Глава 1

Методы решения систем линейных алгебраических уравнений

В данной главе будем рассматривать знакомую, например, из линейной алгебры, задачу — требуется решить систему линейных алгебраических уравнений, записанных в виде матричного уравнения $Ay = f$, где $A = (a_{ij})$ — заданная квадратная матрица размерности $n \times n$ $(i, j = 1, 2, \ldots, n)$, $y =$

$(y_1, y_2, \ldots, y_n)^T$ — вектор-столбец неизвестных, $f = (f_1, f_2, \ldots, f_n)^T$ — заданный вектор–столбец правых частей. Все параметры a_{ij} и f_i — вещественные числа.

Здесь и далее будем предполагать, что у рассматриваемых задач решение существует и единственно. В данном случае, условие, что определитель матрицы A не равен нулю ($\det A \neq 0$), гарантирует существование и единственность решения системы линейных алгебраических уравнений [5].

1.1 Прямые методы решения систем линейных алгебраических уравнений

Рассмотрим несколько прямых методов для решения системы

$$Ay = f. \tag{1.1}$$

Известным, широко используемым методом решения систем линейных алгебраических уравнений с невырожденной матрицей ($\det A \neq 0$), является *метод Гаусса*[1]. Применять метод Гаусса можно тогда и только тогда, когда

[1] Впервые описан К. Гауссом в 1849г.

все угловые миноры матрицы A отличны от нуля ($\det A$ является угловым минором $n-$ого порядка). Кроме того, метод Гаусса в классическом виде является неустойчивым методом [10]. Для того, чтобы обойти эти ограничения, на практике используют расчетные формулы метода Гаусса в сочетании с некоторой схемой выбора главного элемента [3].

Метод Гаусса является прямым методом решения систем линейных алгебраических уравнений. К этой группе относят методы, в которых получают искомый вектор y за конечное число арифметических операций.

Основным показателем при оценки эффективности конкретного метода является количество арифметических операций, необходимых для вычисления y. Общее число операций умножения и деления, более длительных по времени их реализации на вычислительной технике по сравнению с операциями сложения и вычитания, в методе Гаусса равно $n^3/3 + O(n^2)$. Объем числовой информации, которую необходимо хранить при реализации метода Гаусса, составляет $O(n^2)$. Для современных персональных компьютеров при $n \approx 10^4$ число арифметических операций и объем памяти, требующийся для реализации метода Гаусса, вполне приемлемы. Поэтому использование стандартных программ, реализующих метод Гаусса с выбором главного элемента, эффективно при решении систем линейных алгебраических уравнений с числом уравнений $n \leqslant 10^4$. Одной из таких, проверенных вычислительной практикой стандартных программ, имеющейся в свободном доступе, является

программа Y12M. Теоретическое обоснование, использованного в программе Y12M расчетного алгоритма приведено в [9].

Возможны ситуации, когда возникает потребность в использовании для решения систем линейных алгебраических уравнений методов более экономичных по затратам, чем метод Гаусса. Конкурируют по трудоемкости с методом Гаусса только методы, в которых явно учитывается специфика матрицы A, то есть методы пригодные для некоторых частных видов систем линейных алгебраических уравнений. Примером может служить система линейных алгебраических уравнений с трехдиагональной матрицей A. Оптимальным методом решения такой системы является, как известно, метод прогонки [3], для реализации которого требуется $O(n)$ арифметических операций умножения и деления.

Симметричность элементов матрицы A относительно главной диагонали ($A^T = A$) является частным свойством матрицы. Положительность матрицы также является частным свойством матрицы. Напомним, что матрица A называется *положительной* ($A > 0$), если для любого вектора $y \neq 0$ скалярное произведение $(Ay, y) > 0$. Положительность эквивалентна положительности всех угловых миноров матрицы A (критерий Сильвестра) или, для ($A^T = A$), положительности всех собственных чисел $\lambda(A)$ матрицы A (см. [1]). Заметим, что любой из указанных критериев положительности матрицы A затруднительно проверить для матриц большой размерности.

Отметим одно простое и удобное для проверки необходимое условие положительности матрицы и одно достаточное условие положительности матрицы.

Пусть матрица $A > 0$, тогда для вектора $y = (0, \cdots, 0, y_i, 0, \cdots, 0)^T$, где $y_i \neq 0$, $(Ay, y) = a_{ii} y_i^2 > 0$ $(i = 1, \ldots, n)$. Отсюда следует, что у положительной матрицы A все диагональные элементы $a_{ii} > 0$. Эти же неравенства можно получить иначе. Пусть положительны все угловые миноры матрицы A. Следствием этого (см. [1]) является положительность всех главных миноров матрицы A. Так как элементы a_{ii}, находящиеся на главной диагонали матрицы A, являются главными минорами первого порядка, то для них выполнены неравенства $a_{ii} > 0$. Это и есть удобное для проверки необходимое условие положительности матрицы A.

