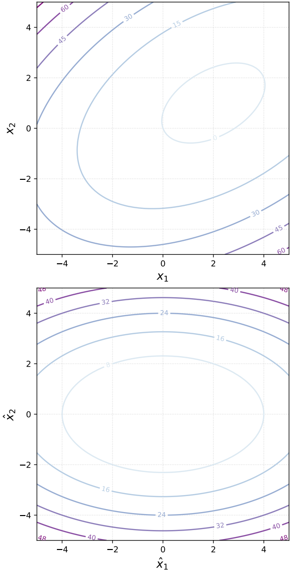
$$
\min_{x \in \mathbb{R}^d} f(x) = \min_{x \in \mathbb{R}^d} \frac{1}{2} x^\top A x - b^\top x + c, \text{ where } A \in \mathbb{S}^d_{++}.
$$

Сдвиг координат

Рассмотрим следующую квадратичную оптимизационную задачу:

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

- Без ограничения общности можно положить $c = 0$, что не повлияет на процесс оптимизации.

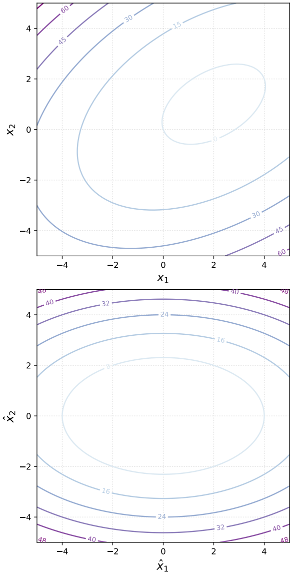


Сдвиг координат

Рассмотрим следующую квадратичную оптимизационную задачу:

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

- Без ограничения общности можно положить $c = 0$, что не повлияет на процесс оптимизации.
- Второй шаг: представим матрицу A в виде $A = Q \Lambda Q^T$.

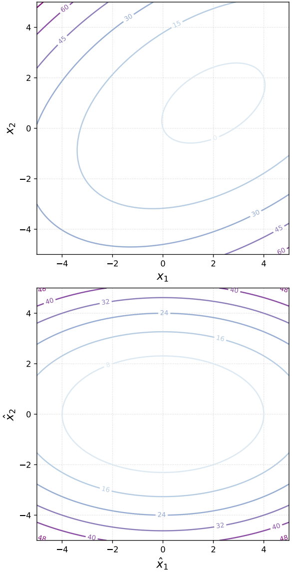


Сдвиг координат

Рассмотрим следующую квадратичную оптимизационную задачу:

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

- Без ограничения общности можно положить $c = 0$, что не повлияет на процесс оптимизации.
- Второй шаг: представим матрицу A в виде $A = Q \Lambda Q^T$.
- Покажем, что мы можем сделать замену 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 $x = Q\hat{x} + x^*$.



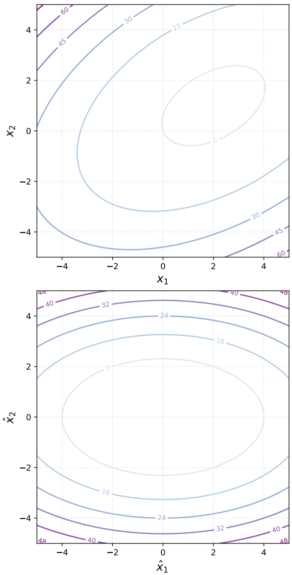
Сдвиг координат

Рассмотрим следующую квадратичную оптимизационную задачу:

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

- Без ограничения общности можно положить $c = 0$, что не повлияет на процесс оптимизации.
- Второй шаг: представим матрицу A в виде $A = Q \Lambda Q^T$.
- Покажем, что мы можем сделать замену 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 $x = Q\hat{x} + x^*$.

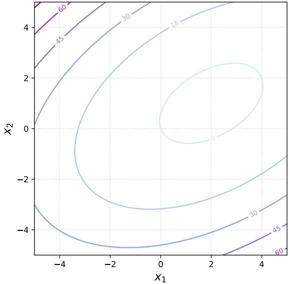
$$f(\hat{x}) = \frac{1}{2}(Q\hat{x} + x^*)^\top A (Q\hat{x} + x^*) - b^\top (Q\hat{x} + x^*)$$



Сдвиг координат

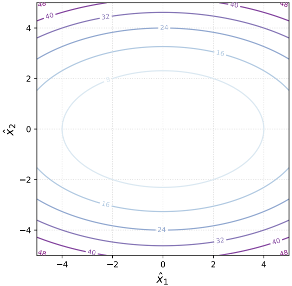
Рассмотрим следующую квадратичную оптимизационную задачу:

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



- Без ограничения общности можно положить $c = 0$, что не повлияет на процесс оптимизации.
- Второй шаг: представим матрицу A в виде $A = Q\Lambda Q^T$.
- Покажем, что мы можем сделать замену координат, чтобы сделать анализ немного проще. Пусть $\hat{x} = Q^T(x - x^*)$, где x^* - точка минимума исходной функции, определяемая как $Ax^* = b$. При этом $x = Q\hat{x} + x^*$.

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



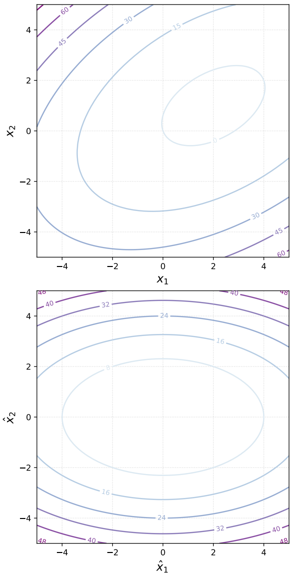
Сдвиг координат

Рассмотрим следующую квадратичную оптимизационную задачу:

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

- Без ограничения общности можно положить $c = 0$, что не повлияет на процесс оптимизации.
- Второй шаг: представим матрицу A в виде $A = Q\Lambda Q^T$.
- Покажем, что мы можем сделать замену 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 $x = Q\hat{x} + x^*$.

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



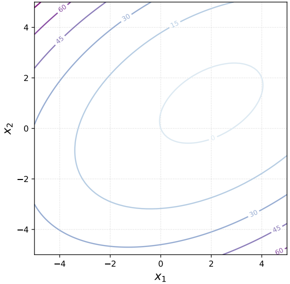
Сдвиг координат

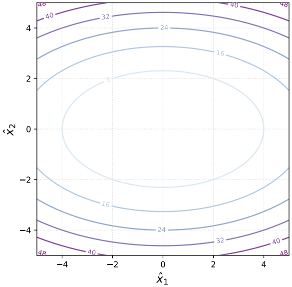
Рассмотрим следующую квадратичную оптимизационную задачу:

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

- Без ограничения общности можно положить $c = 0$, что не повлияет на процесс оптимизации.
- Второй шаг: представим матрицу A в виде $A = Q \Lambda Q^T$.
- Покажем, что мы можем сделать замену 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 $x = Q\hat{x} + x^*$.

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





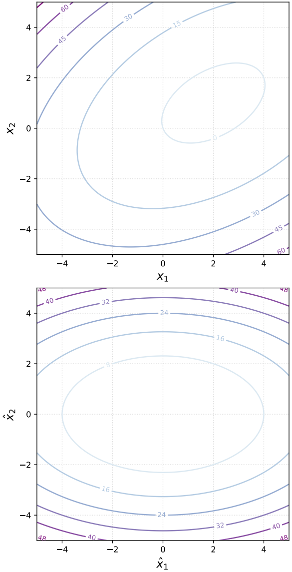
Сдвиг координат

Рассмотрим следующую квадратичную оптимизационную задачу:

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

- Без ограничения общности можно положить $c = 0$, что не повлияет на процесс оптимизации.
- Второй шаг: представим матрицу A в виде $A = Q \Lambda Q^T$.
- Покажем, что мы можем сделать замену 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 $x = Q\hat{x} + x^*$.

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



Сходимость

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

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

Сходимость

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

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

Сходимость

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

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

Сходимость

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

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

Сходимость

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

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

Сходимость

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

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

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

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

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

Сходимость

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

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

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

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

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

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

Сходимость

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

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

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

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

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

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

Сходимость

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

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

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

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

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

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

Сходимость

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

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

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

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

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

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

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

Сходимость

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

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

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

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

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

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

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

Сходимость

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

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

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

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

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

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

Сходимость

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

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

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

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

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

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

Сходимость

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

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

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

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

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

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

$\alpha < \frac{2}{L}$ необходимо для сходимости.

Сходимость

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

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

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

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

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

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

$\alpha < \frac{2}{L}$ необходимо для сходимости.

Теперь мы хотели бы настроить α для выбора лучшего (наименьшего) коэффициента сходимости

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

Сходимость

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

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

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

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

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

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

$\alpha < \frac{2}{L}$ необходимо для сходимости.

