МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Государственное образовательное учреждение высшего профессионального образования
«Новосибирский государственный университет» (НГУ)

Механико-математический факультет

КУРС ЛЕКЦИЙ

«Вычислительные методы линейной алгебры»

Основная образовательная программа подготовки бакалавров
ММФ НИУ-НГУ (Новосибирск, Россия) – Хэйлунцзянский университет (Харбин, Китай)

НАПРАВЛЕНИЕ ПОДГОТОВКИ
010100 «Математика»
ПРОФИЛЬ
«Математика и прикладная математика»

Квалификация (степень) выпускника
Бакалавр

Форма обучения очная

Новосибирск
2013

Курс лекций «Вычислительные методы линейной алгебры» разработан в соответствии с требованиями к обязательному минимуму содержания и уровню подготовки дипломированного бакалавра по профессиональному циклу дисциплин (базовая часть) по направлению подготовки «Математика», а также задачами, стоящими перед Новосибирским государственным университетом по реализации Программы развития НГУ как национального исследовательского института и сотрудничества с Хэйлунцзянским университетом (Харбин, Китайская народная республика).

Автор:
Мацокин Александр Михайлович, д.ф.-м.н., доцент, профессор кафедры Вычислительной математики.

Механико-математический факультет
Кафедра Вычислительной математики

Вычислительные методы линейной алгебры

Содержание

Лекция 1. 6
Традиционные задачи линейной алгебры 6
Векторные и матричные нормы 7
Число обусловленности 10
Лекция 2. Прямые методы решения линейных уравнений 11
Метод исключения Гаусса – схема единственного деления 11
Теорема об LU разложении 12
Разложение Холесского 13
Метод квадратного корня 13
Лекция 3. 15
Метод исключения с выбором главного элемента по столбцу 15
Матрица перестановок 15
Элементарная матрица перестановок 15
Выбор главного элемента по столбцу. 15
Метод вращений решения системы уравнений 17
Элементарная матрица вращения 17
k –ый шаг метода вращений 17
Сведение системы уравнений к системе с двухдиагональной матрицей с помощью матриц вращения 18
1-ый шаг 18
k -ый шаг 18
Система с двухдиагональной матрицей 19
Лекция 4. 20
Метод отражений решения системы уравнений 20
Матрица отражения 20
k –ый шаг метода отражений 20
Решение системы с вырожденной матрицей 21
HR –разложение с перестановками столбцов матрицы A 21
Совместность системы с вырожденной матрицей 22
Применение HR –разложения с перестановками столбцов для решения совместной системы 22
Метод прогонки решения систем с трехдиагональной матрицей 23
Лемма Гершгорина (о локализации собственных значений). 26
Лекция 5. Итерационные методы решения линейных уравнений 28
Пример и основные определения 28
Пример: 28
Одношаговый (двухслойный) итерационный метод решения $Ax = b$: 28
Стационарный одношаговый итерационный метод решения $Ax = b$: 28
Условия сходимости стационарного итерационного метода 29
Достаточные условия: 29
Необходимое и достаточное условие сходимости стационарного итерационного метода 29
Стационарный итерационный метод 29
Асимптотическая скорость сходимости 30
Лекция 6. 33
Метод Якоби 33
Сходимость в случае диагонального преобладания по строкам 33

Сходимость в случае диагонального преобладания по столбцам 33
Необходимое и достаточное условие сходимости метода Якоби в случае симметричной матрицы с положительной главной диагональю 34
Метод Зейделя (Гаусса–Зейделя, Некрасова) 35
Лекция 7. 38
Функционал ошибки 38
Метод полной релаксации 39
Метод неполной релаксации 41
Оценка сходимости методов релаксации 43
Предварительные замечания 43
Пример 45
Лекция 8. 48
Градиент, метод наискорейшего спуска 48
Градиент функции $f(z): R^n \to R$ 48
Градиент функции $f(z) = \| z \|_C^2 \ : R^n \to R$ 48
Выбор матрицы C 49
Метод наискорейшего спуска 49
Метод минимальных невязок 49
Метод простой итерации 50
Предварительные замечания 50
Метод простой итерации 51
Оптимальный выбор параметра τ метода Ричардсона 51
Оценки сходимости МНС и ММН 52
Лекция 9. Метод Ричардсона с чебышевскими параметрами 54
Предварительные замечания 54
Построение полинома $S_{m,\tau^*}(\lambda),\ S_{m,\tau^*}(0) = 1$, наименее уклоняющегося от нуля на интервале $[\alpha, \beta]$ 57
Полином Чебышева 57
Полином $S_{m,\tau^*}(\lambda)$ 58
Норма полинома $S_{m,\tau^*}(\lambda)$ 59
Формулы m-циклического метода Ричардсона с чебышевскими параметрами 60
Численная неустойчивость двучленных формул метода Ричардсона 60
Трехчленные формулы реализации метода Ричардсона 61
Лекция 10. 63
Метод сопряженных градиентов 63
Предварительные замечания 63
Минимизация функционала 64
Построение A–ортогонального базиса 65
Метод сопряженных градиентов 66
Трехслойные формулы метода сопряженных градиентов 67
Лекция 10 (продолжение). 68
Переобусловливатель 68
Метод простой итерации с переобусловливателем 68
Метод наискорейшего спуска с переобусловливателем 70
Метод сопряженных градиентов с переобусловливателем 71
Лекция 11. Проблема собственных значений 73
Корректность задачи на собственные значения 73

Степенной метод вычисления максимального собственного значения матрицы $A = A^* \geq 0$ 74
Степенной метод вычисления минимального собственного значения матрицы $A = A^* \geq 0$ 75
Применение ортогонализации и степенного метода для вычисления очередного собственного значения....75
Степенной метод вычисления границ спектра матрицы $B^{-1}A$ в случае $A = A^* > 0$ и $B = B^* > 0$ 76
Лекция 12. Метод деления пополам (бисекций) 78
Идея метода бисекций вычисления $\lambda_j \in Sp(A)$ 78
Приведение самосопряженной матрицы к трехдиагональному виду ортогональным преобразованием подобия с помощью матриц вращения....79
Якобиевы матрицы....80
О вычислении ЧПЗ....82
О вычислении собственного вектора 82
Лекция 13. Метод вращений (Якоби)....83
Выбор вращения....84
Сходимость собственных значений 85
Сходимость собственных векторов....85
Элементарная матрица унитарного вращения 87
Литература 89

Лекция 1.

Традиционные задачи линейной алгебры

$$A = (a_{ij})_{i,j=1}^{n} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} - \text{матрица}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} - \text{вектор}$$

Задачи:	**Методы** (теория определителей)**:**
вычисление определителя $\Delta = \det A$	по определению $\Delta = \sum\limits_{(i_1,\dots,i_n)} (-1)^r \cdot a_{1i_1} \cdot \ldots \cdot a_{ni_n}$
решение системы уравнений $Ax = b$	метод Крамера: $x_i = \dfrac{\Delta_i}{\Delta}, \quad \Delta_i = \det\{A_i, b\}$
вычисление обратной матрицы: $AX = XA = E$	определение столбца $x^{(j)}$ матрицы X: $Ax^{(j)} = e_j, \quad x_{ij} = \dfrac{\Delta_{ji}}{\Delta}$
спектральная задача: $Ax = \lambda \cdot x$	собственные значения – корни полинома $P_n(\lambda) = \det(A - \lambda \cdot E)$ собственные векторы – решения систем $(A - \lambda \cdot E)x = 0$ r линейно независимых решений, где $r = \dim\{Ker(A - \lambda \cdot E)\}$

Непригодность этих методов:

<table>
<tr>
<td>
количество умножений при вычислении одного определителя:
$(n-1) \cdot n!$
если производительность ЭВМ 10^9 оп/сек, то
<table>
<tr><td>n</td><td>время вычисления</td></tr>
<tr><td>10</td><td>$\approx 10^{-4}$ сек.</td></tr>
<tr><td>20</td><td>> 17 мин.</td></tr>
<tr><td>30</td><td>> 400 тыс. лет</td></tr>
</table>
</td>
<td>
ошибки округления:
$\tilde{a} = a + \varepsilon \cdot |a|, \quad \varepsilon \le 10^{-6}$
если $n = 6$, $|a_{ij}| \ge 10$, $\Delta = 1$,
$a_i = a_{1i_1} \cdot \ldots \cdot a_{ni_n}$, то $a_i - \tilde{a}_i = O(1)$
$\Delta - \tilde{\Delta} = n! \cdot O(1) = O(1)$
т.е. определитель вычисляется с большой ошибкой и, следовательно, решения поставленных задач вычисляются с такой же ошибкой.
</td>
</tr>
</table>

Векторные и матричные нормы

Векторные

$\forall\ x \in R^n\ (C^n)$

$\|x\| > 0, \quad x \neq 0$
$\|x\| = 0, \quad x = 0$

$\|\alpha \cdot x\| = |\alpha| \cdot \|x\|$

$\|x + y\| \leq \|x\| + \|y\|$

Матричные

$\forall\ A \in R^{n\times n}\ (C^{n\times n})$

аксиомы 1. – 3.
– аддитивная

4. $\|AB\| \leq \|A\| \cdot \|B\|$
–мультипликативная

согласованная с векторной, если
$\|Ax\| \leq \|A\| \cdot \|x\|$

подчиненная векторной, если
$$\|A\| = \sup \frac{\|Ax\|}{\|x\|}$$

Примеры:
$\|x\|_\infty = \max |x_i|$
– кубическая или равномерная
$\|x\|_1 = |x_1| + \ldots + |x_n|$
– октаэдрическая
$\|x\|_2 = \sqrt{|x_1|^2 + \ldots + |x_n|^2}$
–сферическая или евклидова

Примеры подчиненных матричных норм:
$$\|A\|_\infty = \sup \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_i \sum_{j=1}^n |a_{ij}|$$
$$\|A\|_1 = \sup \frac{\|Ax\|_1}{\|x\|_1} = \max_j \sum_{i=1}^n |a_{ij}|$$
$$\|A\|_2 = \sup \frac{\|Ax\|_2}{\|x\|_2} = \sqrt{\rho(A^*A)}$$

Вывод формул для некоторых матричных норм, подчиненных векторным нормам.

1) Векторной норме $\|x\|_\infty = \max\limits_{1\leq i\leq n} |x_i|$ подчинена матричная норма $\|A\|_\infty = \max\limits_{1\leq i\leq n} \sum_{j=1}^n |a_{ij}|$.

Действительно:

во-первых; $\|A\|_\infty \leq \max\limits_{1\leq i\leq n} \sum\limits_{j=1}^n |a_{ij}|$:

$$\|A\|_\infty \equiv \sup_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \sup_{x\neq 0} \frac{\max\limits_{1\leq i\leq n} |\sum_{j=1}^n a_{ij} \cdot x_j|}{\max\limits_{1\leq i\leq n} |x_i|} \leq$$
$$\leq \sup_{x\neq 0} \frac{\max\limits_{1\leq i\leq n} (\sum_{j=1}^n |a_{ij}| \cdot \max\limits_{1\leq j\leq n} |x_j|)}{\max\limits_{1\leq i\leq n} |x_i|} = \max_{1\leq i\leq n} \sum_{j=1}^n |a_{ij}|;$$

во-вторых; $\|A\|_\infty \geq \max\limits_{1\leq i\leq n} \sum\limits_{j=1}^n |a_{ij}| \equiv \sum\limits_{j=1}^n |a_{i_0 j}|$, так как для вектора $x^{(i_0)} = \left(x_j^{(i_0)} = \dfrac{\bar{a}_{i_0 j}}{|a_{i_0 j}|}\right)_{j=1}^n$

имеем $\|Ax^{(i_0)}\|_\infty \geq |(Ax^{(i_0)})_{i_0}| = |\sum\limits_{j=1}^n a_{i_0 j} \dfrac{\bar{a}_{i_0 j}}{|a_{i_0 j}|}| = \sum\limits_{j=1}^n |a_{i_0 j}|$ и, поскольку $\|x^{(i_0)}\|_\infty = 1$,

$$\| A \|_{\infty} \equiv \sup_{x \neq 0} \frac{\| Ax \|_{\infty}}{\| x \|_{\infty}} \geq \frac{\| Ax^{(i_0)} \|_{\infty}}{\| x^{(i_0)} \|_{\infty}} \geq \sum_{j=1}^{n} | a_{i_0 j} | \equiv \| A \|_{\infty},$$

что завершает доказательство формулы.

2) Векторной норме $\| x \|_1 = \sum_{i=1}^{n} | x_i |$ подчинена матричная норма $\| A \|_1 = \max\limits_{1 \leq j \leq n} \sum_{i=1}^{n} | a_{ij} |$.

Действительно:

во-первых; $\| A \|_1 \leq \max\limits_{1 \leq j \leq n} \sum\limits_{i=1}^{n} | a_{ij} |$:

$$\| A \|_1 \equiv \sup_{x \neq 0} \frac{\| Ax \|_1}{\| x \|_1} = \sup_{x \neq 0} \frac{\sum_{i=1}^{n} | \sum_{j=1}^{n} a_{ij} \cdot x_j |}{\sum_{i=1}^{n} | x_i |} \leq \sup_{x \neq 0} \frac{\sum_{i=1}^{n} (\sum_{j=1}^{n} | a_{ij} | \cdot | x_j |)}{\sum_{i=1}^{n} | x_i |} =$$

$$= \sup_{x \neq 0} \frac{\sum_{j=1}^{n} (\sum_{i=1}^{n} | a_{ij} |) \cdot | x_j |}{\sum_{i=1}^{n} | x_i |} \leq \max_{1 \leq j \leq n} \sum_{i=1}^{n} | a_{ij} |;$$

во-вторых; $\| A \|_1 \geq \max\limits_{1 \leq j \leq n} \sum\limits_{i=1}^{n} | a_{ij} | \equiv \sum\limits_{i=1}^{n} | a_{i j_0} |$, так как для вектора $x^{(j_0)} = (x_j^{(j_0)} = \delta_{j_0, j})_{j=1}^{n}$

имеем $\| Ax^{(j_0)} \|_1 = \sum\limits_{i=1}^{n} | a_{i j_0} |$ и, поскольку $\| x^{(j_0)} \|_1 = 1$,

$$\| A \|_1 \equiv \sup_{x \neq 0} \frac{\| Ax \|_1}{\| x \|_1} \geq \frac{\| Ax^{(j_0)} \|_1}{\| x^{(j_0)} \|_1} = \sum_{i=1}^{n} | a_{i j_0} | \equiv \| A \|_1,$$

что завершает доказательство формулы.

3) Векторной норме $\| x \|_2 = \sqrt{\sum_{i=1}^{n} | x_i |^2}$ подчинена матричная норма $\| A \|_2 = \sqrt{\rho(A^* A)}$.

Прежде всего, напомним определения скалярного произведения и спектрального радиуса.

а) Отображение $(x, y) : C^n(R^n) \times C^n(R^n) \to R_+$ называется скалярным произведением в векторном пространстве C^n (или R^n), если

1. $(x, x) \geq 0 \quad \forall\, x \in C^n(R^n), \;\; (x, x) = 0 \Leftrightarrow x = 0;$
2. $(x, y) = \overline{(y, x)} \quad \forall\, x, y \in C^n(R^n);$
3. $(\alpha \cdot x, y) = \alpha \cdot (x, y) \quad \forall\, \alpha \in C(R),\; x, y \in C^n(R^n),$
4. $(x + y, z) = (x, z) + (y, z) \quad x, y, z \in C^n(R^n)$

Скалярное произведение определяет векторную норму $\| x \| \equiv \sqrt{(x, x)}$ и имеет место неравенство Коши: $| (x, y) | \leq \| x \| \cdot \| y \|$.

Классическое (евклидово) скалярное произведение $(x, y) = \sum_{i=1}^{n} x_i \cdot \bar{y}_i$ определяет евклидову норму вектора $\| x \|_2 = \sqrt{(x, x)} \equiv \sqrt{\sum_{i=1}^{n} | x_i |^2}$.

б) Обозначим через $Sp(A)$ множество собственных значений матрицы A. Тогда число

$$\rho(A) = \max_{\lambda \in Sp(A)} | \lambda |$$

называется спектральным радиусом матрицы A.

Матрица A^*, сопряженная матрице A относительно скалярного произведения (x,y), определяется тождеством Лагранжа:

$$(Ax,y)=(x,A^*y) \quad \forall\ x,y\in C^n(R^n).$$

Напомним, что матрица $A^*=(\,(A^*)_{i\,j}\equiv\bar{a}_{ji})_{i,j=1}^n$ сопряжена матрице A относительно евклидова скалярного произведения.

Далее, матрица A^*A самосопряжена (эрмитова), все её собственные значения λ_i вещественны и неотрицательны, а из её собственных векторов $x^{(i)}\neq 0:\ A^*Ax^{(i)}=\lambda_i x^{(i)}$ можно образовать ортонормированный базис в $C^n(R^n)$: $(x^{(i)},x^{(j)})=\delta_{i,j},\ i,j=1,\dots,n$.

Теперь докажем равенство $\|A\|_2=\sqrt{\rho(A^*A)}$.

во-первых; $\|A\|_2\le\sqrt{\rho(A^*A)}$, так как

$$\|A\|_2^2\equiv\sup_{x\neq0}\frac{\|Ax\|_2^2}{\|x\|_2^2}=\sup_{x\neq0}\frac{(A^*Ax,x)}{(x,x)}=$$

$$=\sup_{\substack{x\neq0,\\ x=\alpha_1x^{(1)}+\ldots+\alpha_nx^{(n)}}}\frac{\lambda_1\cdot|\alpha_1|^2+\ldots+\lambda_n\cdot|\alpha_n|^2}{|\alpha_1|^2+\ldots+|\alpha_n|^2}\le\max_{1\le i\le n}\lambda_i=\rho(A^*A);$$

во-вторых; $\|A\|_2\ge\sqrt{\rho(A^*A)}$, так как (пусть $\lambda_n=\rho(A^*A)$) для вектора $x^{(n)}$ имеем

$$\|A\|_2^2=\sup_{x\neq0}\frac{(A^*Ax,x)}{(x,x)}\ge\frac{(A^*Ax^{(n)},x^{(n)})}{(x^{(n)},x^{(n)})}\equiv\lambda_n=\rho(A^*A),$$

что завершает доказательство формулы.

Теорема.	Любые две нормы $\|\|x\|\|_*$ и $\|\|x\|\|_{**}$ в конечномерном пространстве эквивалентны: $\exists\,\alpha,\beta:\ \forall\,x\quad \alpha\cdot\|\|x\|\|_*\le\|\|x\|\|_{**}\le\beta\cdot\|\|x\|\|_*$	Примеры: $\forall\,x\in R^n\ (C^n)$ $\|\|x\|\|_\infty\le\|\|x\|\|_1\le n\cdot\|\|x\|\|_\infty$ $\|\|x\|\|_\infty\le\|\|x\|\|_2\le\sqrt{n}\cdot\|\|x\|\|_\infty$ $\|\|x\|\|_2\le\|\|x\|\|_1\le\sqrt{n}\cdot\|\|x\|\|_2$

!!! Константы эквивалентности зависят от размерности пространства **!!!**

При решении системы линейных уравнений $Ax=b$ могут быть неточно заданы либо правая часть $\tilde{b}=b+\delta b$ либо матрица $\tilde{A}=A+\delta A$, где компоненты вектора δb и элементы матрицы δA малы по сравнению с соответствующими элементами исходных вектора и матрицы. Тогда вместо решения x мы получим его приближение $\tilde{x}=x+\delta x$, причем компоненты вектора–ошибки δx могут быть большими.

Оценим норму ошибки через нормы возмущений правой части и матрицы системы, считая, что матричная норма подчинена векторной норме.

Число обусловленности

Определение. $\operatorname{cond} A = \| A \| \cdot \| A^{-1} \|$

1. $\operatorname{cond} A \geq 1$ т.к.

$\| x \| = \| AA^{-1}x \| \leq \| A \| \cdot \| A^{-1} \| \cdot \| x \|$

2. $\operatorname{cond}(AB) \leq \operatorname{cond} A \cdot \operatorname{cond} B$ т.к.

$\| AB \| \cdot \| (AB)^{-1} \| \leq \| A \| \cdot \| B \| \cdot \| B^{-1} \| \cdot \| A^{-1} \|$

Теорема.
$$\begin{matrix} Ax = b, \ \det A \neq 0 \\ A(x + \delta x) = b + \delta b \end{matrix} \quad \Rightarrow \quad \frac{\| \delta x \|}{\| x \|} \leq \operatorname{cond} A \cdot \frac{\| \delta b \|}{\| b \|}$$

Док–во.
$$\begin{matrix} \| \delta x \| = \| A^{-1}\delta b \| \leq \| A^{-1} \| \cdot \| \delta b \| \\ \| A \| \cdot \| x \| \geq \| Ax \| = \| b \| \end{matrix} \quad \Rightarrow \quad \frac{\| \delta x \|}{\| x \|} \leq \operatorname{cond} A \cdot \frac{\| \delta b \|}{\| b \|}.$$

Теорема. $Ax = b, \ \det A \neq 0, \quad (A + \delta A)(x + \delta x) = b + \delta b, \quad \| A^{-1} \| \cdot \| \delta A \| < 1$
$$\Rightarrow \quad \frac{\| \delta x \|}{\| x \|} \leq \frac{\operatorname{cond} A}{1 - \operatorname{cond} A \cdot \frac{\| \delta A \|}{\| A \|}} \cdot \left(\frac{\| \delta b \|}{\| b \|} + \frac{\| \delta A \|}{\| A \|}\right)$$

Док–во.

1. $\exists \ (A + \delta A)^{-1}$ и $\| (A + \delta A)^{-1} \| \leq \dfrac{\| A^{-1} \|}{1 - \| A^{-1} \| \cdot \| \delta A \|}$, т.к. $(A + \delta A) = A(E + A^{-1} \cdot \delta A)$ и

$\| (E + A^{-1}\delta A)z \| \geq \| z \| - \| A^{-1} \cdot \delta A \cdot z \| \geq (1 - \| A^{-1} \| \cdot \| \delta A \|) \cdot \| z \| > 0$, то

$$\exists \ (E + A^{-1} \cdot \delta A)^{-1} \text{ и } \| (E + A^{-1} \cdot \delta A)^{-1} \| = \sup \frac{\| (E + A^{-1} \cdot \delta A)^{-1} x \|}{\| x \|} =$$
$$= \sup \frac{\| z \|}{\| (E + A^{-1} \cdot \delta A)z \|} \leq \frac{1}{1 - \| A^{-1} \| \cdot \| \delta A \|}$$

Замечание. $E + X + X^2 + X^3 + \ldots \to (E - X)^{-1}$ и $\| (E - X)^{-1} \| \leq (1 - \| X \|)^{-1}$, если $\| X \| < 1$. При $X = -A^{-1}\delta A$ получаем еще одно доказательство $\exists \ (E + A^{-1} \cdot \delta A)^{-1}$.)

2. Т.к. $\delta x = (A + \delta A)^{-1}[b + \delta b - (A + \delta A)x] = (A + \delta A)^{-1}[\delta b - \delta A \cdot x]$, то

$$\| \delta x \| \leq \| (A + \delta A)^{-1} \| \cdot (\| \delta b \| + \| \delta A \| \cdot \| x \|) \Rightarrow \frac{\| \delta x \|}{\| x \|} \leq \| (A + \delta A)^{-1} \| \cdot \left(\frac{\| \delta b \|}{\| x \|} + \| \delta A \|\right)$$

т.к. $\| x \| \geq \| b \| / \| A \|$, то

$$\frac{\| \delta x \|}{\| x \|} \leq \| (A + \delta A)^{-1} \| \cdot \left(\| A \| \cdot \frac{\| \delta b \|}{\| b \|} + \| \delta A \| \right) \leq$$
$$\leq \frac{\| A^{-1} \|}{1 - \| A^{-1} \| \cdot \| \delta A \|} \cdot \left(\| A \| \cdot \frac{\| \delta b \|}{\| b \|} + \| \delta A \| \right) = \frac{\| A^{-1} \| \cdot \| A \|}{1 - \| A^{-1} \| \cdot \| A \| \frac{\| \delta A \|}{\| A \|}} \cdot \left(\frac{\| \delta b \|}{\| b \|} + \frac{\| \delta A \|}{\| A \|} \right).$$

Лекция 2. Прямые методы решения линейных уравнений

Метод исключения Гаусса – схема единственного деления

$$Ax = b, \quad \det A_k \equiv \det\begin{bmatrix} a_{11} & \dots & a_{1k} \\ \dots & \dots & \dots \\ a_{k1} & \dots & a_{kk} \end{bmatrix} \neq 0, \;\; k = 1,\dots,n \quad \Rightarrow \quad \begin{matrix} A = LU \\ Ly = b \\ Ux = y \end{matrix}$$

Схема единственного деления на примере системы третьего порядка:

Прямой ход:	Матричная формулировка:
$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1$ $a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2$ $a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3$	$Ax = b,$ $\det A_1 = a_{11} \neq 0, \quad L_1 = \begin{bmatrix} 1/a_{11} & 0 & 0 \\ -a_{21}/a_{11} & 1 & 0 \\ -a_{31}/a_{11} & 0 & 1 \end{bmatrix}$
$x_1 + u_{12}x_2 + u_{13}x_3 = y_1$ $a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 = b_2^{(1)}$ $a_{32}^{(1)}x_2 + a_{33}^{(1)}x_3 = b_3^{(1)}$	$A^{(1)}x = b^{(1)}, \quad A^{(1)} = L_1A, \quad b^{(1)} = L_1b$ $\det A_2^{(1)} \neq 0, \quad L_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/a_{22}^{(1)} & 0 \\ 0 & -a_{32}^{(1)}/a_{22}^{(1)} & 1 \end{bmatrix}$
$x_1 + u_{12}x_2 + u_{13}x_3 = y_1$ $x_2 + u_{23}x_3 = y_2$ $a_{33}^{(2)}x_3 = b_3^{(2)}$	$A^{(2)}x = b^{(2)}, \quad A^{(2)} = L_2A^{(1)}, \quad b^{(2)} = L_2b^{(1)}$ $\det A_3^{(2)} \neq 0, \quad L_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/a_{33}^{(2)} \end{bmatrix}$
$x_1 + u_{12}x_2 + u_{13}x_3 = y_1$ $x_2 + u_{23}x_3 = y_2$ $x_3 = y_3$	$Ux \equiv A^{(3)}x = b^{(3)} \equiv y, \quad A^{(3)} = L_3A^{(2)}, \quad b^{(3)} = L_3b^{(2)}$ $U = \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix}, \quad A = (L_1^{-1} \cdot L_2^{-1} \cdot L_3^{-1}) \cdot U = L \cdot U$

Обратный ход:	Матричная формулировка:
$x_3 = y_3, \quad x_2 = y_2 - u_{23}x_3, \quad x_3 = y_1 - u_{12}x_2 - u_{13}x_3$	$x = U^{-1}y$

Формулы схемы единственного деления (доказать):

k–ый шаг прямого хода:	
$A^{(k-1)}x = b^{(k-1)} \;\Rightarrow\; A^{(k)}x = b^{(k)},$ $A^{(k)} = L_kA^{(k-1)}, \quad b^{(k)} = L_k b^{(k-1)}$ $A^{(n)} = U$ – верхняя треуг. матрица	$L_k = \begin{bmatrix} 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1/a_{kk}^{(k-1)} & 0 & \dots & 0 \\ 0 & \dots & 0 & -a_{k+1k}^{(k-1)}/a_{kk}^{(k-1)} & 1 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & -a_{nk}^{(k-1)}/a_{kk}^{(k-1)} & 0 & \dots & 1 \end{bmatrix}$

Теорема об LU разложении

Если $\forall\, k \;\; \det A_k \neq 0$, то $\exists\, L \;\; \exists\, U: \;\; A = LU$, где L – нижняя, U – верхняя треугольные матрицы.

Доказательство.

Если $A = LU$, то $A_k = L_k U_k$, $\det A_k = \det L_k \cdot \det U_k = l_{11} \cdot \ldots \cdot l_{kk} \cdot u_{11} \cdot \ldots \cdot u_{kk} \neq 0$,

т.к. $\begin{bmatrix} A_k & B_{k,n-k} \\ C_{n-k,k} & A'_{n-k} \end{bmatrix} = \begin{bmatrix} L_k & O_{k,n-k} \\ L_{n-k,k} & L'_{n-k} \end{bmatrix} \cdot \begin{bmatrix} U_k & U_{k,n-k} \\ O_{n-k,k} & U'_{n-k} \end{bmatrix}$.

Предположим, что разложение $A_k = L_k U_k$ найдено ($A_1 \equiv a_{11} = L_1 U_1 \equiv l_{11} \cdot u_{11} \neq 0$).

Вычислим $A_{k+1} = L_{k+1} U_{k+1}$

(т.е. последние строку матрицы L_{k+1} и столбец матрицы U_{k+1}):

т.к. $$\left[\begin{array}{ccc|c} & & & a_{1,k+1} \\ & A_k & & \vdots \\ & & & a_{k,k+1} \\ \hline a_{k+1,1} & \ldots & a_{k+1,k} & a_{k+1,k+1} \end{array}\right] = \left[\begin{array}{ccc|c} & & & 0 \\ & L_k & & \vdots \\ & & & 0 \\ \hline l_{k+1,1} & \ldots & l_{k+1,k} & l_{k+1,k+1} \end{array}\right] \cdot \left[\begin{array}{ccc|c} & & & u_{1,k+1} \\ & U_k & & \vdots \\ & & & u_{k,k+1} \\ \hline 0 & \ldots & 0 & u_{k+1,k+1} \end{array}\right]$$

то $L_k \begin{bmatrix} u_{1,k+1} \\ \vdots \\ u_{k,k+1} \end{bmatrix} = \begin{bmatrix} a_{1,k+1} \\ \vdots \\ a_{k,k+1} \end{bmatrix}$, $\quad \begin{bmatrix} l_{k+1,1} & \ldots & l_{k+1,k} \end{bmatrix} U_k = \begin{bmatrix} a_{k+1,1} & \ldots & a_{k+1,k} \end{bmatrix}$ – системы с треугольными неособенными матрицами (решения $\exists !$), и

$$l_{k+1,k+1} \cdot u_{k+1,k+1} = a_{k+1,k+1} - \begin{bmatrix} l_{k+1,1} & \ldots & l_{k+1,k} \end{bmatrix} \cdot \begin{bmatrix} u_{1,k+1} \\ \vdots \\ u_{k,k+1} \end{bmatrix},$$

очевидно, что решение этого уравнения существует, но не единственно.

(так как $0 \neq \det A_{k+1} = \det L_{k+1} \cdot \det U_{k+1}$, то $\det L_{k+1} \neq 0, \;\; \det U_{k+1} \neq 0$.)

И, наконец, $A \equiv A_n = L_n U_n \equiv LU$.

Объем вычислений.

Так как для решения системы уравнений с треугольной матрицей порядка k достаточно выполнить $k(k+1)/2$ умножений и делений, то полагая на каждом шаге $u_{k+1,k+1} = 1$, получим, что число таких операций для вычисления последних строки и столбца матриц L_{k+1} и U_{k+1} равно $k(k+2)$, а для вычисления матриц L и U достаточно $\sum_1^{n-1} k(k+2) \approx n^3/3$ умножений или делений.

Замечание.

Если построено LU–разложение матрицы A, то ее определитель вычисляется за $2(n-1)$ умножений (перемножаются диагональные (ведущие) элементы).

Теорема (об LDU – разложении).	**Док–во.**
Если $\det A_k \neq 0 \ \ \forall\, k$, то разложение $A = LDU$, где $l_{kk} = u_{kk} = 1 \ \ \forall\, k$, единственно.	Пусть $A = L^{(1)}D^{(1)}U^{(1)} = L^{(2)}D^{(2)}U^{(2)}$, тогда $[L^{(2)}]^{-1}L^{(1)} = D^{(2)}U^{(2)}[D^{(1)}U^{(1)}]^{-1} = \mathrm{diag} = E$, (т.к. $[L^{(2)}]^{-1}L^{(1)}$ – нижняя треуг. м–ца с единицами на диагонали) $\Rightarrow L^{(1)} = L^{(2)}$ $\Rightarrow [D^{(2)}]^{-1}D^{(1)} = U^{(2)}[U^{(1)}]^{-1} = \mathrm{diag} = E$ $\Rightarrow U^{(1)} = U^{(2)} \ \ \&\ \ D^{(1)} = D^{(2)}$.

Разложение Холесского

Теорема.	Если $A = A^* > 0$ (т.е. $(Ax,x) > 0 \quad \forall\, x \neq 0$), то $A = L \cdot D \cdot L^*$, $l_{kk} = 1$, $d_k > 0 \ \ \forall\, k$.
Док–во.	Т.к. $0 < (Ax,x) = (A_k x^k, x^k) \ \ \forall\, x = \begin{pmatrix} x^k \\ 0 \end{pmatrix} \neq 0$, то $\det A_k \neq 0 \ \ \forall\, k$ $\Rightarrow A = LDU = A^* = U^*D^*L^* = LDL^*$. Т.к. $\exists\, y^{(k)} = [L^*]^{-1}e_k \neq 0 \ \ \&\ \ A > 0$, то $(Ay^k, y^k) = (LDL^*y^k, y^k) = (DL^*y^k, L^*y^k) = d_k > 0 \ \ \forall\, k$.

Следствием теоремы о разложении Холлеского является критерий положительной определенности самосопряженной (относительно евклидова скалярного произведения) матрицы A:

Самосопряженная (относительно евклидова скалярного произведения) матрица A положительно определена тогда и только тогда, когда все её главные миноры положительны.

Докажите, что для любой самосопряженной (относительно евклидова скалярного произведения) матрицы A выполняются неравенства:

$$\lambda_{\min}(A)\cdot(x,x) \le (Ax,x) \le \lambda_{\max}(A)\cdot(x,x) \quad \forall\, x.$$

Очевидно, что следствием левого неравенства является утверждение:

Самосопряженная (относительно евклидова скалярного произведения) матрица A положительно определена тогда и только тогда, когда все её собственные значения положительны.

Суммируя эти утверждения, получаем «**критерий Сильвестра»:**

Все собственные значения самосопряженной (относительно евклидова скалярного произведения) матрицы A положительны тогда и только тогда, когда все её главные миноры положительны.