Простым для проверки достаточным условием положительности симметричной матрицы $A = A^T$ является *условие диагонального преобладания*:

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

Покажем это. Пусть выполнено необходимое условие положительности матрицы A, то есть $a_{ii} > 0$, $i = 1, 2, \ldots, n$. Пусть выполнено условие диа-

гонального преобладания, λ — любое из собственных чисел матрицы A и $\xi = (\xi_1, \cdots, \xi_n)^T$ — собственный вектор матрицы A, соответствующий этому собственному числу, то есть $A\xi = \lambda\xi$. Выберем максимальную по модулю компоненту собственного вектора ξ. Пусть $|\xi_i| = \max\limits_{1\le j\le n} |\xi_j|$, тогда компонента с номером i векторного равенства $A\,\xi = \lambda\xi$ имеет вид

$$a_{ii}\xi_i + S = \lambda\xi_i, \text{ где } S = \sum_{\substack{j=1\\ j\neq i}}^{n} a_{ij}\xi_j.$$

Для $|S|$ справедлива следующая оценка:

$$|S| = \Big|\sum_{\substack{j=1\\ j\neq i}}^{n} a_{ij}\,\xi_j\Big| \le \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}||\xi_j| \le |\xi_i| \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| < |\xi_i| a_{ii}.$$

Отсюда следует, что если $\xi_i > 0$, то $-a_{ii}\,\xi_i < S$. Тогда сумма $a_{ii}\,\xi_i + S = \lambda\xi_i > 0$ и, следовательно, $\lambda > 0$. Если $\xi_i < 0$, то $S < -a_{ii}\xi_i$, $a_{ii}\xi_i + S = \lambda\xi_i < 0$ и, как и в предыдущем случае, $\lambda > 0$.

Итак, выполнение условия диагонального преобладания гарантирует положительность всех собственных чисел матрицы A. Следовательно, симметрич-

ная матрица $A = A^T$ обладает свойством положительности (положительной определенности).

1.1.1 Метод квадратного корня (метод Холецкого)

Рассмотрим системы линейных алгебраических уравнений с симметричными и положительно определенными матрицами ($A^T = A > 0$). Для таких матриц возможно представление

$$A = LL^T, \tag{1.2}$$

где L — нижняя треугольная матрица. Такое представление матрицы A называют разложением Холецкого. Подставляя разложение (1.2) матрицы A в уравнение (1.1) и вводя обозначение $L^T y = \tilde{y}$, получим систему уравнений $L\tilde{y} = f$ относительно вектора $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_n)^T$. Поскольку матрица L нижняя треугольная, то решение этой системы осуществляется построчно. Из первого уравнения системы находится компонента $\tilde{y}_1$ вектора $\tilde{y}$. Из второго уравнения, зная $\tilde{y}_1$, находится компонента $\tilde{y}_2$ и так далее. Вычислив все компоненты вектора $\tilde{y}$, получаем для искомого вектора y систему уравнений $L^T y = \tilde{y}$ с верхней треугольной матрицей. Решение этой системы проводится построчно, начиная с последнего уравнения. Описанная процедура реализуется, если найдены элементы матрицы L.

Получим расчетные формулы для вычисления элементов матрицы L. Будем использовать обозначения $L = (l_{ij})$ и $L^T = (\bar{l}_{ij})$, где $\bar{l}_{ij} = l_{ji}$. Найдем элементы этих матриц, исходя из матричного равенства (1.2):

$$\Big(a_{ij}\Big) = \begin{pmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & & \vdots \\ \vdots & & \ddots & 0 \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{pmatrix} \begin{pmatrix} \bar{l}_{11} & \bar{l}_{12} & \dots & \bar{l}_{1n} \\ 0 & \bar{l}_{22} & & \bar{l}_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \bar{l}_{nn} \end{pmatrix}.$$

Имеем

$$a_{ij} = \sum_{m=1}^{n} l_{im}\bar{l}_{mj} = \sum_{m=1}^{\min(i,\,j)} l_{im}\bar{l}_{mj} = \sum_{m=1}^{\min(i,\,j)} l_{im}l_{jm}. \tag{1.3}$$

Используем данное равенство для определения элементов l_{ij} в столбцах матрицы L.

При $j = 1$ соотношение (1.3) принимает вид $a_{i1} = l_{i1}l_{11}$, $i = 1, 2, \dots, n$. Отсюда $a_{11} = l_{11}^2$, где $a_{11} > 0$ в силу положительной определенности матрицы A. Поэтому справедлива формула $l_{11} = \sqrt{a_{11}}$. Определив l_{11}, вычислим остальные элементы первого столбца по формуле $l_{i1} = a_{i1}/l_{11}$, $i = 2, 3, \dots, n$.