Теперь мы хотели бы настроить α для выбора лучшего (наименьшего) коэффициента сходимости

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

Сходимость

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

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

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

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

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

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

$\alpha < \frac{2}{L}$ необходимо для сходимости.

Теперь мы хотели бы настроить α для выбора лучшего (наименьшего) коэффициента сходимости

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

Сходимость

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

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

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

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

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

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

$\alpha < \frac{2}{L}$ необходимо для сходимости.

Теперь мы хотели бы настроить α для выбора лучшего (наименьшего) коэффициента сходимости

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

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

Сходимость

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

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

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

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

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

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

$\alpha < \frac{2}{L}$ необходимо для сходимости.

Теперь мы хотели бы настроить α для выбора лучшего (наименьшего) коэффициента сходимости

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

Сходимость

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

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

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

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

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

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

$\alpha < \frac{2}{L}$ необходимо для сходимости.

Теперь мы хотели бы настроить α для выбора лучшего (наименьшего) коэффициента сходимости

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

Сходимость

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

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

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

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

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

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

$\alpha < \frac{2}{L}$ необходимо для сходимости.

Теперь мы хотели бы настроить α для выбора лучшего (наименьшего) коэффициента сходимости

$$
\begin{aligned}
\rho^* &= \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha \lambda_{(i)}| \\
&= \min_\alpha \left\{ |1 - \alpha\mu|, |1 - \alpha L| \right\} \\
\alpha^* :& \quad 1 - \alpha^*\mu = \alpha^* L - 1 \\
& \alpha^* = \frac{2}{\mu + L} \quad \rho^* = \frac{L - \mu}{L + \mu} \\
x^k_{(i)} &= \left(\frac{L - \mu}{L + \mu} \right)^k x^0_{(i)}
\end{aligned}
$$

Сходимость

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

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

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

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

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

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

$\alpha < \frac{2}{L}$ необходимо для сходимости.

Теперь мы хотели бы настроить α для выбора лучшего (наименьшего) коэффициента сходимости

$$
\begin{aligned}
\rho^* &= \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha\lambda_{(i)}| \\
&= \min_\alpha \{|1 - \alpha\mu|, |1 - \alpha L|\} \\
\alpha^* &: \quad 1 - \alpha^*\mu = \alpha^* L - 1 \\
&\alpha^* = \frac{2}{\mu + L} \quad \rho^* = \frac{L - \mu}{L + \mu} \\
x^k_{(i)} &= \left(\frac{L - \mu}{L + \mu}\right)^k x^0_{(i)} \\
\|x^k\|_2 &\leq \left(\frac{L - \mu}{L + \mu}\right)^k \|x^0\|_2
\end{aligned}
$$

Сходимость

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

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

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

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

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

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

$\alpha < \frac{2}{L}$ необходимо для сходимости.

Теперь мы хотели бы настроить α для выбора лучшего (наименьшего) коэффициента сходимости

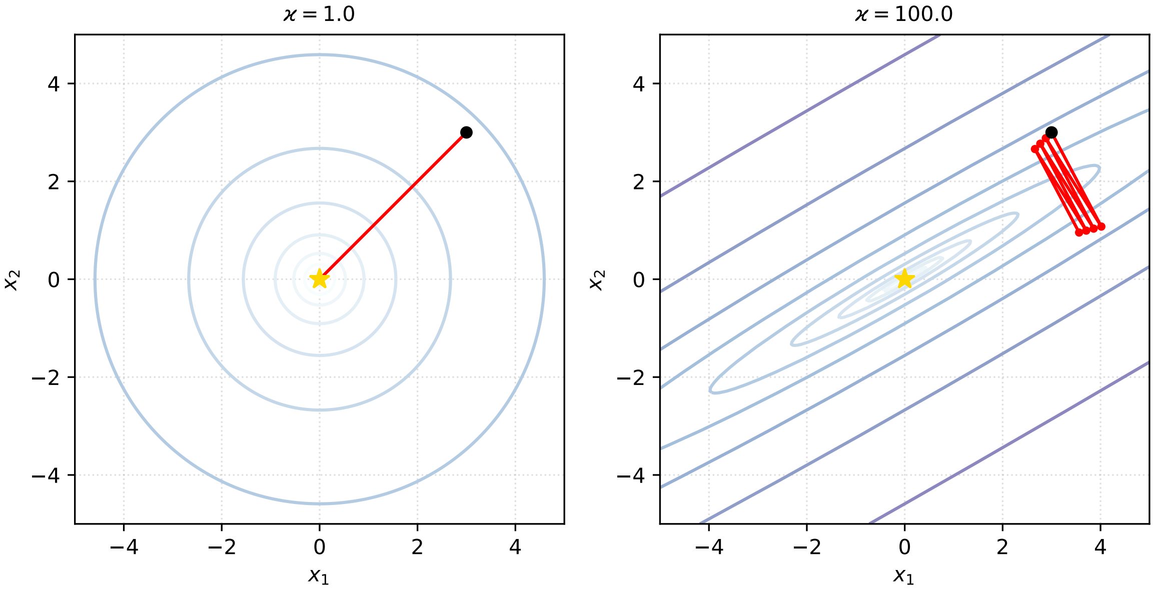
$$
\begin{aligned}
\rho^* &= \min_\alpha \rho(\alpha) = \min_\alpha \max_i |1 - \alpha \lambda_{(i)}| \\
&= \min_\alpha \left\{ |1 - \alpha\mu|, |1 - \alpha L| \right\} \\
\alpha^* &: \quad 1 - \alpha^*\mu = \alpha^* L - 1 \\
&\alpha^* = \frac{2}{\mu + L} \quad \rho^* = \frac{L - \mu}{L + \mu} \\
x^k_{(i)} &= \left(\frac{L - \mu}{L + \mu} \right)^k x^0_{(i)} \\
\|x^k\|_2 &\leq \left(\frac{L - \mu}{L + \mu} \right)^k \|x^0\|_2 \quad f(x^k) \leq \left(\frac{L - \mu}{L + \mu} \right)^{2k} f(x^0)
\end{aligned}
$$

Сходимость

Таким образом, мы имеем линейную сходимость в домене с коэффициентом $\frac{\varkappa-1}{\varkappa+1} = 1 - \frac{2}{\varkappa+1}$, где $\varkappa = \frac{L}{\mu}$ называется *числом обусловленности* квадратичной задачи.

$\varkappa$	ρ	Итерации для уменьшения ошибки в 10 раз	Итерации для уменьшения невязки 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 $\varkappa$



Ускорение для квадратичных функций

Сходимость из первых принципов

$$f(x) = \frac{1}{2}x^T A x - b^T x \qquad x_{k+1} = x_k - \alpha_k \nabla f(x_k).$$

Пусть x^* - единственное решение линейной системы $Ax = b$ и пусть $e_k = x_k - x^*$, где $x_{k+1} = x_k - \alpha_k(Ax_k - b)$ определяется рекурсивно, начиная с некоторого x_0, и α_k - шаг, который мы определим позже.

$$e_{k+1} = (I - \alpha_k A)e_k.$$

Полиномы

Вышеуказанный расчет дает нам $e_k = p_k(A)e_0$,
где p_k - полином

$$p_k(a) = \prod_{i=1}^{k}(1 - \alpha_i a).$$

Мы можем ограничить сверху норму ошибки как

$$\|e_k\| \leq \|p_k(A)\| \cdot \|e_0\|.$$

Сходимость из первых принципов

$$f(x) = \frac{1}{2}x^TAx - b^Tx \qquad x_{k+1} = x_k - \alpha_k \nabla f(x_k).$$

Пусть x^* - единственное решение линейной системы $Ax = b$ и пусть $e_k = x_k - x^*$, где $x_{k+1} = x_k - \alpha_k(Ax_k - b)$ определяется рекурсивно, начиная с некоторого x_0, и α_k - шаг, который мы определим позже.

$$e_{k+1} = (I - \alpha_k A)e_k.$$

Полиномы

Вышеуказанный расчет дает нам $e_k = p_k(A)e_0$, где p_k - полином

$$p_k(a) = \prod_{i=1}^{k}(1 - \alpha_i a).$$

Мы можем ограничить сверху норму ошибки как

$$\|e_k\| \leq \|p_k(A)\| \cdot \|e_0\| .$$

Поскольку A - симметричная матрица с собственными значениями в $[\mu, L]$,:

$$\|p_k(A)\| \leq \max_{\mu \leq a \leq L} |p_k(a)| .$$

Это приводит к интересной проблеме: среди всех полиномов, удовлетворяющих $p_k(0) = 1$, мы ищем полином, величина которого наименьшая в интервале $[\mu, L]$.