Метод квадратного корня

Теорема.	Если $A = A^* > 0$, то $A = BB^*$, где B – нижняя треугольная м–ца, и $\mathrm{cond}_2 B = \mathrm{cond}_2 B^* = \sqrt{\mathrm{cond}_2 A}$.
Док–во.	Из теоремы о разложении Холесского имеем

$$A = LDL^* = L(D^{1/2}D^{1/2})L^* = (LD^{1/2})(LD^{1/2})^* \;\Rightarrow\; B = LD^{1/2}.$$

Т.к. $Sp(B^*B) = Sp(BB^*) \equiv Sp(A)$, то $\| B\|_2 = \| B^*\|_2 = \sqrt{\rho(A)} = \sqrt{\| A\|_2}$.

Аналогично $\| B^{-1}\|_2 = \| (B^*)^{-1}\|_2 = \sqrt{\rho(A^{-1})} = \sqrt{\| A^{-1}\|_2}$.

$\Rightarrow\; cond_2 B = cond_2 B^* = \sqrt{cond_2 A}$.

Решение системы уравнений $Ax = b$ с помощью разложения $A = BB^*$ называется методом квадратного корня. Так как

$$\begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & \bar{b}_{21} & \bar{b}_{31} \\ 0 & b_{22} & \bar{b}_{32} \\ 0 & 0 & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & \bar{a}_{21} & \bar{a}_{31} \\ a_{21} & a_{22} & \bar{a}_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \Rightarrow \begin{array}{l} b_{11}^2 = a_{11},\; b_{21}b_{11} = a_{21},\; b_{31}b_{11} = a_{31} \\ b_{22}^2 = a_{22} - b_{21}\bar{b}_{21},\; b_{32}b_{22} = a_{32} - b_{21}\bar{b}_{31} \\ b_{33}^2 = a_{33} - b_{31}\bar{b}_{31} - b_{32}\bar{b}_{32} \end{array}$$

то элементы матрицы B вычисляются по следующим формулам:

$$\left\{ b_{kk} = \sqrt{a_{kk} - \sum_{j=1}^{k-1} | b_{kj} |^2},\;\; b_{k+i,k} = (a_{k+i,k} - \sum_{j=1}^{k-1} b_{k+i,j}\bar{b}_{k,j}) / b_{kk},\;\; i = 1,\ldots,n-k \right\}_{k=1}^{n}$$

Лекция 3.

Метод исключения с выбором главного элемента по столбцу

Напомним 1–ый шаг схемы единственного деления для решения $Ax = b$:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \quad \Rightarrow \quad \begin{bmatrix} 1 & a_{12}^{(1)} & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(1)} & \dots & a_{2n}^{(1)} \\ \vdots & \vdots & \dots & \vdots \\ 0 & a_{n2}^{(1)} & \dots & aa_{nn}^{(1)} \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1^{(1)} \\ b_2^{(1)} \\ \vdots \\ b_n^{(1)} \end{pmatrix},$$

где $a_{1j}^{(1)} = a_{1j} / a_{11}$, $b_1^{(1)} = b_1 / a_{11}$, $a_{ij}^{(1)} = a_{ij} - a_{i1} \cdot a_{1j} / a_{11}$ $(i, j = 2, ..., n)$.

Эти операции выполнимы, если (главный элемент шага) $a_{11} \neq 0$.

Ошибки округления будут меньше, если $|a_{11}| \geq |a_{1j}|$ или $|a_{11}| \geq |a_{i1}|$.

Матрица перестановок

$$P = (p_{ij})_{i,j=1}^{n}, \quad p_{ij} = \begin{cases} 1, & j = k_i \\ 0, & j \neq k_i \end{cases}, \text{ где } (k_1, k_2, ..., k_n) \text{ – перестановка } (1, 2, ..., n).$$

Доказать, что $PP^* = P^*P = E$, т.е. P – ортогональная матрица.
Доказать, что $\operatorname{cond}_2(P) = 1$.

Элементарная матрица перестановок

P_{kl} – матрица перестановок k и l элементов в n–ке $(1, 2, ..., n)$.

Доказать, что $P_{k,l} = P_{k,l}^* = P_{k,l}^{-1}$.

Доказать, что умножение на матрицу $P_{k,l}$ матрицы A слева $(P_{k,l}A)$ – это перестановка k и l строк, справа $(AP_{k,l})$ –перестановка k и l столбцов матрицы A.

Выбор главного элемента по столбцу.

1–й шаг: находим i_1: $|a_{i_1 1}| \geq \max\limits_{i=1,...,n} |a_{i1}|$ ($\neq 0$, если $\det A \neq 0$);

меняем местами 1 и i_1 строки: $A^{(1/2)} = P_{1,i_1} A$, $b^{(1/2)} = P_{1,i_1} b$;

обнуляем в 1–ом столбце элементы: $A^{(1)} = L_1 A^{(1/2)}$, $b^{(1)} = L_1 b^{(1/2)}$,

$$L_1 = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -\dfrac{a_{21}^{(1/2)}}{a_{11}^{(1/2)}} & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ -\dfrac{a_{n1}^{(1/2)}}{a_{11}^{(1/2)}} & 0 & \dots & 1 \end{bmatrix}, \; A^{(1)} = \left[\begin{array}{c|ccc} a_{11}^{(1/2)} & a_{12}^{(1/2)} & \dots & a_{1n}^{(1/2)} \\ \hline 0 & a_{22}^{(1)} & \dots & a_{2n}^{(1)} \\ \vdots & \vdots & \dots & \vdots \\ 0 & a_{n2}^{(1)} & \dots & a_{nn}^{(1)} \end{array}\right], \; b^{(1)} = \left(\begin{array}{c} b_1^{(1/2)} \\ \hline b_2^{(1)} \\ \vdots \\ b_n^{(1)} \end{array}\right)$$

$\det A \neq 0 \;\Rightarrow\; \det A^{(1)} \neq 0$.

После k шагов имеем $A^{(k)}x = b^{(k)}$, где $\det A^{(k)} \neq 0$, если $\det A \neq 0$

$$A^{(k)} = \left[\begin{array}{c|ccc} U_k & & U_{k,n-k} & \\ \hline & a^{(k)}_{k+1,k+1} & \cdots & a^{(k)}_{k+1,n} \\ 0 & \vdots & \cdots & \vdots \\ & a^{(k)}_{n,k+1} & \cdots & a^{(k)}_{n,n} \end{array}\right], \quad b^{(k)} = \left[\begin{array}{c} y^{(k)} \\ \hline b^{(k)}_{k+1} \\ \vdots \\ b^{(k)}_n \end{array}\right]$$

$(k+1)$–й шаг:

находим i_{k+1}: $|a^{(k)}_{i_{k+1}\,k+1}| \geq \max\limits_{i=k+1,\ldots,n} |a_{i\;k+1}|$;

меняем местами $k+1$ и i_{k+1} строки:

$$P_{k+1,i_{k+1}} = \begin{bmatrix} E_k & 0 \\ 0 & P^{(n-k)}_{1,i_{k+1}-k} \end{bmatrix}, \qquad A^{(k+1/2)} = P_{k+1,i_{k+1}} A^{(k)}, \quad b^{(k+1/2)} = P_{k+1,i_{k+1}} b^{(k)};$$

обнуляем в $(k+1)$–ом столбце элементы:

$$L_{k+1} = \begin{bmatrix} E_k & 0 \\ 0 & L^{(n-k)}_1 \end{bmatrix}, \quad L^{(n-k)}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\dfrac{a^{(k+1/2)}_{k+2,k+1}}{a^{(k+1/2)}_{k+1,k+1}} & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ -\dfrac{a^{(k+1/2)}_{n,k+1}}{a^{(k+1/2)}_{k+1,k+1}} & 0 & \cdots & 1 \end{bmatrix},$$

$A^{(k+1)} = L_{k+1} A^{(k+1/2)}$, $b^{(k+1)} = L_{k+1} b^{(k+1/2)}$:

$$A^{(k+1)} = \left[\begin{array}{c|c:ccc} U_k & & & U_{k,n-k} & \\ \hline & a^{(k+1/2)}_{k+1,k+1} & a^{(k+1/2)}_{k+1,k+2} & \cdots & a^{(k+1/2)}_{k+1,n} \\ \hdashline & 0 & a^{(k+1)}_{k+2,k+2} & \cdots & a^{(k+1)}_{k+2,n} \\ 0 & \vdots & \vdots & \cdots & \vdots \\ & 0 & a^{(k+1)}_{n,k+2} & \cdots & a^{(k+1)}_{n,n} \end{array}\right], \quad b^{(k+1)} = \left[\begin{array}{c} y^{(k)} \\ \hline b^{(k+1/2)}_{k+1} \\ \hdashline b^{(k+1)}_{k+2} \\ \vdots \\ b^{(k+1)}_n \end{array}\right].$$

$\det A^{(k)} \neq 0 \;\Rightarrow\; \det A^{(k+1)} \neq 0$.

Очевидно, что, если $\det A \neq 0$, то выполнив $n-1$ шаг, получим систему с верхней треугольной матрицей: $A^{(n-1)}x \equiv Ux = b^{(n-1)} \equiv y$.

Теорема. Если $\det A \neq 0$, то $PA = LU$, где $P = P_{n-1,i_{n-1}} \cdot \ldots \cdot P_{1,i_1}$, $L^{-1} = \tilde{L}_{n-1} \cdot \ldots \cdot \tilde{L}_1$,

$$\tilde{L}_k = P_{n-1,i_{n-1}} \cdot \ldots \cdot P_{k+1,i_{k+1}} \cdot L_k \cdot P_{k+1,i_{k+1}} \cdot \ldots \cdot P_{n-1,i_{n-1}}$$

Доказать эту теорему в качестве упражнения, проверив, что матрицы L_k и $\tilde{L}_k$ имеют одинаковую структуру.

Метод вращений решения системы уравнений

Элементарная матрица вращения

$$Q_{k,l}=\left[\begin{array}{c|ccc|c} E_{k-1} & & 0 & & 0 \\ \hline & \bar{c}_{k,l} & 0 & -\bar{s}_{k,l} & \\ 0 & 0 & E_{l-k-1} & 0 & 0 \\ & s_{k,l} & 0 & c_{k,l} & \\ \hline 0 & & 0 & & E_{n-l} \end{array}\right]\begin{array}{l} \\ -k-\text{я строка} \\ \qquad c_{k,l}\cdot\bar{c}_{k,l}+s_{k,l}\cdot\bar{s}_{k,l}=1, \qquad (k<l). \\ -l-\text{я строка} \\ \\ \end{array}$$

Доказать, что $Q_{k,l}$ – унитарная матрица, т.е. $Q_{k,l}(Q_{k,l})^*=(Q_{k,l})^*Q_{k,l}=E$.

Доказать, что $\det Q_{k,l}=1$.

Доказать, что при умножении на матрицу $Q_{k,l}$ матрицы A слева ($Q_{k,l}A$) изменяются только k и l строки матрицы A.

k–ый шаг метода вращений

Предположим, что после $k-1$ шага система $Ax=b$ с помощью умножения слева на ортогональную матрицу приведена к виду $A^{(k-1)}x=b^{(k-1)}$, где

$$A^{(k-1)}=\left[\begin{array}{c|ccc} R_{k-1} & & R_{k-1,n-k+1} & \\ \hline & a_{k,k}^{(k-1)} & \dots & a_{k,n}^{(k-1)} \\ 0 & \vdots & \dots & \vdots \\ & a_{n,k}^{(k-1)} & \dots & a_{n,n}^{(k-1)} \end{array}\right],\quad b^{(k-1)}=\left[\begin{array}{c} y^{(k-1)} \\ \hline b_k^{(k-1)} \\ \vdots \\ b_n^{(k-1)} \end{array}\right]\quad (A^{(0)}=A,\ \ b^{(0)}=b).$$

Тогда k–ый шаг состоит из умножения системы $A^{(k-1)}x=b^{(k-1)}$ слева на элементарные матрицы вращений $Q_{k,k+1}, ..., Q_{k,n}$:

$$Q_{k,k+i}=\left[\begin{array}{c|c} E_{k-1} & 0 \\ \hline 0 & Q_{1,1+i}^{(n-k+1)} \end{array}\right],\quad A^{(k-1,i)}=Q_{k,k+i}A^{(k-1,i-1)},\quad b^{(k-1,i)}=Q_{k,k+i}b^{(k-1,i-1)}, \text{ где}$$

$$c_{k,k+i}=\frac{a_{k,k}^{(k-1,i-1)}}{r_{k,k+i}},\quad s_{k,k+i}=-\frac{a_{k+i,k}^{(k-1,i-1)}}{r_{k,k+i}}, \text{ если } r_{k,k+i}=\sqrt{|a_{k,k}^{(k-1,i-1)}|^2+|a_{k+i,k}^{(k-1,i-1)}|^2}\neq 0,$$

$Q_{k,k+i}=E$, если $r_{k,k+i}=0$.

В результате получим $A^{(k)}=Q_kA^{(k-1)}$, $b^{(k)}=Q_kb^{(k-1)}$, где $Q_k=Q_{k,n}...Q_{k,k+2}Q_{k,k+1}$.

Выполнив $n-1$ шаг, получим систему с верхней треугольной матрицей: $A^{(n-1)}x\equiv Rx=b^{(n-1)}\equiv y$ (заметим, что, если $\det A\neq 0$, то и $\det R\neq 0$).

Если определить унитарную матрицу $Q^*=Q_{n-1}\cdot\ldots\cdot Q_1$, то справедлива

Теорема. $\forall A\ \ \exists A=Q\cdot R$.

Доказать, что $\mathrm{cond}_2 A=\mathrm{cond}_2 Q\cdot\mathrm{cond}_2 R=\mathrm{cond}_2 R$.

Сведение системы уравнений к системе с двухдиагональной матрицей с помощью матриц вращения

1-ый шаг

Сначала, умножая последовательно слева систему $Ax = b$ на элементарные матрицы вращений $Q_{1,2}, \dots, Q_{1,n}$, исключим первое неизвестное из всех уравнений, начиная со второго:

$$Q_{1,n} \cdot \ldots \cdot Q_{1,2} \cdot Ax \equiv A^{(1/2)}x = b^{(1)} \equiv Q_{1,n} \cdot \ldots \cdot Q_{1,2} \cdot b,$$

(это первый шаг метода вращений для системы $Ax = b$).

Затем матрицу полученной системы умножим последовательно справа на элементарные матрицы вращений $T_{2,3}, \dots, T_{2,n}$, определяя угол вращения из условия равенства нулю элементов первой строки $(A^{(2,3)} \equiv A^{(1/2)}T_{2,3})_{1,3}, \dots, (A^{(2,n)} \equiv A^{(2,n-1)}T_{2,n})_{1,n}$:

$$(\ldots(A^{(1/2)} \cdot T_{2,3}) \cdot \ldots \cdot T_{2,n-1}) \cdot T_{2,n} = A^{(1)} = \left[\begin{array}{c|cccc} a_{1,1}^{(1/2)} & a_{1,2}^{(1)} & 0 & \ldots & 0 \\ \hline 0 & a_{2,2}^{(1)} & a_{2,3}^{(1)} & \ldots & a_{2,2}^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & a_{n,2}^{(1)} & a_{n,3}^{(1)} & \ldots & a_{n,n}^{(1)} \end{array}\right].$$

И, наконец, перепишем систему $A^{(1/2)}x = b^{(1)}$ в виде

$$A^{(1)} \cdot \underbrace{(T_{2,n}^{*} T_{2,n-1}^{*} \ldots T_{2,3}^{*})x}_{x^{(1)}} = b^{(1)}$$

относительно неизвестного вектора $x^{(1)}$ и $x = T_{2,3} \ldots T_{2,n-1} T_{2,n} \cdot x^{(1)}$, т.е. необходимо сохранить косинусы и синусы углов вращения матриц $T_{2,j}$.

k-ый шаг

Предположим, что после $k-1$ шага система $Ax = b$ с помощью матриц вращения приведена к виду $A^{(k-1)}x^{(k-1)} = b^{(k-1)}$, где

$$A^{(k-1)} = \left[\begin{array}{ccc|cccc} a_{1,1}^{(k-1)} & a_{1,2}^{(k-1)} & & & & & \\ & \ddots & \ddots & & & & \\ & & a_{k-1,k-1}^{(k-1)} & a_{k-1,k}^{(k-1)} & 0 & \ldots & 0 \\ \hline & & 0 & a_{k,k}^{(k-1)} & a_{k,k+1}^{(k-1)} & \ldots & a_{k,n}^{(k-1)} \\ & & 0 & a_{k+1,k}^{(k-1)} & a_{k+1,k+1}^{(k-1)} & \ldots & a_{k+1,n}^{(k-1)} \\ & & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & 0 & a_{n,k}^{(k-1)} & a_{n,k+1}^{(k-1)} & \ldots & a_{n,n}^{(k-1)} \end{array}\right].$$

Тогда k–ый шаг состоит:

из последовательного умножения системы $A^{(k-1)}x^{(k-1)} = b^{(k-1)}$ слева на элементарные матрицы вращений $Q_{k,k+1}, ..., Q_{k,n}$ с тем, чтобы обнулить элементы k-го столбца ниже диагонального:

$$Q_{k,n}\cdot\ldots\cdot Q_{k,k+1}\cdot A^{(k-1)}x^{(k-1)} \equiv A^{(k-1/2)}x^{(k-1)} = b^{(k)} \equiv Q_{k,n}\cdot\ldots\cdot Q_{k,k+1}\cdot b^{(k-1)},$$

затем матрицу полученной системы умножим последовательно справа на элементарные матрицы вращений $T_{k+1,k+2}, ..., T_{k+1,n}$, определяя угол вращения из условия равенства нулю элементов k-ой строки $(A^{(k+1,k+2)} \equiv A^{(k-1/2)}T_{k+1,k+2})_{k,k+2}, \ldots, (A^{(k+1,n)} \equiv A^{(k+1,n-1)}T_{k+1,n})_{k,n}$:

$$(\ldots(A^{(k-1/2)}\cdot T_{k+1,k+2})\cdot\ldots\cdot T_{k+1,n-1})\cdot T_{k+1,n} = A^{(k)} =$$

$$= \left[\begin{array}{ccc|c:cccc}
a_{1,1}^{(k-1)} & a_{1,2}^{(k-1)} & & & & & & \\
 & \ddots & \ddots & & & & & \\
 & & a_{k-1,k-1}^{(k-1)} & a_{k-1,k}^{(k-1)} & 0 & & \ldots & 0 \\
\hline
 & & 0 & a_{k,k}^{(k-1/2)} & a_{k,k+1}^{(k)} & 0 & \ldots & 0 \\
\hdashline
 & & 0 & 0 & a_{k+1,k+1}^{(k)} & a_{k+1,k+2}^{(k)} & \ldots & a_{k+1,n}^{(k)} \\
 & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
 & & 0 & 0 & a_{n,k+1}^{(k)} & a_{n,k+2}^{(k)} & \ldots & a_{n,n}^{(k)}
\end{array}\right].$$

И, наконец, перепишем систему $A^{(k-1/2)}x^{(k-1)} = b^{(k-1)}$ в виде

$$A^{(k)}\cdot\underbrace{(T_{k+1,n}^{*}T_{k+1,n-1}^{*}\ldots T_{k+1,k+2}^{*})x^{(k-1)}}_{x^{(k)}} = b^{(k)}$$

относительно неизвестного вектора $x^{(k)}$ и

$$x^{(k-1)} = T_{k+1,k+2}\ldots T_{k+1,n-1}T_{k+1,n}\cdot x^{(k)}$$

т.е. нужно сохранить косинусы и синусы углов вращения матриц $T_{k+1,j}$.

Система с двухдиагональной матрицей

Выполнив n шагов, мы получим систему с двухдиагональной матрицей:

$$A^{(n)}x^{(n)} = b^{(n)}, \qquad A^{(n)} = \begin{bmatrix}
a_{1,1}^{(n)} & a_{1,2}^{(n)} & & \\
 & \ddots & \ddots & \\
 & & a_{n-1,n-1}^{(n)} & a_{n-1,n}^{(n)} \\
 & & & a_{n,n}^{(n)}
\end{bmatrix}.$$

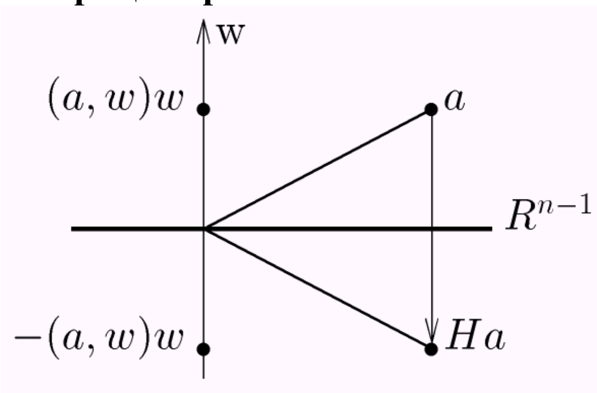
Решив ее, восстановим решение исходной системы по формулам

$$x^{(k-1)} = T_{k+1,k+2}\ldots T_{k+1,n-1}T_{k+1,n}\cdot x^{(k)},\ \ k = n, n-1, ..., 1;\ \ x = x^{(0)}.$$

Лекция 4.

Метод отражений решения системы уравнений

Матрица отражения



$$\mathrm{H}a = a - 2(a,w)\cdot w = (E - 2\cdot w\cdot w^*)a$$

если заданы векторы a и $\mathrm{H}a$, то

$$w = \frac{a - \mathrm{H}a}{\| a - \mathrm{H}a \|_2}$$

Доказать, что $\mathrm{H} = \mathrm{H}^* = \mathrm{H}^{-1}$, $\det \mathrm{H} = -1$.

k–ый шаг метода отражений

Предположим, что после $k-1$ шага система $Ax = b$ с помощью умножения слева на ортогональную матрицу приведена к виду $A^{(k-1)}x = b^{(k-1)}$, где

$$A^{(k-1)} = \left[\begin{array}{c|ccc} R_{k-1} & & R_{k-1,n-k+1} & \\ \hline & a_{k,k}^{(k-1)} & \dots & a_{k,n}^{(k-1)} \\ 0 & \vdots & \cdots & \vdots \\ & a_{n,k}^{(k-1)} & \dots & a_{n,n}^{(k-1)} \end{array}\right], \quad b^{(k-1)} = \left[\begin{array}{c} y^{(k-1)} \\ \hline b_k^{(k-1)} \\ \vdots \\ b_n^{(k-1)} \end{array}\right] \quad (A^{(0)} = A,\ b^{(0)} = b).$$

Тогда k–ый шаг состоит из умножения системы $A^{(k-1)}x = b^{(k-1)}$ слева на ортогональную матрицу вращения H_k:

$$\mathrm{H}_k = \left[\begin{array}{c|c} E_{k-1} & 0 \\ \hline 0 & E_{n-k+1} - 2w_1^{(n-k+1)}\cdot[w_1^{(n-k+1)}]^* \end{array}\right], \quad A^{(k)} = \mathrm{H}_k A^{(k-1)}, \quad b^{(k)} = \mathrm{H}_k b^{(k-1)},$$

где

$w_1^{(n-k+1)} = 0$	если $r_k = \sqrt{\lvert a_{k,k}^{(k-1)}\rvert^2 + \ldots + \lvert a_{n,k}^{(k-1)}\rvert^2} = 0$ или $a_{k+1,k}^{(k-1)} = \ldots = a_{n,k}^{(k-1)} = 0$,
$w_1^{(n-k+1)} = \dfrac{a_1^{(k-1)} - r_k\cdot e_1^{(k-1)}}{\lVert a_1^{(k-1)} - r_k\cdot e_1^{(k-1)}\rVert_2}$	если $r_k \ne 0$ & $a_{k,k}^{(k-1)} = 0$ (здесь $a_1^{(k-1)} = \begin{bmatrix} a_{k,k}^{(k-1)} \\ \vdots \\ a_{n,k}^{(k-1)} \end{bmatrix}$, $e_1^{(k-1)}$ – первый орт),
$w_1^{(n-k+1)} = \dfrac{a_1^{(k-1)} - \beta_k r_k\cdot e_1^{(k-1)}}{\lVert a_1^{(k-1)} - \beta_k r_k\cdot e_1^{(k-1)}\rVert_2}$	если $r_k \ne 0$ & $a_{k,k}^{(k-1)} \ne 0$ (здесь $\beta_k = -\dfrac{a_{k,k}^{(k-1)}}{\lvert a_{k,k}^{(k-1)}\rvert}$).

Выполнив $n-1$ шаг, получим систему с верхней треугольной матрицей: $A^{(n-1)}x \equiv Rx = b^{(n-1)} \equiv y$ (заметим, что, если $\det A \ne 0$, то и $\det R \ne 0$).

Решение системы с вырожденной матрицей

HR **–разложение с перестановками столбцов матрицы** A

1– ый шаг. Определим номер столбца j_1 матрицы $A=[a_1, ..., a_n]$ из условия $\|a_{j_1}\|_2 = \max \|a_j\|_2$ и матрицу перестановок P_{1,j_1}.

Для матрицы $A^{(1/2)} = AP_{1,j_1}$ определим матрицу отражения H_1:

$$A^{(1)} = H_1 A^{(1/2)} = H_1 A P_{1,j_1} = \left[\begin{array}{c|ccc} R_1 & & \tilde{R}_{1,n-1} & \\ \hline & a^{(1)}_{2,2} & \dots & a^{(1)}_{2,n} \\ 0 & \vdots & \dots & \vdots \\ & a^{(1)}_{n,2} & \dots & a^{(1)}_{n,n} \end{array}\right], \quad r_{11} = R_1.$$

Доказать: $|r_{11}| \ge \|a_j\| \ge |a^{(1)}_{ij}| \quad \forall\, i,\, j$

k –ый шаг. После $k-1$ шага имеем

$$A^{(k-1)} = \left[\begin{array}{c|ccc} R_{k-1} & & \tilde{R}_{k-1,n-k+1} & \\ \hline & a^{(k-1)}_{k,k} & \dots & a^{(k-1)}_{k,n} \\ 0 & \vdots & \dots & \vdots \\ & a^{(k-1)}_{n,k} & \dots & a^{(k-1)}_{n,n} \end{array}\right], \quad |r_{k-1,k-1}| \ge \left\| \begin{matrix} a^{(k-1)}_{k-1,j} \\ a^{(k-1)}_{k,j} \\ \vdots \\ a^{(k-1)}_{n,j} \end{matrix} \right\|_2 \ge \max_{k-1\le i,j\le n} |a^{(k-1)}_{i,j}|.$$

Определяем номер столбца j_k из условия $\left\| \begin{matrix} a^{(k-1)}_{k,j_k} \\ \vdots \\ a^{(k-1)}_{n,j_k} \end{matrix} \right\|_2 = \max\limits_{k\le j\le n} \left\| \begin{matrix} a^{(k-1)}_{k,j} \\ \vdots \\ a^{(k-1)}_{n,j} \end{matrix} \right\|_2$

и для $A^{(k-1/2)} = A^{(k-1)} P_{k,j_k}$ определяем матрицу отражения H_k:

$$A^{(k)} = H_k A^{(k-1/2)} = H_k A^{(k-1)} P_{k,j_k} = \left[\begin{array}{c|ccc} R_k & & \tilde{R}_{k,n-k} & \\ \hline & a^{(k)}_{k+1,k+1} & \dots & a^{(k)}_{k+1,n} \\ 0 & \vdots & \dots & \vdots \\ & a^{(k)}_{n,k+1} & \dots & a^{(k)}_{n,n} \end{array}\right].$$

Доказать: $|r_{k,k}| \ge \max\limits_{k\le j\le n} \left\| \begin{matrix} a^{(k)}_{k,j} \\ a^{(k)}_{k+1,j} \\ \vdots \\ a^{(k)}_{n,j} \end{matrix} \right\|_2 \ge \max\limits_{k\le i,j\le n} |a^{(k)}_{i,j}|.$

Ответ: Если $t = \dim(\ker A)$, то после $n-t$ шагов имеем

$$(H_{n-t} \cdot ... \cdot H_1) A (P_{1,j_1} \cdot ... \cdot P_{n-t,j_{n-t}}) = HAP = R = \begin{bmatrix} R_{n-t} & R_{n-t,t} \\ 0 & 0 \end{bmatrix},$$

где H и P – ортогональные матрицы.

Совместность системы с вырожденной матрицей

Система $Ax = b$ называется совместной, если она имеет решение. Следовательно, система совместна $\Leftrightarrow$ $b \in \operatorname{Im} A$.

$\{x^* + y \quad \forall\ y \in \ker A\}$ – общее решение системы, где x^* – любое ее решение.

Теорема. Если система $Ax = b$ совместна ($b \in \operatorname{Im} A$), то $\forall$ B: $\det B \neq 0$ совместна система $(BA)x = (Bb)$ и множества решений этих систем совпадают.

Система $Ax = b$ несовместна, если $b \notin \operatorname{Im} A$.

В этом случае ее обобщенным решением (относительно векторной нормы $\|\cdot\|$) называют вектор x: $\| Ax - b \| = \min\limits_{y} \| Ay - b \|$.

Доказать: общее решение совместной системы совпадает с множеством ее обобщенных решений.

Доказать: множество обобщенных решений $\{x:\ \| Ax - b \|_2 = \min \| Ay - b \|_2\}$ совпадает с общим решением системы $A^* A x = A^* b$.

Применение HR –разложения с перестановками столбцов для решения совместной системы

Выполним эквивалентное преобразование совместной системы $Ax = b$:

$$Ry = g:\quad R = HAP,\quad y = P^* x,\quad g = Hb.$$

Из–за ошибок округления эта система будет иметь вид:

$$\begin{bmatrix} R_{n-t} & R_{n-t,t} \\ 0 & \varepsilon_t \end{bmatrix} \begin{pmatrix} y^{(n-t)} \\ y^{(t)} \end{pmatrix} = \begin{pmatrix} g^{(n-t)} \\ \delta^{(t)} \end{pmatrix},$$

где матрица ε_t и вектор $\delta^{(t)}$ должны иметь малые по модулю элементы. Заменяем их на нулевые матрицу и вектор (диагональные элементы матрицы R по модулю мажорируют все левее и ниже лежащие элементы, как только очередной диагональный элемент стал “намного” меньше предыдущего, то и остальные элементы почти нулевые):

$$\begin{bmatrix} R_{n-t} & R_{n-t,t} \\ 0 & 0 \end{bmatrix} \begin{pmatrix} y^{(n-t)} \\ y^{(t)} \end{pmatrix} = \begin{pmatrix} g^{(n-t)} \\ 0 \end{pmatrix},$$

очевидно, что общее решение этой системы определяется формулой

$$\begin{pmatrix} y^{(n-t)} \\ y^{(t)} \end{pmatrix} = \begin{pmatrix} R_{n-t}^{-1}(g^{(n-t)} - R_{n-t,t} y^{(t)}) \\ y^{(t)} \end{pmatrix} \qquad \forall\ y^{(t)} \in R^t,$$

а решение исходной системы $x = Py$.

Метод прогонки решения систем с трехдиагональной матрицей

Для решения системы линейных алгебраических уравнений

$$\begin{bmatrix} a_1 & b_1 & & & 0 \\ c_2 & a_2 & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & & c_{n-1} & a_{n-1} & b_{n-1} \\ 0 & & 0 & c_n & a_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_{n-1} \\ f_n \end{pmatrix},$$

где диагональные блоки a_i либо числа (метод прогонки), либо квадратные матрицы порядка n_i (метод матричной прогонки), и все ”главные миноры” не равны нулю:

$$A_i = \begin{bmatrix} a_1 & b_1 & & & 0 \\ c_2 & a_2 & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & & c_{i-1} & a_{i-1} & b_{i-1} \\ 0 & & 0 & c_i & a_i \end{bmatrix}, \quad \det A_i \neq 0, \quad i = 1, ..., n,$$

применим следующую схему метода Гаусса – метод (матричной) прогонки.

