Основы матричных вычислений, Экзамен (Теория)

Версия от 27.06.2021 15:35

Содержание

Связь прямой и обратной ошибок через число обусловленности.

Вспомним кто есть ошибки:

$\dfrac{\left\|\widetilde{f}(x) - f(x)\right\|}{\|f(x)\|}$ — прямая ошибка, где f — трушное решение, а $\widetilde{f}$ — алгоритм

$\dfrac{\|\widetilde{x} - x\|}{\|x\|}$ — обратная ошибка, где $\widetilde{x} : f(\widetilde{x}) = \widetilde{f}(x)$

Предположим, что наша задача обратно устойчива, а именно: $f(\widetilde{x}) = \widetilde{f}(x)$, $\dfrac{\|\widetilde{x} - x\|}{\|x\|} = \mathrm{O}(\varepsilon_{machine})$

$$
\begin{aligned}
\text{forward\_err} &= \frac{\|\widetilde{f}(x) - f(x)\|}{\|f(x)\|} \\
&= \frac{\|f(\widetilde{x}) - f(x)\|}{\|f(x)\|} \\
&= \frac{\|f(x + \Delta x) - f(x)\|}{\|f(x)\|} \\
&= \frac{\|f(x) + f'(x) \cdot \Delta x + \mathrm{O}(\|\Delta x\|^2) - f(x)\|}{\|f(x)\|} \\
&= \frac{\|f'(x) \cdot \frac{\Delta x}{\|\Delta x\|} + \mathrm{O}(\|\Delta x\|)\|\|}{\|f(x)\|} \cdot \|\Delta x \\
&\leqslant \frac{\|f'(x) \cdot \frac{\Delta x}{\|\Delta x\|}\| + \mathrm{O}(\|\Delta x\|)}{\|f(x)\|} \cdot \|\Delta x\| \\
&\leqslant \frac{\|f'(x)\| + \mathrm{O}(\|\Delta x\|)}{\|f(x)\|} \cdot \|\Delta x\| \\
&= \frac{\|f'(x)\| \cdot \|x\| + \mathrm{O}(\|\Delta x\|) \cdot \|x\|}{\|f(x)\|} \cdot \frac{\|\Delta x\|}{\|x\|} \\
&= \left(\frac{\|f'(x)\| \cdot \|x\|}{\|f(x)\|} + \underbrace{\frac{\mathrm{O}(\|\Delta x\|) \cdot \|x\|}{\|f(x)\|}}_{\substack{=\mathrm{O}(\|\Delta x\|)\text{ в силу} \\ \text{фиксированности } x}} \right) \cdot \frac{\|\Delta x\|}{\|x\|} \\
&= (\mathrm{cond}(f, x) + \mathrm{O}(\|\Delta x\|)) \cdot \text{backward\_err} \\
&= (\mathrm{cond}(f, x) + \mathrm{O}(\varepsilon_{machine})) \cdot \text{backward\_err} \\
&\approx \mathrm{cond}(f, x) \cdot \text{backward\_err}
\end{aligned}
$$

Критерий сходимости ряда Неймана.

Напомним определение спектрального радиуса $\rho(A) = \max\limits_i |\lambda_i(A)| = \lim\limits_{k\to\infty} \left\|A^k\right\|^{1/k}$

Критерий сходимости ряда Неймана: $\sum\limits_{i=0}^{k} A^k$ сходится $\Longleftrightarrow \rho(A) < 1$.

$\Leftarrow$: $\quad A = UTU^{-1}$ — разложение Шура (здесь T верхнетреугольная).
Введем матрицу $D_\varepsilon = \operatorname{diag}\left(1, \varepsilon, \ldots, \varepsilon^{n-1}\right)$. Оказывается, что $(D_\varepsilon^{-1} T D_\varepsilon)_{ij} = \varepsilon^{j-i} t_{ij}$, то есть каждый элемент верхнего треугольника домножается на эпсилон в какой-то степени. Нижний треугольник — нули, так как T верхнетреугольная. Остаются $|t_{ii}|$ — но это модули собственных значений, а из $\rho(T) < 1$ следует что все меньше единицы.