При $j = 2$ из того же соотношения (1.3) получим, что

$$a_{i2} = l_{i1}l_{21} + l_{i2}l_{22},\ i = 2, 3, \ldots, n. \tag{1.4}$$

Отсюда $l_{22}^2 = a_{22} - l_{21}^2$, где элемент $l_{21} = a_{21}/\sqrt{a_{11}}$ вычислен на предыдущем этапе. Извлекая корень, определим $l_{22} = \sqrt{(a_{11}a_{22} - a_{21}^2)/a_{11}}$. Операция извлечения корня корректна, так как

$$a_{11} > 0,\ a_{11}a_{22} - a_{12}^2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0$$

в силу симметричности и положительной определенности матрицы A. Определив l_{22}, вычислим, используя (1.4), остальные элементы второго столбца по формуле $l_{i2} = (a_{i2} - l_{i1}l_{21})/l_{22},\ i = 3, 4, \ldots, n$.

Далее, исходя из того, что элементы матрицы L в столбцах с номерами $1, 2, \ldots, m-1$ вычислены на предыдущих этапах, получим расчетные формулы для элементов столбца с номером m. При $j = m$ из (1.3) для $i = m, m+1, \ldots, n$ получим:

$$a_{im} = l_{i1}l_{m1} + l_{i2}l_{m2} + \ldots + l_{i(m-1)}l_{m(m-1)} + l_{im}l_{mm}. \tag{1.5}$$

Отсюда, полагая $i = m$, находим

$$l_{mm} = \sqrt{a_{mm} - l_{m1}^2 - \ldots - l_{m(m-1)}^2}.$$

Можно показать, что, как и ранее, под корнем получается отношение положительного углового минора $m-$го порядка матрицы A к положительному минору $(m-1)-$го порядка матрицы A. Определив l_{mm}, вычислим, используя (1.5), остальные элементы $m-$го столбца матрицы L

$$l_{im} = \frac{a_{im} - l_{i1}l_{m1} - \ldots - l_{i(m-1)}l_{m(m-1)}}{l_{mm}},\ i = m+1, m+2, \ldots, n.$$

Указанным способом последовательно определяются все элементы матрицы L. При этом подсчет числа арифметических действий, затрачиваемых на вычисление элементов матрицы L, дает величину $\approx n^3/6$, что в два раза меньше трудоемкости метода Гаусса.

1.1.2 Модифицированный метод квадратного корня

Рассмотрим уравнение $Ay = f$ при условии, что $\det A \neq 0$ и матрица A симметрична $(A^T = A)$. Для таких матриц справедливо представление

$$A = LDL^T, \tag{1.6}$$

где L — нижняя треугольная матрица с единицами на главной диагонали, а D — диагональная матрица. Если указанное представление найдено, то

решение исходной системы уравнений $Ay = f$ сводится к последовательному решению систем $L\hat{y} = f$, $D\tilde{y} = \hat{y}$ и $L^T y = \tilde{y}$ с нижней треугольной, диагональной и верхней треугольной матрицами, соответственно. Решение таких систем не представляет трудностей (см. пункт 1.1.1), и трудоемкость метода фактически сводится к нахождению матриц L и D.

Рассмотрим алгоритм нахождения элементов матриц L и D. Как и ранее, будем использовать обозначения $L = (l_{ij})$ и $L^T = (\bar{l}_{ij})$, где $\bar{l}_{ij} = l_{ji}$, а также $D = (d_{ij})$. Тогда равенство (1.6) примет вид

$$\Big(a_{ij}\Big) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ l_{21} & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ l_{n1} & l_{n2} & \dots & 1 \end{pmatrix} \begin{pmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \dots & d_{nn} \end{pmatrix} \begin{pmatrix} 1 & \bar{l}_{12} & \dots & \bar{l}_{1n} \\ 0 & 1 & & \bar{l}_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix}.$$

Отсюда, используя обозначение $(b_{ij}) = DL^T$, где $b_{ij} = d_{ii}\overline{l_{ij}} = d_{ii}l_{ji}$, получим

$$a_{ij} = \sum_{k=1}^{n} l_{ik}b_{kj} = \sum_{k=1}^{j} l_{ik}b_{kj} = \sum_{k=1}^{j} l_{ik}d_{kk}l_{jk}, \tag{1.7}$$

где $i \geq j$.

При $j = 1$ и $i \geq 1$ равенство (1.7) примет вид $a_{i1} = l_{i1}d_{11}$, так как $l_{11} = 1$. Отсюда находим $d_{11} = a_{11}$ и $l_{i1} = a_{i1}/d_{11}$, $i = 2, 3, \ldots, n$. Формулы корректны при условии, что $a_{11} \neq 0$.