Прямой ход:

1-й шаг: вычисляем матрицу $\beta_1 = a_1^{-1} b_1$ и вектор $z_1 = a_1^{-1} f_1$, т.е. получаем систему

$$\begin{bmatrix} E_1 & \beta_1 & & & 0 \\ c_2 & a_2 & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & & c_{n-1} & a_{n-1} & b_{n-1} \\ 0 & & 0 & c_n & a_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} z_1 \\ f_2 \\ \vdots \\ f_{n-1} \\ f_n \end{pmatrix},$$

где E_1 – единичная матрица порядка n_1;

2-й шаг: сначала из второго уравнения вычитаем первое, умноженное на c_2:

$$\begin{bmatrix} E_1 & \beta_1 & & & 0 \\ 0 & a_2 - c_2\beta_1 & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & & c_{n-1} & a_{n-1} & b_{n-1} \\ 0 & & 0 & c_n & a_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} z_1 \\ f_2 - c_2 z_1 \\ \vdots \\ f_{n-1} \\ f_n \end{pmatrix},$$

затем вычисляем матрицу $\beta_2 = (a_2 - c_2\beta_1)^{-1} b_2$ и вектор $z_2 = (a_2 - c_2\beta_1)^{-1}(f_2 - c_2 z_1)$, т.е. получаем систему

$$
\begin{bmatrix}
E_1 & \beta_1 & & & & 0 \\
0 & E_2 & \beta_2 & & & \\
 & c_3 & a_3 & b_3 & & \\
 & & \ddots & \ddots & \ddots & \\
 & & & c_{n-1} & a_{n-1} & b_{n-1} \\
0 & & & & c_n & a_n
\end{bmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}
=
\begin{pmatrix} z_1 \\ z_2 \\ f_3 \\ \vdots \\ f_{n-1} \\ f_n \end{pmatrix},
$$

где E_2 – единичная матрица порядка n_2;

…

k-й шаг: предыдущие шаги преобразовали исходную систему к системе

$$
\begin{bmatrix}
E_1 & \beta_1 & & & & & & 0 \\
0 & E_2 & \beta_2 & & & & & \\
 & \ddots & \ddots & \ddots & & & & \\
 & & 0 & E_{k-1} & \beta_{k-1} & & & \\
 & & & c_k & a_k & b_k & & \\
 & & & & \ddots & \ddots & \ddots & \\
 & & & & & c_{n-1} & a_{n-1} & b_{n-1} \\
0 & & & & & & c_n & a_n
\end{bmatrix}
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k-1} \\ x_k \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}
=
\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{k-1} \\ f_k \\ \vdots \\ f_{n-1} \\ f_n \end{pmatrix},
$$

сначала из k-го уравнения вычитаем $(k-1)$-ое, умноженное на c_k:

$$
\begin{bmatrix}
E_1 & \beta_1 & & & & & & 0 \\
0 & E_2 & \beta_2 & & & & & \\
 & \ddots & \ddots & \ddots & & & & \\
 & & 0 & E_{k-1} & \beta_{k-1} & & & \\
 & & & 0 & a_k - c_k\beta_{k-1} & b_k & & \\
 & & & & \ddots & \ddots & \ddots & \\
 & & & & & c_{n-1} & a_{n-1} & b_{n-1} \\
0 & & & & & & c_n & a_n
\end{bmatrix}
\begin{pmatrix} x_1 \\ \vdots \\ x_{k-1} \\ x_k \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}
=
\begin{pmatrix} z_1 \\ \vdots \\ z_{k-1} \\ f_k - c_k z_{k-1} \\ \vdots \\ f_{n-1} \\ f_n \end{pmatrix},
$$

затем вычисляем матрицу $\beta_k = (a_k - c_k\beta_{k-1})^{-1} b_k$ и вектор $z_k = (a_k - c_k\beta_{k-1})^{-1}(f_k - c_k z_{k-1})$, т.е. получаем систему

$$\begin{bmatrix} E_1 & \beta_1 & & & & & & 0 \\ 0 & E_2 & \beta_2 & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & 0 & E_k & \beta_k & & & \\ & & & 0 & a_{k+1} & b_{k+1} & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & c_{n-1} & a_{n-1} & b_{n-1} \\ 0 & & & & & & c_n & a_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \\ x_{k+1} \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_k \\ f_{k+1} \\ \vdots \\ f_{n-1} \\ f_n \end{pmatrix},$$

где E_k – единичная матрица порядка n_k;

…

n-й шаг: предыдущие шаги преобразовали исходную систему к системе

$$\begin{bmatrix} E_1 & \beta_1 & & & 0 \\ 0 & E_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & 0 & E_{n-1} & \beta_{n-1} \\ 0 & & & c_n & a_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n-1} \\ f_n \end{pmatrix},$$

сначала из n-го уравнения вычитаем $(n-1)$-ое, умноженное на c_n:

$$\begin{bmatrix} E_1 & \beta_1 & & 0 \\ \ddots & \ddots & \ddots & \\ & 0 & E_{n-1} & \beta_{n-1} \\ 0 & & 0 & a_n - c_n\beta_{n-1} \end{bmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} z_1 \\ \vdots \\ z_{n-1} \\ f_n - c_n z_{n-1} \end{pmatrix},$$

затем вычисляем вектор $z_n = (a_n - c_n\beta_{n-1})^{-1}(f_n - c_n z_{n-1})$, т.е. получаем систему

$$\begin{bmatrix} E_1 & \beta_1 & & & 0 \\ 0 & E_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & 0 & E_{n-1} & \beta_{n-1} \\ 0 & & & 0 & E_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n-1} \\ z_n \end{pmatrix},$$

где E_n – единичная матрица порядка n_n.

Таким образом, на прямом ходе метода прогонки вычисляются матрицы

$$\beta_1 = a_1^{-1} b_1,$$

$$\beta_k = (a_k - c_k\beta_{k-1})^{-1} b_k, \quad k = 2, ..., n-1;$$

и векторы

$$z_1 = a_1^{-1} f_1,$$

$$z_k = (a_k - c_k \beta_{k-1})^{-1} (f_k - c_k z_{k-1}), \quad k = 2, ..., n.$$

Обратный ход метода прогонки элементарен:
решение системы, полученной в результате прямого хода:

$$\begin{bmatrix} E_1 & \beta_1 & & & 0 \\ 0 & E_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & 0 & E_{n-1} & \beta_{n-1} \\ 0 & & & 0 & E_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n-1} \\ z_n \end{pmatrix},$$

вычисляем по формулам:

$$x_n = z_n,$$

$$x_{n-1} = -\beta_{n-1} x_n + z_{n-1},$$

$$...$$

$$x_i = -\beta_i x_{i+1} + z_i,$$

$$...$$

$$x_1 = -\beta_1 x_2 + z_1.$$

Метод прогонки применим для системы, матрица которой имеет строгое диагональное преобладание по всем строкам.

Теорема. Если $\forall\, i \quad |a_i| > |c_i| + |b_i| \quad (c_1 = b_n = 0)$, то $\det A_k \neq 0 \quad \forall\, k$ (т.е. LU–разложение существует и метод прогонки применим).

Док–во. (от противного) Пусть $\exists\, k: \ \det A_k = 0$,
тогда $\exists\, x^{(k)} \neq 0$: $\ A_k x^{(k)} = 0$ и $\exists\, i$: $\ |x_i^{(k)}| = \max\limits_{1 \le j \le k} |x_j^{(k)}| > 0$.

Разделим равенство $c_i x_{i-1}^{(k)} + a_i x_i^{(k)} + b_i x_{i+1}^{(k)} = 0$ на $x_i^{(k)}$ и оценим a_i:

$$|a_i| \le |c_i| \cdot \frac{|x_{i-1}^{(k)}|}{|x_i^{(k)}|} + |b_i| \cdot \frac{|x_{i-1}^{(k)}|}{|x_i^{(k)}|} \le |c_i| + |b_i| \text{ – противоречие условию.}$$

Справедливость этой теоремы следует из более общего утверждения.

Лемма Гершгорина (о локализации собственных значений).

Множество всех собственных значений произвольной квадратной матрицы A принадлежит объединению кругов Гершгорина:

$$Sp(A) \subset \bigcup_{i=1}^{n} S_i, \quad S_i = \{\lambda \in C: \ |\lambda - a_{ii}| \le R_i \equiv \sum_{j=1,\ j \neq i}^{n} |a_{ij}|\}.$$

Доказательство. Пусть λ_k – произвольное собственное число матрицы A, а $x^{(k)} \neq 0$ – соответствующий ему собственный вектор, т.е. $Ax^{(k)} = \lambda_k \cdot x^{(k)}$.

Найдем круг Гершгорина, содержащий точку λ_k на комплексной плоскости.
Определим номер i максимальной по модулю компоненты $x_i^{(k)}$ вектора $x^{(k)}$ и i-ое уравнение системы $Ax^{(k)} = \lambda_k \cdot x^{(k)}$ перепишем в виде

$$(\lambda_k - a_{ii}) \cdot x_i^{(k)} = a_{i1} \cdot x_1^{(k)} + \ldots + a_{i,i-1} \cdot x_{i-1}^{(k)} + a_{i,i+1} \cdot x_{i+1}^{(k)} + \ldots + a_{in} \cdot x_n^{(k)}.$$

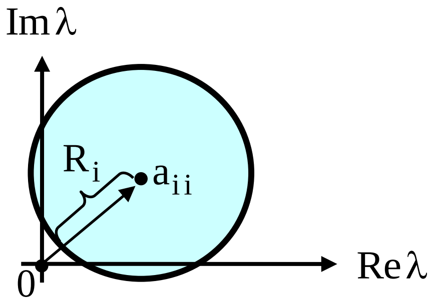
Разделив это равенство $x_i^{(k)} \neq 0$ и, учитывая, что $| x_j^{(k)} / x_i^{(k)} | \leq 1 \ \forall\ j$, после применения неравенства треугольника получим

$$| \lambda_k - a_{ii} | \leq | a_{i1} | \cdot | x_1^{(k)} / x_i^{(k)} | + \ldots + | a_{i,i-1} | \cdot | x_{i-1}^{(k)} / x_i^{(k)} | +$$
$$+ | a_{i,i+1} | \cdot | x_{i+1}^{(k)} / x_i^{(k)} | + \ldots + | a_{in} | \cdot | x_n^{(k)} / x_i^{(k)} | \leq$$
$$\leq | a_{i1} | + \ldots + | a_{i,i-1} | + | a_{i,i+1} | + \ldots + | a_{in} | = R_i,$$

т.е. λ_k принадлежит i-ому кругу Гершгорина.
Итак, для каждого собственного числа матрицы A существует круг Гершгорина, которому оно принадлежит.
Лемма доказана.
Для матрицы A со строгим диагональным преобладанием по всем строкам:

$$| a_{ii} | > R_i \equiv \sum_{j=1,\ j \neq i}^{n} | a_{ij} |, \ i = 1, 2, \ldots, n,$$

все круги Гершгорина не содержат начало координат комплексной плоскости



и, следовательно, $\lambda = 0$ не является её собственным значением, т.е. матрица A – неособенная матрица и её определитель не равен 0.

Лекция 5. Итерационные методы решения линейных уравнений

Мы будем рассматривать только вещественные системы линейных алгебраических уравнений, так как система уравнений $Ax = b$ над полем комплексных чисел сводится (доказать) к системе

$$\begin{bmatrix} \operatorname{Re} A & -\operatorname{Im} A \\ \operatorname{Im} A & \operatorname{Re} A \end{bmatrix} \begin{pmatrix} \operatorname{Re} x \\ \operatorname{Im} x \end{pmatrix} = \begin{pmatrix} \operatorname{Re} b \\ \operatorname{Im} b \end{pmatrix}$$

с вещественными коэффициентами.

Пример и основные определения

Пример:

пусть для матрицы системы $Ax = b$ построена обратная $A^{-1} = (LU)^{-1}$. Из–за ошибок округления мы получим не обратную матрицу, а к ней близкую: $\tilde{A}^{-1}$. Тогда $\tilde{x} = \tilde{A}^{-1}x \neq x$, а для разности $\tilde{x} - x$ имеем уравнение $A(\tilde{x} - x) = A\tilde{x} - b$, приближенное решение которого $(\tilde{\tilde{x}} - x) = \tilde{A}^{-1}(A\tilde{x} - b) \;\Rightarrow\; \tilde{\tilde{x}} = \tilde{x} - \tilde{A}^{-1}(A\tilde{x} - b)$ или итерационное уточнение

$$x^{k+1} = x^k - \tilde{A}^{-1}(Ax^k - b),\;\; k = 0, 1, \ldots .$$

Одношаговый (двухслойный) итерационный метод решения $Ax = b$**:**

$$x^{k+1} = x^k - H_k(Ax^k - b)$$

$$x^0 - \text{задан},\quad k = 0, 1, 2, \ldots;\quad H_k - \text{заданные матрицы.}$$

x^k – k-тое приближение (к решению системы),

$z^k = x^k - x$ – ошибка k-той итерации	$z^{k+1} = z^k - H_k A z^k = (E - H_k A) z^k$ – процесс для ошибки, $S_k = E - H_k A$ – матрица шага для ошибки;
$r^k = Ax^k - b = Az^k$ – невязка k-той итерации	$r^{k+1} = r^k - H_k A r^k = (E - AH_k) r^k$ – процесс для невязки, $T_k = E - AH_k$ – матрица шага для невязки;

Метод называется сходящимся, если $\lim\limits_{k\to\infty} \| z^k \| = 0 \quad \forall\, x^0 \in R^n$**.**

(Так как в R^n все нормы эквивалентны, то определение сходимости от нормы не зависит.)

Стационарный одношаговый итерационный метод решения $Ax = b$**:**

$$x^{k+1} = x^k - H(Ax^k - b)$$

$$x^0 - \text{задан},\quad k = 0, 1, 2, \ldots;\quad H - \text{заданная матрица.}$$

Впредь мы будем предполагать, что $\det A \neq 0$ и $\det H \neq 0$.

Условия сходимости стационарного итерационного метода

$$x^{k+1} = x^k - H(Ax^k - b), \quad x^0 - \text{задан}, \quad k = 0, 1, 2, \ldots; \quad H - \text{заданная матрица.}$$

Достаточные условия:

Теорема. Если $\|S\| = \|E - HA\| < 1$, то $\|z^k\| \to 0$, т.е. $x^k \to x \quad \forall\, x^0 \in R^n$.

Док–во. $\|x^k - x\| = \|z^k\| = \|Sz^{k-1}\| \le \|S\| \cdot \|z^{k-1}\| = \|S\| \cdot \|Sz^{k-2}\| \le$

$$\le \|S\|^2 \cdot \|z^{k-2}\| \le \|S\|^k \cdot \|z^0\| \to 0.$$

Теорема. Если $\|T\| = \|E - AH\| < 1$, то $\|z^k\| \to 0$, т.е. $x^k \to x \quad \forall\, x^0 \in R^n$.

Док–во. $\|Ax^k - b\| = \|r^k\| = \|Tr^{k-1}\| \le \|T\| \cdot \|r^{k-1}\| = \|T\| \cdot \|Tr^{k-2}\| \le$

$$\le \|T\|^2 \cdot \|r^{k-2}\| \le \|T\|^k \cdot \|r^0\| \to 0.$$

$$\Rightarrow \quad \|x^k - x\| = \|z^k\| = \|A^{-1}r^k\| \le \|A^{-1}\| \cdot \|r^k\| \to 0.$$

Необходимое и достаточное условие сходимости стационарного итерационного метода

$$x^{k+1} = x^k - H(Ax^k - b)$$

$$x^0 - \text{задан}, \quad k = 0, 1, 2, \ldots; \quad H - \text{заданная матрица}, \det H \ne 0,$$

для решения системы линейных алгебраических уравнений $Ax = b$, $\det A \ne 0$, сформулировано в следующей теореме.

Теорема.

Стационарный итерационный метод

$$x^{k+1} = x^k - H(Ax^k - b)$$

$$x^0 - \text{задан}, \quad k = 0, 1, 2, \ldots; \quad H - \text{заданная матрица}, \det H \ne 0,$$

для решения системы линейных алгебраических уравнений $Ax = b$, $\det A \ne 0$, сходится тогда и только тогда, когда $\rho(S) < 1$, где $S = E - HA$ – матрица шага для ошибки $z^k = x^k - x$ стационарного итерационного метода: $z^k = Sz^{k-1}$.

Док–во.

Необходимость.

Пусть стационарный итерационный процесс сходится, т.е. $\forall\, z^0 \ne 0$

$$z^k = Sz^{k-1} = S^k z^0 \overset{k\to\infty}{\to} 0.$$

Выберем начальное приближение x^0 стационарного итерационного процесса таким, чтобы вектор начальной ошибки $z^0 = x^0 - x \ne 0$ был собственным вектором матрицы шага S, соответствующим некоторому (любому) ее собственному значению $\lambda \in Sp(S)$, т.е. $Sz^0 = \lambda \cdot z^0$, $z^0 \ne 0$.

Тогда

$$z^k = Sz^{k-1} \equiv S^k z^0 \equiv \lambda^k z^0 \overset{k\to\infty}{\to} 0$$

только при $|\lambda|<1$.

Следовательно, все собственные значения матрицы шага S стационарного итерационного процесса по модулю меньше единицы, т.е.

$$\rho(S)<1$$

– необходимое условие сходимости стационарного итерационного метода.

Достаточность.

Заметим, что для любого $\varepsilon>0$ существует матрица Q_ε, преобразующая матрицу S к жордановой форме:

$$Q_\varepsilon^{-1}SQ_\varepsilon = J_\varepsilon = \mathrm{diag}\{J_{1,\varepsilon}, ..., J_{m,\varepsilon}\},$$

где

$$J_{i,\varepsilon} = \begin{bmatrix} \lambda_i & & & 0 \\ \varepsilon & \lambda_i & & \\ & \ddots & \ddots & \\ 0 & & \varepsilon & \lambda_i \end{bmatrix}, \ \lambda_i \in sP(s).$$

Тогда

$$z^k = Sz^{k-1} \equiv Q_\varepsilon J_\varepsilon Q_\varepsilon^{-1} z^{k-1} = Q_\varepsilon (J_\varepsilon)^k Q_\varepsilon^{-1} z^0,$$

$$\| z^k \|_\infty \le \| Q_\varepsilon \|_\infty \cdot \| Q_\varepsilon^{-1} \|_\infty \cdot (\| J_\varepsilon \|_\infty)^k \cdot \| z^0 \|_\infty \xrightarrow{k\to\infty} 0,$$

если $\| J_\varepsilon \|_\infty = \max\limits_{\lambda\in Sp(S)} \{|\lambda|+\varepsilon\} = \rho(S)+\varepsilon<1$.

Так как $\rho(S)<1$, то $\| J_\varepsilon \|_\infty <1$ для любого $0<\varepsilon<1-\rho(S)$.

Асимптотическая скорость сходимости

Для отношения нормы ошибки z^k к норме начальной ошибки z^0 стационарного итерационного метода $x^{k+1} = x^k - H(Ax^k - b)$ справедлива оценка

$$\frac{\| z^k \|}{\| z^0 \|} \le \| S^k \|, \text{ где } S = E - HA \text{ – матрица шага для ошибки.}$$

Отсюда можно сделать вывод, что за каждую из k итераций ошибка в “среднем” убывает в $1/d_k$, $d_k = \sqrt[k]{\| S^k \|}$, раз. Тогда для уменьшения начальной ошибки в $1/\varepsilon$ раз ($\| z^k \| / \| z^0 \| \le (d_k)^k = \varepsilon$) достаточно выполнить

$$k(\varepsilon) = \frac{-\ln\varepsilon}{-\ln d_k} \equiv \frac{-\ln\varepsilon}{-\ln\sqrt[k]{\| S^k \|}} \text{ итераций.}$$

Эту формулу можно проинтерпретировать следующим образом:

$$\underbrace{-\ln\sqrt[k]{\| S^k \|}}_{\text{скорость}} \cdot \underbrace{k(\varepsilon)}_{\text{время}} = \underbrace{-\ln\varepsilon}_{\text{растояние}}.$$

Величина $R_k = -\ln \sqrt[k]{\| S^k \|}$ **называется средней скоростью сходимости итерационного метода за** k **итераций.** Очевидно, что средняя скорость сходимости R_k зависит от выбора матричной нормы.

Теорема. Если $\| S \| = \| E - HA \| < 1$, то для уменьшения начальной ошибки в $1/\varepsilon$ раз достаточно выполнить $k(\varepsilon) = \left[\dfrac{-\ln \varepsilon}{-\ln \| S \|} \right] + 1$ итераций.

Док–во. При $k \ge k(\varepsilon)$ имеем

$$\| z^k \| = \| S^k z^0 \| \le \| S \|^k \cdot \| z^0 \| \le$$

$$\le \| S \|^{k(\varepsilon)} \cdot \| z^0 \| = e^{k(\varepsilon) \cdot \ln \| S \|} \cdot \| z^0 \| =$$

$$= e^{\ln \varepsilon} \cdot \| z^0 \| = \varepsilon \cdot \| z^0 \|.$$

Доказать: $R_k \ge -\ln \| S \|$.

Величина $R_\infty = \lim\limits_{k \to \infty} R_k = -\ln \{ \lim\limits_{k \to \infty} \sqrt[k]{\| S^k \|} \}$ **называется асимптотической скоростью сходимости стационарного итерационного метода.**

Теорема. Если $\rho(S) = \rho(E - HA) < 1$, то $R_\infty = -\ln \rho(S)$.

Доказательство. Достаточно доказать, что $\exists \lim\limits_{k \to \infty} \sqrt[k]{\| S^k \|} = \rho(S)$.

1. $\varliminf\limits_{k \to \infty} \sqrt[k]{\| S^k \|} \ge \rho(S)$.

   Действительно, так как $\| S \| \ge \rho(S)$ для любой матричной нормы $\| S \|$, подчиненной векторной норме $\| z \|$, то $\| S^k \| \ge \rho(S^k) \equiv [\rho(S)]^k$. Следовательно,

$$\sqrt[k]{\| S^k \|} \ge \rho(S) \;\Rightarrow\; \varliminf\limits_{k \to \infty} \sqrt[k]{\| S^k \|} \ge \rho(S).$$

2. $\varlimsup\limits_{k \to \infty} \sqrt[k]{\| S^k \|_\infty} \le \rho(S)$.

   Действительно, при доказательстве теоремы о необходимом и достаточном условии сходимости стационарного итерационного метода для любого достаточно малого $\varepsilon \in (0, \varepsilon_0)$ мы установили неравенство

$$\| S^k z^0 \|_\infty \le \underbrace{\| Q_\varepsilon \|_\infty \cdot \| Q_\varepsilon^{-1} \|_\infty}_{c_\varepsilon} \cdot [\rho(S) + \varepsilon]^k \cdot \| z^0 \|_\infty .$$

   Следствием этого неравенства является оценка

$$\| S^k \|_\infty \le c_\varepsilon \cdot [\rho(S) + \varepsilon]^k \;\Rightarrow\; \sqrt[k]{\| S^k \|_\infty} \le \sqrt[k]{c_\varepsilon} \cdot [\rho(S) + \varepsilon].$$

   Из последнего неравенства следует, что

$$\overline{\lim_{k\to\infty}} \sqrt[k]{\| S^k \|_\infty} \le \underbrace{\lim_{k\to\infty} \sqrt[k]{c_\varepsilon}}_{\equiv 1} \cdot [\rho(S)+\varepsilon] \equiv \rho(S)+\varepsilon \quad \forall\ \varepsilon \in (0,\varepsilon_0).$$

Переходя в этом неравенстве к пределу при $\varepsilon \to 0$ и учитывая, что левая часть неравенства от ε не зависит, получим $\overline{\lim\limits_{k\to\infty}} \sqrt[k]{\| S^k \|_\infty} \le \rho(S)$.

3. $\overline{\lim\limits_{k\to\infty}} \sqrt[k]{\| S^k \|} \le \rho(S)$.

Действительно, так как все нормы в конечномерном пространстве эквивалентны, то $\exists\ \beta :\ \| S^k \| \le \beta \cdot \| S^k \|_\infty$ и

$$\overline{\lim_{k\to\infty}} \sqrt[k]{\| S^k \|} \le \overline{\lim_{k\to\infty}} \sqrt[k]{\beta \cdot \| S^k \|_\infty} \le \underbrace{\lim_{k\to\infty} \sqrt[k]{\beta}}_{\equiv 1} \cdot \overline{\lim_{k\to\infty}} \sqrt[k]{\| S^k \|_\infty} \le \rho(S).$$

4. Из полученных неравенств $\underline{\lim\limits_{k\to\infty}} \sqrt[k]{\| S^k \|} \ge \rho(S)$ и $\overline{\lim\limits_{k\to\infty}} \sqrt[k]{\| S^k \|} \le \rho(S)$ следует, что $\underline{\lim\limits_{k\to\infty}} \sqrt[k]{\| S^k \|} = \overline{\lim\limits_{k\to\infty}} \sqrt[k]{\| S^k \|} = \rho(S) = \lim\limits_{k\to\infty} \sqrt[k]{\| S^k \|}$.

Замечание. Асимптотическая скорость сходимости стационарного итерационного метода в отличие от средних скоростей сходимости не зависит от выбора матричной нормы. Обычно полагают, что метод с большей асимптотической скоростью “лучше” метода с меньшей асимптотической скоростью сходимости. Но такое мнение не всегда оправдано, поскольку определить количество итераций, необходимое для уменьшения начальной ошибки в $1/\varepsilon$ раз в конкретной норме, зная только $R_\infty = \rho(S)$ невозможно.

Лекция 6.

Один из способов построения итерационного метода решения системы линейных алгебраических уравнений $Ax = b$ состоит из представления матрицы в виде $A = B - C$, переписи системы в виде $Bx = Cx + b$ и определении очередного приближения x^{k+1} по известному приближению x^k из решения системы $Bx^{k+1} = Cx^k + b$.

Доказать: $Bx^{k+1} = Cx^k + b \quad \Rightarrow \quad x^{k+1} = x^k - B^{-1}(Ax^k - b)$.

Метод Якоби

Если $D = \operatorname{diag} A = \operatorname{diag}\{a_{11}, a_{22}, ..., a_{nn}\}$, то итерационный процесс

$$x^{k+1} = x^k - D^{-1}(Ax^k - b)$$

называется методом Якоби для решения системы $Ax = b$.

Сходимость в случае диагонального преобладания по строкам

Теорема. Если $|a_{ii}| > \sum\limits_{j=1,\ j\neq i}^{n} |a_{ij}| \quad \forall\, i$, то метод Якоби сходится.

Док–во. i–тая строка матрицы $S = E - D^{-1}A$: $\left[\dfrac{a_{i1}}{a_{ii}}, ..., \dfrac{a_{i,i-1}}{a_{ii}}, 0, \dfrac{a_{i,i+1}}{a_{ii}}, ..., \dfrac{a_{in}}{a_{ii}}\right]$.

Из условия теоремы $\Rightarrow$ $\left|\dfrac{a_{i1}}{a_{ii}}\right| + ... + \left|\dfrac{a_{i,i-1}}{a_{ii}}\right| + \left|\dfrac{a_{i,i+1}}{a_{ii}}\right| + ... + \left|\dfrac{a_{in}}{a_{ii}}\right| < 1 \Rightarrow$

$\|S\|_\infty < 1$, т.е. выполняется достаточное условие сходимости.

Сходимость в случае диагонального преобладания по столбцам

Теорема. Если $|a_{jj}| > \sum\limits_{i=1,\ i\neq j}^{n} |a_{ij}| \quad \forall\, j$, то метод Якоби сходится.

Док–во. j–ый столбец матрицы $T = E - AD^{-1}$: $\left[\dfrac{a_{1j}}{a_{jj}}, ..., \dfrac{a_{j-1,j}}{a_{jj}}, 0, \dfrac{a_{j+1,j}}{a_{jj}}, ..., \dfrac{a_{nj}}{a_{jj}}\right]$.

Из условия теоремы $\Rightarrow$ $\left|\dfrac{a_{1j}}{a_{jj}}\right| + ... + \left|\dfrac{a_{j-1,j}}{a_{jj}}\right| + \left|\dfrac{a_{j+1,j}}{a_{jj}}\right| + ... + \left|\dfrac{a_{nj}}{a_{jj}}\right| < 1 \Rightarrow$

$\rho(S) = \rho(E - D^{-1}A) = \rho(DTD^{-1}) = \rho(T) \le \|T\|_1 < 1$, т.е. выполняется необходимое условие сходимости.

Необходимое и достаточное условие сходимости метода Якоби в случае симметричной матрицы с положительной главной диагональю

Теорема.

Если матрица A системы $Ax = b$ самосопряжена: $A = A^*$, а ее диагональ $D \equiv D^* = \operatorname{diag} A = \operatorname{diag}\{a_{11}, a_{22}, \ldots, a_{nn}\}$ положительно определена: $D > 0$, то метод Якоби:

$$\begin{cases} x^0 \ - \text{ задан}, \\ x^{k+1} = x^k - D^{-1}(Ax^k - b), \\ k = 0, 1, \ldots, \end{cases}$$

сходится к решению x системы $Ax = b$ тогда и только тогда, когда положительно определены матрицы A и $2D - A$.

Доказательство. Прежде всего, отметим, что

- для самосопряженной, положительно определенной матрицы D можно построить ее квадратный корень:
$$D^{1/2} = (D^{1/2})^* > 0: \quad D^{1/2}D^{1/2} = D,$$
- собственные значения $\lambda(S) = 1 - \lambda(D^{-1}A)$ матрицы $S = E - D^{-1}A$ совпадают с собственными значениями $\lambda(S) = 1 - \lambda(D^{-1/2}AD^{-1/2})$ матрицы $D^{1/2}SD^{-1/2} = E - D^{-1/2}AD^{-1/2}$, которая самосопряжена ($A = A^*$) и, следовательно, ее собственные значения вещественны:
$$-\rho(S) \le 1 - \lambda(D^{-1/2}AD^{-1/2}) \le \rho(S),$$
- так как необходимым и достаточным условием сходимости метода Якоби является условие $\rho(S) \equiv \rho(E - D^{-1}A) < 1$, то нам нужно найти **необходимые и достаточные условия выполнения неравенств**:
$$-2 < -\lambda(D^{-1/2}AD^{-1/2}) < 0 \Leftrightarrow 0 < \lambda(D^{-1/2}AD^{-1/2}) < 2.$$

Необходимость: метод Якоби сходится, т.е. $\rho(S) < 1$ и, следовательно,

$$0 < \lambda(D^{-1/2}AD^{-1/2}) < 2,$$

т.е. самосопряженные матрицы $D^{-1/2}AD^{-1/2}$ и $2E - D^{-1/2}AD^{-1/2}$ положительно определены:

$$(D^{-1/2}AD^{-1/2}y, y) \ge \lambda_{\min}(D^{-1/2}AD^{-1/2}) \cdot (y, y) > 0 \quad \forall\ y \ne 0 \in R^n,$$
$$([2E - D^{-1/2}AD^{-1/2}]y, y) \ge [2 - \lambda_{\min}(D^{-1/2}AD^{-1/2})] \cdot (y, y) > 0 \quad \forall\ y \ne 0 \in R^n.$$

Тогда $A > 0$, так как

$$\forall\ z \ne 0 \quad (Az, z) = (A\underbrace{D^{-1/2}y}_{z}, \underbrace{D^{-1/2}y}_{z}) = (D^{-1/2}AD^{-1/2}y, y) > 0.$$

Тогда $2D - A > 0$, так как

$$\forall\ z \ne 0 \quad ([2D - A]z, z) = ([2E - D^{-1/2}AD^{-1/2}]\underbrace{D^{1/2}z}_{y}, \underbrace{D^{1/2}z}_{y}) > 0.$$

Достаточность: самосопряженные матрицы A и $2D - A$ положительно определены, следовательно:

$$\lambda_{\min}(A) > 0, \quad \lambda_{\min}(2D - A) > 0.$$

Тогда

положительно определена матрица $D^{-1/2}AD^{-1/2}$, так как

$$\forall\ z \neq 0 \quad (D^{-1/2}AD^{-1/2}z, z) = (A\underbrace{D^{-1/2}z}_{y}, \underbrace{D^{-1/2}z}_{y}) > 0,$$

и, следовательно, $\forall\ \lambda(D^{-1/2}AD^{-1/2}) > 0$;

положительно определена матрица $2E - D^{-1/2}AD^{-1/2}$, так как

$$\forall\ z \neq 0 \quad ([2E - D^{-1/2}AD^{-1/2}]z, z) = ([2D - A]\underbrace{D^{-1/2}z}_{y}, \underbrace{D^{-1/2}z}_{y}) > 0,$$

и, следовательно, $\forall\ \lambda(2E - D^{-1/2}AD^{-1/2}) \equiv 2 - \lambda(D^{-1/2}AD^{-1/2}) > 0$.

Таким образом, $0 < \lambda(D^{-1/2}AD^{-1/2}) < 2$ и, следовательно, любое собственное значение матрицы шага для ошибки метода Якоби меньше 1:

$$|\lambda(S)| = |1 - \lambda(D^{-1/2}AD^{-1/2})| < 1,$$

т.е. выполняется условие сходимости метода: $\rho(S) < 1$.

Пример матрицы $A = D - L - L^* > 0$ & $D > 0$, но матрица $2D - A$ не положительно определена, т.е. метод Якоби не будет сходиться:

$$A = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}, \quad \lambda_1(A) = 2, \ \lambda_2(A) = \lambda_3(A) = 0.5 \ \Rightarrow \ A > 0,$$

$$2D - A = \begin{bmatrix} 1 & -0.5 & -0.5 \\ -0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{bmatrix}, \quad \lambda_1(2D - A) = 0 \ \Rightarrow \ 2D - A \not> 0.$$

Метод Зейделя (Гаусса–Зейделя, Некрасова)

Если матрицу системы $Ax = b$ представить в виде суммы $A = -L + D - R$, где $D = \operatorname{diag} A = \operatorname{diag}\{a_{11}, a_{22}, \ldots, a_{nn}\}$,

$$L = -\begin{bmatrix} 0 & & & & \\ a_{21} & 0 & & 0 & \\ \ddots & \ddots & \ddots & & \\ \ddots & \ddots & \ddots & 0 & \\ a_{n1} & \ddots & \ddots & a_{n,n-1} & 0 \end{bmatrix}, \quad R = -\begin{bmatrix} 0 & a_{12} & \ddots & \ddots & a_{1n} \\ & 0 & \ddots & \ddots & \ddots \\ & & \ddots & \ddots & \ddots \\ & 0 & & 0 & a_{n-1,n} \\ & & & & 0 \end{bmatrix}$$

то итерационный процесс

$$x^{k+1} = x^k - (D - L)^{-1}(Ax^k - b)$$

называется методом Зейделя для решения системы $Ax = b$.

Доказать: $S = E - (D-L)^{-1}A = E - (A+R)^{-1}A = (A+R)^{-1}R$.

Теорема (о необходимом и достаточном условии сходимости метода Зейделя в случае симметричной матрицы с положительной главной диагональю)

<u>Если</u> $A = A^* \quad \& \quad D > 0$,

<u>то</u> $\rho(S) = \rho(E - (D-L)^{-1}A) < 1$ (т.е. метод Зейделя сходится)

<u>тогда и только тогда, когда</u> $A > 0$.

Доказательство.

Необходимость:

Пусть метод Зейделя сходится, т.е. $\rho(S) = \rho(E - (D-L)^{-1}A) < 1$.

Предположим, что $A \not> 0$ (**доказать**, что $\Rightarrow \lambda_{min}(A) < 0$).

Тогда $\exists\ \varphi:\ (A\varphi, \varphi) < 0$ (например, $\varphi:\ A\varphi = \lambda_{min}(A)\cdot\varphi$).

Зададим $z^0 = \varphi$ и оценим (Az^{k+1}, z^{k+1}):

$$(Az^{k+1}, z^{k+1}) = (A\underbrace{(E - [A+R]^{-1}A)z^k}_{z^{k+1}},\ \underbrace{(E - [A+R]^{-1}A)z^k}_{z^{k+1}}) =$$

$$= (Az^k - A\underbrace{[A+R]^{-1}Az^k}_{y^k \neq 0},\ z^k - \underbrace{[A+R]^{-1}Az^k}_{y^k}) =$$

$$= (Az^k, z^k) - (Az^k, y^k) - (Ay^k, z^k) + (Ay^k, y^k) =$$

$$= (Az^k, z^k) - ([A+R]y^k, y^k) - (y^k, [A+R]y^k) + (Ay^k, y^k) =$$

$$= (Az^k, z^k) - ([A+R]y^k, y^k) - ([A+L]y^k, y^k) + (Ay^k, y^k) =$$

$$= (Az^k, z^k) - \underbrace{(Dy^k, y^k)}_{>0} < (Az^k, z^k) < \ldots < (Az^0, z^0) = \text{const} < 0.$$

$\Rightarrow \quad \lim \| z^{k+1} \| \le \text{const} < 0$, метод не сходится, что противоречит $\rho(S) < 1$, следовательно, предположение, что $A \not> 0$, неверно и $A > 0$.