Из этого следует, что $\exists \varepsilon : \left\|D_\varepsilon^{-1} T D_\varepsilon\right\|_1 < 1$. То есть $\sum\limits_{k=0}^{\infty} \left(D_\varepsilon^{-1} T D_\varepsilon\right)^k = \sum\limits_{k=0}^{\infty} D_\varepsilon^{-1} T^k D_\varepsilon$ сходится.

Вынесем по матрице слева и справа, сходимость не сломается: $\sum\limits_{k=0}^{\infty} T^k$ сходится.

Теперь занесем по матрице слева и справа, но уже другие, тогда $\sum\limits_{k=0}^{\infty} UT^kU^{-1} = \sum\limits_{k=0}^{\infty} \left(UTU^{-1}\right)^k$ сходится.

$\Rightarrow$: $\quad$ пусть $\rho(A) \geqslant 1 \implies \exists |\lambda_i| > 1$, но ряд сходится.
$\exists \|x\|_2 = 1 : Ax = \lambda_i x \implies A^k x = \lambda_i^k x$.
$\left\|A^k\right\|_2 = \left\|A^k\right\|_2 \|x\|_2 \geqslant \left\|A^k x\right\|_2 = \left\|\lambda^k x\right\|_2 = |\lambda^k| \|x\|_2 = |\lambda^k| \geqslant 1$, то есть $\left\|A^k\right\|_2 \nrightarrow 0$, получается ряд не сходится.

Существование и единственность LU и LDL разложений.

Определение. Пусть есть матрица $A \in \mathbb{R}^{n\times n}$. Разложение $A = LU$ называется LU-разложением, если матрица L нижнетреугольная с единицами на диагонали, а матрица U верхнетреугольная.

Определение. Матрица A называется **строго регулярной**, если все её ведущие подматрицы невырождены. (ведущие подматрицы — верхние левые $k \times k$ блоки).

Теорема (Существование LU-разложения)**.** Пусть $\det(A) \neq 0$. Тогда A имеет LU-разложение $\iff$ A строго регулярна.

Доказательство.
$\Rightarrow$
A имеет LU-разложение, то есть $A = LU$.
Мы знаем, что матрица невырождена, то есть

$$0 \neq \det(A) = \underbrace{\det(L)}_{1}\det(U) = u_{11} \cdot \ldots \cdot u_{nn}.$$

Следовательно, $u_{kk} \neq 0$ для любого $k \in \{1, \ldots, n\}$. Дальше нам надо убедиться, что матрица A строго регулярна. То есть, надо проверить что ведущие подматрицы тоже невырождены. Запишем для ведущих подматриц:

$$A = \begin{pmatrix} L_k & 0 \\ * & * \end{pmatrix}\begin{pmatrix} U_k & * \\ 0 & * \end{pmatrix} = \begin{pmatrix} L_kU_k & * \\ * & * \end{pmatrix}.$$

Обозначим $A_k := L_kU_k$. Тогда $\det(A_k) = \underbrace{\det(L_k)}_{1}\det(U_k) = u_{11} \cdot \ldots \cdot u_{kk} \neq 0$, аналогично случаю с полной матрицей.

$\Leftarrow$
Доказываем по индукции. Мы считаем, что пусть для $n-1$ уже доказано утверждение. Докажем для n.
Пусть матрица $A = \begin{pmatrix} a & c^t \\ b & D \end{pmatrix}$. Здесь a — число, не равное нулю в силу строгой регулярности, D — матрица $n-1 \times n-1$.
Мы считаем, что для D мы уже умеем искать LU — разложение. Давайте попробуем преобразовать нашу матрицу, чтобы она привелась к блочно-верхнетреугольному виду:

$$\begin{pmatrix} 1 & 0 \\ -\dfrac{1}{a}b & I \end{pmatrix}\begin{pmatrix} a & c^t \\ b & D \end{pmatrix} = \begin{pmatrix} a & c^T \\ 0 & D - \dfrac{1}{a}bc^T \end{pmatrix} =: A'.$$