При $j = 2$ и $i \geq 2$ из (1.7) получим $a_{i2} = l_{i1}d_{11}l_{21} + l_{i2}d_{22}$, так как $l_{22} = 1$. Отсюда

$$d_{22} = a_{22} - l_{21}^2 d_{11} = a_{22} - \frac{a_{21}^2}{a_{11}} = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11}} = \frac{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}{d_{11}}.$$

В правой части этого выражения находится отношения угловых миноров второго и первого порядка матрицы A. Определив d_{22}, вычислим элементы $l_{i2} = (a_{i2} - l_{i1}d_{11}l_{21})/d_{22}$, $i = 3, 4, \ldots, n$. Формулы корректны при $d_{22} \neq 0$, то есть помимо углового минора первого порядка a_{11} должен быть отличен от нуля угловой минор второго порядка матрицы A.

Далее последовательно вычисляются элементы следующих столбцов матриц D и L. Приведем формулы для вычисления элементов столбца с номером

m:

$$d_{mm} = a_{mm} - \left(l_{m1}^2 d_{11} + \ldots + l_{mm-1}^2 d_{m-1m-1}\right) = \frac{\begin{vmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{vmatrix}}{d_{11}\, d_{22} \cdots d_{m-1m-1}},$$

$$l_{im} = (a_{im} - l_{i1} d_{11} l_{m1} - \ldots - l_{im-1} d_{m-1m-1} l_{mm-1})/d_{mm},$$

$$i = m+1, m+2, \ldots, n.$$

Отметим, что эти формулы применимы в случае, когда угловые миноры всех порядков матрицы A отличны от нуля. То есть, условия применимости модифицированного метода квадратного корня совпадают с условиями применимости обычного метода Гаусса.

Подсчитав количество операций умножения и деления, необходимых для получения разложения (1.6) и для поиска вектора y, получим сложность данного метода, равную $\approx n^3/6$.

1.2 Итерационные методы решения систем линейных алгебраических уравнений

К итерационным методам относят те методы, в которых искомое решение системы линейных алгебраических уравнений строится как предел некоторой сходящейся последовательности. Эффективными итерационные методы оказываются при решении систем большой размерности (систем содержащих тысячи и более уравнений) с разреженными матрицами.

В итерационных методах, задав некоторый вектор y^0, называемый *начальным приближением*, строят по некоторому правилу последовательность векторов $y^1,\ y^2,\ \ldots,\ y^k,\ \ldots$. Верхний индекс k называют номером итерационного приближения. В общем случае правило вычисления итерационного приближения может зависть от номера k и очередное $(k+1)$-ое итерационное приближение может строиться по всем предыдущим итерационным приближениям, то есть

$$y^{k+1} = G_{k+1}(y^0,\ y^1,\ \ldots,\ y^k).$$

Выбор правила вычисления итерационных приближений определяет конкретный итерационный метод.

Определение. Итерационный метод называется *$m-$шаговым*, если каждое последующее итерационное приближение строится лишь по m предыдущим:

$$y^{k+1} = G_{k+1}(y^{k-m+1},\ \ldots,\ y^{k-1},\ y^k).$$

Рассмотрим одношаговые ($m = 1$) и двухшаговые ($m = 2$) итерационные методы. На примере одношаговых итерационных методов удобно обсуждать математический аппарат, используемый для исследования итерационных методов. В классе двухшаговых итерационных методов существуют достаточно эффективные, широко используемые на практике итерационные методы решения систем линейных алгебраических уравнений.

Определение. Если G_{k+1} — линейная функция своих аргументов, то такой итерационный метод называется *линейным*.

1.2.1 Линейные одношаговые итерационные методы

Согласно введенным определениям, любой линейный одношаговый итерационный метод имет вид:

$$y^{k+1} = S_{k+1}y^k + \psi_{k+1}, \tag{1.8}$$

где S_{k+1} — матрица, а ψ_{k+1} — вектор, задание которых определяет конкретный итерационный метод (размерность векторов и порядок матриц считаются одинаковыми).

Будем требовать от итерационного метода, чтобы вектор $y = A^{-1}f$ (искомое точное решение исходной задачи (1.1)) при подстановке вместо y^{k+1} и y^k обращал бы (1.8) в тождество:

$$A^{-1}f = S_{k+1}A^{-1}f + \psi_{k+1}.$$

Тогда вектор ψ_{k+1} можно представить в виде $\psi_{k+1} = Q_{k+1}f$. Здесь введено обозначение $Q_{k+1} = A^{-1} - S_{k+1}A^{-1}$. Формулу для Q_{k+1} домножим справа на матрицу A. Тогда для матрицы S_{k+1} получим представление $S_{k+1} = E - Q_{k+1}A$, где E — единичная матрица. Выражения для вектора ψ_{k+1} и матрицы S_{k+1} подставим в (1.8):