Необходимость утверждения теоремы доказана.

Отметим на будущее, что

$$(Az^{k+1}, z^{k+1}) < (Az^k, z^k), \text{ если } z^k \neq 0.$$

Достаточность.

Нужно доказать, что из условия $A = A^* > 0 \Rightarrow \rho(S) = \rho([A+R]^{-1}R) < 1$.

Во-первых, заметим, что из условия $A = A^* > 0 \Rightarrow$

1. $\lambda_{min}(A) > 0,\ (A\varphi, \varphi) \ge \lambda_{min}(A)\cdot(\varphi, \varphi) \quad \forall\ \varphi$;

2. $d_{min} \equiv \min\limits_{1 \le i \le n} a_{i,i} \ge \lambda_{min}(A)$, т.к.

$$d_{min} = (D\vec{e}_{min}, \vec{e}_{min}) \equiv (A\vec{e}_{min}, \vec{e}_{min}) \ge \lambda_{min}(A) > 0, \text{ т.е. } D > 0.$$

Теперь, пусть $\lambda \in Sp(S)$ – собственное значение, а φ – собственный вектор:

$$S\varphi \equiv [A+R]^{-1}R\varphi = \lambda(S)\cdot\varphi \Rightarrow R\varphi = \lambda(S)\cdot[A+R]\cdot\varphi, \quad \|\varphi\|_2 = 1.$$

Умножим последнее равенство скалярно на собственный вектор:

$$\underbrace{(R\varphi,\ \varphi)}_{r+\underline{i}\cdot\mu} = \lambda(S)\cdot[\ \underbrace{(A\varphi,\ \varphi)}_{a \ge \lambda_{min}(A) > 0} + \underbrace{(R\varphi,\ \varphi)}_{r+\underline{i}\cdot\mu}]$$

тогда

$$|\lambda(S)|^2 \equiv \frac{r^2+\mu^2}{(a+r)^2+\mu^2} \equiv \frac{r^2+\mu^2}{a(a+2r)+r^2+\mu^2} < 1,$$

если $a + 2r = (A\varphi,\ \varphi) + 2\cdot\mathrm{Re}[(R\varphi,\ \varphi)] > 0$.

Действительно,

$$a + 2r = (A\varphi,\ \varphi) + 2\cdot\mathrm{Re}[(R\varphi,\ \varphi)] =$$

$$= \underbrace{(A\varphi,\ \varphi)}_{(D\varphi,\ \varphi)-(R^*\varphi,\ \varphi)-(R\varphi,\ \varphi)} + 2\cdot\mathrm{Re}[(R\varphi,\ \varphi)] =$$

$$= (D\varphi,\ \varphi) \ge d_{min} \ge \lambda_{min}(A) > 0$$

Следовательно,

$$|\lambda(S)|^2 = \frac{r^2+\mu^2}{a(a+2r)+r^2+\mu^2} \le \frac{r^2+\mu^2}{\lambda_{min}^2(A)+r^2+\mu^2} < 1, \text{ т.к. } \lambda_{min}(A) > 0.$$

Теорема доказана, но в дополнение можно вывести оценку $\rho(S)$:

$$\rho(S) \le \max_{\lambda(S)\in Sp(S)} |\lambda(S)|.$$

Т.к. $0 \le r^2 + \mu^2 = |(R\varphi,\varphi)|^2 \le \|R\|_2^2 \cdot \|\varphi\|_2^2 = \|R\|_2^2 \equiv \rho(R^*\cdot R)$, то

$$|\lambda(S)|^2 \le \frac{x}{\lambda_{min}^2(A)+x} \le \max_{0\le x\le\rho(R^*R)} \frac{x}{\lambda_{min}^2(A)+x} = \frac{\rho(R^*R)}{\lambda_{min}^2(A)+\rho(R^*R)}, \text{ т.е.}$$

$$\rho(S) \le \sqrt{\frac{\rho(R^*R)}{\lambda_{min}^2(A)+\rho(R^*R)}} < 1.$$

Лекция 7.

Функционал ошибки

Второй из способов построения итерационного метода решения системы линейных алгебраических уравнений $Ax = b$ ($\det A \neq 0$) состоит из построения последовательности приближений $\{x^k\}_{k=0}^{\infty}$ такой, что $\| z^{k+1} \| < \| z^k \|$, т.е. строгого убывания на каждом шаге **функционала ошибки** $f(x^k - x) \equiv \| z^k \|$, где $z^k = x^k - x$.

Теорема. Если $\| z^{k+1} \| < \| z^k \| \quad \forall\, z^k \neq 0$ и отображение $S: R^n \to R^n$ (оператор шага для ошибки: $z^{k+1} = S(z^k)$) непрерывно при $z \neq 0$, то $\| z^k \| \to 0$, т.е. $x^k \to x$.

Доказательство.

Так как последовательность чисел $\{ \| z^k \| \}_{k=0}^{\infty}$ ограничена снизу 0 и монотонно убывает, то она имеет предел: $\lim\limits_{k\to\infty} \| z^k \| = \alpha \geq 0$.

Если $\alpha = 0$, то теорема доказана.

Предположим, что $\alpha > 0$.

Так как последовательность векторов $\{ z^k \}_{k=0}^{\infty}$ лежит в шаре $\{ z: \| z \| \leq \| z^0 \| \}$ конечномерного пространства R^n, то из нее можно извлечь сходящуюся подпоследовательность $\{ z^{k_m} \}$:

$$z^{k_m} \to z, \quad \| z \| = \alpha > 0.$$

Так как $z^{k_m} \neq 0$ и $z \neq 0$, то к этим векторам применимо отображение S: $\exists \; z^{k_m+1} \equiv S(z^{k_m})$, $\exists \; S(z)$: $\| S(z) \| < \| z \| = \alpha > 0$.

Тогда, выполнив предельный переход в соотношениях

$$\begin{array}{ccccc} \| z^{k_m+1} \| & = & \| S(z^{k_m}) \| & < & \| z^{k_m} \| \\ \downarrow & & \downarrow & & \downarrow \\ \alpha & = & \| S(z) \| & \leq & \| z \|, \end{array}$$

и, учитывая, что $\| S(z) \| < \| z \| = \alpha$, получим противоречие: $\alpha < \alpha$.

Следовательно, предположение $\lim\limits_{k\to\infty} \| z^k \| = \alpha > 0$ неверно и, стало быть, $\lim\limits_{k\to\infty} \| z^k \| = 0$. Теорема доказана.

Метод полной релаксации

Вспомним, что в стационарном итерационном методе Зейделя для решения системы $Ax = b$ компоненты $\mathbf{x_i^{k+1}}$ приближения x^{k+1} определялись последовательно ($i = 1, 2, ..., n$) как решения уравнений

$$a_{i1} \cdot x_1^{k+1} + ... + a_{i,i-1} \cdot x_{i-1}^{k+1} + a_{ii} \cdot \mathbf{x_i^{k+1}} + a_{i,i+1} \cdot x_{i+1}^{k} + ... a_{in} \cdot x_{in}^{k} = b_i,$$

т.е. $\mathbf{x_i^{k+1}} = (a_{ii})^{-1}\{[a_{i1} \cdot x_1^{k+1} + ... + a_{i,i-1} \cdot x_{i-1}^{k+1} + a_{i,i+1} \cdot x_{i+1}^{k} + ... a_{in} \cdot x_{in}^{k}] - b_i\}$.

Если через $x^{k+i/n}$ обозначить вектор

$$x^{k+i/n} = (x_1^{k+1}, \quad ..., \quad x_{i-1}^{k+1}, \quad \mathbf{x_i^{k+1}}, \quad x_{i+1}^{k}, \quad ..., \quad x_n^{k})^T,$$

то очередное приближение x^{k+1} стационарного итерационного метода Зейделя для решения системы $Ax = b$ вычисляется по формулам

$$\begin{cases} x^{k+i/n} = x^{k+(i-1)/n} - \alpha_{k,i} \cdot e_i, \\ \alpha_{k,i} = x^{k+(i-1)/n} - (a_{ii})^{-1}\{[a_{i1} \cdot x_1^{k+1} + ... + a_{i,i-1} \cdot x_{i-1}^{k+1} + a_{i,i+1} \cdot x_{i+1}^{k} + ... a_{in} \cdot x_{in}^{k}] - b_i\} \end{cases}$$

$$i = 1, 2, ..., n,$$

где e_i – i-ый орт.

Но в формуле $x^{k+i/n} = x^{k+(i-1)/n} - \alpha_{k,i} \cdot e_i$ параметр $\alpha_{k,i}$ можно выбирать из условия минимизации функционала ошибки $f(x^{k+i/n} - x) \equiv \| z^{k+i/n} \|$.

Предположим, что матрица системы $Ax = b$ симметрична и положительно определена: $A = A^* > 0$. Определим (энергетические) скалярное произведение и норму в R^n:

$$(x, y)_A = (Ax, y), \ \| x \|_A = \sqrt{(x, x)_A},$$

и параметр $\alpha_{k,i}$ будем выбирать из условия минимума $\| z^{k+i/n} \|_A$.

Теорема. Если матрица системы $Ax = b$ симметрична и положительно определена: $A = A^* > 0$, а параметры $\alpha_{k,i}$ итерационного метода (метод полной релаксации)

$x^0 \in R^n$ – задан,

$$\begin{cases} x^{k+i/n} = x^{k+(i-1)/n} - \alpha_{k,i} \cdot e_i, \\ \alpha_{k,i} : \ \| z^{k+i/n} \|_A = \inf\limits_{\alpha} \| z^{k+(i-1)/n} - \alpha \cdot e_i \|_A, \ i = 1, 2, ..., n, \end{cases}$$

$k = 0, 1, ...,$

то

1) $\alpha_{k,i} = (a_{ii})^{-1}[r^{k+(i-1)/n}]_i$,

2) $\| z^{k+1} \|_A < \| z^k \|_A$, если $z^k \neq 0$,

3) метод полной релаксации можно записать в виде

$x^0 \in R^n$ – задан,

$$\begin{cases} x_i^{k+1} = x_i^k - (a_{ii})^{-1}[a_{i,1}x_1^{k+1} + \ldots + a_{i,i-1}x_{i-1}^{k+1} + a_{i,i}x_i^k + \ldots + a_{i,n}x_n^k - b_i], \\ i = 1, 2, \ldots, n, \end{cases}$$

$k = 0, 1, \ldots,$

или ($A = D - L - R$)

$$x^0 \in R^n \text{ – задан,}$$

$$x^{k+1} = x^k - D^{-1}[-Lx^{k+1} + (D-R)x^k - b] \equiv$$

$$\equiv x^k - (D-L)^{-1}[Ax^k - b],$$

$$k = 0, 1, \ldots,$$

(т.е. метод полной релаксации совпадает с методом Зейделя!),

4) отображение $S \equiv E - (D-L)^{-1}A: \quad R^n \to R^n$ (оператор шага для ошибки: $z^{k+1} = S(z^k)$) непрерывно и из теоремы о строгом убывании функционала ошибки следует сходимость метода полной релаксации.

Доказательство.

1) Решим задачу $\alpha_{k,i}: \ \| z^{k+i/n} \|_A = \inf_{\alpha} \| z^{k+(i-1)/n} - \alpha \cdot e_i \|_A$.

Так как

$$\| z^{k+(i-1)/n} - \alpha \cdot e_i \|_A^2 = (A[z^{k+(i-1)/n} - \alpha \cdot e_i], [z^{k+(i-1)/n} - \alpha \cdot e_i]) =$$

$$= \| z^{k+(i-1)/n} \|_A^2 - 2\alpha \cdot \underbrace{(Az^{k+(i-1)/n}, e_i)}_{[r^{k+(i-1)/n}]_i} + \alpha^2 \cdot \underbrace{(Ae_i, e_i)}_{=a_{ii}>0} \equiv$$

$$\equiv \| z^{k+(i-1)/n} \|_A^2 - \frac{([r^{k+(i-1)/n}]_i)^2 - (\alpha \cdot a_{i,i} - [r^{k+(i-1)/n}]_i)^2}{a_{i,i}},$$

то, очевидно, что при $\alpha \cdot a_{i,i} - [r^{k+(i-1)/n}]_i = 0$ будет максимальное уменьшение ошибки (полная релаксация):

$$\| z^{k+i/n} \|_A^2 = \| z^{k+(i-1)/n} \|_A^2 - \frac{(r_i^{k+(i-1)/n})^2}{a_{i,i}},$$

т.е. $\alpha_{k,i} = (a_{ii})^{-1}[r^{k+(i-1)/n}]_i$.

2) Теперь докажем, что $\| z^{k+1} \|_A < \| z^k \|_A$, если $z^k \neq 0$.

Действительно, так как

$$\| z^{k+1} \|_A^2 = \| z^k \|_A^2 - \frac{(r_1^k)^2}{a_{1,1}} - \frac{(r_2^{k+1/n})^2}{a_{2,2}} - \ldots - \frac{(r_n^{k+(n-1)/n})^2}{a_{n,n}}$$

и хотя бы одна из компонент $r_i^{k+(i-1)/n} \neq 0$ (в противном случае $x^{k+i/n} = x^{k+(i-1)/n} = x^k$, $r^{k+i/n} = r^k$, т.е. $r^k = 0 \Rightarrow z^k = 0$), то функционал ошибки строго убывает.

3) Выпишем формулы реализации метода полной релаксации. Так как

$$x^{k+i/n} = (x_1^{k+1}, \quad \ldots, \quad x_{i-1}^{k+1}, \quad \mathbf{x_i^{k+1}}, \quad x_{i+1}^k, \quad \ldots, \quad x_n^k)^T =$$
$$= x^{k+(i-1)/n} - \alpha_{k,i} \cdot \mathbf{e_i}, \ \ \alpha_{k,i} = (a_{ii})^{-1}[r^{k+(i-1)/n}]_i,$$
$$[r^{k+(i-1)/n}]_i = [Ax^{k+(i-1)/n} - b]_i = [-Lx^{k+1} + Dx^k - Rx^k - b]_i,$$

то

$$x_i^{k+1} \equiv x_i^{k+i/n} = x_i^{k+(i-1)/n} - (a_{ii})^{-1}[r^{k+(i-1)/n}]_i =$$
$$= x_i^{k+(i-1)/n} - (a_{ii})^{-1}[a_{i,1}x_1^{k+1} + \ldots + a_{i,i-1}x_{i-1}^{k+1} + a_{i,i}x_i^k + \ldots + a_{i,n}x_n^k - b_i],$$

или

$$x^{k+1} = x^k - D^{-1}[-Lx^{k+1} + Dx^k - Rx^k - b] =$$
$$= D^{-1}\{Dx^k + Lx^{k+1} - (A+L)x^k + b]\} =$$
$$= D^{-1}\{Lx^{k+1} + (D-L)x^k - Ax^k + b]\}$$
$$\Rightarrow$$
$$x^{k+1} = x^k - (D-L)^{-1}[Ax^k - b].$$

4) Очевидно, что мы получили стационарный итерационный метод с матрицей шага для ошибки $S \equiv E - (D-L)^{-1}A$, которая является непрерывным отображением из R^n в R^n, и из теоремы о строгом убывании функционала ошибки следует сходимость метода полной релаксации.

Метод неполной релаксации

Изменим метод полной релаксации решения системы $Ax = b$ с матрицей $A = A^* > 0$, введя вещественный параметр ω:

$x^0 \in R^n$ – задан,

$$\begin{cases} x^{k+i/n} = x^{k+(i-1)/n} - \omega \cdot \alpha_{k,i} \cdot e_i, \\ \alpha_{k,i} = (a_{ii})^{-1}[r^{k+(i-1)/n}]_i : \ \| z^{k+i/n} \|_A = \inf\limits_{\alpha} \| z^{k+(i-1)/n} - \alpha \cdot e_i \|_A, \ i = 1, 2, ..., n, \end{cases}$$

$k = 0, 1, ...,$

т.е. на каждом дробном шаге энергетическая норма ошибки уменьшается меньше, чем в методе полной релаксации ($\omega = 1$):

$$\| z^{k+i/n} \|_A^2 = \| z^{k+(i-1)/n} \|_A^2 - \frac{(r_i^{k+(i-1)/n})^2 - (\omega \cdot \alpha_{k,i} \cdot a_{i,i} - r_i^{k+(i-1)/n})^2}{a_{i,i}} =$$

$$= \| z^{k+(i-1)/n} \|_A^2 - [1 - (\omega - 1)^2] \frac{(r_i^{k+(i-1)/n})^2}{a_{ii}}.$$

Очевидно, что норма ошибки строго уменьшится, если $r_i^{k+(i-1)/n} \neq 0$ и

$$0 < 1 - (\omega - 1)^2, \text{ т.е. } \omega \in (0, 2).$$

Расчетные формулы вычисления компонент очередного приближения x^{k+1} по заданному приближению x^k имеют вид:

$$x_i^{k+1} = x_i^k - \alpha_{k,i} = x_i^k - \omega \frac{a_{i,1}x_1^{k+1} + ... + a_{i,i-1}x_{i-1}^{k+1} + a_{i,i}x_i^k + ... + a_{i,n}x_n^k - b_i}{a_{ii}}$$

$i = 1, 2, ..., n,$

а матричная запись метода неполной релаксации имеет вид:

$$x^{k+1} = x^k - \omega D^{-1}(-Lx^{k+1} + (D - R)x^k - b)$$

$$\Rightarrow \ (D - \omega L)x^{k+1} = Dx^k - \omega Dx^k + \omega(Rx^k + b) =$$

$$= Dx^k - \omega Lx^k + \omega Lx^k - \omega Dx^k + \omega(Rx^k + b) =$$

$$= (D - \omega L)x^k - \omega(Ax^k - b)$$

$$\Rightarrow x^{k+1} = x^k - \omega(D - \omega L)^{-1}(Ax^k - b),$$

с матрицей шага для ошибки $S \equiv E - \omega(D - \omega L)^{-1}A$, которая является непрерывным отображением из R^n в R^n.

Из теоремы о строгом убывании функционала ошибки следует сходимость метода неполной релаксации.

Теорема. Если матрица системы $Ax = b$ симметрична и положительно определена: $A = A^* > 0$, то для любого $\omega \in (0, 2)$ метод неполной релаксации:

$$x^0 \in R^n \text{ – задан,}$$
$$x^{k+1} = x^k - \omega(D - \omega L)^{-1}(Ax^k - b),$$
$$k = 0, 1, ...,$$

сходится, а энергетическая норма его ошибки убывает:

$$\| z^{k+1} \|_A < \| z^k \|_A, \text{ если } z^k \neq 0.$$

Оценка сходимости методов релаксации

Предварительные замечания

Итак, ошибка $z^k = x^k - x$ метода релаксации

$$x^0 \in R^n \text{ задан,}$$
$$x^{k+1} = x^k - \omega(D - \omega L)^{-1}(Ax^k - b)$$
$$k = 0, 1, ...$$

строго монотонно убывает в норме $\| z \|_A = \sqrt{(Az, z)}$.

Т.к. $A = -L + D - R = (-L + 0.5 \cdot D) + (0.5 \cdot D - R) \equiv R_1 + R_1^*$ и $R_1 > 0$ в R^n, то “переобусловливатель” $D - \omega L$ метода релаксации целесообразно представить в виде суммы положительно определьных матриц:

$$D - \omega L = (1 - 0.5\omega)D + \omega(0.5 \cdot D - L) = (1 - 0.5\omega)D + \omega R_1 =$$

$$= (1 - 0.5\omega)(D + \frac{\omega}{(1 - 0.5\omega)} R_1) \equiv \frac{\omega}{\tau}(D + \tau \cdot R_1) \equiv \frac{\omega}{\tau} B > 0$$

где $\tau = \dfrac{\omega}{1 - 0.5\omega} \in (0, \infty)$ при $\omega \in (0, 2)$.

Тогда метод релаксации может быть переписан в виде

$$x^0 \in R^n \text{ задан,}$$
$$x^{k+1} = x^k - \tau \cdot B^{-1}(Ax^k - b), \; B = D + \tau \cdot R_1 > 0 \;\; \forall \; \tau > 0,$$
$$k = 0, 1, ...$$

и он сходится при любом $\tau > 0$.

Очевидно, что

$$z^{k+1} = [E - \tau \cdot B^{-1}A]z^k \equiv Sz^k$$

Замечание. Мы знаем, что для любого вектора $z \neq 0$

$$\| Sz \|_A < \| z \|_A \Rightarrow \| S \|_A = \sup_{\|z\|_A = 1} \| Sz \|_A = \| Sz_{\sup} \|_A < 1.$$

Здесь мы воспользовались теоремой о достижении максимального значения непрерывной функции на сфере конечномерного пространства.
Это замечание является еще одним доказательством сходимости метода релаксации, следовательно $\| S \|_A < 1$, а параметр τ нужно выбирать из условия минимальности $\| S \|_A = \| E - \tau \cdot B^{-1} A \|_A$ или верхней оценки для этой нормы.

Оценим $\| S \|_A^2 = \max\limits_{z \neq 0} \dfrac{(ASz, Sz)}{(Az, z)}$, где $S = E - \tau B^{-1} A$.

$$\Rightarrow \; \| S \|_A^2 = \max_{z \neq 0} \left[1 - 2\tau \frac{(B^{-1}Az, Az)}{(Az, z)} + \tau^2 \frac{(AB^{-1}Az, B^{-1}Az)}{(Az, z)} \right].$$

Т.к. $\tau(Ay, y) = \tau([R_1 + R_1^*]y, y) = 2(\tau R_1 y, y) = 2([B - D]y, y)$, то

$$\tau^2 (A[B^{-1}Az], [B^{-1}Az]) = 2\tau (B[B^{-1}Az], [B^{-1}Az]) - 2\tau(D[B^{-1}Az], [B^{-1}Az])$$

$$= 2\tau (B^{-1}Az, Az) - 2\tau(D B^{-1}Az, B^{-1}Az)$$

$$\Rightarrow \; \| S \|_A^2 = \max_{z \neq 0} \left[1 - 2\tau \frac{(DB^{-1}Az, B^{-1}Az)}{(Az, z)} \right] = 1 - 2\tau \cdot \min_{z \neq 0} \frac{(DB^{-1}Az, B^{-1}Az)}{(Az, z)} =$$

$$= 1 - 2\tau \cdot \min_{z \neq 0} \frac{([A^{1/2}(B^{-1})^* DB^{-1} A^{1/2}] A^{1/2} z, A^{1/2} z)}{(A^{1/2} z, A^{1/2} z)} = 1 - 2\tau \gamma,$$

где $\gamma = \lambda_{\min}(A^{1/2}(B^{-1})^* DB^{-1} A^{1/2})$.

Пусть $A^{1/2}(B^{-1})^* DB^{-1} A^{1/2} y = \gamma \cdot y \quad \Rightarrow \quad Av = \gamma \cdot BD^{-1}B^* v, \quad v = A^{-1/2} y$.
Т.к. $BD^{-1}B^* = (D + \tau R_1) D^{-1} (D + \tau R_1^*) = D + \tau A + \tau^2 R_1 D^{-1} R_1^*$,
то

$$\gamma = \frac{(Av, v)}{(Dv, v) + \tau(Av, v) + \tau^2 (R_1 D^{-1} R_1^* v, v)} \geq$$

$$\geq \frac{1}{\max\limits_{v \neq 0} \dfrac{(Dv, v)}{(Av, v)} + \tau + \tau^2 \max\limits_{v \neq 0} \dfrac{(R_1 D^{-1} R_1^* v, v)}{(Av, v)}} = \frac{1}{\rho(A^{-1}D) + \tau + \tau^2 \rho(A^{-1} R_1 D^{-1} R_1^*)}$$

$$\geq \frac{1}{\beta(A^{-1}D) + \tau + \tau^2 \beta(A^{-1} R_1 D^{-1} R_1^*)}, \text{ где } \begin{cases} \beta(A^{-1}D) \geq \rho(A^{-1}D), \\ \beta(A^{-1} R_1 D^{-1} R_1^*) \geq \rho(A^{-1} R_1 D^{-1} R_1^*). \end{cases}$$

Теорема.

Если известны оценки $\begin{cases} 1/\delta \ge \rho(A^{-1}D) \\ \Delta \ge \rho(A^{-1}R_1D^{-1}R_1^*) \end{cases}$,

то

$$\|S\|_A^2 = 1 - 2\tau\gamma \le 1 - 2\tau\frac{1}{1/\delta + \tau + \tau^2\Delta} = \frac{1 - \tau\delta + \tau^2\delta\Delta}{1 + \tau\delta + \tau^2\delta\Delta} < 1 \quad \forall\ \tau > 0.$$

Док–во очевидно.

Замечание. Обычно постоянные $\delta > 0$ и $\Delta > 0$ определяют из неравенств

$$\delta \cdot (Dv, v) \le (Av, v), \quad (R_1D^{-1}R_1^*v, v) \le \Delta \cdot (Av, v) \quad \forall\ v$$

$(\delta D \le A, \quad \delta \le \lambda_{min}(D^{-1}A),\ R_1D^{-1}R_1^* \le \Delta \cdot A, \quad \Delta \ge \lambda_{max}(A^{-1}R_1D^{-1}R_1^*))$.

Задача. Найти минимум функции

$$g(\tau) = \frac{1 - \tau\delta + \tau^2\delta\Delta}{1 + \tau\delta + \tau^2\delta\Delta} \ge \|S\|_A^2$$

Ответ. $\min\limits_{\tau>0} g(\tau) = g(\tau_*) = \dfrac{1 - \sqrt{\delta/(4\Delta)}}{1 + \sqrt{\delta/(4\Delta)}}, \qquad \tau_* = \dfrac{1}{\sqrt{\delta\Delta}}$.

Задача. Найти оптимальный параметр ω_*, если $\tau_* = 1/\sqrt{\delta\Delta}$

Ответ. $\omega_* = \dfrac{2}{1 + 2\sqrt{\delta\Delta}}$.

Пример

$$A = \begin{bmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & & & 0 \\ -1 & 1 & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 1 & 0 \\ 0 & & & -1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & -1 & & & 0 \\ 0 & 1 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & 0 & 1 & -1 \\ 0 & & & 0 & 1 \end{bmatrix} = R_1 + R_1^*$$

Собственные значения матрицы A известны: $\lambda_k = 4 \cdot \sin^2 \dfrac{k \cdot \pi}{2(n+1)}$, $k = 1, \ldots, n$,

$$\lambda_{min}(A) = 4 \cdot \sin^2 \frac{\pi}{2(n+1)} \approx \frac{\pi^2}{(n+1)^2},\ \lambda_{max}(A) = 4 \cdot \sin^2 \frac{n \cdot \pi}{2(n+1)} \approx 4 - \frac{\pi^2}{(n+1)^2}.$$

Оценим постоянные $\delta \cdot D \le A$, $R_1D^{-1}R_1^* \le \Delta \cdot A$:

$$\delta = \lambda_{min}(D^{-1}A) = 0.5 \cdot \lambda_{min}(A) \approx \frac{\pi^2}{2(n+1)^2} \ll 1,$$

т.к. $A = 2 \cdot R_1D^{-1}R_1^* + \mathrm{diag}\{1, 0, \ldots, 0\}$, а $\mathrm{diag}\{1, 0, \ldots, 0\} \ge 0$, то

$$A \ge 2 \cdot R_1D^{-1}R_1^* \text{ и } \Delta = 0.5.$$

Для метода полной релаксации ($\omega = 1,\ \tau = \dfrac{\omega}{1-0.5\omega} = 2$):

$$x^{k+1} = x^k - (D-L)^{-1}(Ax^k - b)$$

$$\| z^{k+1} \|_A \le \| S \|_A \cdot \| z^k \|_A \le \sqrt{\frac{1-2\delta+4\delta\Delta}{1+2\delta+4\delta\Delta}} \cdot \| z^k \|_A$$

$$= \sqrt{\frac{1}{1+4\delta}} \cdot \| z^k \|_A \approx (1-2\delta) \cdot \| z^k \|_A \approx (1-\frac{\pi^2}{(n+1)^2}) \cdot \| z^k \|_A$$

Для метода неполной релаксации с “оптимальным” параметром $\omega_* = \dfrac{2}{1+2\sqrt{\delta\Delta}}$,

$\tau_* = \omega_* /(1-0.5\omega_*) = 1/\sqrt{\delta\Delta}$:

$$\| z^{k+1} \|_A \le \| S \|_A \cdot \| z^k \|_A \le \sqrt{\frac{1-\tau_*\delta+\tau_*^2\delta\Delta}{1+\tau_*\delta+\tau_*^2\delta\Delta}} \cdot \| z^k \|_A \le \sqrt{\frac{1-\sqrt{\delta/(4\Delta)}}{1+\sqrt{\delta/(4\Delta)}}} \cdot \| z^k \|_A$$

$$= \sqrt{\frac{1-\sqrt{\delta/2}}{1+\sqrt{\delta/2}}} \cdot \| z^k \|_A \approx (1-\sqrt{\delta/2}) \cdot \| z^k \|_A \approx (1-\frac{\pi}{2(n+1)}) \cdot \| z^k \|_A$$

Для метода Якоби

$$x^{k+1} = x^k - D^{-1}(Ax^k - b)$$

$$\| z^{k+1} \|_A \le \| S \|_A \cdot \| z^k \|_A =$$

$$= \max | 1 - 0.5 \cdot \lambda(A) | \cdot \| z^k \|_A =$$

$$\approx (1 - 0.5 \cdot \frac{\pi^2}{(n+1)^2}) \cdot \| z^k \|_A =$$

так как

$$\| S \|_A^2 = \sup \frac{(A(E-D^{-1}A)z, (E-D^{-1}A)z)}{(Az,z)} = \sup \frac{\| (E-A^{1/2}D^{-1}A^{1/2})A^{1/2}z \|^2}{(A^{1/2}z, A^{1/2}z)} =$$

$$= \rho^2(E-A^{1/2}D^{-1}A^{1/2}) = \rho^2(E-D^{-1}A) = \rho^2(E-0.5\cdot A)$$

Тогда

верхняя релаксация	$\tau_* = \dfrac{1}{\sqrt{\delta\Delta}} \approx 2(n+1)/\pi, \quad \omega_* = \dfrac{2}{1+\pi/(n+1)} > 1,$ $g(\tau_*) \approx 1 - \dfrac{\pi}{n+1},\ \| S \|_A \le \sqrt{1-\dfrac{\pi}{n+1}} \approx 1 - \dfrac{\pi}{2(n+1)}$, $k(\varepsilon) = \dfrac{2(n+1)}{\pi} \cdot \ln\dfrac{1}{\varepsilon}$

полная релаксация	$g(2) \approx 1 - \frac{2\pi^2}{(n+1)^2}$, $\Vert S \Vert_A \le \sqrt{g(2)} \approx 1 - \frac{\pi^2}{(n+1)^2}$, $k(\varepsilon) = \frac{(n+1)^2}{\pi^2} \cdot \ln \frac{1}{\varepsilon}$
метод Якоби	$\Vert S \Vert_A = \max \vert 1 - 0,5 \cdot \lambda(A) \vert \approx 1 - \frac{\pi^2}{2(n+1)^2}$ $k(\varepsilon) = \frac{2(n+1)^2}{\pi^2} \cdot \ln \frac{1}{\varepsilon}$

Лекция 8.

Градиент, метод наискорейшего спуска

Как выбирать вектор y при построении итерационного метода $x^{k+1} = x^k + \alpha \cdot y$ из условия минимизации ошибки:

$$\| z^{k+1} \|^2 = \min_{\alpha} \| z^k + \alpha \cdot y \|^2 ?$$

Градиент функции $f(z) : R^n \to R$

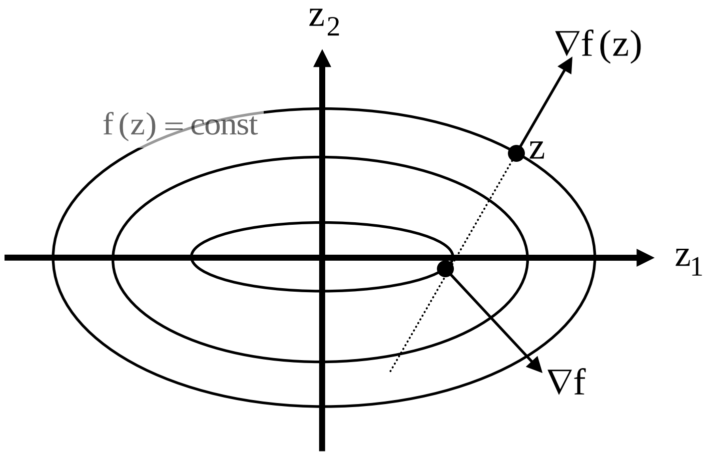
$$f(z + \alpha \cdot y) = f(z) + \left. \frac{d f(z + \alpha \cdot y)}{d\alpha} \right|_{\alpha = 0} \cdot \alpha + O(\alpha^2) =$$

$$= f(z) + \left[\frac{\partial f(z)}{\partial z_1} y_1 + \ldots + \frac{\partial f(z)}{\partial z_n} y_n \right] \cdot \alpha + O(\alpha^2) =$$

$$= f(z) + (\nabla f, y) \cdot \alpha + O(\alpha^2) \approx f(z) + (\nabla f, y) \cdot \alpha.$$

Так как $(\nabla f, y) = \cos(\nabla f, y) \cdot \| \nabla f \| \cdot \| y \|$ принимает минимальное или максимальное значения, когда вектора ∇f, y параллельны, то очевидно, что

- $y = \nabla f$ – направление максимального возрастания $f(z)$,
- $y = -\nabla f$ – направление максимального убывания $f(z)$.



Пример спуска по антиградиенту.