Блок $D - \dfrac{1}{a}bc^T$ называется дополнением по Шуру матрицы A. Обозначим $A_1 := D - \dfrac{1}{a}bc^T$.
Докажем, что A_1 строго регулярна (было в ДЗ). A' получилась из A с помощью $n-1$-го элементарного преобразования первого типа (вычесть из строки другую, домноженную на коэффициент). Помним, что такие элементарные преобразования не меняют определитель матрицы, поэтому $0 \neq \Delta_k(A) = \Delta_k(A')$ для любого $k \in \{1, \ldots, n\}$. (здесь $\Delta_k(A)$ — главный угловой минор матрицы A). Но в матрице A' мы видим угол нулей, поэтому $0 \neq \Delta_k(A') = a \cdot \Delta_{k-1}(A_1)$ для любого $k \in \{1, \ldots, n\}$. Следовательно, $\Delta_i(A_1) \neq 0$ для любого $i \in \{1, \ldots, n-1\} \implies A_1$ строго регулярна.
Продолжим доказательство теоремы. По предположению индукции тогда считаем, что A_1 имеет LU-разложение: $D - \dfrac{1}{a}bc^T = A_1 = L_1U_1$. Этого уже достаточно для того, чтобы построить LU-разложение самой матрицы A:

$$\begin{pmatrix} 1 & 0 \\ \dfrac{1}{a}b & L1 \end{pmatrix}\begin{pmatrix} a & c^T \\ 0 & U_1 \end{pmatrix} = \begin{pmatrix} a & c^T \\ b & \dfrac{1}{a}bc^T + L_1U_1 \end{pmatrix} = \begin{pmatrix} a & c^T \\ b & D \end{pmatrix}.$$

■

Утверждение. LU-разложение определяется единственным образом.

Доказательство.
Предположим, что есть два разложения:

$$A = L_1U_1 = L_2U_2.$$

Преобразуем равенство:

$$L_2^{-1}L_1 = U_2U_1^{-1}.$$

Обратная к нижнетреугольной матрице — нижнетреугольная матрица, и произведение нижнетреугольных — тоже нижнетреугольная. Для верхнетреугольных то же самое. Значит, $L_2^{-1}L_1$ — диагональная матрица. Более того, это единичная матрица (в силу того, что на диагонали матриц L_2 и L_1 стоят 1). Значит,

$$L_1 = L_2;$$
$$U_1 = U_2.$$

■

Следствие (LDL-разложение)**.** Пусть $A \in \mathbb{C}^{n\times n}$ является строго регулярной и $A = A^*$. Тогда $\exists L$ — нижнетреугольная и D — диагональная, такие что

$$A = LDL^*.$$

Доказательство.
A — строго регулярна, следовательно,

$$A = LU = L\underbrace{D}_{\mathrm{diag}(u_{11},\ldots,u_{nn})}D^{-1}U = A^* = (U^*D^{-*})(D^*L^*)$$

Но LU-разложение единственно, следовательно, $L = U^*D^{-*}$. Значит, $U = DL^*$. Доказали.

■

Теорема о сходимости градиентного спуска для линейной системы с симметричной положительно определенной матрицей.

Теорема (Сходимость наискорейшего спуска)**.** Пусть $A = A^T > 0$ и x_k сгеннерированного с помощью метода наискорейшего спуска, тогда для ошибки $e_k := x_* - x_k$ справедливо:

$$||e_{k+1}||_A \leqslant \frac{cond_2(A) - 1}{cond_2(A) + 1}||e_k||_A$$

Доказательство. Заметим, что оценку лучше действительно нельзя получить, но данным доказательством мы не отвечаем на вопрос почему это так.