Наивный полиномиальный подход

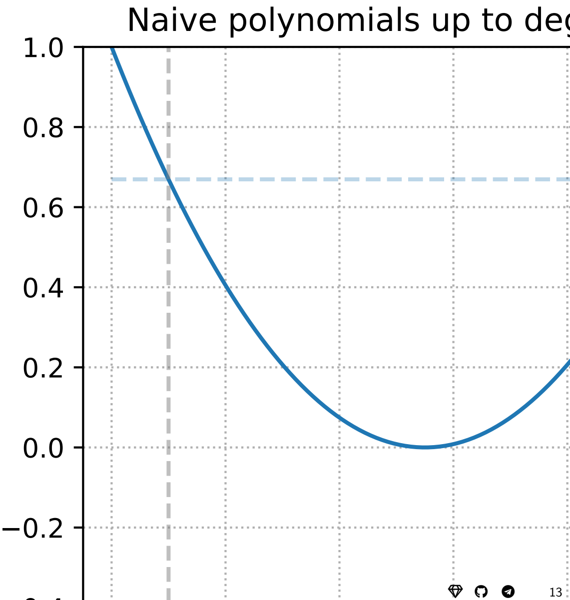
Наивный подход состоит в выборе равномерного шага $\alpha_k = \frac{2}{\mu+L}$ в выражении. Этот выбор делает $|p_k(\mu)| = |p_k(L)|$.

$$\|e_k\| \leq \left(\frac{L-\mu}{L+\mu}\right)^k \|e_0\|$$

Это точно такой же результат, который мы доказали для сходимости градиентного спуска в случае квадратичной функции.

Давайте взглянем на этот полином поближе. На правом рисунке мы выбрали $\alpha = 1$ и $\beta = 10$ так, что $\kappa = 10$. Соответствующий интервал, таким образом, равен $[1, 10]$.

Можем ли мы сделать лучше? Ответ - да.



Наивный полиномиальный подход

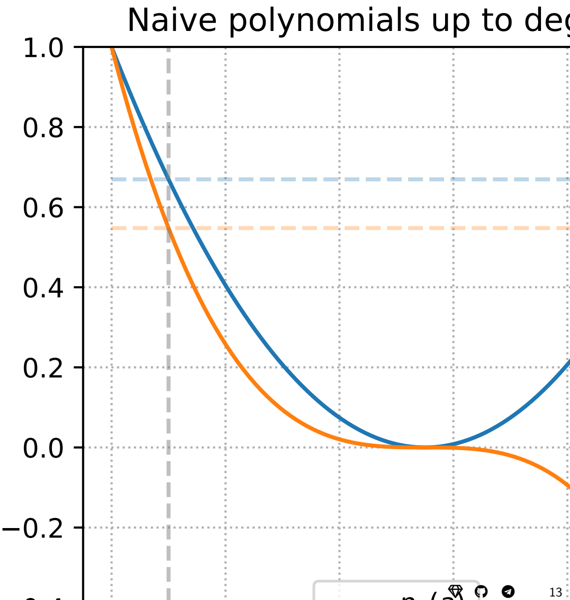
Наивный подход состоит в выборе равномерного шага $\alpha_k = \frac{2}{\mu+L}$ в выражении. Этот выбор делает $|p_k(\mu)| = |p_k(L)|$.

$$
\|e_k\| \leq \left(\frac{L-\mu}{L+\mu}\right)^k \|e_0\|
$$

Это точно такой же результат, который мы доказали для сходимости градиентного спуска в случае квадратичной функции.

Давайте взглянем на этот полином поближе. На правом рисунке мы выбрали $\alpha = 1$ и $\beta = 10$ так, что $\kappa = 10$. Соответствующий интервал, таким образом, равен $[1, 10]$.

Можем ли мы сделать лучше? Ответ - да.

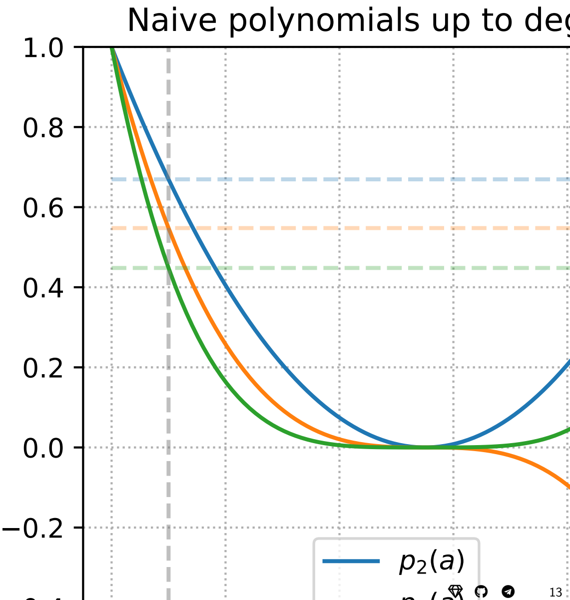


Наивный полиномиальный подход

Наивный подход состоит в выборе равномерного шага $\alpha_k = \frac{2}{\mu+L}$ в выражении. Этот выбор делает $|p_k(\mu)| = |p_k(L)|$.

$$\|e_k\| \leq \left(\frac{L-\mu}{L+\mu}\right)^k \|e_0\|$$

Это точно такой же результат, который мы доказали для сходимости градиентного спуска в случае квадратичной функции.
Давайте взглянем на этот полином поближе. На правом рисунке мы выбрали $\alpha = 1$ и $\beta = 10$ так, что $\kappa = 10$. Соответствующий интервал, таким образом, равен $[1, 10]$.
Можем ли мы сделать лучше? Ответ - да.



Наивный полиномиальный подход

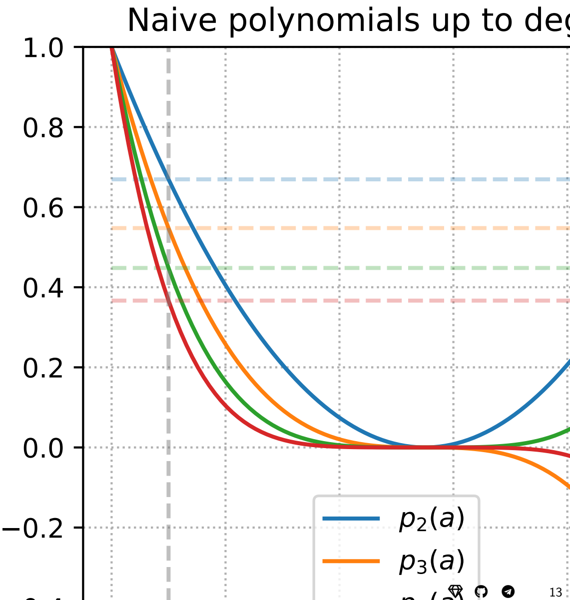
Наивный подход состоит в выборе равномерного шага $\alpha_k = \frac{2}{\mu+L}$ в выражении. Этот выбор делает $|p_k(\mu)| = |p_k(L)|$.

$$\|e_k\| \leq \left(\frac{L-\mu}{L+\mu}\right)^k \|e_0\|$$

Это точно такой же результат, который мы доказали для сходимости градиентного спуска в случае квадратичной функции.

Давайте взглянем на этот полином поближе. На правом рисунке мы выбрали $\alpha = 1$ и $\beta = 10$ так, что $\kappa = 10$. Соответствующий интервал, таким образом, равен $[1, 10]$.

Можем ли мы сделать лучше? Ответ - да.

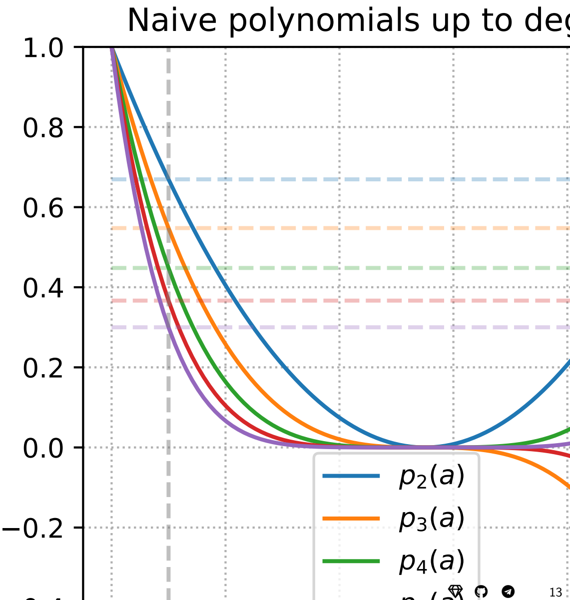


Наивный полиномиальный подход

Наивный подход состоит в выборе равномерного шага $\alpha_k = \frac{2}{\mu+L}$ в выражении. Этот выбор делает $|p_k(\mu)| = |p_k(L)|$.

$$\|e_k\| \leq \left(\frac{L-\mu}{L+\mu}\right)^k \|e_0\|$$

Это точно такой же результат, который мы доказали для сходимости градиентного спуска в случае квадратичной функции.
Давайте взглянем на этот полином поближе. На правом рисунке мы выбрали $\alpha = 1$ и $\beta = 10$ так, что $\kappa = 10$. Соответствующий интервал, таким образом, равен $[1, 10]$.
Можем ли мы сделать лучше? Ответ - да.