Градиент функции $f(z) = \| z \|_C^2 \ : R^n \to R$

Если $f(z) \equiv \| z \|_C^2 = (Cz, z), \quad C = C^* > 0$, то

$$\left. \begin{aligned} f(z + \alpha \cdot y) &= f(z) + (\nabla f, y) \cdot \alpha + O(\alpha^2) = \\ &= f(z) + (Cz, y) \cdot 2\alpha + \alpha^2 \cdot (Cy, y) \end{aligned} \right\} \Rightarrow \nabla f(z) = 2Cz$$

Следовательно, итерационный процесс имеет вид ($y = -0.5 \cdot Cz$)

$$x^{k+1} = x^k - \alpha_k \cdot Cz^k, \quad z^{k+1} = z^k - \alpha_k \cdot Cz^k,$$

$$\alpha_k : \| z^{k+1} \|_C^2 = \min_\alpha [\| z^k \|_C^2 - 2\alpha \cdot (Cz^k, Cz^k) + \alpha^2 \cdot (CCz^k, Cz^k)],$$

Так как

$$\alpha(z^k) \equiv \alpha_k = \frac{(Cz^k, z^k)}{(CCz^k, Cz^k)} \Rightarrow \| z^{k+1} \|_C^2 = \| z^k \|_C^2 - \frac{(Cz^k, Cz^k)^2}{(CCz^k, Cz^k)} < \| z^k \|_C^2$$

и отображение $S(z) = z - \alpha(z) \cdot Cz$ непрерывно при $z \neq 0$, то по теореме о строгом убывании функционала итерационный процесс сходится.

Выбор матрицы C

Очевидно, что условием реализуемости метода является вычислимость вектора $Cz^k = Cx^k - Cx$. Если $C = HA = C^* > 0$ (H – заданная матрица), то вектор $Cz^k = H(Ax^k - b) \equiv Hr^k$ вычислим.

Метод наискорейшего спуска

Если матрица системы $Ax = b$ симметрична и положительно определена, то, выбирая $C = A$, получаем метод наискорейшего спуска

$$x^0 \text{ – задан,}$$

$$x^{k+1} = x^k - \tau_k (Ax^k - b), \quad \tau_k = \frac{(r^k, r^k)}{(Ar^k, r^k)},$$

$$k = 0, 1, \ldots$$

энергетическая норма $\| z^k \|_A$ ошибки которого строго убывает.

Метод минимальных невязок

В итерационном процессе $x^{k+1} = x^k - \tau_k (Ax^k - b)$ параметр τ_k будем выбирать из условия минимизации невязки:

$$(r^{k+1}, r^{k+1}) = \min_\tau (r^k - \tau Ar^k, \ r^k - \tau Ar^k).$$

Теорема. Если матрица вещественной системы $Ax = b$ положительно определена, то метод минимальных невязок

$$x^0 \text{ – задан,}$$

$$x^{k+1} = x^k - \tau_k (Ax^k - b), \quad \tau_k = \frac{(Ar^k, r^k)}{(Ar^k, Ar^k)},$$

$$k = 0, 1, \ldots$$

сходится, а норма $\| r^k \|_2 = \sqrt{(r^k, r^k)}$ его невязок строго убывает.

Док-во. Во-первых; $\min\limits_{\tau} (r^k - \tau A r^k,\ r^k - \tau A r^k) =$

$$= \min_{\tau} [(r^k, r^k) - \tau \cdot (r^k, Ar^k) - \tau \cdot (Ar^k, r^k) + \tau^2 \cdot (Ar^k, Ar^k)]$$

$$= \min_{\tau} [(r^k, r^k) - 2\tau \cdot (r^k, Ar^k) + \tau^2 \cdot (Ar^k, Ar^k)]$$

достигается при $\tau_k = (Ar^k, r^k)/(Ar^k, A\,r^k)$ и

$$\| r^{k+1} \|_2^2 = \| r^k \|_2^2 - \frac{|(Ar^k, r^k)|^2}{(Ar^k, A\,r^k)} < \| r^k \|_2^2,$$

если $r^k \neq 0$ & $(Ar^k, r^k) \neq 0$, то. невязка строго убывает при $A > 0$.

Во-вторых; из строгого убывания нормы невязки $r^k = Az^k$ следует

$$\| z^{k+1} \|_{A^*A}^2 \equiv (r^{k+1}, r^{k+1}) < (r^k, r^k) \equiv \| z^k \|_{A^*A}^2,$$

т.е. A^*A-норма ошибки строго убывает.

Кроме того, очевидно, что оператор S:

$$z^{k+1} = S(z^k) = z^k - \tau_k(z^k) \cdot Az^k$$

непрерывен всюду за (исключением, быть может, 0) и по теореме о строгом убывании функционала итерационный процесс сходится.

Метод простой итерации

В методах наискорейшего спуска и минимальных невязок для определения параметра τ_k на каждом шаге нужно вычислять два скалярных произведения (с умножением невязки на матрицу системы). Использование постоянного параметра $\tau_k \equiv \tau$ существенно уменьшает объем вычислений на каждом шаге.

Предварительные замечания

Выясним, при каких условиях стационарный итерационный метод

$$x^0 \text{ – задан,}$$

$$x^{k+1} = x^k - \tau \cdot (Ax^k - b),$$

$$k = 0, 1, \ldots$$

сходится к решению системы $Ax = b$.

Мы знаем, что необходимым и достаточным условием сходимости стационарного итерационного метода является условие $\rho(S) = \rho(E - \tau \cdot A) < 1$.

Если $\lambda(A)$ – собственное значение матрицы A, то $\lambda(S) = 1 - \tau \cdot \lambda(A)$ и

$$|\lambda(S)|^2 = [1 - \tau \cdot \mathrm{Re}\,\lambda(A)]^2 + \tau^2 \cdot [\mathrm{Im}\,\lambda(A)]^2 =$$

$$= 1 - 2\tau \cdot \mathrm{Re}\,\lambda(A) + \tau^2 \cdot |\lambda(A)|^2$$

Очевидно, что, если $\tau \cdot \mathrm{Re}\,\lambda(A) \leq 0$, то $|\lambda(S)| \geq 1$ и метод не сходится.

Следовательно, необходимым условием сходимости метода является условие

$$\tau \cdot \operatorname{Re}\lambda(A) > 0 \quad \forall\ \lambda \in \operatorname{Sp} A$$

или знакоопределенность $\operatorname{Re}\lambda(A)\ \ \forall\ \lambda \in \operatorname{Sp} A$.

Пусть $\operatorname{Re}\lambda(A) > 0\ \ \forall\ \lambda \in \operatorname{Sp} A$. Тогда

$$\max_{\lambda(S)\in \operatorname{Sp} S} |\lambda(S)|^2 = \max_{\lambda(A)\in \operatorname{Sp} A} |1-\tau\cdot\lambda(A)|^2 \ < \ 1, \text{ если}$$

$$0 < \tau < \min_{\lambda(A)\in \operatorname{Sp} A} \frac{2\cdot \operatorname{Re}\lambda(A)}{|\lambda(A)|^2}.$$

Метод простой итерации

Теорема. Если матрица системы $Ax = b$ симметрична и положительно определена, то метод простой итерации (Ричардсона)

$$x^0 \ - \text{ задан},$$
$$x^{k+1} = x^k - \tau\cdot(Ax^k - b),$$
$$k = 0, 1, \ldots$$

сходится при $\forall\ \tau \in (0, \dfrac{2}{\rho(A)})$.

Док-во. Так как $A = A^* > 0$, то $\lambda(A) = \operatorname{Re}\lambda(A) > 0$ и (как следует из предварительных замечаний) метод сходится, если

$$0 < \tau < \min_{\lambda(A)\in \operatorname{Sp} A} \frac{2\cdot \operatorname{Re}\lambda(A)}{|\lambda(A)|^2} = \min_{\lambda(A)\in \operatorname{Sp} A} \frac{2}{|\lambda(A)|} = \frac{2}{\rho(A)}.$$

Оптимальный выбор параметра τ метода Ричардсона

Пусть границы спектра матрицы $A = A^* > 0$ известны:

$$0 < \lambda_{\min}(A) \le \lambda(A) \le \lambda_{\max}(A) = \rho(A)$$

Определим $\tau_{\text{опт}}$:

$$\rho(S_{\tau_{\text{опт}}}) = \min_{\tau} \rho(S_\tau)\,| = \min_{\tau} \max_{\lambda(A)\in \operatorname{Sp} A} |1-\tau\cdot\lambda(A)| \le \min_{\tau} \max_{\lambda_{\min}\le\lambda\le\lambda_{\max}} |1-\tau\cdot\lambda| =$$

$$= \min_{\tau} \max\{|1-\tau\cdot\lambda_{\min}(A)|, |1-\tau\cdot\lambda_{\max}(A)|\} =$$

$$= |1-\tau_{\text{опт}}\cdot\lambda_{\min}(A)| = |1-\tau_{\text{опт}}\cdot\lambda_{\max}(A)|$$

$$\Rightarrow \qquad \tau_{\text{опт}} = \frac{2}{\lambda_{\min}+\lambda_{\max}}, \qquad \rho(S_{\tau_{\text{опт}}}) = \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}} < 1.$$

Теорема. Если матрица системы $Ax = b$ симметрична и положительно определена, то для ошибки метода простой итерации (Ричардсона) с оптимальным параметром $\tau_{\text{опт}} = 2/[\lambda_{\max}(A)+\lambda_{\min}(A)]$ справедливы оценки

$$\| z^k \|_2 \le \rho^k(S_{\tau_{опт}}) \cdot \| z^0 \|_2,$$

$$\| z^k \|_A \le \rho^k(S_{\tau_{опт}}) \cdot \| z^0 \|_A, \text{ где } \rho(S_{\tau_{опт}}) = \frac{\lambda_{max}(A) - \lambda_{min}(A)}{\lambda_{max}(A) + \lambda_{min}(A)} < 1.$$

Док-во. Так как $A = A^*$, то $S_\tau = S_\tau^*$ и, следовательно, $\| S_\tau \|_2 = \rho(S_\tau)$.

Далее,

$$\| S_\tau \|_A^2 = \max_{z \ne 0} \frac{(AS_\tau z, S_\tau z)}{(Az, z)} = \max_{z \ne 0} \frac{(S_\tau[A^{1/2}z], S_\tau[A^{1/2}z])}{([A^{1/2}z],[A^{1/2}z])} = \| S_\tau \|_2^2 = \rho^2(S_\tau),$$

а из неравенств $\| z^k \| \le \| S_\tau \| \cdot \| z^{k-1} \|$ и вышеизложенного следуют оценки теоремы.

Оценки сходимости МНС и ММН

Теорема. Если $A = A^* > 0$, то для ошибки z^k метода наискорейшего спуска:

$$x^{k+1} = x^k - \tau_k (Ax^k - b), \quad \tau_k = \frac{(r^k, r^k)}{(Ar^k, r^k)}, \quad k = 0, 1, \dots,$$

справедливы оценки:

$$\| z^k \|_A \le \left[\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right]^k \| z^0 \|_A, \| z^k \|_2 \le \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \left[\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right]^k \| z^0 \|_2$$

Док-во. Так как

$$\| z^{k+1} \|_A = \inf_\tau \| z^k - \tau A z^k \|_A \le \| S_{\tau_{опт}} z^k \|_A \le$$

$$\le \| S_{\tau_{опт}} \|_A \cdot \| z^k \|_A = \rho(S_{\tau_{опт}}) \| z^k \|_A$$

то $\| z^k \|_A \le [\rho_{опт}]^k \| z^0 \|_A$, где

$$\rho_{опт} = \rho(S_{\tau_{опт}}) = \frac{\lambda_{max}(A) - \lambda_{min}(A)}{\lambda_{max}(A) + \lambda_{min}(A)}.$$

Так как $\lambda_{min}(z, z) \le (Az, z) \le \lambda_{max}(z, z)$ и, следовательно,

$$\sqrt{\lambda_{min}} \| z \|_2 \le \| z \|_A \le \sqrt{\lambda_{max}} \| z \|_2,$$

то $\| z^k \|_2 \le \sqrt{\lambda_{max} / \lambda_{min}} \, [\rho_{опт}]^k \| z^0 \|_A$.

Теорема. Если $A = A^* > 0$, то для ошибки z^k метода минимальных невязок:

$$x^{k+1} = x^k - \tau_k (Ax^k - b), \quad \tau_k = \frac{(Ar^k, r^k)}{(Ar^k, Ar^k)}, \quad k = 0, 1, \dots,$$

справедливы оценки:

$$\| r^k \|_2 \le \left[\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right]^k \| r^0 \|_2 , \quad \| z^k \|_2 \le \frac{\lambda_{max}}{\lambda_{min}} \left[\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right]^k \| z^0 \|_2$$

Док-во. Так как

$$\| r^{k+1} \|_2 = \inf_{\tau} \| r^k - \tau A r^k \|_2 \le \| r^k - \tau_{опт} A r^k \|_2 \le$$

$$\le \| S_{\tau_{опт}} \|_2 \cdot \| r^k \|_2 = \rho(S_{\tau_{опт}}) \| z^k \|_2$$

то $\| r^k \|_2 \le [\rho_{опт}]^k \| r^0 \|_2$, где

$$\rho_{опт} = \rho(S_{\tau_{опт}}) = \frac{\lambda_{max}(A) - \lambda_{min}(A)}{\lambda_{max}(A) + \lambda_{min}(A)}.$$

Так как $\| r^k \|_2 = \| A z^k \|_2$ и

$$|\lambda_{min}|^2 (z, z) \le (Az, Az) \le |\lambda_{max}|^2 (z, z),$$

то из неравенств $\lambda_{min} \| z^k \|_2 \le \| r^k \|_2 \le \lambda_{max} \| z^k \|_2$ следует оценка

$$\| z^k \|_2 \le \frac{\lambda_{max}}{\lambda_{min}} \rho_{опт}^k \| z^0 \|_2 .$$

Лекция 9. Метод Ричардсона с чебышевскими параметрами

Предварительные замечания

В предыдущем разделе для решения системы $Ax = b$ с матрицей $A = A^* > 0$ мы рассмотрели стационарный метод Ричардсона (простой итерации)

$$x^0 \text{ – задан},$$
$$x^{k+1} = x^k - \tau \cdot (Ax^k - b),$$
$$k = 0, 1, \dots$$

и определили оптимальный параметр $\tau = \tau_{опт} = 2/(\lambda_{min} + \lambda_{max})$ такой, что

$$\rho(S_{\tau_{опт}}) = \inf_{\tau} \rho(S_\tau) = (\lambda_{max} - \lambda_{min})/(\lambda_{max} + \lambda_{min})$$

где $S_\tau = E - \tau \cdot A$, $0 < \lambda_{min} < \lambda_{max}$ – минимальное и максимальное собственные значения матрицы A.

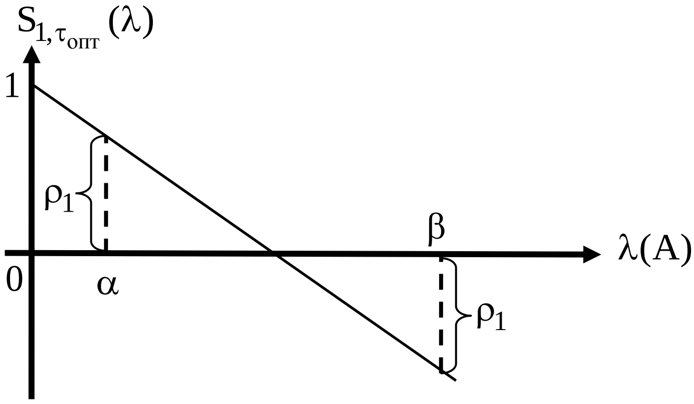
Если вместо собственных значений $0 < \lambda_{min} < \lambda_{max}$ известны их оценки

$$0 < \alpha \le \lambda_{min} < \lambda_{max} \le \beta,$$

то оптимальным параметром $\tau = \tau_{опт}$ метода простой итерации называют параметр, при котором минимизируется оценка для $\rho(S_\tau)$:

$$\inf_{\tau} \rho(S_\tau) \equiv \inf_{\tau} [\max_{\lambda \in Sp(A)} |1 - \tau \cdot \lambda|] \le \inf_{\tau} [\max_{\alpha \le \lambda \le \beta} |1 - \tau \cdot \lambda|] \equiv \inf_{\tau} \| S_{1,\tau}(\lambda) \|_{C[\alpha,\beta]} =$$
$$= \inf_{\tau} [\max\{|1 - \tau \cdot \alpha|, |1 - \tau \cdot \beta|\}] = (\beta - \alpha)/(\beta + \alpha) \equiv \rho_1,$$
$$\tau_{опт} = 2/(\alpha + \beta).$$

Решение этой минимаксной задачи иллюстрируется на следующем графике:



Теперь сделаем две итерации метода Ричардсона (2-циклический метод Ричардсона), но с разными параметрами:

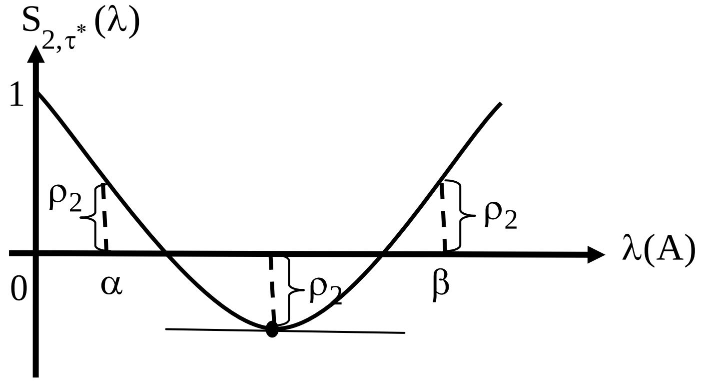
$$x^{2k+1} = x^{2k} - \tau_1 \cdot (Ax^{2k} - b),$$
$$x^{2k+2} = x^{2k+1} - \tau_2 \cdot (Ax^{2k+1} - b).$$

Будем выбирать параметры τ_1 и τ_2 из условия минимизации оценки для спектрального радиуса матрицы $S_{2,\tau=[\tau_1,\tau_2]} = (E - \tau_2 A)(E - \tau_1 A)$:

$$\inf_{\tau=(\tau_1,\tau_2)} \rho(S_{2,\tau}) \equiv \inf_{\tau} [\max_{\lambda\in Sp(A)} |(1-\tau_1\cdot\lambda|)\cdot(1-\tau_2\cdot\lambda|)] \le$$

$$\le \inf_{\tau} [\max_{\alpha\le\lambda\le\beta} (1-\tau_1\cdot\lambda|)\cdot(1-\tau_2\cdot\lambda|)] \equiv \inf_{\tau} \| S_{2,\tau}(\lambda) \|_{C[\alpha,\beta]} = \rho_2.$$

Можно доказать (**докажите!**), что оптимальные значения параметров $\tau^* = (\tau_1^*, \tau_2^*)$ определяются из условий, показанных на следующем графике:



Практически очевидно, что $\rho_2 < \rho_1$, т.е. 2-циклический метод Ричаздсона сходится “быстрее” метода простой итерации.

Тогда, очевидно, что оптимальные параметры $\tau^* = \{\tau_i = \tau_i^*\}_{i=1}^m$ m-циклического метода Ричардсона:

$$\begin{cases} x^{mk+1} = x^{mk} - \tau_1\cdot(Ax^{mk} - b), \\ \dots \\ x^{mk+m} = x^{mk+m-1} - \tau_m\cdot(Ax^{mk+m-1} - b), \end{cases}$$

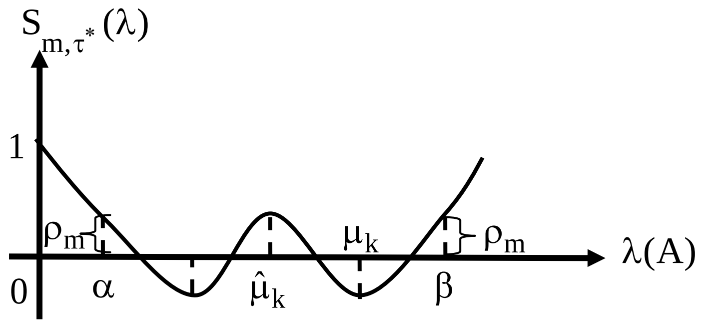
$$k = 0, 1, \dots$$

следует выбирать из условия минимизации оценки для спектрального радиуса матрицы $S_{m,\tau=[\tau_1, ,..., \tau_m]} = (E-\tau_m A)\dots(E-\tau_1 A)$

$$\inf_{\tau=(\tau_1, \dots, \tau_m)} \rho(S_{m,\tau}) \equiv \inf_{\tau} [\max_{\lambda\in Sp(A)} |(1-\tau_1\cdot\lambda|)\cdot\ldots\cdot(1-\tau_m\cdot\lambda|)] \le$$

$$\le \inf_{\tau} [\max_{\alpha\le\lambda\le\beta} |(1-\tau_1\cdot\lambda)\cdot\ldots\cdot(1-\tau_m\cdot\lambda)|] \equiv \inf_{\tau} \| S_{m,\tau}(\lambda) \|_{C[\alpha,\beta]} = \rho_m.$$

а решение этой задачи (предположительно) изображено на следующем графике



где изображен “чебышевский альтернанс”: на интервале $[\alpha,\beta]$ полином $S_{m,\tau^*}(\lambda)=(1-\tau_1^*\cdot\lambda)\cdot\ldots\cdot(1-\tau_m^*\cdot\lambda)$ в точках $\alpha=\hat{\mu}_0<\ldots<\hat{\mu}_k<\ldots<\hat{\mu}_m=\beta$ имеет $m+1$ чередующихся экстремумов $S_{m,\tau^*}(\hat{\mu}_k)=(-1)^k\rho_m$.

Тогда $S_{m,\tau^*}(\lambda)$ имеет m попарно различных положительных корней $0<\alpha\le\hat{\mu}_{k-1}<\mu_k<\hat{\mu}_k\le\beta$, и $\tau_k^*=(\mu_k)^{-1}$.

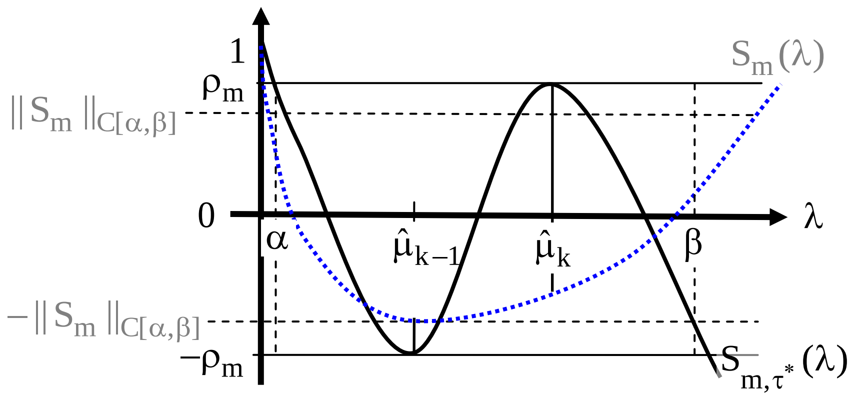
Предположим, что существует полином $S_{m,\tau^*}(\lambda)$, $S_{m,\tau^*}(0)=1$, имеющий “чебышевский альтернанс” на интервале $[\alpha,\beta]$.

Покажем, что этот полином наименее уклоняется от нуля на интервале $[\alpha,\beta]$ среди всех полиномов $S_m(\lambda)$, $S_m(0)=1$.

Теорема. Если $\alpha>0$, то $\|S_{m,\tau^*}(\lambda)\|_{C[\alpha,\beta]}=\inf\limits_{S_m(\lambda),\ S_m(0)=1}\|S_m(\lambda)\|_{C[\alpha,\beta]}$.

Док-во. Пусть $\|S_{m,\tau^*}(\lambda)\|_{C[\alpha,\beta]}>\inf\limits_{S_m(\lambda),\ S_m(0)=1}\|S_m(\lambda)\|_{C[\alpha,\beta]}$,

тогда $\exists S_m(\lambda)$: $\|S_{m,\tau^*}(\lambda)\|_{C[\alpha,\beta]}>\|S_m(\lambda)\|_{C[\alpha,\beta]}$.



Т.к. последовательность $\{S_{m,\tau^*}(\hat{\mu}_k)\}_{k=0}^m$ знакопеременна и $\rho_m=|S_{m,\tau^*}(\hat{\mu}_k)|>|S_m(\hat{\mu}_k)|$,

то послед-ность $\{R_m(\hat{\mu}_k)=S_{m,\tau^*}(\hat{\mu}_k)-S_m(\hat{\mu}_k)\}_{k=0}^m$ знакопеременна, т.е. полином $R_m(\lambda)\equiv S_{m,\tau^*}(\lambda)-S_m(\lambda)$ в каждом интервале $(0<\hat{\mu}_{k-1},\hat{\mu}_k)$ имеет положительный корень.

Т.к имеем m таких интервалов, то полином $R_m(\lambda)$ $[\alpha,\beta]$ имеет m попарно различных положительных корней.

Но $R_m(0)=S_{m,\tau^*}(0)-S_m(0)=1-1=0$ – $(m+1)$-й корень:

у полинома $R_m(\lambda)$ степени m разных корней больше, чем его степень, т.е. $R_m(\lambda)\equiv 0$ – противоречие предположению $\|S_{m,\tau^*}(\lambda)\|_{C[\alpha,\beta]}>\|S_m(\lambda)\|_{C[\alpha,\beta]}$.

Построение полинома $S_{m,\tau^*}(\lambda)$, $S_{m,\tau^*}(0)=1$, наименее уклоняющегося от нуля на интервале $[\alpha, \beta]$

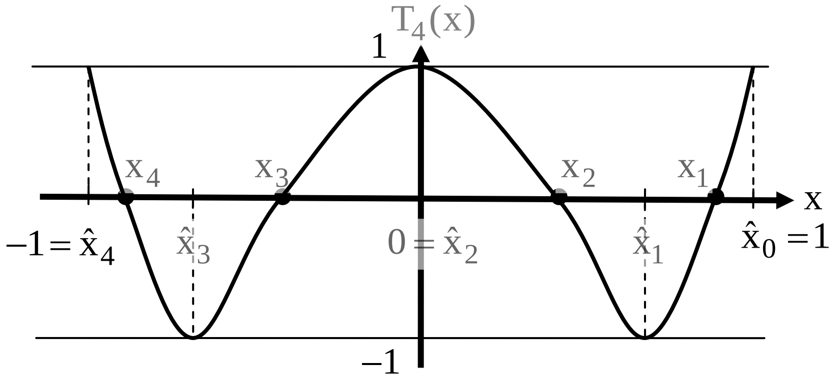
Полином Чебышева

На интервале [0, 1] рассмотрим функцию $T_m(x) = \cos(m \cdot \arccos(x))$.
Очевидно, что

- $\| T_m(x) \|_{C[-1,1]} = 1$,
- $T_m(\hat{x}) = 1$, если $m \cdot \arccos(\hat{x}) = 2k\pi$, т.е. $\hat{x} = \hat{x}_{2k} = \cos(2k\pi / m)$,
- $T_m(\hat{x}) = -1$, если $m \cdot \arccos(\hat{x}) = (2k+1)\pi$, т.е. $\hat{x} = \hat{x}_{2k+1} = \cos(\frac{(2k+1)\pi}{m})$

  $\forall\ k = 0, \pm 1, \ldots$,
- $-1 = \hat{x}_m < \hat{x}_{m-1} < \ldots < \hat{x}_1 < \hat{x}_0 = 1$, $\qquad T_m(\hat{x}_k) = (-1)^k$,
- $T_m(x) = 0$, если $m \cdot \arccos(x) = k\pi - \pi/2$, т.е. $x = x_k = \cos(\frac{(2k-1)\pi}{2m})$

  $-1 \le \hat{x}_k < x_k < \hat{x}_{k-1} \le 1,\ k = 1, 2, \ldots, m$.



Лемма. Функция $T_m(x) = \cos(m \cdot \arccos(x))$ – полином (Чебышева) степени m.

Док-во. $T_0(x) = 1$, $T_1(x) = x$ – полиномы.

Т.к. $\cos((k+1)\varphi) + \cos((k-1)\varphi) = 2\cos(\varphi) \cdot \cos(k\varphi)$,

то при $\varphi = \arccos x$ имеем

$T_{k+1}(x) = 2 \cdot T_1(x) \cdot T_k(x) - T_{k-1}(x)$ – полином степени $k+1\ \forall\ k > 1$.

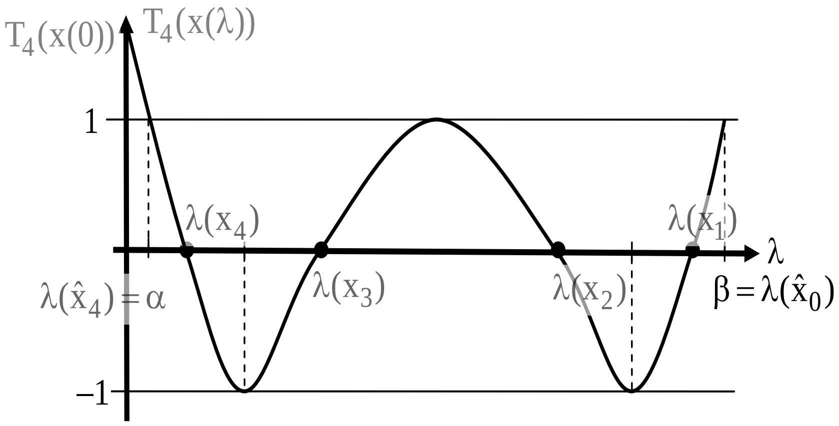
Следствие. $T_m(x) \ne 0 \quad \forall\ x \notin [-1, 1]$.

Замечание. Очевидно, что полином Чебышева $T_m(x)$ имеет на интервале $[-1, 1]$ альтернанс: в $(m+1)$-й точке $-1 = \hat{x}_m < \hat{x}_{m-1} < \ldots < \hat{x}_1 < \hat{x}_0 = 1$ принимает чередующиеся экстремальные значения $T_m(\hat{x}_k) = (-1)^k$, равные по модулю.

Полином $S_{m,\tau^*}(\lambda)$

Построим полином $S_{m,\tau^*}(\lambda)$, $S_{m,\tau^*}(0)=1$ с чебышевским альтернансом на $[\alpha,\beta]$, $\alpha>0$. Т.к. полином Чебышева $T_m(x)$ имеет на интервале $[-1,1]$ альтернанс, то поступим следующим образом:

1. интервал $[-1,1]$ длины 2 отмасштабируем до интервала $[-\frac{\beta-\alpha}{2},\frac{\beta-\alpha}{2}]$ длины $\beta-\alpha$: график полинома $T_m(x)$ сожмется по горизонтали в график полинома $T_m(x(y))$, где $y=\frac{\beta-\alpha}{2}x$;
2. интервал $[-\frac{\beta-\alpha}{2},\frac{\beta-\alpha}{2}]$ сдвинем на $-\frac{\beta-\alpha}{2}+\alpha$ в интервал $[\alpha,\beta]$: сжатый по горизонтали график полинома $T_m(x)$ сдвинется и будет графиком полинома $T_m(x(\lambda))$, где $\lambda(x)=y+\alpha+\frac{\beta-\alpha}{2}=\frac{\beta-\alpha}{2}x+\frac{\beta+\alpha}{2}$ и $x(\lambda)=\frac{2\lambda-(\beta+\alpha)}{\beta-\alpha}$, но, это очевидно, альтернанс у полинома $T_m(x(\lambda))$ на интервале $\lambda\in[\alpha,\beta]$ сохранится;



3. отнормируем полином $T_m(x(\lambda))$: $S_{m,\tau^*}(\lambda)=\frac{1}{T_m(x(0))}T_m(x(\lambda))$, т.е. $S_{m,\tau^*}(0)=1$, и, снова заметим, что у полученного полинома есть чебышевский альтернанс на интервале $\lambda\in[\alpha,\beta]$.

Следовательно, мы построили нужный нам полином (наименее уклоняющийся от 0 на интервале $\lambda\in[0<\alpha,\beta]$ и равный 1 при $\lambda=0$):

$$S_{m,\tau^*}(\lambda)=\frac{1}{T_m(x(0))}T_m(x(\lambda)),\ \lambda\in[\alpha,\beta]$$

с корнями (занумерованными в порядке убывания)

$$\mu_k=\lambda(x_k)=\frac{\beta-\alpha}{2}x_k+\frac{\beta+\alpha}{2},\ x_k=\cos(\frac{(2k-1)\pi}{2m}),\ k=1,2,\dots,m.$$

Замечание. Очевидно, что $\rho_m = \| S_{m,\tau^*}(\lambda) \|_{C[\alpha,\beta]} = | T_m(x(0)) |^{-1}$.

Норма полинома $S_{m,\tau^*}(\lambda)$

Осталось вычислить $\rho_m = \| S_{m,\tau^*}(\lambda) \|_{C[\alpha,\beta]}$.

Теорема.

Если $\alpha > 0$, то $\rho_m = \| S_{m,\tau^*}(\lambda) \|_{C[\alpha,\beta]} = \dfrac{2\gamma^m}{1+\gamma^{2m}} < 1$, где $\gamma = \dfrac{\sqrt{\beta} - \sqrt{\alpha}}{\sqrt{\beta} + \sqrt{\alpha}}$.

Док-во.

Очевидно, $\rho_m = \| S_{m,\tau^*}(\lambda) \|_{C[\alpha,\beta]} = | T_m(x(0)) |^{-1}$, $x(0) = -\dfrac{\beta+\alpha}{\beta-\alpha} < -1$.