1. Уже обсуждалось, что нам без разницы: минимизировать $f(x)$ или честную A-норму ошибки, $||x - x_*||_A^2$. Результат не должен поменяться, поэтому для теоретических выкладок мы будем минимизировать именно честную A-норму ошибки.
$$2f(x) + const = ||x - x_*||_A^2 \Rightarrow$$
$$\Rightarrow \tau_k = \underset{\tau}{\operatorname{argmin}} f(x_k + \tau r_k) = \underset{\tau}{\operatorname{argmin}} ||x_k + \tau r_k - x_*||_A^2$$

2. Теперь запишем A-норму ошибки, сдвиг вдоль градиента на параметр τ хотим выбрать оптимальным образом из условия минимизации $||x_k + \tau r_k - x_*||_A^2$.
$$||e_{k+1}||_A^2 \overset{1.}{=} \min_{\tau} ||x_k + \tau r_k - x_*||_A^2 \overset{\forall t}{\leqslant} ||x_k + t r_k - x_*||_A^2 = \Big[\text{т.к. } x_k - x_* = e_k \text{ и } r_k = -Ae_k\Big] = ||(I - tA)e_k||_A^2 =$$
$$= \Big[\text{теперь вспомним про } A = A^T \text{ и заметим, что } ||y||_A^2 = y^T A y = y^T A^{1/2} A^{1/2} y = (A^{1/2}y)^T A^{1/2} y = ||A^{1/2}y||_2^2\Big] =$$
$$= ||A^{1/2}(I - tA)e_k||_2^2 = ||(I - tA)A^{1/2}e_k||_2^2 \leqslant ||I - tA||_2^2 \cdot ||e_k||_A^2$$

3. В предыдущем пункте мы свели все ко второй норме, для которой уже доказывали оценку из теоремы. Остается только выбрать $t = \dfrac{2}{\lambda_1 + \lambda_n}$, для которого как раз работает необходимая оценка.

■

Оценка сходимости метода сопряженных градиентов для линейной системы с произвольной симметричной положительно определенной матрицей. Случай $\lambda_1 >> \lambda_2$.

Предложение. Для $A = A^\top > 0$ и любого многочлена $h(\lambda) : h(0) = 1$ степени k верно:

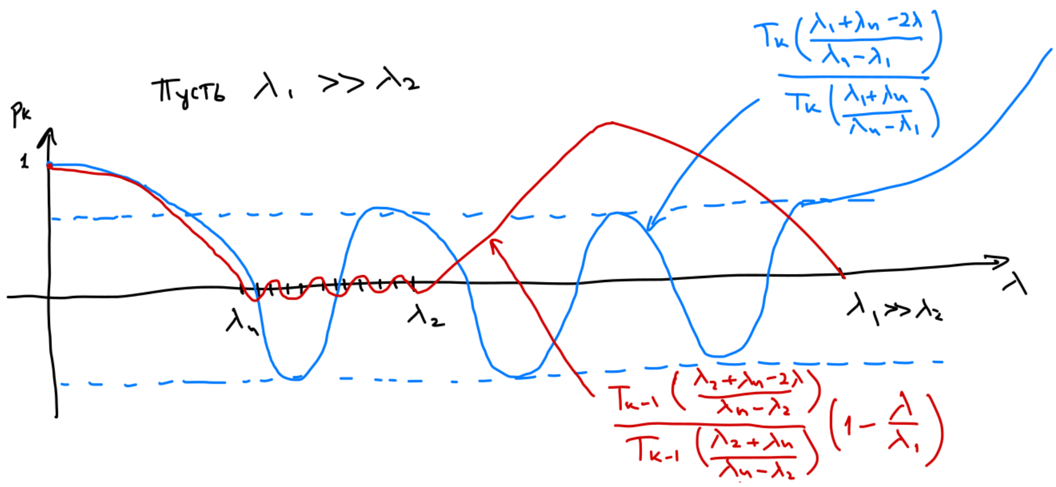
$$\|e_k\|_A \leq \max_i |h(\lambda_i)| \|e_0\|_A$$