$$y^{k+1} = y^k - Q_{k+1}Ay^k + Q_{k+1}f.$$

Отсюда получим

$$(Q_{k+1})^{-1}\,\tau_{k+1}\frac{y^{k+1} - y^k}{\tau_{k+1}} + Ay^k = f,$$

где $\tau_{k+1} > 0$ — некоторое вещественное число. Вводя обозначение $B_{k+1} = (Q_{k+1})^{-1}\tau_{k+1}$, приходим к так называемой *канонической форме* записи одношагового линейного итерационного метода:

$$B_{k+1}\frac{y^{k+1} - y^k}{\tau_{k+1}} + Ay^k = f. \tag{1.9}$$

Конкретный линейный одношаговый метод определяется заданием матриц B_{k+1} и числовых параметров τ_{k+1}. В дальнейшем будем пользоваться следующей терминологией.

Определение. Если матрица $B_{k+1} = E$, то соответствующий итерационный метод называется *явным*, в противном случае — *неявным*.

В явных итерационных методах вектор y^{k+1} находится без решения вспомогательной системы линейных алгебраических уравнений с матрицей B_{k+1}.

Определение. Если $B_{k+1} = B$ и $\tau_{k+1} = \tau$, то метод называется *стационарным*, в противном случае — *нестационарным*.

Определение. Вектор $z^k = y^k - y$ (отклонение итерационного приближения y^k от точного решения y) будем называть *погрешностью итерационного приближения* на $k-$ой итерации.

Определение. Метод называется *сходящимся*, если $\|z^k\| \xrightarrow[k\to\infty]{} 0$ для некоторой выбранной нормы $\|\cdot\|$.

Точное решение y исходной системы ((1.1)) является пределом последовательности итерационных приближений и в большинстве случаев не может быть получено за конечное число арифметических действий. Пусть для начального итерационного приближения y^0, очередного итерационного приближения y^k и достаточно малой величиной $\varepsilon > 0$ выполняется неравенство

$$\|y^k - y\| \leqslant \varepsilon \|y^0 - y\|.$$

Неравенство означает, что на $k-$*ой* итерации погрешность итерационного приближения y^k не превышает погрешности начального приближения y^0, уменьшенной в $1/\varepsilon$ раз. Тогда итерационное приближение y^k будем считать приближенным решением, полученным с точностью ε. Для итерационных методов естественно ожидать, что, если неравенство выполняется для некоторого $k_0 = k_0(\varepsilon)$, то оно должно выполняться и для любого $k > k_0(\varepsilon)$. Число $k_0(\varepsilon)$ будем называть *минимальным числом итераций, необходимым для достижения заданной точности* ε. Чем меньше при прочих равных условиях $k_0(\varepsilon)$, тем эффективнее итерационный метод.

Следует отметить, что использовать указанное неравенство для контроля достигнутой точности на $k-$*ой* итерации и завершения итерационного про-

цесса не представляется возможным, поскольку точное решение y неизвестно. Поэтому, для конкретных итерационных методов, используются априорные оценки для $k_0(\varepsilon)$.

1.2.2 Примеры одношаговых линейных итерационных методов

Запишем систему уравнений $Ay = f$ в виде:

$$\begin{cases} a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n = f_1, \\ a_{21}y_1 + a_{22}y_2 + \ldots + a_{2n}y_n = f_2, \\ \ldots \\ a_{n1}y_1 + a_{n2}y_2 + \ldots + a_{nn}y_n = f_n. \end{cases} \tag{1.10}$$

Используя такую форму записи исходной системы линейных алгебраических уравнений, линейный одношаговый итерационный метод можно задать, расставляя итерационные индексы у компонент вектора y.

Метод Якоби.

Для построения итерационного метода Якоби припишем неизвестным y_i, имеющим в качестве сомножителей коэффициенты a_{ii}, $i = 1, 2, \ldots, n$, индекс следующего итерационного приближения $k+1$, а прочим неизвестным — индекс k:

$$\begin{cases} a_{11}y_1^{k+1} + a_{12}y_2^k \quad + \ldots + a_{1n}y_n^k \quad = f_1, \\ a_{21}y_1^k \quad + a_{22}y_2^{k+1} + \ldots + a_{2n}y_n^k \quad = f_2, \\ \ldots \\ a_{n1}y_1^k \quad + a_{n2}y_2^k \quad + \ldots + a_{nn}y_n^{k+1} = f_n. \end{cases}$$

Тогда, расчетная формула для вычисления компонент $k+1$ итерационного приближения вектора y примет вид:

$$y_i^{k+1} = \frac{1}{a_{ii}} \left(f_i - \sum_{j=1}^{i-1} a_{ij}y_j^k - \sum_{j=i+1}^{n} a_{ij}y_j^k \right), \; i = 1, 2, \ldots, n.$$