Для вычисления $T_m(x(0))$ воспользуемся формулой

$$T_m(x) = \frac{(x+\sqrt{x^2-1})^m + (x-\sqrt{x^2-1})^m}{2} \text{ при } |x| > 1.$$

Заметим, что

$$x(0) + \sqrt{x^2(0)-1} = \frac{-(\sqrt{\beta})^2 + 2\sqrt{\beta}\sqrt{\alpha} - (\sqrt{\alpha})^2}{(\sqrt{\beta}+\sqrt{\alpha})\cdot(\sqrt{\beta}-\sqrt{\alpha})} = -\frac{\sqrt{\beta}-\sqrt{\alpha}}{\sqrt{\beta}+\sqrt{\alpha}} = -\gamma,$$

$$x(0) - \sqrt{x^2(0)-1} = \frac{1}{x(0)+\sqrt{x^2(0)-1}} = -\frac{1}{\gamma}.$$

Тогда $T_m(x(0)) = \dfrac{(-\gamma)^m + (-\gamma^{-1})^m}{2} = (-1)^m \dfrac{1+\gamma^{2m}}{2\gamma^m} \Rightarrow \rho_m = \dfrac{2\gamma^m}{1+\gamma^{2m}} < 1$.

Док-во формулы $T_k(x) = 0.5[(x+\sqrt{x^2-1})^k + (x-\sqrt{x^2-1})^k]$при $|x| > 1$.

Действительно, $T_0(x) = 1$ и $T_1(x) = x$.

Осталось проверить, что $T_{k+1}(x) = 2\cdot T_1(x)\cdot T_k(x) - T_{k-1}(x)$ или

$$(x+\sqrt{x^2-1})^{k+1} + (x-\sqrt{x^2-1})^{k+1} =$$
$$= 2x[(x+\sqrt{x^2-1})^k + (x-\sqrt{x^2-1})^k] - [(x+\sqrt{x^2-1})^{k-1} + (x-\sqrt{x^2-1})^{k-1}].$$

Пусть $y = \sqrt{x^2-1}$, тогда

$$(x+y)^{k+1} + (x-y)^{k+1} = x\cdot[(x+y)^k + (x-y)^k] + y\cdot[(x+y)^k - (x-y)^k] =$$
$$= x\cdot[(x+y)^k + (x-y)^k] +$$
$$+ y\cdot[x\cdot\{(x+y)^{k-1} - (x-y)^{k-1}\} + y\cdot\{(x+y)^{k-1} + (x-y)^{k-1}\}] =$$

$$= x\cdot[(x+y)^k+(x-y)^k]+$$
$$+y\cdot x\cdot\{(x+y)^{k-1}-(x-y)^{k-1}\}+(x^2-1)\cdot\{(x+y)^{k-1}+(x-y)^{k-1}\}=$$
$$=2x\cdot[(x+y)^k+(x-y)^k]-\{(x+y)^{k-1}+(x-y)^{k-1}\}+$$
$$+(-x^2-xy+yx+x^2)\cdot(x+y)^{k-1}+(-x^2+xy-yx+x^2)\cdot(x-y)^{k-1}=$$

$=2x\cdot[(x+y)^k+(x-y)^k]-\{(x+y)^{k-1}+(x-y)^{k-1}\}$, что и тр. док.

Формулы m-циклического метода Ричардсона с чебышевскими параметрами

Из вышеизложенного следует, что для решения системы $Ax=b$, $A=A^*>0$, $Sp(A)\subset[\alpha,\beta]$, $\alpha>0$, двучленные формулы m-циклического метода Ричардсона с чебышевскими параметрами имеют вид

$$\begin{cases} x^{k+1}=x^k-\tau_{k+1}(Ax^k-b), \\ \quad k=0,1,\ldots, \\ \tau_1,\ldots,\tau_m,\ \tau_{m+1}=\tau_1,\ldots,\tau_{2m}=\tau_m,\ldots\ \ (\tau_{m+j}=\tau_j), \end{cases}$$

$$\tau_j\equiv\tau_j^*=2/[(\beta+\alpha)+(\beta-\alpha)\cdot\cos\frac{(2j-1)\pi}{2m}],\ j=1,2,\ldots,m,$$

а матрица перехода для его ошибки за полный цикл ($z^{(t+1)\cdot m}=S_{m,\tau^*}\cdot z^{t\cdot m}$)

$S_{m,\tau^*}(A)=(E-\tau_m^*A)\cdot\ldots\cdot(E-\tau_1^*A)$ симметрична, и, следовательно,

$$\|S_{m,\tau^*}(A)\|_2=\|S_{m,\tau^*}(A)\|_A=\rho_m.$$

Численная неустойчивость двучленных формул метода Ричардсона

Из–за ошибок округления реализация формул $x^{k+1}=x^k-\tau_{k+1}(Ax^k-b)$ неустойчива, т.к. норма оператора шага $\|E-\tau_{k+1}A\|_2$ для ошибки может быть значительно больше 1 (в методе простой итерации эта норма меньше 1), в то время как $\|(E-\tau_1A)\cdot\ldots\cdot(E-\tau_mA)\|_2\le\rho_m<1$. Поэтому, если несколько первых итераций выполняется с параметрами τ_{k+1}, для которых нормы операторов шага $\|E-\tau_{k+1}A\|_2>>1$, то процесс вычислений прервется из-за “переполнения”.

Выходом из такого положения является специальная нумерация параметров τ_{k+1}, при которой перемешиваются операторы шага с большими и малыми нормами. Но мы на этом остонавливаться не будем. а построим устойчивые трехчленные формулы реализации метода Ричардсона.

Трехчленные формулы реализации метода Ричардсона

Итак, наименее уклоняющийся от 0 на интервале $\lambda\in[0<\alpha,\beta]$ и равный 1 при $\lambda=0$ полином имеет вид:

$$S_{m,\tau^*}(\lambda)=t_m^{-1}\cdot T_m(x(\lambda)),\ x(\lambda)=\frac{2\lambda-(\beta+\alpha)}{\beta-\alpha},\ \ t_m=T_m(x(0)),\ x(0)=-\frac{\beta+\alpha}{\beta-\alpha},$$

а для полиномов Чебышева $T_m(x)$ и констант t_m имеем трехчленные формулы:

$$\begin{cases}T_0(x)=1,\ \ T_1(x)=x,\\ T_{k+1}(x)=2\cdot T_1(x)\cdot T_k(x)-T_{k-1}(x),\end{cases}\quad \begin{cases}t_0=1,\ \ t_1=-(\beta+\alpha)/(\beta-\alpha),\\ t_{k+1}=2\cdot t_1\cdot t_k-t_{k-1}.\end{cases}$$

В последних формулах сделаем замену $T_k(x)=t_k\cdot S_{k,\tau^*}(\lambda)$:

$$\begin{cases}t_0\cdot S_{0,\tau^*}(\lambda)=1,\ \ t_1\cdot S_{1,\tau^*}(\lambda)=\dfrac{2}{\beta-\alpha}\lambda-\dfrac{\beta+\alpha}{\beta-\alpha}=\dfrac{2}{\beta-\alpha}\lambda+t_1,\\ t_{k+1}\cdot S_{k+1,\tau^*}(\lambda)=2t_1t_k\cdot S_{1,\tau^*}(\lambda)\cdot S_{k,\tau^*}(\lambda)-t_{k-1}\cdot S_{k-1,\tau^*}(\lambda).\end{cases}$$

Или, учитывая формулы для t_k:

$$S_{0,\tau^*}(\lambda)=1,\ \ S_{1,\tau^*}(\lambda)=1-\frac{2}{\beta+\alpha}\lambda\equiv 1-\tau_{\text{опт}}\cdot\lambda,$$

$$S_{k+1,\tau^*}(\lambda)=\frac{t_{k+1}+t_{k-1}}{t_{k+1}}\cdot(1-\tau_{\text{опт}}\cdot\lambda)\cdot S_{k,\tau^*}(\lambda)-\frac{t_{k-1}}{t_{k+1}}\cdot S_{k-1,\tau^*}(\lambda)=$$

$$=S_{k,\tau^*}(\lambda)+\frac{t_{k-1}}{t_{k+1}}\cdot[S_{k,\tau^*}(\lambda)-S_{k-1,\tau^*}(\lambda)]-\tau_{\text{опт}}\cdot(1+\frac{t_{k-1}}{t_{k+1}})\cdot\lambda\cdot S_{k,\tau^*}(\lambda).$$

В этих формулах λ заменим на матрицу системы $Ax=b$:

$$S_{0,\tau^*}(A)=E,\ \ S_{1,\tau^*}(A)=E-\tau_{\text{опт}}\cdot A,$$

$$S_{k+1,\tau^*}(A)=S_{k,\tau^*}(A)+\frac{t_{k-1}}{t_{k+1}}\cdot[S_{k,\tau^*}(A)-S_{k-1,\tau^*}(A)]-\tau_{\text{опт}}\cdot(1+\frac{t_{k-1}}{t_{k+1}})\cdot A\cdot S_{k,\tau^*}(A).$$

Тогда, если $z^k\equiv x^k-x=S_{k,\tau^*}(A)\cdot z^0$ – k-я ошибка итерационного метода, т.е.

$$z^0=S_{0,\tau^*}(A)\cdot z^0=z^0,\ \ z^1=S_{1,\tau^*}(A)\cdot z^0=z^0-\tau_{\text{опт}}\cdot Az^0,$$

$$z^{k+1}=z^k+\frac{t_{k-1}}{t_{k+1}}\cdot[z^k-z^{k-1}]-\tau_{\text{опт}}\cdot(1+\frac{t_{k-1}}{t_{k+1}})\cdot A\cdot z^k,$$

то для приближений x^k справедливы формулы (**проверьте!**):

$$x^0 \text{ – задан},\ \ x^1=x^0-\tau_{\text{опт}}\cdot(Ax^0-b),$$

$$x^{k+1}=x^k+\frac{t_{k-1}}{t_{k+1}}\cdot(x^k-x^{k-1})-\tau_{\text{опт}}\cdot(1+\frac{t_{k-1}}{t_{k+1}})\cdot(Ax^0-b).$$

Обычно вводят обозначение $\dfrac{t_{k-1}}{t_{k+1}} = \omega_k \omega_{k+1}$, $\omega_k = t_{k-1} / t_k$, из формулы $t_{k+1} = 2t_1 t_k - t_{k-1}$ получают рекуррентное соотношение

$$\omega_1 = \frac{1}{t_1} \equiv -\frac{\beta-\alpha}{\beta+\alpha}, \quad \omega_{k+1} = \frac{1}{2t_1 - \omega_k} = \frac{1}{2(\omega_1)^{-1} - \omega_k}, \quad k = 1, 2, \dots .$$

Тогда

$$x^{k+1} = x^k + \omega_k \omega_{k+1}(x^k - x^{k-1}) - \frac{2}{\beta+\alpha}(1 + \omega_k \omega_{k+1}) \cdot (Ax^k - b), \ \ k = 1, 2, \dots ,$$

– двухшаговый (трехслойный) итерационный процесс.

Лекция 10.

Метод сопряженных градиентов

Предварительные замечания

Циклический метод Ричардсона $x^{k+1} = x^k - \tau^*_{k+1}(Ax^k - b)$ с чебышевскими параметрами $\tau^*_{k+m} = \tau^*_k$ (длина цикла равна m) для решения системы $Ax = b$ с симметричной и положительно определенной матрицей: $A = A^* > 0$, спектр которой лежит в интервале $[\alpha, \beta]$, $\alpha > 0$, характеризуется тем, что за $t \cdot m$ итераций (t циклов) для ошибки $z^{t\cdot m} = S_{m,\tau^*}(A) \cdot z^{(t-1)\cdot m} = [S_{m,\tau^*}(A)]^t \cdot z^0$ справедливо неравенство

$$\| z^{t\cdot m} \|_A \le \rho_m \cdot \| z^{(t-1)\cdot m} \|_A \le [\rho_m]^t \cdot \| z^0 \|_A \quad \forall\ z^0,$$

где $\rho_m = \| S_{m,\tau^*}(\lambda) \|_{C[\alpha,\beta]} \equiv 2\gamma^m / (1 + \gamma^{2m}) < 1,\ \gamma = (\sqrt{\beta} - \sqrt{\alpha})/(\sqrt{\beta} + \sqrt{\alpha})$.

Подчеркнем, что эта оценка справедлива для всех z^0, а для вычисления $\tau^*_{k+m} = \tau^*_k$ необходимы оценки $0 < \alpha \le \lambda_{\min}(A)$ и $\lambda_{\max}(A) \le \beta$.

Но, если параметры $\tau^*_{(t-1)\cdot m+1}, \ldots, \tau^*_{(t-1)\cdot m+m}$ заменить на параметры $\tau^{(t)}_{(t-1)\cdot m+1}, \ldots, \tau^{(t)}_{(t-1)\cdot m+m}$, зависящие от ошибки $\hat{z}^{(t-1)\cdot m}$ ($\hat{z}^0 = z^0$), при которых достигается минимум нормы ошибки $\hat{z}^{t\cdot m} = S_{m,\tau^{(t)}}(A) \cdot \hat{z}^{(t-1)\cdot m}$:

$$\| \hat{z}^{t\cdot m} \|_A = \| S_{m,\tau^{(t)}}(A) \cdot \hat{z}^{(t-1)\cdot m} \|_A = \min_{\tau} \| S_{m,\tau}(A) \cdot \hat{z}^{(t-1)\cdot m} \|_A,$$

то очевидно, что

$$\min_{\tau} \| S_{m,\tau}(A) \cdot \hat{z}^{(t-1)\cdot m} \|_A \le \| S_{m,\tau^*}(A) \cdot \hat{z}^{(t-1)\cdot m} \|_A \le \rho_m \cdot \| \hat{z}^{(t-1)\cdot m} \|_A,$$

т.е. имеем оценку сходимости нового итерационного метода вариационного типа (его параметры выбираются из условия минимизации функционала):

$$\| \hat{z}^{t\cdot m} \|_A \le \rho_m \cdot \| \hat{z}^{(t-1)\cdot m} \|_A \le [\rho_m]^t \cdot \| \hat{z}^0 \|_A,$$

совпадающую с оценкой сходимости m-циклического метода Ричардсона с чебышевскими параметрами.

Для определения параметров τ_{k+1} m-циклического метода Ричардсона $x^{k+1} = x^k - \tau_{k+1}(Ax^k - b)$ для решения системы $Ax = b$ необходимо предварительное вычисление (точное или приближенное) границ спектра матрицы A, чего не требуется в методах наискорейшего спуска и минимальных невязок.

Минимизация функционала

Итак, по вектору $z^{(t-1)\cdot m}$ нам нужно построить вектор $z^{(t)\cdot m} = S_{m,\tau^{(t)}}(A)\cdot z^{(t-1)\cdot m}$ такой, что

$$\| z^{t\cdot m} \|_A = \min_{\tau} \| S_{m,\tau}(A)\cdot z^{(t-1)\cdot m} \|_A .$$

Решим эту задачу при $t=1$ (т.к. при других t решение задачи будет таким же с точностью до обозначений).

Т.к.

$$z^m \equiv (E-\tau_m A)...(E-\tau_1 A)z^0 = z^0 - q_1(\tau)Az^0 - ... - q_m(\tau)A^m z^0 =$$
$$= z^0 - \alpha_1(\tau)g_1 - ... - \alpha_m(\tau)g_m \in L_m,$$

где $L_m = L\{Az^0, ..., A^m z^0\} \equiv L\{g_1, ..., g_m\}$, а $\{g_1, ..., g_m\}$ – некоторая система векторов (например, “ортогональный” базис в L_m), то

$$\| z^m \|_A = \min_{\tau} \| (E-\tau_m A)...(E-\tau_1 A)z^0 \|_A = \min_{\alpha} \| z^0 - \alpha_1 g_1 - ... - \alpha_m g_m \|_A .$$

Параметры $\alpha_1, ..., \alpha_m$ удовлетворяют системе уравнений

$$\frac{1}{2}\frac{\partial(Az^m, z^m)}{\partial\alpha_i} = (A\frac{\partial z^m}{\partial\alpha_i}, z^m) \equiv (Ag_i, z^0 - \alpha_1 g_1 - ... - \alpha_m g_m) = 0, \quad i=1, ..., m,$$

или

$$\begin{bmatrix} (Ag_1,g_1) & (Ag_1,g_2) & ... & (Ag_1,g_m) \\ (Ag_2,g_1) & (Ag_2,g_2) & ... & (Ag_2,g_m) \\ \vdots & \vdots & ... & \vdots \\ (Ag_m,g_1) & (Ag_m,g_2) & ... & (Ag_m,g_m) \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} (Az^0,g_1) \\ (Az^0,g_2) \\ \vdots \\ (Az^0,g_m) \end{pmatrix}.$$

Матрица этой системы – матрица Грамма в скалярном произведении $(x,y)_A = (Ax,y)$ базиса $\{g_i\}_{i=1}^m$ в L_m, ее определитель не равен нулю, следовательно решение α существует и единственно.

Если базис $\{g_i\}_{i=1}^m$ является A–ортогональным, т.е. $(Ag_i, g_j) = 0, \ i \ne j$, то

$$\alpha_k = \frac{(Az^0,g_k)}{(Ag_k,g_k)} \equiv \frac{(r^0,g_k)}{(Ag_k,g_k)}, \qquad \begin{cases} z^m = z^0 - \alpha_1 g_1 - \alpha_2 g_2 - ... - \alpha_m g_m, \\ x^m = x^0 - \alpha_1 g_1 - \alpha_2 g_2 - ... - \alpha_m g_m. \end{cases}$$

Важное замечание. В случае A–ортогонального базиса $\{g_i\}_{i=1}^k$ в L_k, $k=\overline{1,m}$,

вектор $z^k = z^0 - \alpha_1 g_1 - ... - \alpha_k g_k$ является решением задачи

$$\| z^k \|_A = \min_{\tau} \| (E-\tau_k A)...(E-\tau_1 A)z^0 \|_A = \min_{\tau} \| S_{k,\tau}(A)z^0 \|_A$$

т.е., в отличие от метода Ричардсона, минимизация нормы ошибки осуществляется на каждом внутреннем шаге цикла, и, следовательно,

$$\| z^k \|_A \le \rho_k \cdot \| z^0 \|_A \quad \forall\, k=1, ..., m.$$

Построение A–ортогонального базиса

A–ортогонализуем систему векторов $\{Az^0, ..., A^m z^0\}$.

Зададим $g_1 = Az^0 \equiv r^0 \in L_1$, тогда $\alpha_1 = \dfrac{(r^0, g_1)}{(Ag_1, g_1)}$, $z^1 = z^0 - \alpha_1 g_1$.

Вычислим вектор $r^1 = Az^1 = Az^0 - \alpha_1 Ag_1 = Az^0 - \alpha_1 A^2 z^0 \in L_2$. Легко проверить, что $(r^1, g_1) = 0$, т.е. $r^1 \perp L_1$ относительно обычного скалярного произведения. Далее применяем метод матиндукции.

Предположим, что мы построили A-ортогональный базис $\{g_i\}_{i=1}^k$ в $L_k = L\{Az^0, ..., A^k z^0\}$ такой, что

1. $\{g_i\}_{i=1}^j$ – A-ортогональный базис в $L_j = L\{Az^0, ..., A^j z^0\}$, $j = \overline{1, k}$,
2. $z^j = z^0 - \alpha_1 \cdot g_1 - ... - \alpha_j \cdot g_j$, $\alpha_i = \dfrac{(r^0, g_i)}{(Ag_i, g_i)}$,
3. $r^j = Az^j \in L_{j+1}$ и $r^j \perp L_j$, т.е. $(r^j, g_i) = 0 \;\; \forall\, i = \overline{1, j}$.

Определим $g_{k+1} = r^k - \gamma_1 g_1 - ... - \gamma_k g_k \in L_{k+1}$ из условий

$$(Ag_{k+1}, g_i) \equiv (Ar^k, g_i) - \gamma_1 \cdot (Ag_1, g_i) - ... - \gamma_k \cdot (Ag_k, g_i) = 0,$$
$$i = 1, ..., k.$$

- Так как $(Ag_j, g_i) = 0,\ i \neq j$, то $\gamma_i = (Ar^k, g_i)/(Ag_i, g_i)$,
- так как $r^k \perp L_k$ и $Ag_i \in L_{i+1} \subset L_k$ при $i+1 \le k$, то $(Ar^k, g_i) = (r^k, Ag_i) = 0$ и $\gamma_i = 0,\ i = \overline{1, k-1}$.

Тогда

1. $g_{k+1} = r^k - \gamma_k g_k$, $\gamma_k = (Ar^k, g_k)/(Ag_k, g_k)$, и
   $\{g_i\}_{i=1}^{k+1}$ – A-ортогональный базис в $L_{k+1} = L\{Az^0, ..., A^{k+1} z^0\}$,
2. $z^{k+1} = z^0 - \alpha_1 \cdot g_1 - ... - \alpha_{k+1} \cdot g_{k+1} = z^k - \alpha_{k+1} \cdot g_{k+1}$, $\alpha_{k+1} = \dfrac{(r^0, g_{k+1})}{(Ag_{k+1}, g_{k+1})}$,
3. $r^{k+1} = Az^{k+1} = r^0 - \alpha_1 \cdot Ag_1 - ... - \alpha_{k+1} \cdot Ag_{k+1} \perp L_{k+1}$, т.к. легко проверяются равенства $(r^{k+1}, g_i) = 0,\ i = \overline{1, k+1}$;
   $r^{k+1} \in L_{k+1}$, и, очевидно, что
   $r^{k+1} = Az^{k+1} \in L\{r^0, L\{A^2 z^0, ..., A^{k+2} z^0\}\} \equiv L_{k+2}$.

Таким образом построен очередной базисный вектор g_{k+1} и все предположения метода матиндукции для него выполняются, из которых следует, что

$$\alpha_{k+1} = \frac{(r^0, g_{k+1})}{(Ag_{k+1}, g_{k+1})} = \frac{(r^1 + \alpha_1 Ag_1, g_{k+1})}{(Ag_{k+1}, g_{k+1})} = \frac{(r^1, g_{k+1})}{(Ag_{k+1}, g_{k+1})} = ... = \frac{(r^k, g_{k+1})}{(Ag_{k+1}, g_{k+1})}.$$

Метод сопряженных градиентов

Итак, для решения x системы $Ax = b$ $(A = A^* > 0)$ задано начальное приближение x^0 и пусть $z^0 = x^0 - x$ – вектор начальной ошибки, $r^0 = Az^0 = Ax^0 - b$ – вектор (известный) начальной невязки.

Выпишем, построенные в предыдущем разделе, формулы A–ортогонального базиса $\{g_i\}_{i=1}^m$ в $L_m = L\{Az^0, ..., A^m z^0\}$ и добавим к ним выражение для очередного приближения $x^k = x^0 - \alpha_1 g_1 - ... - \alpha_k g_k = x^{k-1} - \alpha_k g_k$:

1 шаг:

$$g_1 = r^0 \equiv Ax^0 - b,$$

$$\alpha_1 = (r^0, g_1)/(Ag_1, g_1),$$

$$x^1 = x^0 - \alpha_1 \cdot g_1,$$

$$r^1 = r^0 - \alpha_1 \cdot Ag_1 \equiv Ax^1 - b;$$

2 шаг: если $r^1 \neq 0$, то

$$g_2 = r^1 - \gamma_1 g_1, \;\; \gamma_1 = (Ar^1, g_1)/(Ag_1, g_1)$$

$$\alpha_2 = (r^1, g_2)/(Ag_2, g_2),$$

$$x^2 = x^1 - \alpha_2 \cdot g_2,$$

$$r^2 = r^1 - \alpha_2 \cdot Ag_2 \equiv Ax^2 - b;$$

...

$(k+1)$-й шаг: если $r^k \neq 0$, то

$$g_{k+1} = r^k - \gamma_k g_k, \;\; \gamma_k = (Ar^k, g_k)/(Ag_k, g_k)$$

$$\alpha_{k+1} = (r^k, g_{k+1})/(Ag_{k+1}, g_{k+1}),$$

$$x^{k+1} = x^k - \alpha_{k+1} \cdot g_{k+1},$$

$$r^{k+1} = r^k - \alpha_{k+1} \cdot Ag_{k+1} \equiv Ax^{k+1} - b.$$

Этот итерационный процесс решения системы $Ax = b$ $(A = A^* > 0)$ называется методом сопряженных градиентов, если $k = 0, 1, ...$.

Теорема.

Если $A = A^* > 0$, то метод сопряженных градиентов продолжается до получения решения системы $Ax = b$ за $m \le n$ итераций (пока $r^k \neq 0$) и

$$\| z^k \|_A \le \frac{2\gamma^k}{1+\gamma^{2k}} \| z^0 \|_A, \quad \gamma = \frac{\sqrt{\lambda_{max}} - \sqrt{\lambda_{min}}}{\sqrt{\lambda_{max}} + \sqrt{\lambda_{min}}}.$$

Доказать теорему в качестве упражнения.

Трехслойные формулы метода сопряженных градиентов

Напомним двухслойные формулы метода сопряженных градиентов для решения x системы $Ax=b$ $(A=A^*>0)$, A-ортогональность векторов $\{g_i\}$, ортогональность невязок $r^k \in L_{k+1}$ подпространству L_k:

$$\left.\begin{array}{ll} x^1 = x^0 - \alpha_1 \cdot g_1, & g_1 = r^0, \\ & \alpha_1 = (r^0, g_1)/(Ag_1, g_1); \\ k \ge 1: & \\ x^{k+1} = x^k - \alpha_{k+1} \cdot g_{k+1}, & g_{k+1} = r^k - \gamma_k g_k, \\ & \gamma_k = (Ar^k, g_k)/(Ag_k, g_k), \\ & \alpha_{k+1} = (r^k, g_{k+1})/(Ag_{k+1}, g_{k+1}) \end{array}\right| \begin{array}{l} (Ag_i, g_j) = 0,\ i \ne j, \\ (r^k, g_i) = 0,\ i \le k, \\ (r^k, r^i) = 0,\ i < k. \end{array}$$

Так как $g_k = -(x^k - x^{k-1})/\alpha_k$, то эти формулы можно переписать в виде:

$$x^1 = x^0 - \alpha_1 \cdot r^0,\ \ \alpha_1 = (r^0, r^0)/(Ar^0, r^0);$$

$$k > 0:$$

$$\begin{aligned} x^{k+1} &= x^k - \alpha_{k+1} \cdot (r^k + \gamma_k (x^k - x^{k-1})/\alpha_k) = \\ &= x^k - \underbrace{(\alpha_{k+1}/\alpha_k \cdot \gamma_k)}_{\beta_{k+1}} \cdot (x^k - x^{k-1}) - \alpha_{k+1} \cdot r^k. \end{aligned}$$

Параметры α_{k+1} и β_{k+1} можно определять из условий ортогональности невязки r^{k+1} невязкам r^{k-1} и r^k:

$$\begin{cases} (r^{k+1}, r^{k-1}) \equiv (r^k, r^{k-1}) - \beta_{k+1} \cdot (r^k - r^{k-1}, r^{k-1}) - \alpha_{k+1} \cdot (Ar^k, r^{k-1}) = 0, \\ (r^{k+1}, r^k) \equiv (r^k, r^k) - \beta_{k+1} \cdot (r^k - r^{k-1}, r^k) - \alpha_{k+1} \cdot (Ar^k, r^k) = 0, \end{cases}$$

или из условия (**докажите!**) минимизации A-нормы ошибки:

$$\| z^{k+1} \|_A^2 = \min_{\alpha,\beta} (A[z^k - \beta \cdot (z^k - z^{k-1}) - \alpha \cdot r^k], [z^k - \beta \cdot (z^k - z^{k-1}) - \alpha \cdot r^k]),$$

что сводится (**докажите!**) к решению другой системы:

$$\begin{cases} \beta_{k+1} \cdot (r^k - r^{k-1}, x^k - x^{k-1}) + \alpha_{k+1} \cdot (r^k - r^{k-1}, r^k) = (r^k, x^k - x^{k-1}), \\ \beta_{k+1} \cdot (r^k - r^{k-1}, r^k) + \alpha_{k+1} \cdot (Ar^k, r^k) = (r^k, r^k). \end{cases}$$

Тогда трехслойные формулы метода сопряженных градиентов имеют вид

$$x^1 = x^0 - \alpha_1 \cdot r^0,\ \ \alpha_1 = (r^0, r^0)/(Ar^0, r^0);$$

$$x^{k+1} = x^k - \beta_{k+1} \cdot (x^k - x^{k-1}) - \alpha_{k+1} \cdot r^k, \quad k > 0,$$

где параметры α_{k+1} и β_{k+1} являются решением одной из выше сформулированных систем, зависящие только от x^{k-1} и x^k.

Лекция 10 (продолжение).

Переобусловливатель

Если систему уравнений $Ax = b$ преобразовать к системе уравнений $B^{-1}Ax = B^{-1}b$, то обусловленность матрицы $B^{-1}A$ новой системы может оказаться значительно меньше обусловленности матрицы A исходной системы и тогда влияние ошибок округления на решение системы уменьшится.

Матрица B называется переобусловливателем для матрицы A (для системы уравнений $Ax = b$).

Положительно определенные матрицы A и B называются эквивалентными по спектру с постоянными $\gamma_1 \ge \gamma_0 > 0$, если

$$\gamma_0(Bv, v) \le (Av, v) \le \gamma_1(Bv, v) \quad \forall\ v \in R^n.$$

Теорема. Если матрицы A и B самосопряжены и положительно определены, то все собственные значения матрицы $B^{-1}A$ вещественны, положительны и принадлежат интервалу $[\gamma_0, \gamma_1]$.

Док–во. Так как $B = B^* > 0$, то существует $B^{1/2} = (B^{1/2})^* > 0$.

Матрицы $B^{-1}A$ и $B^{1/2}(B^{-1}A)B^{-1/2}$ подобны, и, следовательно, имеют одинаковые собственные значения.

Матрица $B^{1/2}(B^{-1}A)B^{-1/2} = B^{-1/2}AB^{-1/2}$ самосопряжена и положительно определена, следовательно, ее собственные значения $\lambda(B^{-1}A)$ вещественны и положительны, а собственные векторы w вещественны: $[B^{-1/2}AB^{-1/2}]w = \lambda(B^{-1}A)w$.

Очевидно, что $v = B^{-1/2}w$ – собственный вектор матрицы $B^{-1}A$: $B^{-1}Av = \lambda(B^{-1}A)v$, и $\lambda(B^{-1}A) = (Av, v)/(Bv, v)$.

Так как $\gamma_0(Bv, v) \le (Av, v) \le \gamma_1(Bv, v) \quad \forall\ v \in R^n$, то $\gamma_0 \le \lambda(B^{-1}A) \le \gamma_1$.

Метод простой итерации с переобусловливателем

Теорема. Если $A = A^* > 0$ & $B = B^* > 0$, то метод простой итерации

$$\begin{cases} x^0 \ - \text{задан}, \\ x^{k+1} = x^k - \tau \cdot B^{-1}(Ax^k - b), \quad k = 0, 1, ..., \end{cases}$$

сходится при любом $\tau \in (0, 2/\gamma_1)$, а для его ошибки $z^{k+1} = S_\tau z^k$, $S_\tau \equiv E - \tau B^{-1}A$, справедливы оценки

$$\| z^{k+1} \|_A \le \| S_\tau \|_A \cdot \| z^k \|_A \le \rho_\tau \cdot \| z^k \|_A,$$

$$\| z^{k+1} \|_B \le \| S_\tau \|_B \cdot \| z^k \|_B \le \rho_\tau \cdot \| z^k \|_B, \quad \rho_\tau = \max\{|1-\tau\cdot\gamma_0|, |1-\tau\cdot\gamma_1|\} < 1,$$

при оптимальном параметре $\tau = \tau_{\text{опт}} = 2/(\gamma_1+\gamma_0)$ имеем $\rho_{\tau_{\text{опт}}} = \dfrac{\gamma_1-\gamma_0}{\gamma_1+\gamma_0}$.

Док-во. Мы знаем, что, если все собственные значения матрицы $B^{-1}A$ вещественны и положительны, то условием сходимости метода простой итерации является выбор параметра τ из интервала $(0, 2/\rho(S_\tau))$, а спектральный радиус $\rho(S_\tau)$ оценивается сверху величиной $\rho_\tau = \max\{|1-\tau\cdot\gamma_0|, |1-\tau\cdot\gamma_1|\}$, где γ_0 и γ_1 оценки спектра матрицы $B^{-1}A$. Следовательно, метод простой итерации сходится при любом $\tau\in(0, 2/\rho_\tau)\subset(0, 2/\rho(S_\tau))$ и ρ_τ минимально, если $\tau = \tau_{\text{опт}} = 2/(\gamma_1+\gamma_0)$.

Для доказательства оценок для ошибки $z^{k+1} = S_\tau z^k$ достаточно установить равенство энергетических A- и B-норм матрицы перехода $S_\tau \equiv E - \tau B^{-1}A$ ее спектральному радиусу:

$$\| S_\tau \|_A^2 = \inf_{z\ne 0} \frac{(A(E-\tau B^{-1}A)z, (E-\tau B^{-1}A)z)}{(Az, z)} =$$

$$= \inf_{v=A^{1/2}z\ne 0} \frac{((E-\tau A^{1/2}B^{-1}A^{1/2})v, (E-\tau A^{1/2}B^{-1}A^{1/2})v)}{(v, v)} =$$

$$= \| E-\tau A^{1/2}B^{-1}A^{1/2} \|_2^2 = \rho^2(E-\tau A^{1/2}B^{-1}A^{1/2}) = \rho^2(A^{1/2}S_\tau A^{-1/2}) = \rho^2(S_\tau),$$

так как матрица $E-\tau A^{1/2}B^{-1}A^{1/2}$ симметрична.
Аналогично,

$$\| S_\tau \|_B^2 = \inf_{z\ne 0} \frac{(B(E-\tau B^{-1}A)z, (E-\tau B^{-1}A)z)}{(Bz, z)} =$$

$$= \inf_{v=B^{1/2}z\ne 0} \frac{((E-\tau B^{-1/2}AB^{-1/2})v, (E-\tau B^{-1/2}AB^{-1/2})v)}{(v, v)} =$$

$$= \| E-\tau B^{-1/2}AB^{-1/2} \|_2^2 = \rho^2(E-\tau B^{-1/2}AB^{-1/2}) = \rho^2(B^{1/2}S_\tau B^{-1/2}) = \rho^2(S_\tau),$$

так как матрица $E-\tau B^{-1/2}AB^{-1/2}$ симметрична.

Теорема. Если $A = A^* > 0$ & $B > 0.5\tau A$ ($\tau > 0$), то метод простой итерации

$$\begin{cases} x^0 - \text{задан}, \\ x^{k+1} = x^k - \tau\cdot B^{-1}(Ax^k - b), \quad k = 0, 1, \ldots, \end{cases}$$

сходится.