В общем случае мы использовали многочлен вида:

$$t_k(\lambda) = \frac{T_k\left(\frac{\lambda_1+\lambda_n-2\lambda}{\lambda_1-\lambda_n}\right)}{T_k\left(\frac{\lambda_1+\lambda_n}{\lambda_1-\lambda_n}\right)}$$

Где T_k - многочлен Чебышева. Такой t_k меньше всего отклоняется от 0 на отрезке $[\lambda_n, \lambda_1]$, однако чем больше отрезок тем больше мы отклоняемся. Вообще говоря нас интересует отклонение только в точках $\lambda \in \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. Давайте рассмотрим другой многочлен:

$$p_k(\lambda) = \frac{T_{k-1}\left(\frac{\lambda_2+\lambda_n-2\lambda}{\lambda_2-\lambda_n}\right)}{T_{k-1}\left(\frac{\lambda_2+\lambda_n}{\lambda_2-\lambda_n}\right)}\left(1 - \frac{\lambda}{\lambda_1}\right)$$



(Заметка: на лекции путаница с знаком знаменателя аргумента T_i)

Тоесть мы уменьшили отрезок, а также пожертвовав степенью многочлена Чебышева, мы добавили множитель который обращается в 0 при $\lambda = \lambda_1$
Нам нужно оценить $\max_i |p_k(\lambda_i)|$, заметим, что:

$$\max_i |p_k(\lambda_i)| \leq \max_{\lambda \in [\lambda_n, \lambda_2] \cup \{\lambda_1\}} |p_k(\lambda)| = \max_{\lambda \in [\lambda_n, \lambda_2]} |p_k(\lambda)|$$

Предложение. С лекции 14 нам известно:

$$\frac{T_k\left(\frac{a+b-2\lambda}{a-b}\right)}{T_k\left(\frac{a+b}{a-b}\right)} \leq 2\left(\frac{\sqrt{\frac{a}{b}}-1}{\sqrt{\frac{a}{b}}+1}\right)^k$$

Теперь оценим p_k на множестве $[\lambda_n, \lambda_2]$:

$$p_k(\lambda) = \frac{T_{k-1}\left(\frac{\lambda_2+\lambda_n-2\lambda}{\lambda_2-\lambda_n}\right)}{T_{k-1}\left(\frac{\lambda_2+\lambda_n}{\lambda_2-\lambda_n}\right)}\left(1 - \frac{\lambda}{\lambda_1}\right) \leq 2\left(1 - \frac{\lambda}{\lambda_1}\right)\left(\frac{\sqrt{\frac{\lambda_2}{\lambda_n}}-1}{\sqrt{\frac{\lambda_2}{\lambda_n}}+1}\right)^{k-1} \leq 2\left(1 - \frac{\lambda_n}{\lambda_1}\right)\left(\frac{\sqrt{\frac{\lambda_2}{\lambda_n}}-1}{\sqrt{\frac{\lambda_2}{\lambda_n}}+1}\right)^{k-1}$$

Осталось заметить, что $\left|1 - \frac{\lambda_n}{\lambda_1}\right| \leq 1$, тогда:

$$p_k(\lambda) \leq 2\left(1 - \frac{\lambda_n}{\lambda_1}\right)\left(\frac{\sqrt{\frac{\lambda_2}{\lambda_n}}-1}{\sqrt{\frac{\lambda_2}{\lambda_n}}+1}\right)^{k-1} \leq 2\left(\frac{\sqrt{\frac{\lambda_2}{\lambda_n}}-1}{\sqrt{\frac{\lambda_2}{\lambda_n}}+1}\right)^{k-1}$$

Заметим, что:

$$0 \leq 2\left(\frac{\sqrt{\frac{\lambda_2}{\lambda_n}}-1}{\sqrt{\frac{\lambda_2}{\lambda_n}}+1}\right)^{k-1} \implies |p_k(\lambda)| \leq 2\left(\frac{\sqrt{\frac{\lambda_2}{\lambda_n}}-1}{\sqrt{\frac{\lambda_2}{\lambda_n}}+1}\right)^{k-1}$$

Используя первое предложение можно сделать вывод:

$$\|e_k\|_A \leq 2\left(\frac{\sqrt{\frac{\lambda_2}{\lambda_n}}-1}{\sqrt{\frac{\lambda_2}{\lambda_n}}+1}\right)^{k-1} \|e_0\|_A$$

Сходимость степенного метода для диагонализуемых матриц.