Запишем метод Якоби в канонической форме. Для это представим матрицу A в виде $A = L + D + R$, где $L = (l_{ij})$ — нижняя треугольная, $D = (d_{ij})$ —

диагональная, $R = (r_{ij})$ — верхняя треугольная матрицы.

$$l_{ij} = \begin{cases} a_{ij}, & i > j, \\ 0, & i \leqslant j; \end{cases} \quad d_{ij} = \begin{cases} a_{ij}, & i = j, \\ 0, & i \neq j; \end{cases} \quad r_{ij} = \begin{cases} a_{ij}, & i < j, \\ 0, & i \geqslant j. \end{cases}$$

Тогда, система уравнений, определяющая метод Якоби, в матричной форме примет вид

$$Ly^k + Dy^{k+1} + Ry^k = f.$$

Прибавляя и вычитая в левой части Dy^k, получим

$$D(y^{k+1} - y^k) + Ay^k = f.$$

Отсюда следует, что в канонической форме записи (1.9) методу Якоби соответствует выбор $B_{k+1} = D$ и $\tau_{k+1} = 1$. Таким образом, метод Якоби является стационарным неявным одношаговым линейным итерационным методом с легко обратимой матрицей B_{k+1}.

Метод Зейделя.

Метод Зейделя получается, если в развернутой записи системы уравнений (1.10) приписать неизвестным y_i, имеющим в качестве сомножителей коэффициенты a_{ij} при $i \geqslant j$, индекс следующего итерационного приближения $k + 1$,

а прочим неизвестным — индекс k:

$$\begin{cases} a_{11}y_1^{k+1} + a_{12}y_2^k \quad + \ldots + a_{1n}y_n^k \quad = f_1, \\ a_{21}y_1^{k+1} + a_{22}y_2^{k+1} + \ldots + a_{2n}y_n^k \quad = f_2, \\ \ldots \\ a_{n1}y_1^{k+1} + a_{n2}y_2^{k+1} + \ldots + a_{nn}y_n^{k+1} = f_n. \end{cases}$$

Относительно y^{k+1} эта система решается следующим образом. Сначала из первого уравнения находится y_1^{k+1}, потом из второго — y_2^{k+1} и так далее. Расчетная формула для вычисления компонент $(k+1)$—го итерационного приближения вектора y имеет вид:

$$y_i^{k+1} = \frac{f_i - \sum\limits_{j=1}^{i-1} a_{ij}y_j^{k+1} - \sum\limits_{j=i+1}^{n} a_{ij}y_j^k}{a_{ii}}, \; i = 1, 2, \ldots, n.$$

Используя ранее введенное представление матрицы $A = L + D + R$, запишем метод Зейделя в матричной форме

$$Ly^{k+1} + Dy^{k+1} + Ry^k = f.$$

Добавляя и вычитая в левой части этого соотношения комбинацию $Ly^k + Dy^k$, получим:

$$(L + D)(y^{k+1} - y^k) + Ay^k = f.$$

Следовательно, в канонической форме записи (1.9) методу Зейделя соответствует выбор матрицы $B_{k+1} = L + D$ и итерационного параметра $\tau_{k+1} = 1$. То есть, метод Зейделя является стационарным неявным итерационным методом. Нижняя треугольная матрица $L + D$ легко обратима.

Метод релаксации.

Этот итерационный метод определяется выбором $B_{k+1} = D + \omega L$ и $\tau_{k+1} = \omega$ в канонической форме записи (1.9), где ω — заданный числовой параметр:

$$(D + \omega L)\frac{y^{k+1} - y^k}{\omega} + Ay^k = f.$$

Данный метод является стационарным и неявным с легко обратимой нижней треугольной матрицей B_{k+1}. Заметим, что рассмотренный выше метод Зейделя является частным случаем метода релаксации при $\omega = 1$. Выбирая параметр ω из диапазона $0 < \omega < 1$ получаем метод, который часто называют методом нижней релаксации. Выбирая $1 < \omega < 2$, получим метод верхней релаксации.

Метод простой итерации.

Этот метод является примером явного стационарного итерационного метода. В канонической форме записи (1.9) ему соответствует выбор $B_{k+1} = E$ и $\tau_{k+1} = \tau$:

$$\frac{y^{k+1} - y^k}{\tau} + Ay^k = f.$$

Метод Ричардсона.

Примером явного нестационарного итерационного метода является метод Ричардсона, который имеет вид:

$$\frac{y^{k+1} - y^k}{\tau_{k+1}} + Ay^k = f.$$

Здесь матрица $B_{k+1} = E$, а итерационные параметры τ_{k+1} рассчитываются на каждой итерации по некоторым формулам, к которым вернемся позже.