Док–во. Энергетическая норма ошибки $z^{k+1} = S_\tau z^k$ метода простой итерации строго убывает, если $\tau \cdot ([B - 0.5\tau A]w, w) \; \forall \; w \neq 0$:

$$\| z^{k+1} \|_A^2 = (Az^{k+1}, z^{k+1}) = (Az^k, z^k) - 2\tau(A \underbrace{B^{-1}Az^k}_{w^k}, z^k) + \tau^2(A \underbrace{B^{-1}Az^k}_{w^k}, \underbrace{B^{-1}Az^k}_{w^k}) =$$

$$= (Az^k, z^k) - 2\tau([B - 0.5\tau A]w^k, w^k) < \| z^k \|_A^2$$

и, т.к. оператор S_τ непрерывен, то итерационный процесс сходится.

Метод наискорейшего спуска с переобусловливателем

На каждом шаге метода простой итерации параметр τ можно выбирать из условия минимизации энергетической нормы ошибки $z^{k+1} = S_\tau z^k$.

Теорема. Если $A = A^* > 0 \;\; \& \;\; B > 0$, то итерационный метод

$$\begin{cases} x^0 - \text{задан}, \\ x^{k+1} = x^k - \tau_{k+1} \cdot B^{-1}(Ax^k - b), \quad \tau_{k+1} = \dfrac{(B^{-1}r^k, r^k)}{(AB^{-1}r^k, B^{-1}r^k)}, \\ \quad k = 0, 1, \ldots, \end{cases}$$

сходится.

Если, кроме всего прочего, $B = B^*$ и известны границы γ_0 и γ_1 спектра матрицы $B^{-1}A$, то

$$\| z^{k+1} \|_A \leq \rho_{\tau_{опт}} \cdot \| z^k \|_A \leq (\rho_{\tau_{опт}})^{k+1} \cdot \| z^0 \|_A, \; \rho_{\tau_{опт}} = \frac{\gamma_1 - \gamma_0}{\gamma_1 + \gamma_0}.$$

Док–во. Очевидно, что минимум энергетической нормы ошибки

$$\| z^{k+1} \|_A^2 = \inf_\tau [(Az^k, z^k) - 2\tau \cdot (AB^{-1}Az^k, z^k) + \tau^2(AB^{-1}Az^k, B^{-1}Az^k)] =$$

$$= (Az^k, z^k) - \frac{(B^{-1}r^k, r^k)^2}{(AB^{-1}r^k, B^{-1}r^k)} \text{ при } \tau = \tau_{k+1} = \frac{(B^{-1}r^k, r^k)}{(AB^{-1}r^k, B^{-1}r^k)}$$

и норма ошибки строго убывает, так как $(B^{-1}r^k, r^k) \equiv ([B^{-1}r^k], B[B^{-1}r^k]) > 0$.

Так как оператор $S_{\tau_{k+1}} z^k = z^k - \tau_{k+1}(z^k) \cdot B^{-1}Az^k$ непрерывен всюду (кроме быть может 0, то итерационный процесс сходится.

Оценка нормы ошибки при $B = B^* > 0$ устанавливается элементарно:

$$\| z^{k+1} \|_A = \inf_\tau \| S_\tau z^k \|_A \leq \| S_{\tau_{опт}} z^k \|_A \leq \| S_{\tau_{опт}} \|_A \cdot \| z^k \|_A \leq \rho_{\tau_{опт}} \cdot \| z^k \|_A,$$

что и требовалось доказать.

Замечание. Очевидно, что при $B = E$ мы получим формулы метода наискорейшего спуска.

Метод сопряженных градиентов с переобусловливателем

Напомним, что метод сопряженных градиентов для решения системы $Ax=b$, $A=A^*>0$, был построен из условия минимизации энергетической нормы ошибки за m шагов. Проделаем аналогичную процедуру для решения системы $B^{-1}Ax=B^{-1}b$ с $A=A^*>0$, $B=B^*>0$, т.е. решим задачу построения ошибки z^m такой, что

$$\| z^m \|_A = \min_{\tau} \| (E-\tau_m B^{-1}A)\ldots(E-\tau_1 B^{-1}A)z^0 \|_A = \min_{\alpha} \| z^0-\alpha_1 g_1-\ldots-\alpha_m g_m \|_A,$$

где $L_m = L\{B^{-1}Az^0, \ldots, (B^{-1}A)^m z^0\} \equiv L\{g_1, \ldots, g_m\}$, а систему векторов $\{g_i\}_{i=1}^m$ предстоит построить.

Как и ранее, параметры $\alpha_1, \ldots, \alpha_m$ удовлетворяют системе уравнений

$$\frac{1}{2}\frac{\partial (Az^m, z^m)}{\partial \alpha_i} = (A\frac{\partial z^m}{\partial \alpha_i}, z^m) \equiv (Ag_i,\, z^0-\alpha_1 g_1-\ldots-\alpha_m g_m) = 0, \ \ i=1, \ldots, m,$$

или

$$\begin{bmatrix} (Ag_1,g_1) & (Ag_1,g_2) & \ldots & (Ag_1,g_m) \\ (Ag_2,g_1) & (Ag_2,g_2) & \ldots & (Ag_2,g_m) \\ \vdots & \vdots & \ldots & \vdots \\ (Ag_m,g_1) & (Ag_m,g_2) & \ldots & (Ag_m,g_m) \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix} = \begin{pmatrix} (Az^0,g_1) \\ (Az^0,g_2) \\ \vdots \\ (Az^0,g_m) \end{pmatrix}.$$

Матрица этой системы – матрица Грамма в скалярном произведении $(x,y)_A=(Ax,y)$ базиса $\{g_i\}_{i=1}^m$ в L_m, ее определитель не равен нулю, следовательно решение α существует и единственно.

Если базис $\{g_i\}_{i=1}^m$ является A–ортогональным, т.е. $(Ag_i,g_j)=0, \ \ i\neq j$, то

$$\alpha_k = \frac{(Az^0,g_k)}{(Ag_k,g_k)} \equiv \frac{(r^0,g_k)}{(Ag_k,g_k)}, \qquad \begin{cases} z^m = z^0-\alpha_1 g_1-\alpha_2 g_2-\ldots-\alpha_m g_m, \\ x^m = x^0-\alpha_1 g_1-\alpha_2 g_2-\ldots-\alpha_m g_m. \end{cases}$$

Построение A–ортогонального базиса

A–ортогонализуем систему векторов $\{B^{-1}Az^0, \ldots, (B^{-1}A)^m z^0\}$.

Зададим $g_1 = B^{-1}Az^0 \equiv B^{-1}r^0 \in L_1$, тогда $\alpha_1 = \dfrac{(r^0,g_1)}{(Ag_1,g_1)}$, $z^1=z^0-\alpha_1 g_1$.

Вычислим вектор $r^1 = Az^1 = Az^0-\alpha_1 Ag_1 = Az^0-\alpha_1 A(B^{-1}Az^0)$.

Легко проверить, что $B^{-1}r^1 = B^{-1}Az^0-\alpha_1(B^{-1}A)^2 z^0 \in L_2$ и $(r^1,g_1)=0$, т.е. $r^1 \perp L_1$ относительно обычного скалярного произведения.

Далее применяем метод матиндукции.

Предположим, что мы построили A-ортогональный базис $\{g_i\}_{i=1}^k$ в $L_k = L\{(B^{-1}A)z^0, \ldots, (B^{-1}A)^k z^0\}$ такой, что

1. $\{g_i\}_{i=1}^{j}$ – A-ортогональный базис в $L_j = L\{(B^{-1}A)z^0, ..., (B^{-1}A)^j z^0\}$, $j \le k$,
2. $z^j = z^0 - \alpha_1 \cdot g_1 - ... - \alpha_j \cdot g_j$, $\alpha_i = \dfrac{(r^0, g_i)}{(Ag_i, g_i)}$,
3. $r^j = Az^j$: $B^{-1}r^j \in L_{j+1}$ и $r^j \perp L_j$, т.е. $(r^j, g_i) = 0 \;\; \forall\, i = \overline{1, j}$.

Определим $g_{k+1} = B^{-1}r^k - \gamma_1 g_1 - ... - \gamma_k g_k \in L_{k+1}$ из условий

$$(Ag_{k+1}, g_i) \equiv (AB^{-1}r^k, g_i) - \gamma_1 \cdot (Ag_1, g_i) - ... - \gamma_k \cdot (Ag_k, g_i) = 0,$$
$$i = 1, ..., k.$$

- Так как $(Ag_j, g_i) = 0,\ i \ne j$, то $\gamma_i = (AB^{-1}r^k, g_i)/(Ag_i, g_i)$,
- так как $r^k \perp L_k$ и $B^{-1}Ag_i \in L_{i+1} \subset L_k$ при $i + 1 \le k$, то $(AB^{-1}r^k, g_i) = (r^k, B^{-1}Ag_i) = 0$ и $\gamma_i = 0,\ i = \overline{1, k-1}$,

  (здесь потребовалась симметричность матрицы B^{-1}).

Тогда

1. $g_{k+1} = B^{-1}r^k - \gamma_k g_k$, $\gamma_k = (B^{-1}r^k, Ag_k)/(Ag_k, g_k)$, и $\{g_i\}_{i=1}^{k+1}$ – A-ортогональный базис в $L_{k+1} = L\{(B^{-1}A)z^0, ..., (B^{-1}A)^{k+1} z^0\}$,
2. $z^{k+1} = z^0 - \alpha_1 \cdot g_1 - ... - \alpha_{k+1} \cdot g_{k+1} = z^k - \alpha_{k+1} \cdot g_{k+1}$, $\alpha_{k+1} = \dfrac{(r^0, g_{k+1})}{(Ag_{k+1}, g_{k+1})}$,
3. $r^{k+1} = Az^{k+1} = r^0 - \alpha_1 \cdot Ag_1 - ... - \alpha_{k+1} \cdot Ag_{k+1} \perp L_{k+1}$, т.к. легко проверяются равенства $(r^{k+1}, g_i) = 0,\ i = \overline{1, k+1}$;

   и, очевидно, что

$$B^{-1}r^{k+1} = B^{-1}Az^{k+1} \in L\{\ B^{-1}r^0,\ L\{(B^{-1}A)^2 z^0, ..., (B^{-1}A)^{k+2} z^0\}\ \} \equiv L_{k+2}.$$

Таким образом построен очередной базисный вектор g_{k+1} и все предположения метода матиндукции для него выполняются, из которых следует, что

$$\alpha_{k+1} = \frac{(r^0, g_{k+1})}{(Ag_{k+1}, g_{k+1})} = \frac{(r^1 + \alpha_1 Ag_1, g_{k+1})}{(Ag_{k+1}, g_{k+1})} = \frac{(r^1, g_{k+1})}{(Ag_{k+1}, g_{k+1})} = ... = \frac{(r^k, g_{k+1})}{(Ag_{k+1}, g_{k+1})}.$$

Формулы метода сопряженных градиентов с переобусловливателем

$$\vec{g}_1 = B^{-1}\vec{r}^{\,0} \equiv B^{-1}(A\vec{x}^0 - \vec{b},)$$
$$\vec{x}^1 = \vec{x}^0 - \alpha_1 \cdot \vec{g}_1, \qquad \alpha_1 = (\vec{r}^{\,0}, \vec{g}_1) / (A\vec{g}_1, \vec{g}_1)$$
$$\vec{g}_{k+1} = B^{-1}\vec{r}^{\,k} - \gamma_k \vec{g}_k, \qquad \gamma_k = (B^{-1}\vec{r}^{\,k}, A\vec{g}_k) / (A\vec{g}_k, \vec{g}_k)$$
$$\vec{x}^{k+1} = \vec{x}^k - \alpha_{k+1} \cdot \vec{g}_{k+1}, \quad \alpha_{k+1} = (\vec{r}^{\,k}, \vec{g}_{k+1}) / (A\vec{g}_{k+1}, \vec{g}_{k+1})$$

Лекция 11. Проблема собственных значений

Для матрицы $A=\{a_{ij}\}_{i,j=1}^{n}$ нужно найти числа λ и ненулевые векторы x такие, что $Ax=\lambda x$: λ – собственное значение, x – собственный вектор.

Корректность задачи на собственные значения

Известно, что все собственные значения матрицы A являются корнями характеристического полинома

$$P_n(\lambda)=\det(A-\lambda E)=(-1)^n\lambda^n+p_{n-1}\lambda^{n-1}+\ldots+p_1\lambda+p_0,$$

а коэффициенты $p_0,\ldots,p_{n-1}$ – непрерывные функции элементов матрицы A.

Пусть δA – матрица с “малыми” по величине элементами, $P_{\delta,n}(\lambda)$ – характеристический полином матрицы $A+\delta A$. Следствием непрерывности $\det(A+\delta A)$ как функции элементов матрицы $A+\delta A$ является

Лемма 1. $\lim\limits_{\delta A\to 0} P_{\delta,n}(\lambda)=P_n(\lambda) \quad \forall\,\lambda\in C$.

Лемма 2. В любом круге на комплексной плоскости с центром в точке λ_c и радиуса $\sqrt[n]{|P_n(\lambda_c)|}$ лежит хотя бы один корень полинома $P_n(\lambda)$.

Док–во. Разложим $P_n(\lambda)$ в ряд Тейлора в точке λ_c:

$$P_n(\lambda)=P_n(\lambda_c)+\frac{P_n'(\lambda_c)}{1!}(\lambda-\lambda_c)+\ldots+\frac{P_n^{(n)}(\lambda_c)}{n!}(\lambda-\lambda_c)^n\equiv Q(z),$$ где $z=\lambda-\lambda_c$.

Пусть $z_1,\ldots,z_n$ – корни полинома $Q(z)$, среди которых корень с минимальной абсолютной величиной имеет номер $\min$.

Так как $|P_n(\lambda_c)|=|Q(0)|=|z_1\cdot\ldots\cdot z_n|\ge|z_{\min}|^n=|\lambda_{\min}-\lambda_c|^n$, то $\lambda_{\min}$ (корень полинома $P_n(\lambda)$) лежит в круге радиуса $\sqrt[n]{|P_n(\lambda_c)|}$.

Лемма 3. Если $\lambda_1,\ldots,\lambda_n$ – корни полинома $P_n(\lambda)$, то $\exists$ нумерация корней $\lambda_{\delta,1},\ldots,\lambda_{\delta,n}$ полинома $P_{\delta,n}(\lambda)$: $\lambda_{\delta,k}\to\lambda_k \quad \forall\,k$ при $\delta A\to 0$.

Док–во методом матиндукции по степени полинома.

$n=1 \Rightarrow \lambda_{\delta,1}=p_{\delta,0}\to p_0=\lambda_1$.

Пусть лемма верна при $n<k$.

$n=k$: из леммы 2 $\Rightarrow \exists\,\lambda_{\delta,1}: |\lambda_{\delta,1}-\lambda_1|\le\sqrt[n]{|P_{\delta,n}(\lambda_1)|}\to 0$.

Т.к. $P_n(\lambda)=(\lambda-\lambda_1)R_{n-1}(\lambda)$, $P_{\delta,n}(\lambda)=(\lambda-\lambda_{\delta,1})R_{\delta,n-1}(\lambda)$

и $R_{\delta,n-1}(\lambda)\to R_{n-1}(\lambda)$, то $\lambda_{\delta,2}\to\lambda_2,\ldots,\lambda_{\delta,n}\to\lambda_n$.

Степенной метод вычисления максимального собственного значения матрицы $A = A^* \geq 0$

Идея метода: для заданного вектора x^0 рассмотрим его k–ю итерацию $A^k x^0$,
если $0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{n-1} < \lambda_n = \rho(A)$ – собственные значения,
$q^{(1)}, q^{(2)}, \ldots, q^{(n)}$ – соответствующие им собственные векторы, то

$$A^k x^0 = \rho^k [\alpha_n q^{(n)} + \underbrace{(\frac{\lambda_1}{\rho})^k \alpha_1 q^{(1)} + \ldots + (\frac{\lambda_{n-1}}{\rho})^k \alpha_{n-1} q^{(n-1)}}_{\to 0}] \approx \rho^k \alpha_n q^{(n)},$$

$$\frac{\| A^{k+1} x^0 \|}{\| A^k x^0 \|} \approx \rho, \qquad \frac{1}{\| A^k x^0 \|} A^{k+1} x^0 \approx \rho \cdot q^{(n)},$$

где $\alpha_1, \alpha_2, \ldots, \alpha_n$ – коэффициенты (неизвестные!)
разложения вектора x^0 по базису $q^{(1)}, q^{(2)}, \ldots, q^{(n)}$.

Итерационный процесс

$$x^0 \neq 0, \quad x^{k+1} = A \frac{x^k}{\| x^k \|}, \qquad k = 0, 1, \ldots,$$

называется степенным методом вычисления максимального собственного значения матрицы $A = A^* \geq 0$**:**

$$\| x^k \| \to \rho(A), \qquad x^k \to x: \quad Ax = \rho \cdot x,$$

если проекция начального вектора x^0 на линейную оболочку собственных векторов, соответствующих $\rho(A)$, не равна 0.

Док–во. Пусть $0 \leq \lambda_1 \leq \ldots \leq \lambda_r < \lambda_{r+1} = \ldots = \lambda_n = \rho$ – собственные значения,
$q^{(1)}, \ldots, q^{(r)},\ q^{(r+1)}, \ldots, q^{(n)}$ – собственные векторы матрицы A, и

$$x^0 = \alpha_1 q^{(1)} + \ldots + \alpha_r q^{(r)} + \alpha_{r+1} q^{(r+1)} + \ldots + \alpha_n q^{(n)} =$$
$$= \alpha_1 q^{(1)} + \ldots + \alpha_r q^{(r)} + y, \quad y \neq 0.$$

Тогда $A^k x^0 = \rho^k [x + (\lambda_1/\rho)^k \alpha_1 q^{(1)} + \ldots + (\lambda_r/\rho)^k \alpha_r q^{(r)}]$ и,

т.к. $x^k = \dfrac{Ax^{k-1}}{\| x^{k-1} \|} = \dfrac{A^2 x^{k-2}}{\| Ax^{k-2} \|} = \ldots = \dfrac{A^k x^0}{\| A^{k-1} x^0 \|}$, $0 \leq \dfrac{\lambda_1}{\rho} \leq \ldots \leq \dfrac{\lambda_r}{\rho} < 1$,

то $\| x^k \| = \rho \dfrac{\| y + (\lambda_1/\rho)^k \alpha_1 q^{(1)} + \ldots + (\lambda_r/\rho)^k \alpha_r q^{(r)} \|}{\| y + (\lambda_1/\rho)^{k-1} \alpha_1 q^{(1)} + \ldots + (\lambda_r/\rho)^{k-1} \alpha_r q^{(r)} \|} \to \rho$,

$$x^k = \rho \frac{y + (\lambda_1/\rho)^k \alpha_1 q^{(1)} + \ldots + (\lambda_r/\rho)^k \alpha_r q^{(r)}}{\| y + (\lambda_1/\rho)^{k-1} \alpha_1 q^{(1)} + \ldots + (\lambda_r/\rho)^{k-1} \alpha_r q^{(r)} \|} \to x = \rho \frac{y}{\| y \|}.$$

Замечание. Сходимость степенного метода не зависит от выбора в нем векторной нормы, т.к. все нормы в R^n эквивалентны.

Степенной метод вычисления минимального собственного значения матрицы $A = A^* \geq 0$

Задача вычисления минимального собственного значения матрицы $A = A^* \geq 0$ легко сводится к задаче вычисления максимального собственного значения матрицы $\beta \cdot E - A \geq 0$, где $\beta \geq \rho(A)$, так как $\rho(\beta \cdot E - A) = \beta - \lambda_{\min}(A)$.

Оценку для $\rho(A)$ легко найти: $\beta = \| A \|_\infty \geq \rho(A)$. Тогда

итерационный процесс

$$x^0 \neq 0, \quad x^{k+1} = (\| A \|_\infty E - A)\frac{x^k}{\| x^k \|}, \quad k = 0, 1, \ldots,$$

называется степенным методом вычисления минимального собственного значения матрицы $A = A^* \geq 0$**:** $(\| A \|_\infty - \| x^k \|) \to \lambda_{\min}(A)$**,**

если проекция начального вектора x^0 на линейную оболочку собственных векторов, соответствующих $\lambda_{\min}(A)$, не равна 0.

Справедливость этого утверждения является следствием сходимости степенного метода вычисления спектрального радиуса матрицы $B = \| A \|_\infty E - A$.

Применение ортогонализации и степенного метода для вычисления очередного собственного значения

Предположим, что собственное значение $\lambda_n = \rho(A)$ и соответствующий ему собственный вектор (какой–то!) $q^{(n)}$ матрицы $A = A^* \geq 0$ мы приближенно (например степенным методом) вычислили: $\tilde{\lambda}_n \approx \lambda_n$, $\tilde{q}^{(n)} \approx q^{(n)}$.

Построим симметричную положительно определенную матрицу $\tilde{A}_{n-1} = \tilde{P}_n A \tilde{P}_n$, где матрица $\tilde{P}_n = E - \tilde{q}^{(n)}[\tilde{q}^{(n)}]^T$ – ортогональный проектор на подпространство $(L\{\tilde{q}^{(n)}\})^\perp$, ортогональное вектору $\tilde{q}^{(n)}$.

Докажите, что спектр матрицы A_{n-1} (т.е. $\tilde{\lambda}_n = \lambda_n$, $\tilde{q}^{(n)} = q^{(n)}$) состоит из собственных значений $\lambda_1 \leq \ldots \leq \lambda_{n-1}$ матрицы A и нуля (вектор $q^{(n)}$ принадлежит ее ядру).

Отсюда следует, что, если $\tilde{q}^{(n)} \to q^{(n)}$ (а степенной метод такую сходимость гарантирует), то $\rho(\tilde{A}_{n-1}) \to \rho(A_{n-1}) = \lambda_{n-1}(A)$.

Следовательно, применяя степенной метод для матрицы $\tilde{A}_{n-1}$, мы получим приближение к $\lambda_{n-1}(A)$ и $q^{(n-1)}$ – очередным собственным значению и вектору матрицы A.

Эту процедуру можно продолжать до тех пор, пока мы не получим все собственные значения.

Степенной метод вычисления границ спектра матрицы $B^{-1}A$ в случае $A = A^* > 0$ и $B = B^* > 0$

Знание оценок спектра матрицы $B^{-1}A$ в случае $A = A^* > 0$ и $B = B^* > 0$ необходимо для построения параметров циклического метода Ричардсона (простой итерации) для решения системы уравнений $Ax = b$ с переобусловливателем B.

Лемма 4. Все собственные значения матрицы $B^{-1}A$ в случае $A = A^* > 0$ и $B = B^* > 0$ положительны, а соответствующие им собственные векторы A-ортогональны.

Док-во. Задача на собственные значения $B^{-1}A \cdot x = \lambda \cdot x$ эквивалентна задаче

$$(A^{1/2}B^{-1}A^{1/2})[A^{1/2}x] = \lambda \cdot [A^{1/2}x]$$

на собственные значения самосопряженной положительно определенной матрицы $C = A^{1/2}B^{-1}A^{1/2} = C^* > 0$, имеющей положительные собственные значения $\{\lambda_i\}_{i=1}^n$ и систему собственных векторов $\{y^{(i)} = A^{1/2}x^{(i)}\}_{i=1}^n$, образующую ортонормированный базис:

$$(y^{(i)}, y^{(j)}) \equiv (A^{1/2}x^{(i)}, A^{1/2}x^{(j)}) = \delta_{i,j}, \ i, j = 1, 2, ..., n.$$

Очевидно, что $\{\lambda_i\}_{i=1}^n$ и $\{x^{(i)} = A^{-1/2}y^{(i)}\}_{i=1}^n$ являются системой собственных значений и векторов матрицы $B^{-1}A$ и собственные векторы A-ортогональны:

$$(Ax^{(i)}, x^{(j)}) \equiv (A^{1/2}x^{(i)}, A^{1/2}x^{(j)}) = \delta_{i,j}, \ i, j = 1, 2, ..., n.$$

Теорема 1. Если $A = A^* > 0$ и $B = B^* > 0$, то степенной метод

$$x^0 \neq 0 - \text{задан}, \quad x^{k+1} = B^{-1}A\frac{x^k}{\| x^k \|_A}, \ k = 0, 1, ...$$

сходится к решению задачи $B^{-1}A \cdot x = \rho \cdot x$,

где $\rho = \max\limits_{\lambda \in Sp(B^{-1}A)} \lambda(B^{-1}A) = \lim\limits_{k\to\infty} \| x^k \|_A$ – максимальное собственное значение, а $x = \lim\limits_{k\to\infty} x^k$ – соответствующий ему собственный вектор, если $(Ax^0, x) \neq 0$.

Док-во. Пусть $0 \le \lambda_1 \le ... \le \lambda_r < \lambda_{r+1} = ... = \lambda_n = \rho$ – собственные значения, $\{x^{(i)}\}_{i=1}^n$ – A-ортонормальной система собственных векторов из леммы 4 матрицы $B^{-1}A$. Представим начальное приближение степенного метода в виде разложения:

$$x^0 = \underbrace{\alpha_1 \cdot x^{(1)} + ... + \alpha_r \cdot x^{(r)}}_{\varepsilon^0} + \underbrace{\alpha_{r+1} \cdot x^{(r+1)} + ... + \alpha_n \cdot x^{(n)}}_{y \neq 0}$$

и предположим, что $y \neq 0$. Очевидно, что $B^{-1}A \cdot y = \rho \cdot y$,

$$x^k = \frac{(B^{-1}A)x^{k-1}}{\| x^{k-1} \|_A} = \frac{(B^{-1}A)^2 x^{k-2}}{\| (B^{-1}A)x^{k-2} \|_A} = \ldots = \frac{(B^{-1}A)^k x^0}{\| (B^{-1}A)^{k-1} x^0 \|_A},$$

$$(B^{-1}A)^k x^0 = \rho^k \cdot [\underbrace{\alpha_1(\frac{\lambda_1}{\rho})^k \cdot x^{(1)} + \ldots + \alpha_r(\frac{\lambda_r}{\rho})^k \cdot x^{(r)}}_{\varepsilon^k \to 0} + y],$$

$$\| (B^{-1}A)^{k-1} x^0 \|_A = \rho^{k-1} \cdot \sqrt{\underbrace{[\alpha_1(\frac{\lambda_1}{\rho})^{k-1}]^2 + \ldots + [\alpha_r(\frac{\lambda_r}{\rho})^{k-1}]^2}_{\|\varepsilon^{k-1}\|_A^2 \to 0} + \| y \|_A^2}.$$

Тогда $x^k \to \rho \cdot \frac{y}{\| y \|_A}$, а $\| x^k \|_A \to \rho$. Кроме того,

$$0 \le \frac{\rho - \| x^k \|_A}{\rho} = \frac{\rho - \rho\sqrt{\| y \|_A^2 + \| \varepsilon^k \|_A^2} / \sqrt{\| y \|_A^2 + \| \varepsilon^{k-1} \|_A^2}}{\rho} =$$

$$= \frac{\| \varepsilon^{k-1} \|_A^2 - \| \varepsilon^k \|_A^2}{(\sqrt{\| y \|_A^2 + \| \varepsilon^{k-1} \|_A^2} + \sqrt{\| y \|_A^2 + \| \varepsilon^k \|_A^2}) \cdot \sqrt{\| y \|_A^2 + \| \varepsilon^{k-1} \|_A^2}} \le$$

$$\le \frac{\| \varepsilon^{k-1} \|_A^2}{2 \cdot \| y \|_A^2} \le \frac{[\alpha_1]^2 + \ldots + [\alpha_r]^2}{2 \cdot \| y \|_A^2} (\frac{\lambda_r}{\rho})^{2(k-1)} < \frac{\| x^0 \|_A^2}{2 \cdot \| y \|_A^2} (\frac{\lambda_r}{\rho})^{2(k-1)}$$

что и требовалось доказать.

Теорема 2. Если $A = A^* > 0$, $B = B^* > 0$ и известна постоянная $\gamma_1 \ge \rho(B^{-1}A)$, то степенной метод

$$x^0 \ne 0 - \text{задан}, \quad x^{k+1} = (\gamma_1 \cdot E - B^{-1}A) \frac{x^k}{\| x^k \|_A}, \; k = 0, 1, \ldots$$

сходится к решению задачи $B^{-1}A \cdot x = (\gamma_1 - \lambda_{min}) \cdot x$,

где $\lambda_{min} = \min\limits_{\lambda \in Sp(B^{-1}A)} \lambda(B^{-1}A) = \gamma_1 - \lim\limits_{k \to \infty} \| x^k \|_A$ – минимальное собственное значение, а $x = \lim\limits_{k \to \infty} x^k$ – соответствующий ему собственный вектор, если $(Ax^0, x) \ne 0$.

Док-во. Повторяя доказательство теоремы 1, получим, что сформулированный метод определяет спектральный радиус и соответствующий собственный вектор матрицы $\gamma_1 \cdot E - B^{-1}A$. Так как $\lambda(\gamma_1 \cdot E - B^{-1}A) = \gamma_1 - \lambda(B^{-1}A)$ и $\lambda(B^{-1}A) > 0$, то $\beta \equiv \rho(\gamma_1 \cdot E - B^{-1}A) = \gamma_1 - \lambda_{min}(B^{-1}A)$, т.е. $\lambda_{min}(B^{-1}A) = \gamma_1 - \beta$.

Лекция 12. Метод деления пополам (бисекций)

Для самосопряженной матрицы $A=\{a_{ij}\}_{i,j=1}^{n}$ имеет место **закон инерции:** если матрицу A конгруэнтным преобразованием привести к диагональному виду: $D=T^{*}AT$, где $\det T\neq 0$, то от матрицы T (способа преобразования) не зависит

- $\sigma_{-}(A)$ – количество отрицательных элементов,
- $\sigma_{0}(A)$ – количество нулевых элементов,
- $\sigma_{+}(A)$ – количество положительных элементов на диагонали D.

Нам известно (из теоремы и алгоритма LDU–разложения), что если все $\det A_k\neq 0$, то $A=LDL^{*},\ D=\mathrm{diag}\{d_1,\ \dots,\ d_n\},\ \det A_k=d_1\cdot\ldots\cdot d_k$.
Следовательно, в этом случае за конечное число действий мы можем определить $\sigma(A)=\{\sigma_{-}(A),\ \sigma_{0}(A),\ \sigma_{+}(A)\},\ \sigma_{0}(A)=0$.

Матрица $A=A^{*}$ преобразованием подобия ортогональной матрицей Q (конгруэнтным преобразованием) из собственных векторов приводится к диагональному виду $\Lambda=\mathrm{diag}\{\lambda_1,\ \dots,\ \lambda_n\}=Q^{*}AQ$. Следовательно,

- $\sigma_{-}(A)$ = количеству отрицательных,
- $\sigma_{0}(A)$ = количеству нулевых,
- $\sigma_{+}(A)$ = количеству положительных собственных значений матрицы A,

и, используя LDL^{*}–разложение, мы можем эти числа определить.
Подытожим эти рассуждения в виде следующей леммы.

Лемма 1. Если матрица $A=A^{*}$ и $\det A_k\neq 0\ \ \forall\ k$,
то количество ее отрицательных собственных значений

$$\sigma_{-}(A)=\text{ЧПЗ}\{1,\ \det A_1,\ \det A_2,\ \dots,\ \det A_n\}$$

– число перемен знака.

Док–во леммы оставляется в виде упражнения.

Идея метода бисекций вычисления $\lambda_j\in Sp(A)$

$\lambda_j\in[a_0,b_0]=[-\|A\|_\infty,\ \|A\|_\infty]$, т.к. $\rho(A)\leq\|A\|_\infty$, т.е. все собственные значения $\lambda_1\leq\lambda_2\leq\ldots\leq\lambda_n$ матрицы $A=A^{*}$ лежат в этом интервале.
Определим в какой половине интервала $[a_0,b_0]$ лежит λ_j. Для этого вычислим $\sigma_{-}(A-c_0E)$ – количество собственных значений меньших $c_0=(a_0+b_0)/2$.
Если $\sigma_{-}(A-c_0E)\geq j$, то $\lambda_j\in[a_0,c_0]\equiv[a_1,b_1]$, иначе $\lambda_j\in[c_0,b_0]\equiv[a_1,b_1]$.
Через k таких шагов получим: $\lambda_j\in[a_k,b_k],\ b_k-a_k=\|A\|_\infty/2^{k-1}\to 0$, т.е. мы можем получить оценку искомого собственного числа с любой точностью.

Приведение самосопряженной матрицы к трехдиагональному виду ортогональным преобразованием подобия с помощью матриц вращения

Как и раньше, через $Q_{i,i+k}$ будем обозначать элементарную матрицу вращения, отличающуюся от единичной матрицы
двумя диагональными элементами: $(Q_{i,i+k})_{i,i} = \overline{c}_{i,i+k}$, $(Q_{i,i+k})_{i+k,i+k} = c_{i,i+k}$, и двумя внедиагональными элементами: $(Q_{i,i+k})_{i,i+k} = -\overline{s}_{i,i+k}$, $(Q_{i,i+k})_{i+k,i} = s_{i,i+k}$, $|c_{i,i+k}|^2 + |s_{i,i+k}|^2 = 1$.
Выполним и

1–й шаг. Исключение элементов **1–го столбца** матрицы A, начиная с **3–его**, с помощью последовательного умножения на унитарные матрицы $Q_{2,3}, \ldots, Q_{2,n}$: $A_1 = (Q_{2,n} \cdot \ldots \cdot Q_{2,3})A(Q_{2,n} \cdot \ldots \cdot Q_{2,3})^* \equiv Q_1 A Q_1^*$.

2–й шаг. Исключение элементов **2–го столбца** матрицы A_1, начиная с **4–ого**, с помощью последовательного умножения на унитарные матрицы $Q_{3,4}, \ldots, Q_{3,n}$: $A_2 = (Q_{3,n} \cdot \ldots \cdot Q_{3,4})A_1(Q_{3,n} \cdot \ldots \cdot Q_{3,4})^* \equiv Q_2 A_1 Q_2^*$.