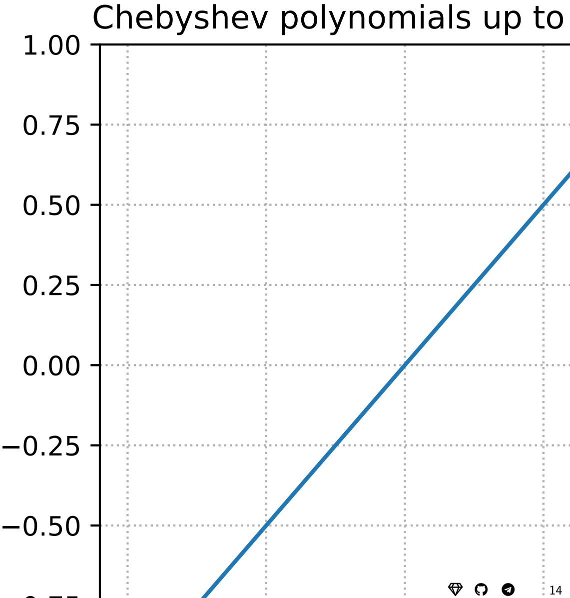


Полиномы Чебышёва

Полиномы Чебышёва оказываются оптимальным ответом на вопрос, который мы задавали. Соответствующим образом масштабированные, они минимизируют абсолютное значение в желаемом интервале $[\mu, L]$ при условии, что значение равно 1 в начале.

$$
\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_k(x) &= 2xT_{k-1}(x) - T_{k-2}(x), \qquad k \geq 2.
\end{aligned}
$$

Давайте построим стандартные полиномы Чебышёва (без масштабирования):

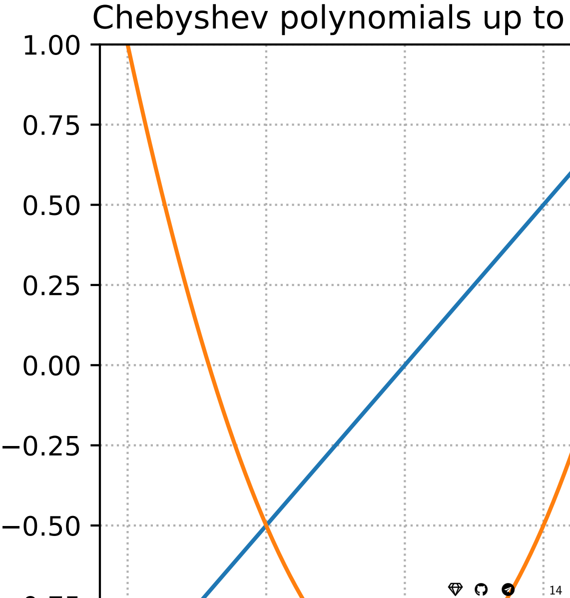


Полиномы Чебышёва

Полиномы Чебышёва оказываются оптимальным ответом на вопрос, который мы задавали. Соответствующим образом масштабированные, они минимизируют абсолютное значение в желаемом интервале $[\mu, L]$ при условии, что значение равно 1 в начале.

$$
\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_k(x) &= 2xT_{k-1}(x) - T_{k-2}(x), \qquad k \geq 2.
\end{aligned}
$$

Давайте построим стандартные полиномы Чебышёва (без масштабирования):

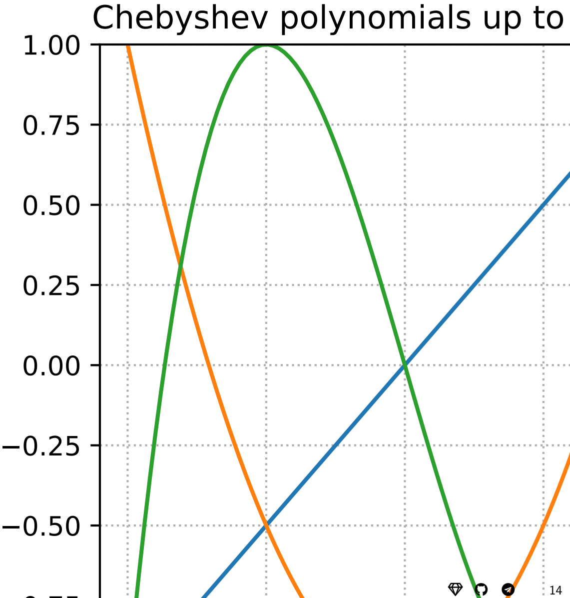


Полиномы Чебышёва

Полиномы Чебышёва оказываются оптимальным ответом на вопрос, который мы задавали. Соответствующим образом масштабированные, они минимизируют абсолютное значение в желаемом интервале $[\mu, L]$ при условии, что значение равно 1 в начале.

$$
\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_k(x) &= 2xT_{k-1}(x) - T_{k-2}(x), \qquad k \geq 2.
\end{aligned}
$$

Давайте построим стандартные полиномы Чебышёва (без масштабирования):

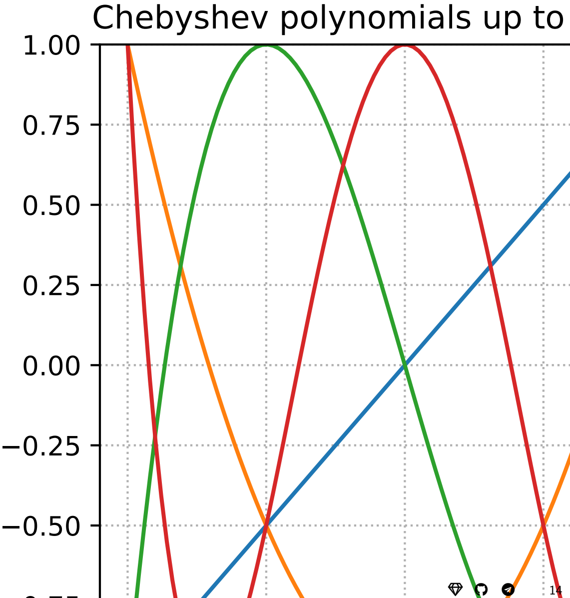


Полиномы Чебышёва

Полиномы Чебышёва оказываются оптимальным ответом на вопрос, который мы задавали. Соответствующим образом масштабированные, они минимизируют абсолютное значение в желаемом интервале $[\mu, L]$ при условии, что значение равно 1 в начале.

$$
\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_k(x) &= 2xT_{k-1}(x) - T_{k-2}(x), \qquad k \geq 2.
\end{aligned}
$$

Давайте построим стандартные полиномы Чебышёва (без масштабирования):

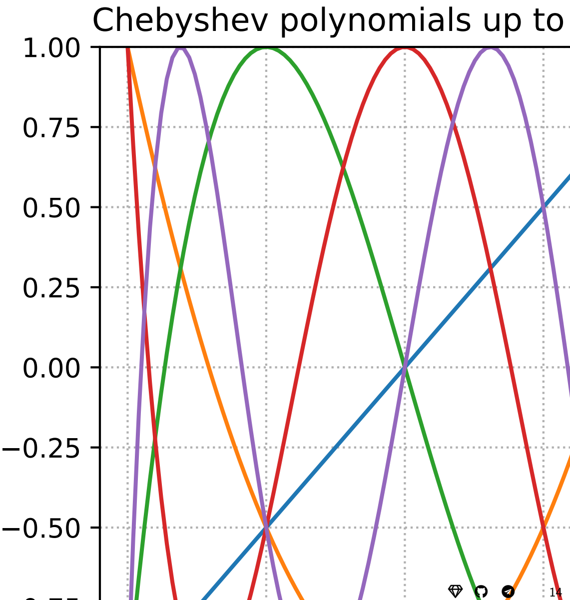


Полиномы Чебышёва

Полиномы Чебышёва оказываются оптимальным ответом на вопрос, который мы задавали. Соответствующим образом масштабированные, они минимизируют абсолютное значение в желаемом интервале $[\mu, L]$ при условии, что значение равно 1 в начале.

$$
\begin{aligned}
T_0(x) &= 1 \\
T_1(x) &= x \\
T_k(x) &= 2xT_{k-1}(x) - T_{k-2}(x), \qquad k \geq 2.
\end{aligned}
$$

Давайте построим стандартные полиномы Чебышёва (без масштабирования):



Масштабированные полиномы Чебышёва

Исходные полиномы Чебышёва определяются на интервале $[-1, 1]$. Чтобы использовать их для наших целей, нам нужно их масштабировать на интервал $[\mu, L]$.