Дано: A — диагонализуемая и $\lambda^{(1)}$ — простое ($|\lambda^{(1)}| > |\lambda^{(2)}| \geqslant \ldots \geqslant |\lambda^{(n)}|$).

Доказать: отношение Релея $R(Ax_k, x_k) = \dfrac{(Ax_k, x_k)}{(x_k, x_k)} = (Ax_k, x_k)$ сходится к $\lambda^{(1)}$ — старшему собственному значению.

Доказательство. Поскольку A диагонализуемая, то существует n линейно независимых собственных векторов $v^{(i)}$, далее считаем, что $\|v^{(i)}\| = 1$. В таком случае $x_0 = \alpha_1 v^{(1)} + \cdots + \alpha_n v^{(n)}$. **Ключевой момент:** считаем, что $\alpha_1 \neq 0$. Мы используем это в доказательстве, а в противоположном случае метод вообще сойдётся не пойми куда, но точно не туда, куда надо было.

$$\begin{aligned}
x_k &= \frac{Ax_{k-1}}{\|Ax_{k-1}\|_2} = \frac{A^k x_0}{\|A^k x_0\|_2} = \frac{\alpha_1(\lambda^{(1)})^k v^{(1)} + \cdots + \alpha_n(\lambda^{(n)})^k v^{(n)}}{\left\|\alpha_1(\lambda^{(1)})^k v^{(1)} + \cdots + \alpha_n(\lambda^{(n)})^k v^{(n)}\right\|_2} \\
&= \left(\frac{\lambda^{(1)}}{|\lambda^{(1)}|}\right)^k \cdot \left(\frac{\alpha_1}{|\alpha_1|}\right)^k \cdot \frac{v^{(1)} + \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k v^{(2)} + \cdots + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda^{(n)}}{\lambda^{(1)}}\right)^k v^{(n)}}{\left\|v^{(1)} + \frac{\alpha_2}{\alpha_1}\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k v^{(2)} + \cdots + \frac{\alpha_n}{\alpha_1}\left(\frac{\lambda^{(n)}}{\lambda^{(1)}}\right)^k v^{(n)}\right\|_2} \\
&= e^{i\varphi} \cdot \frac{v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right)}{\left\|v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right)\right\|_2} \ominus
\end{aligned}$$

Знаменатель равен $1 + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right)$, разложим $\dfrac{1}{1 + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right)}$ по Тейлору и получим $1 + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right)$, откуда:

$$\begin{aligned}
&\ominus e^{i\varphi} \cdot (v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right))(1 + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right)) \\
&= e^{i\varphi} \cdot (v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right))
\end{aligned}$$

Отсюда имеем:

$$\begin{aligned}
\mathrm{R}(x_k) &= (Ax_k, x_k) = (A \cdot (e^{i\varphi} \cdot (v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right))))^T \cdot (e^{i\varphi} \cdot (v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right))) \\
&= (e^{i\varphi} \cdot (v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right)))^T A(e^{i\varphi} \cdot (v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right))) \\
&= (e^{i\varphi} \cdot (v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right)))^T (e^{i\varphi} \cdot (\lambda^{(1)} v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right))) \\
&= \lambda^{(1)} \cdot \left\|e^{i\varphi} \cdot (v^{(1)} + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right))\right\|_2 \\
&= \lambda^{(1)} \cdot (1 + \mathrm{O}\left(\left(\frac{\lambda^{(2)}}{\lambda^{(1)}}\right)^k\right))
\end{aligned}$$

■

Вывод двух основных свойств QR алгоритма.