1.3 Условия сходимости одношаговых стационарных итерационных методов

1.3.1 Необходимые и достаточные условия сходимости

Для установления факта сходимости и проведения сравнительного анализа свойств различных итерационных методов необходим соответствующий аналитический аппарат. В данном пункте рассмотрим некоторые элементы такого математического аппарата.

При выяснении возможной области применения конкретного стационарного одношагового итерационного метода может быть полезно следующее утверждение [3].

Теорема 1.1 (Достаточное условие сходимости стационарного одношагового итерационного метода). *Пусть A — симметричная и положительно определенная матрица ($A^T = A > 0$), B — положительно определенная матрица ($B > 0$) и числовой параметр $\tau > 0$. Тогда итерационный процесс*

$$B\frac{y^{k+1} - y^k}{\tau} + Ay^k = f \tag{1.11}$$

сходится для любого начального приближения y^0, если выполнено матричное неравенство $B > \frac{\tau}{2}A$.

(Доказательство см. [3].)

Следствием из теоремы являются, например, два следующих утверждения, определяющие достаточные условия сходимости итерационного метода Якоби и метода релаксации. Напомним (см. пункт 1.2.2), что, если представить матрицу A в виде $A = L + D + R$, где L — нижняя треугольная, D — диагональная, R — верхняя треугольная матрицы, то методу Якоби в канонической форме записи (1.11) соответствует выбор $B = D$, $\tau = 1$, а методу релаксации $B = D + \omega L$, $\tau = \omega$.

Теорема 1.2. *Пусть* $A^T = A$ *и выполнено условие диагонального преобладания*

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

Тогда, при любом начальном приближении итерационный метод Якоби сходится.

▼ Доказательство. С учетом диагонального преобладания и симметричности матрицы A для $\forall\, y \neq 0$ выполнено:

$$0 < (Ay, y) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} y_i y_j \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}||y_i y_j| \leqslant \left\{2|y_i y_j| \leqslant y_i^2 + y_j^2\right\} \leqslant$$

$$\leqslant \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}| y_i^2 + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}| y_j^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}| y_i^2 =$$

$$= \sum_{i=1}^{n} y_i^2 \Big(a_{ii} + \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|\Big) < 2\sum_{i=1}^{n} y_i^2 a_{ii} = 2(Dy, y).$$

Таким образом, показано, что $((2D - A)y, y) > 0\ \forall\, y \neq 0$. Это условие эквивалентно матричному неравенству $D > \frac{1}{2}A$, которое в силу теоремы 1.1 достаточно для сходимости метода Якоби, поскольку $A^T = A > 0$ и $B = D > 0$.

▲ Утверждение доказано.

Теорема 1.3. *Пусть $A^T = A > 0$. Тогда метод релаксации с итерационным параметром $0 < \omega < 2$, является сходящимся итерационным методом для любого начального приближения.*

▼ Доказательство. Запишем достаточное условие сходимости $B > \frac{\tau}{2}A$ при $B = D + \omega L$, $\tau = \omega$ в виде

$$((2D + 2\omega L - \omega A)y, y) > 0 \quad (y \neq 0).$$

Так как $A^T = A$, то $A = L + D + L^T$ и, следовательно

$$(Ay, y) = (Ly.y) + (Dy, y) + (L^T y, y) = (Dy, y) + 2(Ly, y).$$

Тогда достаточное условие сходимости принимает вид:

$$2(Dy, y) + 2\omega(Ly, y) - 2\omega(Ly, y) - \omega(Dy, y) > 0.$$

В результате имеем неравенство $(2 - \omega)(Dy, y) > 0$, которое выполнено, так как в силу условий теоремы $(2 - \omega) > 0$ и $(Dy, y) > 0$.

▲ Утверждение доказано.

Сформулированные выше теоремы являются примерами достаточных условий сходимости. В случае, когда условия теорем не выполнены, сделать вывод о сходимости или расходимости соответствующих методов не представляется возможным. Полностью решить вопрос о сходимости конкретного итерационного метода можно лишь на основе какого-либо критерия его сходимости.

Рассмотрим пример критерия сходимости стационарных одношаговых итерационных методов (1.11). Предварительно получим уравнение для погрешности z^k. Подставляя в (1.11) представление итерационного приближения $y^k = z^k + y$, где y — точное решение исходного уравнения $Ay = f$, приходим к однородному уравнению $B\dfrac{z^{k+1} - z^k}{\tau} + Az^k = 0$, из которого следует, что $z^{k+1} = (E - \tau B^{-1}A)z^k$. Введем обозначение $S = E - \tau B^{-1}A$. Матрица S называется *матрицей перехода*. Тогда уравнение для погрешности примет вид:

$$z^{k+1} = Sz^k. \tag{1.12}$$

Теорема 1.4 (Критерий сходимости одношагового стационарного итерационного метода)**.** *Итерационный метод (1.11) сходится для любого начального приближения y^0 тогда и только тогда, когда для всех собственных значений $\lambda(S)$ матрицы S выполнено неравенство $|\lambda(S)| < 1$.*

▼ Доказательство.