Масштабированные полиномы Чебышёва

Исходные полиномы Чебышёва определяются на интервале $[-1, 1]$. Чтобы использовать их для наших целей, нам нужно их масштабировать на интервал $[\mu, L]$.

Мы будем использовать следующую аффинную трансформацию:

$$
x = \frac{L + \mu - 2a}{L - \mu}, \quad a \in [\mu, L], \quad x \in [-1, 1].
$$

Обратите внимание, что $x = 1$ соответствует $a = \mu$, $x = -1$ соответствует $a = L$ и $x = 0$ соответствует $a = \frac{\mu+L}{2}$. Эта трансформация гарантирует, что поведение полинома Чебышёва на интервале $[-1, 1]$ отражается в интервал $[\mu, L]$

Масштабированные полиномы Чебышёва

Исходные полиномы Чебышёва определяются на интервале $[-1, 1]$. Чтобы использовать их для наших целей, нам нужно их масштабировать на интервал $[\mu, L]$.

Мы будем использовать следующую аффинную трансформацию:

$$x = \frac{L + \mu - 2a}{L - \mu}, \quad a \in [\mu, L], \quad x \in [-1, 1].$$

Обратите внимание, что $x = 1$ соответствует $a = \mu$, $x = -1$ соответствует $a = L$ и $x = 0$ соответствует $a = \frac{\mu+L}{2}$. Эта трансформация гарантирует, что поведение полинома Чебышёва на интервале $[-1, 1]$ отражается в интервал $[\mu, L]$

В нашем анализе ошибок мы требуем, чтобы полином был равен 1 в 0 (т.е., $p_k(0) = 1$). После применения трансформации значение T_k в точке, соответствующей $a = 0$, может не быть 1. Таким образом, мы нормируем полином T_k, деля его на значение $T_k\left(\frac{L+\mu}{L-\mu}\right)$:

$$\frac{L + \mu}{L - \mu}, \qquad \text{гарантируя, что} \qquad P_k(0) = T_k\left(\frac{L + \mu - 0}{L - \mu}\right) \cdot T_k\left(\frac{L + \mu}{L - \mu}\right)^{-1} = 1.$$

Масштабированные полиномы Чебышёва

Исходные полиномы Чебышёва определяются на интервале $[-1, 1]$. Чтобы использовать их для наших целей, нам нужно их масштабировать на интервал $[\mu, L]$.

Мы будем использовать следующую аффинную трансформацию:

$$x = \frac{L + \mu - 2a}{L - \mu}, \quad a \in [\mu, L], \quad x \in [-1, 1].$$

Обратите внимание, что $x = 1$ соответствует $a = \mu$, $x = -1$ соответствует $a = L$ и $x = 0$ соответствует $a = \frac{\mu+L}{2}$. Эта трансформация гарантирует, что поведение полинома Чебышёва на интервале $[-1, 1]$ отражается в интервал $[\mu, L]$

В нашем анализе ошибок мы требуем, чтобы полином был равен 1 в 0 (т.е., $p_k(0) = 1$). После применения трансформации значение T_k в точке, соответствующей $a = 0$, может не быть 1. Таким образом, мы нормируем полином T_k, деля его на значение $T_k\left(\frac{L+\mu}{L-\mu}\right)$:

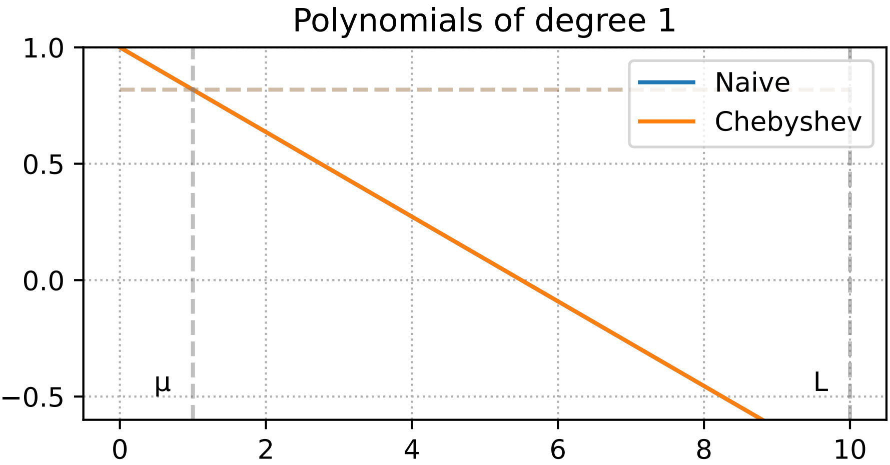
$$\frac{L+\mu}{L-\mu}, \qquad \text{гарантируя, что} \qquad P_k(0) = T_k\left(\frac{L+\mu-0}{L-\mu}\right) \cdot T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1} = 1.$$

Давайте построим масштабированные полиномы Чебышёва

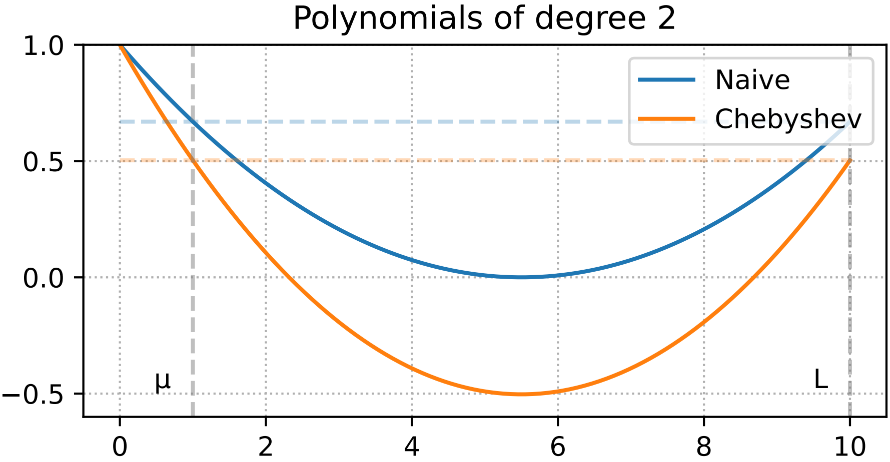
$$P_k(a) = T_k\left(\frac{L+\mu-2a}{L-\mu}\right) \cdot T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1}$$

и наблюдаем, что они значительно лучше ведут себя в интервале $[\mu, L]$ по сравнению с наивными полиномами.