Теорема Леви-Деспланка и первая теорема Гершгорина.

Теорема Леви-Деспланка.

Определение. Матрица A обладает строгим строчным диагональным преобладанием, если $\forall i\ |a_{ii}| > \sum\limits_{j=1, i\neq j}^{n} |a_{ij}|$.

Определение. Матрица A обладает строгим столбцовым диагональным преобладанием, если $\forall j\ |a_{jj}| > \sum\limits_{i=1, i\neq j}^{n} |a_{ij}|$.

Теорема (Леви-Деспланка)**.** Матрица, обладающая строгим строчным (столбцовым) диагональным преобладанием является невырожденной.

Доказательство.
Докажем для строгого строчного, для столбцового аналогично.

Представим A в следующем виде: $A = diag(A)(I + diag(A)^{-1})(A - diag(A))$, где матрица $diag(A)$ – это диагональная матрица, у которой на диагонали стоят диагональные элементы матрицы A. Раскрыв скобки, можно проверить, что равенство действительно выполняется.

Вспомним немного из курса линала:

- Матрица A – обратима $\Leftrightarrow$ матрица A – невырожденна
- Если $A = BC$, то $det(A) = det(B) \cdot det(C)$

Таким образом, нам необходимо и достаточно доказать обратимость матриц $diag(A)$ и $(I + diag(A)^{-1})(A - diag(A))$.

Обратимость первой почти очевидна: если бы на диагонали могли стоять нулевые элементы, то матрица A не обладала бы строгим строчным диагональным пребладанием.

Теперь заметим, что $(I + diag(A)^{-1})(A - diag(A))$ обратима $\Leftrightarrow \exists \sum\limits_{k=0}^{\infty}(-diag(A)^{-1})(A - diag(A))^k$ (вспоминаем про ряды Неймана).

Осталось доказать, что такой ряд Неймана сходится. Не будем использовать критерий, а используем признак: докажем, что $||diag(A)^{-1})(A - diag(A)|| < 1$ для некоторой нормы.

Рассмотрим бесконечную норму: $||diag(A)^{-1})(A - diag(A)||_\infty < 1$. Пусть это неравенство не выполняется, тогда: $max_i \frac{\sum_{j=1, j\neq i}^{n} |a_{ij}|}{|a_{ii}|} \geqslant 1$ (это я просто руками записала бесконечную норму для матрицы). Но тогда выходит, что A не обладает строгим строчным диагональным преобладанием – противоречие $\Rightarrow ||diag(A)^{-1})(A - diag(A)||_\infty < 1 \Rightarrow \sum\limits_{k=0}^{\infty}(-diag(A)^{-1})(A - diag(A))^k$ – сходится $\Rightarrow (I + diag(A)^{-1})(A - diag(A))$ обратима.
Таким образом, A представляет собой произведение обратимых матриц $\Rightarrow A$ и сама обратима, то есть, невырожденна.
■

Теорема Гершгорина.

Теорема (1-я теорема Гершгорина)**.** Пусть $A \in \mathbb{C}$, тогда собственные значения матрицы A находятся внутри

$$D = D_1 \cup D_2 \cup \ldots \cup D_n$$

где

$$D_k = \{z \in \mathbb{C} : |a_{kk} - z| \leqslant \sum_{i\neq k} |a_{ki}|\}$$

Доказательство.
Пусть $\lambda \notin D \Rightarrow A - I\lambda$ обладает строгим строчным диагональным преобладанием (просто посмотрите на то как мы определяем D_k, на то, что условие в D_k для данной λ не выполняются, и на строение матрицы A) $\Rightarrow$ (по предыдущей теореме) $A - \lambda I$ – невырожденна, но тогда λ не может являться собственным значением A (вспоминаем курс линала: характеристический многочлен и его корни). Получили противоречние. ■

Определение. Матрица D_k – круги Гершгорина.

Теорема (2-я теорема Гершгорина без доказательства)**.** Если есть m кругов Гершгорина, образующих область G, то в G находится ровно m собственных значений.