… …………………..

k–й шаг. Исключение элементов **k–го столбца** матрицы A_{k-1}, начиная с **(k+2)–ого**, с помощью последовательного умножения на матрицы $Q_{k+1,k+2}, \ldots, Q_{k+1,n}$:

$$A_k = (Q_{k+1,n} \cdot \ldots \cdot Q_{k+1,k+2})A_{k-1}(Q_{k+1,n} \cdot \ldots \cdot Q_{k+1,k+2})^* \equiv Q_k A_{k-1} Q_k^*.$$

… …………………..

(n–2)–й шаг. Исключение последнего элемента **(n-2)–го столбца** матрицы A_{n-3} с помощью умножения на матрицу $Q_{n-1,n}$:

$$A_{n-2} = (Q_{n-1,n})A_{n-3}(Q_{n-1,n})^* \equiv Q_{n-2} A_{n-3} Q_{n-2}^*.$$

$$T = A_{n-2} = (Q_{n-2} \cdot \ldots \cdot Q_1)A(Q_{n-2} \cdot \ldots \cdot Q_1)^* = \begin{bmatrix} \alpha_1 & \beta_1 & & & \\ \overline{\beta}_1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \overline{\beta}_{n-2} & \alpha_{n-1} & \beta_{n-1} \\ & & & \overline{\beta}_{n-1} & \alpha_n \end{bmatrix},$$

$$Sp(A) = Sp(T).$$

Если $\beta_k = 0$, то $T = \begin{bmatrix} T_k & 0 \\ 0 & \hat{T}_{n-k} \end{bmatrix} \Rightarrow Sp(T) = Sp(T_k) \bigcup Sp(\hat{T}_{n-k})$,

т.е. поиск собственных значений самосопряженной матрицы сводится к задаче на собственные значения якобиевых трехдиагональных матриц.

Лемма 2. Самосопряженная матрица подобна трехдиагональной вещественной матрице.

Док–во. Только что мы привели самосопряженную матрицу A к трехдиагональному виду ортогональным преобразованием подобия: $T=(Q_{n-2}\cdot\ldots\cdot Q_1)A(Q_{n-2}\cdot\ldots\cdot Q_1)^*=\mathrm{tridiag}\{\overline{\beta}_{i-1},\alpha_i,\beta_i\}$.

Определим матрицу $D=\mathrm{diag}\{d_1,\ \ldots,\ d_n\}$: (предполагая $\forall\ \beta_i\neq 0$)

$d_1=1,\ d_2=\beta_1/|\beta_1|,\ \ldots,\ d_n=\beta_1/|\beta_1|\cdot\ldots\cdot\beta_{n-1}/|\beta_{n-1}|$.

Тогда $D^{-1}=D^*,\ B=DTD^{-1}=\mathrm{tridiag}\{|\beta_{i-1}|,\alpha_i,|\beta_i|\}$.

Якобиевы матрицы

Вещественная матрица

$$B=\begin{bmatrix} a_1 & b_1 & & & \\ c_2 & a_2 & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & & c_{n-1} & a_{n-1} & b_{n-1} \\ & & & c_n & a_n \end{bmatrix},\ b_1\cdot c_2>0,\ b_2\cdot c_3>0,\ \ldots,\ b_{n-1}\cdot c_n>0,$$

называется **якобиевой** (у нас $c_i=b_{i-1}$).

Лемма 3. Пусть $B=\mathrm{tridiag}\{b_{i-1},a_i,b_i\}$ – якобиева матрица, тогда

1. $\det B_0\equiv 1,\ \det B_1=a_1,$
   $\det B_{i+1}=a_{i+1}\cdot\det B_i-b_i^2\cdot\det B_{i-1},$
   $i=1,\ \ldots,\ n-1.$
2. если $\det B_i=0\ \ (i<n)$, то $\det B_{i-1}\cdot\det B_{i+1}<0$, если $\det B_n=0$, то $\det B_{n-1}\neq 0$.

Док–во оставляется читателю в качестве упражнения.

Лемма 4. Собственные значения якобиевой матрицы B попарно различные (простые).

Док–во. Т.к. размерность ядра симметричной матрицы $B_\lambda=B-\lambda E$ совпадает с кратностью $\lambda\in Sp(B)$, а из леммы 3 следует, что у вырожденной якобиевой матрицы B_λ минор $[\det B_\lambda]_{n-1}\neq 0$, то $\mathrm{rang}\,B_\lambda=n-1,\ \dim\mathrm{Ker}\,B_\lambda=1$ и λ простое собственное значение матрицы B.

Теорема. Пусть $B = \mathrm{tridiag}\{b_{i-1}, a_i, b_i\}$ – якобиева матрица, тогда

$$\sigma\_(B) = \text{ЧПЗ}\{1,\ \det B_1,\ \det B_2, \ldots,\ \det B_n\},$$

если $\det B_k = 0$ приписать знак $\det B_{k-1}$.

Док–во. 1. Если $\det B_k \neq 0\ \forall\ k$, то это лемма 1.

2. Пусть $\exists\ k:\ \det B_k = 0$. Пусть $\mathrm{sign}(\det B_k) = \mathrm{sign}(\det B_{k-1})$.

Определим $\varepsilon_0 = \min\limits_{\lambda \in Sp(B_i),\ \lambda \neq 0,\ i=1,\ldots,n} |\lambda| > 0$ и рассмотрим якобиевы матрицы $B_{\pm\varepsilon} = B \pm \varepsilon E,\ \varepsilon \in (0,\ \varepsilon_0)$.

Т.к. $\lambda([B_{\pm\varepsilon}]_i) \equiv \lambda(B_i) \pm \varepsilon \neq 0\quad \forall\ i = 1, \ldots, n$, то

а) $\det[B_{\pm\varepsilon}]_i \neq 0\quad \forall\ i = 1, \ldots, n,$ (т.к. определитель матрицы равен произведению ее собственных значений),

б) $\mathrm{sign}(\det[B_\varepsilon]_i) = \mathrm{sign}(\det[B_{-\varepsilon}]_i) = \mathrm{sign}(\det B_i)\quad \forall\ \det B_i \neq 0$,

в) $\mathrm{sign}(\det[B_\varepsilon]_k) \cdot \mathrm{sign}(\det[B_{-\varepsilon}]_k) < 0\quad \forall\ \det B_k = 0$,

(т.к. из леммы 4 следует, что $\lambda(B_k) = 0$ простое и отрицательных собственных значений у матрицы $[B_{-\varepsilon}]_k$ на одно больше, чем у матрицы $[B_\varepsilon]_k$),

г) $\sigma\_(B_\varepsilon) = \sigma\_(B),\quad \sigma\_(B_{-\varepsilon}) = \sigma\_(B) + \sigma_0(B)$.

Из леммы 1, а) и г) следует, что

$$\sigma\_(B_\varepsilon) = \text{ЧПЗ}\{1,\ \det[B_\varepsilon]_1,\ \det[B_\varepsilon]_2, \ldots,\ \det[B_\varepsilon]_n\} = \sigma\_(B),$$

$$\sigma\_(B_{-\varepsilon}) = \text{ЧПЗ}\{1,\ \det[B_{-\varepsilon}]_1,\ \det[B_{-\varepsilon}]_2, \ldots,\ \det[B_{-\varepsilon}]_n\} = \sigma\_(B) + \sigma_0(B),$$

$$\text{ЧПЗ}\{1,\ \det B_1,\ \det B_2, \ldots,\ \det B_n\} = ?$$

Подсчитаем эти числа:

Из б) следует, что если $\det B_j \neq 0$ и $\det B_{j+1} \neq 0$, то перемена знака происходит (или нет) одновременно в этих последовательностях.

Случай $\det B_k = 0,\ k \neq n$.

Из леммы 3 имеем $\det B_{k-1} \cdot \det B_{k+1} < 0$, отсюда и из б) следует $\det[B_{\pm\varepsilon}]_{k-1} \cdot \det[B_{\pm\varepsilon}]_{k+1} < 0$ и на участках

$$\det[B_{\pm\varepsilon}]_{k-1},\quad \det[B_{\pm\varepsilon}]_k,\quad \det[B_{\pm\varepsilon}]_{k+1}$$

$$\det B_{k-1},\qquad \det B_k,\qquad \det B_{k+1}$$

по одной перемене знака.

Случай $\det B_n = 0,\ \sigma_0(B) = 1$. Отсюда, из в) и г) следует, что

$$\det[B_\varepsilon]_{n-1} \cdot \det[B_\varepsilon]_n > 0,\qquad \det[B_{-\varepsilon}]_{n-1} \cdot \det[B_{-\varepsilon}]_n < 0,$$

$$\mathrm{sign}(\det B_{n-1}) \cdot \mathrm{sign}(\det B_n) > 0.$$

Следовательно, (если $\det B_k = 0$ приписать знак $\det B_{k-1}$) последовательности миноров матриц B_ε и B имеют одинаковые знаки. Теорема доказана.

О вычислении ЧПЗ

Для вычисления $\sigma_-(B) = \text{ЧПЗ}\{1,\ \det B_1,\ \det B_2, \dots,\ \det B_n\}$ якобиевой матрицы $B = \text{tridiag}\{b_{i-1}, a_i, b_i\}$ достаточно знать знак каждого $\det B_k$. Если

$$\begin{aligned}
&d_0 = 1,\ d_1 = \det B_1,\\
&d_{i+1} = a_{i+1} \cdot d_i - b_i^2 \cdot d_{i-1},\ i = 1, \dots,\ k-1,\\
&d_{k-1} := d_{k-1} / |t_k|,\ \ d_k := d_k / |t_k|,\\
&d_{i+1} = a_{i+1} \cdot d_i - b_i^2 \cdot d_{i-1},\ i = k, \dots,\ n-1,
\end{aligned}$$

(обычно выбирают $t_k = \max\{|d_{k-1}|, |d_k|\}$), то $\text{sign}\{d_i\} = \text{sign}\{\det B_i\}\ \forall\ i$ и $\sigma_-(B) = \text{ЧПЗ}\{1,\ d_1, \dots,\ d_n\}$. Нормировку можно применять неоднократно, что позволит избежать быстрого роста (переполнения) чисел $\{d_i\}$.

О вычислении собственного вектора

Лемма 5. Последняя компонента собственного вектора x якобиевой матрицы $B = \text{tridiag}\{b_{i-1}, a_i, b_i\}$ не равна нулю.

Док–во. Пусть $Bx = \lambda x,\ x \neq 0$. Предположим, что $x_n = 0$. Тогда

$$\begin{aligned}
&x_{n-1} = -(a_n - \lambda) \cdot x_n / b_{n-1} = 0\\
&x_{n-i} = -[(a_{n-i+1} - \lambda) \cdot x_{n-i+1} + b_{n-i+1} \cdot x_{n-i+2}] / b_{n-i} = 0,\\
&\qquad i = 2, \dots,\ n-1,
\end{aligned}$$

$\Rightarrow\ x = 0$ – противоречие, значит $x_n \neq 0$.

Собственный вектор x якобиевой матрицы $B = \text{tridiag}\{b_{i-1}, a_i, b_i\}$ мы можем, положив $x_n = 1$, вычислить по формулам

$$\begin{aligned}
&x_{n-1} = -(a_n - \lambda) \cdot x_n / b_{n-1}\\
&x_{n-2} = -[(a_{n-1} - \lambda) \cdot x_{n-1} + b_{n-1} \cdot x_n] / b_{n-2}\\
&\quad \dots\dots\dots\dots\dots\\
&x_1 = -[(a_2 - \lambda) \cdot x_2 + b_2 \cdot x_3] / b_1
\end{aligned}$$

или решив систему

$$\begin{bmatrix} a_1 - \lambda & b_1 & & & \\ b_1 & a_2 - \lambda & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & & b_{n-3} & a_{n-2} - \lambda & b_{n-2} \\ & & & b_{n-2} & a_{n-1} - \lambda \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-2} \\ x_{n-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -b_{n-1} \end{pmatrix}$$

с неособенной матрицей.

Лекция 13. Метод вращений (Якоби)

Для самосопряженной матрицы $A=\{a_{ij}\}_{i,\,j=1}^{n}$ существует унитарная матрица Q (столбцы которой – собственные векторы матрицы A):

$$Q^{*}AQ=\Lambda\equiv \mathrm{diag}\{\lambda_1,\,...,\,\lambda_n\}$$

$\Rightarrow\ \Phi(Q^{*}AQ)=\min\limits_{T^{*}T=E}\Phi(T^{*}AT)$, где $\Phi(A)=\sum\limits_{i\neq j}|\,a_{ij}\,|^2$.

Идея:

построить $\{A_k=Q_k^{*}A_{k-1}Q_k : Q_k^{*}Q_k=E,\ A_0=A\}$: $\Phi(A_{k-1})>\Phi(A_k)\to 0$, тогда на диагональные элементы A_k будут приближать собственные значения, а столбцы $(Q_0...Q_{k+1})$ – собственные векторы матрицы A.

Определим $S(A)=\sum_{i,\,j=1}^{n}|\,a_{ij}\,|^2$.

Лемма 1. Для любых квадратной матрицы A и унитарной матрицы T имеем

$$S(TA)=S(AT)=S(A).$$

Док–во. Если $A=\left[a_1\ ...\ a_n\right]$, то

$$S(TA)=S([Ta_1\ \ ...\ \ Ta_n])=(Ta_1,Ta_1)+...+(Ta_n,Ta_n)=$$
$$=(a_1,a_1)+...+(a_n,a_n)=S(A).$$

В качестве матриц Q_k будем выбирать элементарные матрицы вращения.

Лемма 2. Пусть $A=A^{*}$, $\tilde{A}=Q_{ij}^{*}AQ_{ij}=\{\tilde{a}_{kl}\}$,

где Q_{ij} – элементарная матрица вращения, тогда

$$\Phi(\tilde{A})=\Phi(A)+\left[|\,a_{ii}\,|^2+|\,a_{jj}\,|^2-|\,\tilde{a}_{ii}\,|^2-|\,\tilde{a}_{jj}\,|^2\right].$$

Док–во. Заметим, что изменились только строки и столбцы с номерами i, j. Тогда, используя лемму 1, получим

$$S(A)\equiv\Phi(A)+\sum_{\substack{k=1\\ k\neq i,\,k\neq j}}^{n}|\,a_{kk}\,|^2+|\,a_{ii}\,|^2+|\,a_{jj}\,|^2\equiv$$

$$\equiv\Phi(\tilde{A})+\sum_{\substack{k=1\\ k\neq i,\,k\neq j}}^{n}|\,a_{kk}\,|^2+|\,\tilde{a}_{ii}\,|^2+|\,\tilde{a}_{jj}\,|^2\equiv S(\tilde{A})$$

откуда следует утверждение леммы.

Выбор вращения

Для простоты будем полагать, что матрица A вещественная. Выразим разность $\Phi(A)-\Phi(\tilde{A}) = |\tilde{a}_{ii}|^2 + |\tilde{a}_{jj}|^2 - |a_{ii}|^2 - |a_{jj}|^2$ через элементы матрицы A.

Лемма 3. Пусть $A = A^*$, $\tilde{A} = Q_{ij}^* A Q_{ij} = \{\tilde{a}_{kl}\}$, где Q_{ij} – элементарная матрица вращения (α – угол вращения), тогда

$$\Phi(A)-\Phi(\tilde{A}) = 2|a_{ij}|^2 - \frac{1}{2}\Big[(a_{ii}-a_{jj})\sin 2\alpha + 2a_{ij}\cos 2\alpha\Big]^2 =$$

$$= 2|a_{ij}|^2 - 2|\tilde{a}_{ij}|^2,$$

Док–во. Требуемые равенства выводятся из соотношения

$$\begin{bmatrix} \tilde{a}_{ii} & \tilde{a}_{ij} \\ \tilde{a}_{ij} & \tilde{a}_{jj} \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ij} & a_{jj} \end{bmatrix} \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}.$$

Лемма 4. Пусть $A = A^*$, $\tilde{A} = Q_{ij}^* A Q_{ij} = \{\tilde{a}_{kl}\}$,
где Q_{ij} – элементарная матрица вращения такая, что

$$|a_{ij}| = \max_{k\neq l}|a_{kl}|, \quad (a_{ii}-a_{jj})\sin 2\alpha + 2a_{ij}\cos 2\alpha = 0,$$

то $\Phi(\tilde{A}) \le \big[1 - 2/(n(n-1))\big]\cdot\Phi(A)$.

Док–во. Требуемое неравенство следует из равенства $\Phi(\tilde{A}) = \Phi(A) - 2|a_{ij}|^2$ и оценки $\Phi(A) \le n(n-1)|a_{ij}|^2$.

Следующая лемма обеспечивает существование для леммы 4 матрицы Q_{ij}.

Лемма 5. Решением уравнения $a\cdot\sin 2\alpha + 2b\cdot\cos 2\alpha = 0$ при $b \neq 0$ является угол α такой, что

$$\cos\alpha = \sqrt{\frac{1}{2}\Big(1-\frac{a}{r}\Big)}, \quad r = \sqrt{|a|^2 + 4|b|^2}, \text{ т.к. } \cos 2\alpha = 2\cos^2\alpha - 1 = -\frac{a}{r},$$

$$\sin\alpha = \frac{b}{r\cdot\cos\alpha} = \operatorname{sign} b\cdot\sqrt{\frac{1}{2}\Big(1+\frac{a}{r}\Big)}, \text{ т.к. } \sin 2\alpha = 2\cos\alpha\cdot\sin\alpha = \frac{2b}{r}.$$

Из последних двух лемм следует справедливость теоремы сходимости метода.

Теорема 1. Последовательность матриц $\{A_k\}_{k=0}^{\infty}$ **метода вращений**:
$A_0 = A$, $A_k = Q_k^* A_{k-1} Q_k$, где $Q_k = Q_{i(k),\, j(k)}$ – матрица вращения, определяемая по формулам лемм 4 и 5,
для решения полной проблемы на собственные значения $A = A^*$, сходится к диагональному виду, т.е. $\Phi(A_k) \to 0$, причем

$$\Phi(A_k) \le \big[1 - 2/(n(n-1))\big]^k \cdot \Phi(A).$$

Из теоремы 1 $\Rightarrow\ \forall\, \varepsilon > 0\ \exists\, k$:

$$\tilde{A} \equiv A_k = (Q_1 ... Q_k)^* A (Q_1 ... Q_k) \equiv \tilde{Q}^* A \tilde{Q}, \qquad \Phi(\tilde{A}) \le \varepsilon^2.$$

Пусть $\quad \tilde{\Lambda} = \operatorname{diag} \tilde{A} = \operatorname{diag}\{\tilde{\lambda}_1, ..., \tilde{\lambda}_n\},$

$$\Lambda = Q^* A\, Q = \operatorname{diag}\{\lambda_1, ..., \lambda_n\}.$$

Сходимость собственных значений

Лемма 6. $\tilde{P}(\lambda) \equiv \det(\tilde{\Lambda} - \lambda E) \ \to\ P(\lambda) \equiv \det(A - \lambda E)$ при $\varepsilon \to 0$.

Док–во. Т.к.

$$\det(\tilde{\Lambda} - \lambda E) = \det(\tilde{Q}\tilde{\Lambda}\tilde{Q}^* - \lambda E), \qquad \tilde{Q}\tilde{\Lambda}\tilde{Q}^* = A - \tilde{Q}(\tilde{A} - \tilde{\Lambda})\tilde{Q}^*,$$

$$S(\tilde{Q}(\tilde{A} - \tilde{\Lambda})\tilde{Q}^*) = S(\tilde{A} - \tilde{\Lambda}) = \Phi(\tilde{A}) \le \varepsilon^2 \to 0$$

то $\tilde{Q}\tilde{\Lambda}\tilde{Q}^* \to A \ \Rightarrow\ \det(\tilde{\Lambda} - \lambda E) \to \det(A - \lambda E)$.

Теорема 2 (оценка приближения собственных значений).

а) $\forall\, \lambda_i \ \ \exists\, \tilde{\lambda}_{j(i)}:\ |\lambda_i - \tilde{\lambda}_{j(i)}| \le \sqrt{n} \cdot \varepsilon,$

б) $\forall\, \tilde{\lambda}_j \ \ \exists\, \lambda_{i(j)}:\ |\tilde{\lambda}_j - \tilde{\lambda}_{i(j)}| \le \sqrt{n} \cdot \varepsilon.$

Док–во. Т.к. $Q\Lambda Q^* = A = \tilde{Q}\tilde{\Lambda}\tilde{Q}^* + \tilde{Q}(\tilde{A} - \tilde{\Lambda})\tilde{Q}^*$, то

$$\Lambda(Q^*\tilde{Q}) - (Q^*\tilde{Q})\tilde{\Lambda} = (Q^*\tilde{Q})(\tilde{A} - \tilde{\Lambda}) \equiv \mathcal{E}, \qquad |\varepsilon_{i\,j}|^2 \le S(\mathcal{E}) \le \varepsilon^2.$$

$\Rightarrow\ \lambda_i \cdot r_{i\,j} - r_{i\,j} \cdot \tilde{\lambda}_j = \varepsilon_{i\,j}$, где $\{r_{i\,j}\} = R = Q^*\tilde{Q}$ – ортогональная м–ца.

а) $\forall\, i\ \exists\, j(i): |r_{i\,j(i)}|^2 = \max\limits_k |r_{i\,k}|^2 \ge 1/n$, т.к. $|r_{i1}|^2 + ... + |r_{in}|^2 = 1$.

$\Rightarrow |\lambda_i - \tilde{\lambda}_{j(i)}| = |\varepsilon_{i\,j} / r_{i\,j(i)}| \le \sqrt{n} \cdot \varepsilon.$

б) доказывается аналогично.

Сходимость собственных векторов

Будем предполагать, что $\lambda_1 < \lambda_2 < ... < \lambda_n$ и $\tilde{\lambda}_1 \le \tilde{\lambda}_2 \le ... \le \tilde{\lambda}_n$ (этого всегда можно добиться, переставив столбцы матриц Q и $\tilde{Q}$).

Лемма 7. Если $\lambda_1 < \lambda_2 < ... < \lambda_n$, $\tilde{\lambda}_1 \le \tilde{\lambda}_2 \le ... \le \tilde{\lambda}_n$, $\sqrt{n} \cdot \varepsilon < 0.5 \cdot a$,

$$a = \min_{i \ne j} |\lambda_i - \lambda_j|,$$

то $|\lambda_i - \tilde{\lambda}_i| \le \sqrt{n} \cdot \varepsilon, \qquad |\lambda_i - \tilde{\lambda}_j| > 0.5 \cdot a \ \ \forall\, i \ne j$.

Док–во оставляется в качестве упражнения.

Т.к. собственные векторы $Q=\left[q_1 \ldots q_n\right]$ матрицы A определяются с точностью до их направления, будем считать, что $(q_i, \tilde{q}_i)\geq 0$ ($\left[\tilde{q}_1 \ldots \tilde{q}_n\right]=\tilde{Q}$ – приближения к собственным векторам матрицы A), т.е. диагональные элементы матрицы $R=Q^*\tilde{Q}$ неотрицательны.

Теорема 3 (оценка приближения собственных векторов).

В условиях леммы 7 $S(Q-\tilde{Q})\leq \dfrac{8}{a^2}\,\varepsilon^2$.

Док–во. Т.к. $S(Q-\tilde{Q})=S(E-Q^*\tilde{Q})\equiv S(E-R)$ и из доказательства теоремы 2 ($\mathcal{E}=\Lambda R-R\tilde{\Lambda}=R(\tilde{A}-\tilde{\Lambda})$) и леммы 7 следует, что $|r_{i\,j}|=\dfrac{|\varepsilon_{i\,j}|}{|\lambda_i-\tilde{\lambda}_j|}<\dfrac{|\varepsilon_{i\,j}|}{0.5\cdot a}$ $\forall\, i\neq j$, то $\Phi(E-R)<\dfrac{4}{a^2}\,S(\mathcal{E})\leq\dfrac{4}{a^2}\,\varepsilon^2$.

Осталось оценить $\sum_{i=1}^{n}(1-r_{i\,i})^2=\sum_{i=1}^{n}\left(1-\sqrt{1-\sum_{\substack{j=1\\ j\neq i}}^{n}|r_{i\,j}|^2}\right)^2$ (здесь мы воспользовались условием $r_{i\,i}\geq 0$).

Т.к. $(1-x)^2\leq 1-x\ \ \forall\, x\in[0,1]$, то

$$\sum_{i=1}^{n}\left(1-\sqrt{1-\sum_{\substack{j=1\\ j\neq i}}^{n}|r_{i\,j}|^2}\right)^2\leq\sum_{i=1}^{n}\left(1-\sqrt{1-\sum_{\substack{j=1\\ j\neq i}}^{n}|r_{i\,j}|^2}\right)=$$

$$=\sum_{i=1}^{n}\frac{\sum_{\substack{j=1\\ j\neq i}}^{n}|r_{i\,j}|^2}{1+\sqrt{1-\sum_{\substack{j=1\\ j\neq i}}^{n}|r_{i\,j}|^2}}\leq\Phi(R)=\Phi(E-R)\leq\frac{4}{a^2}\,\varepsilon^2.$$

Подводя итог, имеем

$$S(Q-\tilde{Q})=S(E-R)=\Phi(E-R)+\sum_{i=1}^{n}(1-r_{i\,i})^2\leq\frac{8}{a^2}\,\varepsilon^2.$$

Элементарная матрица унитарного вращения

$$T^* = \begin{bmatrix} \cos\alpha \cdot e^{\underline{i}\cdot\eta} & -\sin\alpha \cdot e^{\underline{i}\cdot\gamma} \\ \sin\alpha \cdot e^{\underline{i}\cdot\delta} & \cos\alpha \cdot e^{\underline{i}\cdot\mu} \end{bmatrix}, \ \eta+\mu-\gamma-\delta = 2\pi\cdot k \quad \forall\ k = 0, \pm 1, \ldots,$$

так как

$$T^*T = \begin{bmatrix} \cos\alpha \cdot e^{\underline{i}\cdot\eta} & -\sin\alpha \cdot e^{\underline{i}\cdot\gamma} \\ \sin\alpha \cdot e^{\underline{i}\cdot\delta} & \cos\alpha \cdot e^{\underline{i}\cdot\mu} \end{bmatrix} \begin{bmatrix} \cos\alpha \cdot e^{-\underline{i}\cdot\eta} & \sin\alpha \cdot e^{-\underline{i}\cdot\delta} \\ -\sin\alpha \cdot e^{-\underline{i}\cdot\gamma} & \cos\alpha \cdot e^{-\underline{i}\cdot\mu} \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & \cos\alpha\cdot\sin\alpha\cdot(e^{\underline{i}\cdot(\eta-\delta)} - e^{\underline{i}\cdot(\gamma-\mu)}) = 0 \\ \cos\alpha\cdot\sin\alpha\cdot(e^{\underline{i}\cdot(\delta-\eta)} - e^{\underline{i}\cdot(\mu-\gamma)}) = 0 & 1 \end{bmatrix}$$

$$\Rightarrow e^{\underline{i}\cdot(\eta-\delta)} = e^{\underline{i}\cdot(\gamma-\mu)} \Rightarrow \eta-\delta = \gamma-\mu+2\pi\cdot k \quad \forall\ k = 0, \pm 1, \ldots .$$

Выберем $\gamma = -\eta,\ \ \delta = \eta,\ \ \mu = -\eta$:

$$T^* = \begin{bmatrix} \cos\alpha \cdot e^{\underline{i}\cdot\eta} & -\sin\alpha \cdot e^{-\underline{i}\cdot\eta} \\ \sin\alpha \cdot e^{\underline{i}\cdot\eta} & \cos\alpha \cdot e^{-\underline{i}\cdot\eta} \end{bmatrix} \equiv \begin{bmatrix} c & -s \\ \bar{s} & \bar{c} \end{bmatrix}.$$

Пусть $A = A^*,\ \tilde{A} = Q^*_{i\,j} A Q_{i\,j} = \{\tilde{a}_{k\,l}\}$,

где $Q_{i\,j}$ – элементарная матрица унитарного вращения вращения:

$$Q^*_{i,j} = \begin{array}{r} \\ \\ \\ i\ \text{строка} \\ \\ \\ \\ j\ \text{строка} \\ \\ \\ \\ \end{array} \begin{bmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & c & & & & -s & & & \\ & & & & 1 & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & 1 & & & & \\ & & & \bar{s} & & & & \bar{c} & & & \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix},$$

тогда (см. лемму 2):

$$\Phi(\tilde{A}) = \Phi(A) + [\underbrace{|a_{i\,i}|^2 + |a_{j\,j}|^2}_{S(A_{ij})-\Phi(A_{ij})} - (\underbrace{|\tilde{a}_{i\,i}|^2 + |\tilde{a}_{j\,j}|^2}_{S(\tilde{A}_{ij})-\Phi(\tilde{A}_{ij})})] =$$

$$= \Phi(A) - \Phi(A_{ij}) + \Phi(\tilde{A}_{ij}) = \Phi(A) - [2\cdot|a_{ij}|^2 - 2\cdot|\tilde{a}_{ij}|^2],$$

где

$$\tilde{A}_{ij} \equiv \begin{bmatrix} \tilde{a}_{i\,i} & \tilde{a}_{i\,j} \\ \bar{\tilde{a}}_{i\,j} & \tilde{a}_{j\,j} \end{bmatrix} = \begin{bmatrix} c & -s \\ \bar{s} & \bar{c} \end{bmatrix} \begin{bmatrix} a_{i\,i} & a_{i\,j} \\ \bar{a}_{i\,j} & a_{j\,j} \end{bmatrix} \begin{bmatrix} \bar{c} & s \\ -\bar{s} & c \end{bmatrix} \equiv T^* A_{ij} T,$$

а параметры вращения c и s нужно выбрать из условия $|\tilde{a}_{ij}|^2 = 0$.

Легко вычислить $\tilde{a}_{ij}$:

$$\tilde{a}_{ij} = \underbrace{(a_{ii} - a_{jj})}_{a=\mathrm{Re}(a)} \cdot c \cdot s + a_{ij} \cdot c^2 - \bar{a}_{ij} \cdot s^2 =$$

$$= a \cdot \cos\alpha \cdot \sin\alpha + a_{ij} \cdot \cos^2\alpha \cdot e^{\underline{i} \cdot 2\eta} - \bar{a}_{ij} \cdot \sin^2\alpha \cdot e^{\underline{i} \cdot (-2\eta)}$$

Осталось для заданных $a = \mathrm{Re}(a)$ и комплексного $b = a_{ij} \equiv |b| \cdot e^{\underline{i} \cdot \varphi} \neq 0$ решить уравнение

$$a \cdot \cos\alpha \cdot \sin\alpha + |b| \cdot e^{\underline{i} \cdot \varphi} \cdot \cos^2\alpha \cdot e^{\underline{i} \cdot 2\eta} - |b| \cdot e^{-\underline{i} \cdot \varphi} \cdot \sin^2\alpha \cdot e^{\underline{i} \cdot (-2\eta)} = 0.$$

Положим $\eta = -\varphi/2$, тогда наше уравнение примет вид

$$a \cdot \cos\alpha \cdot \sin\alpha + |b| \cdot (\cos^2\alpha - \sin^2\alpha) \equiv$$

$$\equiv \frac{1}{2}[a \cdot \sin 2\alpha + 2|b| \cdot \cos 2\alpha] = 0,$$

решение которого является удовлетворяет соотношениям

$$\cos 2\alpha = -\frac{a}{\sqrt{a^2 + 4 \cdot |b|^2}} \equiv -\frac{a}{r}, \quad \sin 2\alpha = \frac{2|b|}{\sqrt{a^2 + 4 \cdot |b|^2}} \equiv \frac{2|b|}{r} > 0 \Rightarrow \alpha \in (0, \pi/2).$$

В принципе угол $\alpha \in (0, \pi/2)$ определен: $\alpha = \frac{1}{2} \operatorname{arcctg} \frac{a}{2|b|}$, но нам нужен не угол, а значения его косинуса и синуса, и мы не хотим использовать для их вычисления соответствующие функции (дорого).

Так как $\cos 2\alpha = 2\cos^2\alpha - 1 = -a/r$, то

$$\cos\alpha = \sqrt{0.5 \cdot (1 - a/r)}.$$

Так как $\sin 2\alpha = 2\cos\alpha \cdot \sin\alpha = 2|b|/r$, то

$$\sin\alpha = |b|/\cos\alpha = \sqrt{0.5 \cdot (1 - a/r)}.$$

И, последнее, по известному $a_{ij} \equiv |b| \cdot e^{\underline{i} \cdot \varphi} \equiv |b| \cdot (\frac{\mathrm{Re}(a_{ij})}{|b|} + \underline{i} \cdot \frac{\mathrm{Im}(a_{ij})}{|b|})$ выпишем формулы для $\cos(\varphi/2)$ и $\sin(\varphi/2)$:

$$\sin(\varphi/2) = \sqrt{\frac{1}{2}(1 - \frac{\mathrm{Re}(a_{ij})}{|b|})}, \quad \cos(\varphi/2) = \mathrm{Sign}(\mathrm{Im}(a_{ij}))\sqrt{\frac{1}{2}(1 + \frac{\mathrm{Re}(a_{ij})}{|b|})},$$

тогда

$$e^{\underline{i} \cdot \eta} = \cos(\varphi/2) - \underline{i} \cdot \sin(\varphi/2), \quad e^{\underline{i} \cdot (-\eta)} = \cos(\varphi/2) + \underline{i} \cdot \sin(\varphi/2)$$

и матрица унитарного вращения полностью вычислена без использования тригонометрических функций.

Литература

1. Фаддеев Д.К., Фаддеева В.Н. Вычислительные методы линейной алгебры. - М.: Л.: Физматгиз, 1963.
2. Коновалов А.Н. Введение в вычислительные методы линейной алгебры.- Новосибирск: ВО "Наука", Сибирская издательская фирма, 1993.
3. Воеводин В.В. Вычислительнные основы линейной алгебры. - М.: Наука, 1977.
4. Годунов С.К. Решение систем линейных уравнений. - Новосибирск: Наука, Сиб. отд-ние, 1980.
5. Бахвалов Н.С. Численные методы. - М.: Наука, 1975.
6. Самарский А.А., Гулин А.В. Численные методы. - М.: Наука, 1989.