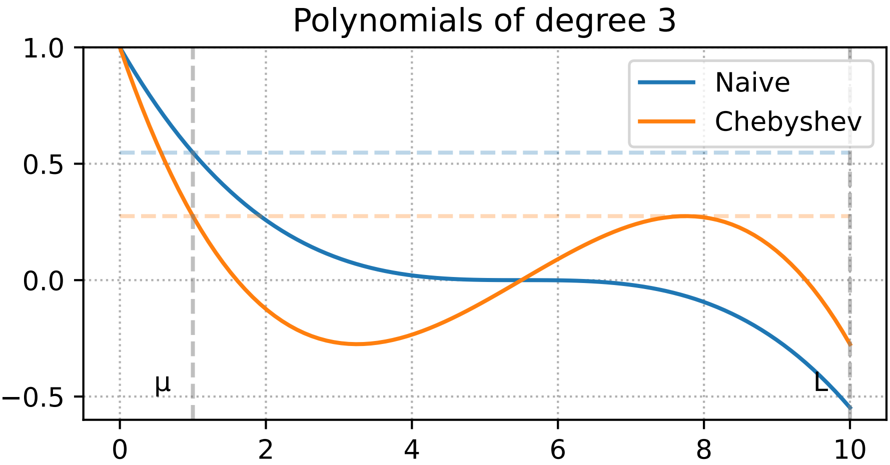
Масштабированные полиномы Чебышёва



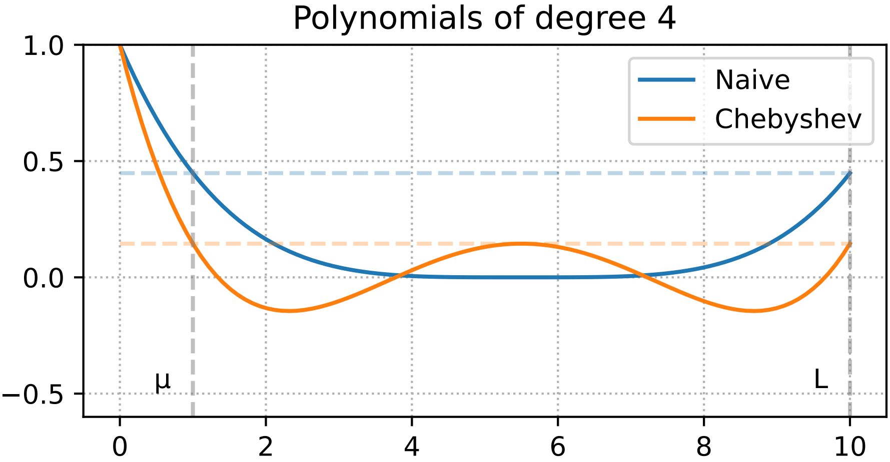
Масштабированные полиномы Чебышёва



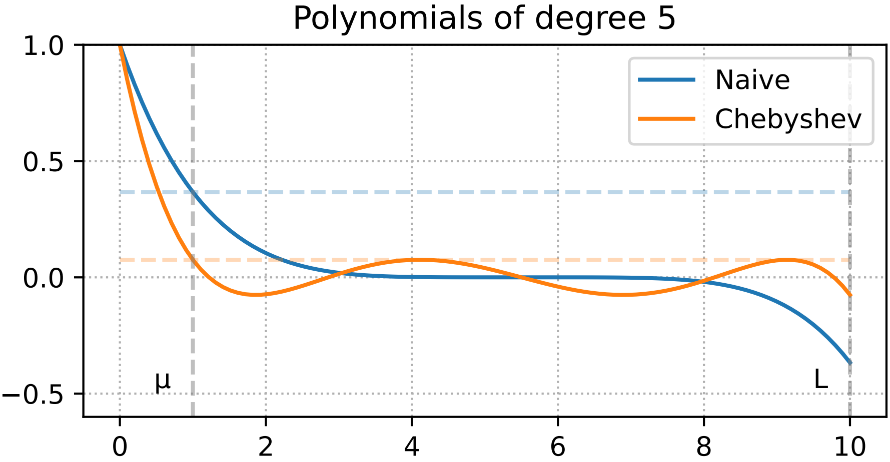
Масштабированные полиномы Чебышёва



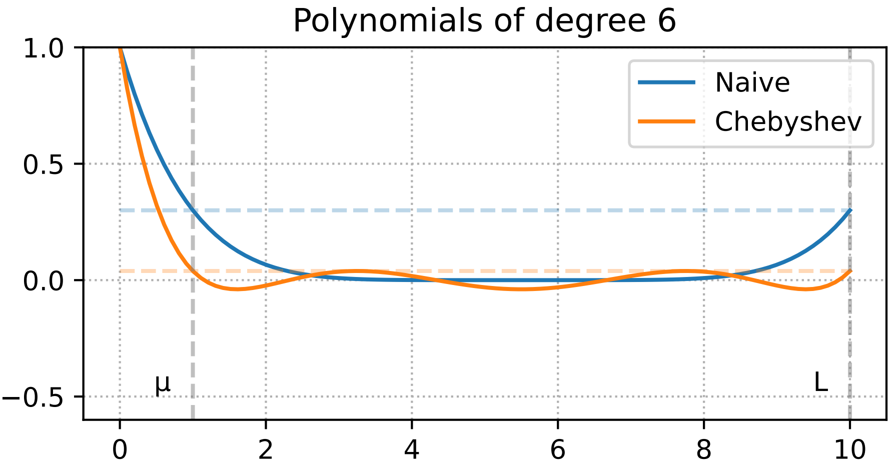
Масштабированные полиномы Чебышёва



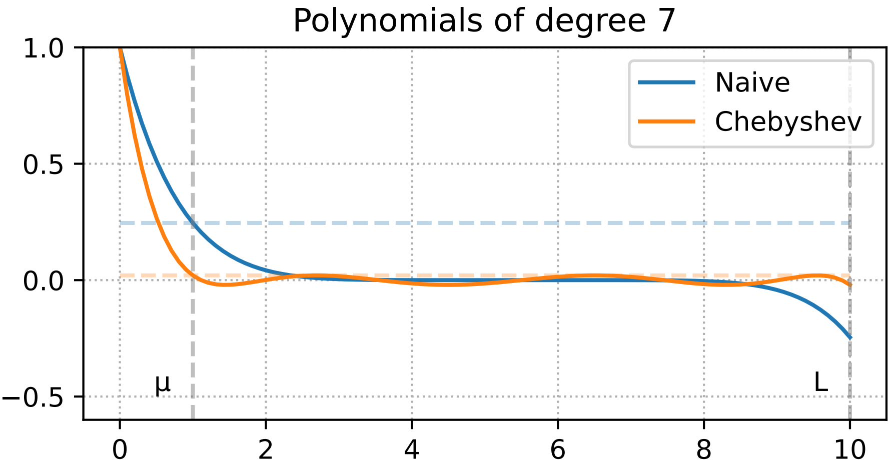
Масштабированные полиномы Чебышёва



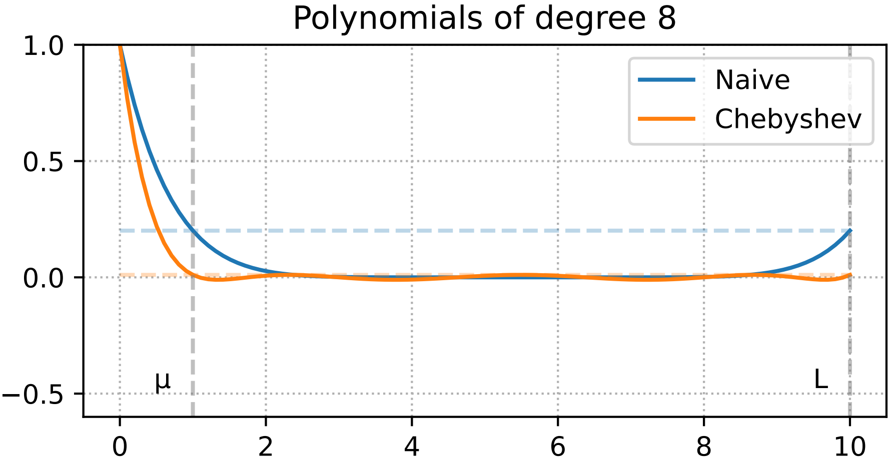
Масштабированные полиномы Чебышёва



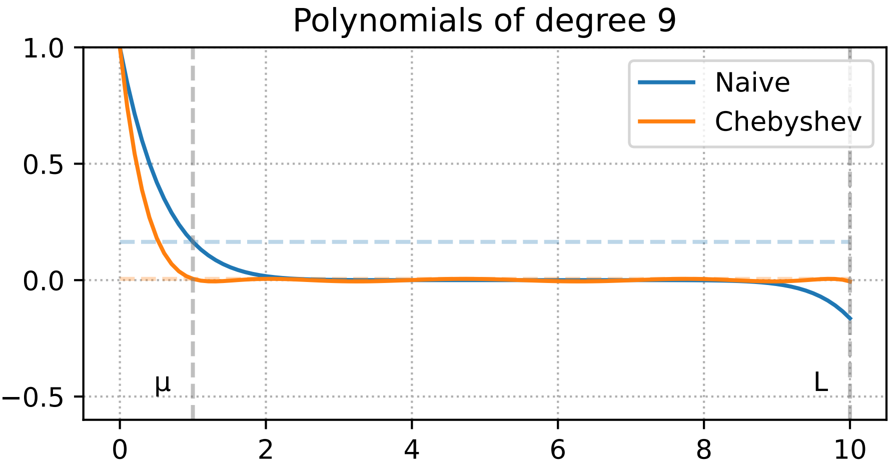
Масштабированные полиномы Чебышёва



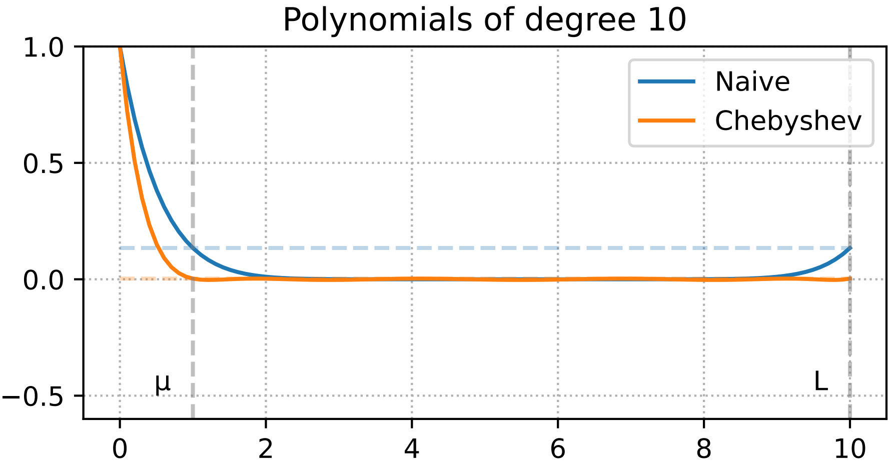
Масштабированные полиномы Чебышёва



Масштабированные полиномы Чебышёва



Масштабированные полиномы Чебышёва



Верхняя граница ошибки с помощью полиномов Чебышёва

Мы видим, что максимальное значение полинома Чебышёва на интервале $[\mu, L]$ достигается в точке $a = \mu$. Следовательно, мы можем использовать следующую верхнюю границу:

$$\|P_k(A)\|_2 \leq P_k(\mu) = T_k\left(\frac{L+\mu-2\mu}{L-\mu}\right) \cdot T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1} = T_k(1) \cdot T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1} = T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1}$$

Верхняя граница ошибки с помощью полиномов Чебышёва

Мы видим, что максимальное значение полинома Чебышёва на интервале $[\mu, L]$ достигается в точке $a = \mu$. Следовательно, мы можем использовать следующую верхнюю границу:

$$
\|P_k(A)\|_2 \leq P_k(\mu) = T_k\left(\frac{L+\mu-2\mu}{L-\mu}\right) \cdot T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1} = T_k(1) \cdot T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1} = T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1}
$$

Используя определение числа обусловленности $\varkappa = \frac{L}{\mu}$, мы получаем:

$$
\|P_k(A)\|_2 \leq T_k\left(\frac{\varkappa+1}{\varkappa-1}\right)^{-1} = T_k\left(1+\frac{2}{\varkappa-1}\right)^{-1} = T_k(1+\epsilon)^{-1}, \quad \epsilon = \frac{2}{\varkappa-1}.
$$

Верхняя граница ошибки с помощью полиномов Чебышёва

Мы видим, что максимальное значение полинома Чебышёва на интервале $[\mu, L]$ достигается в точке $a = \mu$. Следовательно, мы можем использовать следующую верхнюю границу:

$$\|P_k(A)\|_2 \leq P_k(\mu) = T_k\left(\frac{L+\mu-2\mu}{L-\mu}\right) \cdot T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1} = T_k(1) \cdot T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1} = T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1}$$

Используя определение числа обусловленности $\varkappa = \frac{L}{\mu}$, мы получаем:

$$\|P_k(A)\|_2 \leq T_k\left(\frac{\varkappa+1}{\varkappa-1}\right)^{-1} = T_k\left(1 + \frac{2}{\varkappa-1}\right)^{-1} = T_k(1+\epsilon)^{-1}, \quad \epsilon = \frac{2}{\varkappa-1}.$$

Следовательно, нам нужно только понять значение T_k в $1+\epsilon$. Это то, откуда берется ускорение. Мы будем ограничивать это значение сверху величиной $\mathcal{O}\left(\frac{1}{\sqrt{\epsilon}}\right)$.

Верхняя граница ошибки с помощью полиномов Чебышёва

Чтобы ограничить $|P_k|$ сверху, нам нужно оценить $|T_k(1+\epsilon)|$ снизу.

Верхняя граница ошибки с помощью полиномов Чебышёва

Чтобы ограничить $|P_k|$ сверху, нам нужно оценить $|T_k(1+\epsilon)|$ снизу.

1. Для любого $x \geq 1$, полином Чебышёва первого рода может быть записан как

$$
\begin{aligned}
T_k(x) &= \cosh\left(k \operatorname{arccosh}(x)\right) \\
T_k(1+\epsilon) &= \cosh\left(k \operatorname{arccosh}(1+\epsilon)\right).
\end{aligned}
$$

Верхняя граница ошибки с помощью полиномов Чебышёва

Чтобы ограничить $|P_k|$ сверху, нам нужно оценить $|T_k(1+\epsilon)|$ снизу.

1. Для любого $x \geq 1$, полином Чебышёва первого рода может быть записан как

$$
\begin{aligned}
T_k(x) &= \cosh\left(k \operatorname{arccosh}(x)\right) \\
T_k(1+\epsilon) &= \cosh\left(k \operatorname{arccosh}(1+\epsilon)\right).
\end{aligned}
$$

2. Помним, что:

$$
\cosh(x) = \frac{e^x + e^{-x}}{2} \quad \operatorname{arccosh}(x) = \ln(x+\sqrt{x^2-1}).
$$

Верхняя граница ошибки с помощью полиномов Чебышёва

Чтобы ограничить $|P_k|$ сверху, нам нужно оценить $|T_k(1+\epsilon)|$ снизу.

1. Для любого $x \geq 1$, полином Чебышёва первого рода может быть записан как

$$
\begin{aligned}
T_k(x) &= \cosh\left(k \operatorname{arccosh}(x)\right) \\
T_k(1+\epsilon) &= \cosh\left(k \operatorname{arccosh}(1+\epsilon)\right).
\end{aligned}
$$

2. Помним, что:

$$
\cosh(x) = \frac{e^x + e^{-x}}{2} \qquad \operatorname{arccosh}(x) = \ln(x + \sqrt{x^2 - 1}).
$$

3. Пусть $\phi = \operatorname{arccosh}(1+\epsilon)$,

$$
e^{\phi} = 1 + \epsilon + \sqrt{2\epsilon + \epsilon^2} \geq 1 + \sqrt{\epsilon}.
$$

Верхняя граница ошибки с помощью полиномов Чебышёва

Чтобы ограничить $|P_k|$ сверху, нам нужно оценить $|T_k(1+\epsilon)|$ снизу.

1. Для любого $x \geq 1$, полином Чебышёва первого рода может быть записан как

$$
\begin{aligned}
T_k(x) &= \cosh\left(k \operatorname{arccosh}(x)\right) \\
T_k(1+\epsilon) &= \cosh\left(k \operatorname{arccosh}(1+\epsilon)\right).
\end{aligned}
$$

2. Помним, что:

$$
\cosh(x) = \frac{e^x + e^{-x}}{2} \qquad \operatorname{arccosh}(x) = \ln(x + \sqrt{x^2 - 1}).
$$

3. Пусть $\phi = \operatorname{arccosh}(1+\epsilon)$,

$$
e^{\phi} = 1 + \epsilon + \sqrt{2\epsilon + \epsilon^2} \geq 1 + \sqrt{\epsilon}.
$$

4. Следовательно,

$$
\begin{aligned}
T_k(1+\epsilon) &= \cosh\left(k \operatorname{arccosh}(1+\epsilon)\right) \\
&= \cosh\left(k\phi\right) \\
&= \frac{e^{k\phi} + e^{-k\phi}}{2} \geq \frac{e^{k\phi}}{2} \\
&= \frac{\left(1 + \sqrt{\epsilon}\right)^k}{2}.
\end{aligned}
$$

Верхняя граница ошибки с помощью полиномов Чебышёва

Чтобы ограничить $|P_k|$ сверху, нам нужно оценить $|T_k(1+\epsilon)|$ снизу.

1. Для любого $x \geq 1$, полином Чебышёва первого рода может быть записан как
$$
\begin{aligned}
T_k(x) &= \cosh\left(k \operatorname{arccosh}(x)\right) \\
T_k(1+\epsilon) &= \cosh\left(k \operatorname{arccosh}(1+\epsilon)\right).
\end{aligned}
$$
2. Помним, что:
$$
\cosh(x) = \frac{e^x + e^{-x}}{2} \quad \operatorname{arccosh}(x) = \ln(x + \sqrt{x^2 - 1}).
$$
3. Пусть $\phi = \operatorname{arccosh}(1+\epsilon)$,
$$
e^{\phi} = 1 + \epsilon + \sqrt{2\epsilon + \epsilon^2} \geq 1 + \sqrt{\epsilon}.
$$
4. Следовательно,
$$
\begin{aligned}
T_k(1+\epsilon) &= \cosh\left(k \operatorname{arccosh}(1+\epsilon)\right) \\
&= \cosh\left(k\phi\right) \\
&= \frac{e^{k\phi} + e^{-k\phi}}{2} \geq \frac{e^{k\phi}}{2} \\
&= \frac{\left(1+\sqrt{\epsilon}\right)^k}{2}.
\end{aligned}
$$
5. Наконец, мы получаем:
$$
\begin{aligned}
\|e_k\| &\leq \|P_k(A)\| \|e_0\| \leq \frac{2}{\left(1+\sqrt{\epsilon}\right)^k} \|e_0\| \\
&\leq 2\left(1 + \sqrt{\frac{2}{\varkappa - 1}}\right)^{-k} \|e_0\| \\
&\leq 2\exp\left(-\sqrt{\frac{2}{\varkappa - 1}} k\right) \|e_0\|
\end{aligned}
$$

Ускоренный метод [1/2]

Из-за рекурсивного определения полиномов Чебышёва, мы непосредственно получаем итерационную схему ускорения. Переформулируя рекуррентное соотношение в терминах наших масштабированных полиномов Чебышёва, мы получаем:

$$T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$$

Поскольку $x = \frac{L+\mu-2a}{L-\mu}$, и:

$$P_k(a) = T_k\left(\frac{L+\mu-2a}{L-\mu}\right) T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1}$$

$$T_k\left(\frac{L+\mu-2a}{L-\mu}\right) = P_k(a) T_k\left(\frac{L+\mu}{L-\mu}\right)$$

Ускоренный метод [1/2]

Из-за рекурсивного определения полиномов Чебышёва, мы непосредственно получаем итерационную схему ускорения. Переформулируя рекуррентное соотношение в терминах наших масштабированных полиномов Чебышёва, мы получаем:

$$T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$$

Поскольку $x = \frac{L+\mu-2a}{L-\mu}$, и:

$$P_k(a) = T_k\left(\frac{L+\mu-2a}{L-\mu}\right) T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1}$$

$$T_k\left(\frac{L+\mu-2a}{L-\mu}\right) = P_k(a) T_k\left(\frac{L+\mu}{L-\mu}\right)$$

$$T_{k-1}\left(\frac{L+\mu-2a}{L-\mu}\right) = P_{k-1}(a) T_{k-1}\left(\frac{L+\mu}{L-\mu}\right)$$

$$T_{k+1}\left(\frac{L+\mu-2a}{L-\mu}\right) = P_{k+1}(a) T_{k+1}\left(\frac{L+\mu}{L-\mu}\right)$$

$$P_{k+1}(a)t_{k+1} = 2\frac{L+\mu-2a}{L-\mu}P_k(a)t_k - P_{k-1}(a)t_{k-1}, \text{ where } t_k = T_k\left(\frac{L+\mu}{L-\mu}\right)$$

$$P_{k+1}(a) = 2\frac{L+\mu-2a}{L-\mu}P_k(a)\frac{t_k}{t_{k+1}} - P_{k-1}(a)\frac{t_{k-1}}{t_{k+1}}$$

Ускоренный метод [1/2]

Из-за рекурсивного определения полиномов Чебышёва, мы непосредственно получаем итерационную схему ускорения. Переформулируя рекуррентное соотношение в терминах наших масштабированных полиномов Чебышёва, мы получаем:

$$
T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)
$$

Поскольку $x = \frac{L+\mu-2a}{L-\mu}$, и:

$$
P_k(a) = T_k\left(\frac{L+\mu-2a}{L-\mu}\right) T_k\left(\frac{L+\mu}{L-\mu}\right)^{-1}
$$

$$
T_k\left(\frac{L+\mu-2a}{L-\mu}\right) = P_k(a) T_k\left(\frac{L+\mu}{L-\mu}\right)
$$

$$
T_{k-1}\left(\frac{L+\mu-2a}{L-\mu}\right) = P_{k-1}(a) T_{k-1}\left(\frac{L+\mu}{L-\mu}\right)
$$

$$
T_{k+1}\left(\frac{L+\mu-2a}{L-\mu}\right) = P_{k+1}(a) T_{k+1}\left(\frac{L+\mu}{L-\mu}\right)
$$

$$
P_{k+1}(a) t_{k+1} = 2\frac{L+\mu-2a}{L-\mu} P_k(a) t_k - P_{k-1}(a) t_{k-1}, \text{ where } t_k = T_k\left(\frac{L+\mu}{L-\mu}\right)
$$

$$
P_{k+1}(a) = 2\frac{L+\mu-2a}{L-\mu} P_k(a) \frac{t_k}{t_{k+1}} - P_{k-1}(a) \frac{t_{k-1}}{t_{k+1}}
$$

Поскольку мы имеем $P_{k+1}(0) = P_k(0) = P_{k-1}(0) = 1$, мы можем записать метод в следующей форме:

$$
P_{k+1}(a) = (1 - \alpha_k a) P_k(a) + \beta_k \left(P_k(a) - P_{k-1}(a)\right).
$$

Ускоренный метод [2/2]

Перегруппируя члены, мы получаем:

$$P_{k+1}(a) = (1 + \beta_k)P_k(a) - \alpha_k a P_k(a) - \beta_k P_{k-1}(a),$$

$$P_{k+1}(a) = 2\frac{L+\mu}{L-\mu}\frac{t_k}{t_{k+1}}P_k(a) - \frac{4a}{L-\mu}\frac{t_k}{t_{k+1}}P_k(a) - \frac{t_{k-1}}{t_{k+1}}P_{k-1}(a)$$

Ускоренный метод [2/2]

Перегруппируя члены, мы получаем:

$$P_{k+1}(a) = (1+\beta_k)P_k(a) - \alpha_k a P_k(a) - \beta_k P_{k-1}(a),$$

$$P_{k+1}(a) = 2\frac{L+\mu}{L-\mu}\frac{t_k}{t_{k+1}}P_k(a) - \frac{4a}{L-\mu}\frac{t_k}{t_{k+1}}P_k(a) - \frac{t_{k-1}}{t_{k+1}}P_{k-1}(a)$$

$$\begin{cases} \beta_k = \dfrac{t_{k-1}}{t_{k+1}}, \\ \alpha_k = \dfrac{4}{L-\mu}\dfrac{t_k}{t_{k+1}}, \\ 1+\beta_k = 2\dfrac{L+\mu}{L-\mu}\dfrac{t_k}{t_{k+1}} \end{cases}$$

Ускоренный метод [2/2]

Перегруппируя члены, мы получаем:

$$P_{k+1}(a) = (1+\beta_k)P_k(a) - \alpha_k a P_k(a) - \beta_k P_{k-1}(a),$$

$$P_{k+1}(a) = 2\frac{L+\mu}{L-\mu}\frac{t_k}{t_{k+1}}P_k(a) - \frac{4a}{L-\mu}\frac{t_k}{t_{k+1}}P_k(a) - \frac{t_{k-1}}{t_{k+1}}P_{k-1}(a)$$

$$\begin{cases} \beta_k = \dfrac{t_{k-1}}{t_{k+1}}, \\ \alpha_k = \dfrac{4}{L-\mu}\dfrac{t_k}{t_{k+1}}, \\ 1+\beta_k = 2\dfrac{L+\mu}{L-\mu}\dfrac{t_k}{t_{k+1}} \end{cases}$$

Мы почти закончили ☻. Помним, что $e_{k+1} = P_{k+1}(A)e_0$. Также отметим, что мы работаем с квадратичной задачей, поэтому мы можем предположить $x^* = 0$ без потери общности. В этом случае $e_0 = x_0$ и $e_{k+1} = x_{k+1}$.

$$\begin{aligned} x_{k+1} &= P_{k+1}(A)x_0 = (I - \alpha_k A)P_k(A)x_0 + \beta_k\left(P_k(A) - P_{k-1}(A)\right)x_0 \\ &= (I-\alpha_k A)x_k + \beta_k\left(x_k - x_{k-1}\right) \end{aligned}$$

Ускоренный метод [2/2]

Перегруппируя члены, мы получаем:

$$
\begin{cases}
\beta_k = \dfrac{t_{k-1}}{t_{k+1}}, \\
\alpha_k = \dfrac{4}{L-\mu}\dfrac{t_k}{t_{k+1}}, \\
1 + \beta_k = 2\dfrac{L+\mu}{L-\mu}\dfrac{t_k}{t_{k+1}}
\end{cases}
$$

$$P_{k+1}(a) = (1+\beta_k)P_k(a) - \alpha_k a P_k(a) - \beta_k P_{k-1}(a),$$

$$P_{k+1}(a) = 2\frac{L+\mu}{L-\mu}\frac{t_k}{t_{k+1}}P_k(a) - \frac{4a}{L-\mu}\frac{t_k}{t_{k+1}}P_k(a) - \frac{t_{k-1}}{t_{k+1}}P_{k-1}(a)$$

Мы почти закончили ☻. Помним, что $e_{k+1} = P_{k+1}(A)e_0$. Также отметим, что мы работаем с квадратичной задачей, поэтому мы можем предположить $x^* = 0$ без потери общности. В этом случае $e_0 = x_0$ и $e_{k+1} = x_{k+1}$.

$$
\begin{aligned}
x_{k+1} &= P_{k+1}(A)x_0 = (I - \alpha_k A)P_k(A)x_0 + \beta_k\left(P_k(A) - P_{k-1}(A)\right)x_0 \\
&= (I - \alpha_k A)x_k + \beta_k\left(x_k - x_{k-1}\right)
\end{aligned}
$$

Для квадратичной задачи мы имеем $\nabla f(x_k) = Ax_k$, поэтому мы можем переписать обновление как:

$$\boxed{x_{k+1} = x_k - \alpha_k \nabla f(x_k) + \beta_k\left(x_k - x_{k-1}\right)}$$

Ускорение из первых принципов