Предположим, что метод сходится при любом выборе начального приближения y^0. Пусть μ — собственный вектор матрицы S, отвечающий собственному значению λ. Рассмотрим вектор начального приближения $y^0 = \mu + y$,

тогда $z^0 = \mu$. Из уравнения для погрешности (1.12) получим

$$z^k = Sz^{k-1} = S^2 z^{k-2} = \ldots = S^k z^0 = S^k \mu = S^{k-1}\Big(S\mu\Big) = \lambda S^{k-1}\mu =$$
$$\ldots = \lambda^k \mu \;\Rightarrow \|z^k\| = |\lambda|^k \|\mu\|.$$

По предположению о сходимости $\|z^k\| = |\lambda|^k\|\mu\| \xrightarrow[k\to\infty]{} 0$. Поэтому приходим к неравенству $|\lambda| < 1$.

Доказательство достаточности проведем при дополнительном предположении, что матрица S имеет n линейно независимых собственных векторов μ_l, образующих базис n мерного пространства.

Пусть для любого собственного значения матрицы S выполнено неравенство $|\lambda| < 1$. Погрешность произвольного начального приближения z^0 представим в виде разложения по базису μ_l с коэффициентами c_l. Из уравнения (1.12) получим:

$$z^k = S^k z^0 = S^k \sum_{l=1}^{n} c_l \mu_l = \sum_{l=1}^{n} c_l \lambda_l^k \mu_l \;\Rightarrow \|z^k\| \leqslant \sum_{l=1}^{n} |c_l|\,|\lambda_l|^k \|\mu_l\| \leqslant \bar{\lambda}^k M.$$

Здесь $\bar{\lambda} = \max\limits_{1\le l\le n} |\lambda_l|$, $M = \sum\limits_{l=1}^{n} |c_l|\,\|\mu_l\|$. Так как $\bar{\lambda} < 1$, а $M = \text{const}$, то из последней оценки вытекает сходимость итерационного метода.

▲ Утверждение доказано.

Замечание. Дополнительное предположение существенно упрощает доказательство достаточности в предыдущей теореме. Однако доказательство можно провести (см. [3]) и для матрицы S общего вида на основе ее приведения к жордановой форме.

Наличие критерия сходимости казалось бы полностью решает вопрос об исследовании итерационного метода на сходимость. Однако применение рассмотренного критерия требует поиска всех собственных значений матрицы перехода. Данная задача может оказаться сложнее, чем решение исходной системы линейных алгебраических уравнений. Поэтому необходимы и другие, проще реализуемые на практике, способы исследования сходимости итерационных методов.

При практическом использовании итерационных методов важен не только факт сходимости, но и оценка количества итераций, необходимых для достижения заданной точности. Именно по этому показателю в совокупности с вычислительными затратами на осуществление одной итерации целесообразно оценивать качество метода.

Рассмотрим специальный класс удобных для исследования итерационных методов.

Определение. Итерационный метод *сходится со скоростью геометрической прогрессии* со знаменателем $\rho \in (0;1)$, если для погрешности итерационного приближения справедливо неравенство

$$||y^k - y|| \leqslant \rho^k ||y^0 - y||. \tag{1.13}$$

Пусть $\varepsilon > 0$ — некоторое достаточное малое число. Потребуем, чтобы погрешность итерационного приближения на итерации с номером k была бы не больше, чем погрешность начального итерационного приближения уменьшенная в $1/\varepsilon$ раз. Это означает, что $\|y^k - y\| \leqslant \varepsilon \|y^0 - y\|$. Для того, чтобы для итерационного метода, сходящегося со скоростью геометрической прогрессии, выполнялось это неравенство достаточно выполнения неравенства $\rho^k \leqslant \varepsilon$ верного при

$$k > k_0(\varepsilon) = \left[\frac{\ln(1/\varepsilon)}{\ln(1/\rho)} \right].$$

Определение. Число $k_0(\varepsilon)$ называется *минимальным числом итераций, необходимым для достижения заданной точности* ε.

Определение. Выражение $\ln(1/\rho)$ называется *скоростью сходимости* итерационного метода.

Чем больше скорость сходимости, тем меньше итераций необходимо выполнить для достижения требуемой точности вычисления итерационного приближения и тем лучше соответствующий итерационный метод.

1.3.2 Оценка скорости сходимости одношаговых стационарных методов

Рассмотрим пример утверждения позволяющего для конкретного итерационного метода, сходящегося со скоростью геометрической прогрессии, определить значение параметра ρ в неравенстве (1.13).

Далее будут использоваться следующие определения и утверждения (см. [3]).

1. Если $A^T = A$, то $A > 0 \Leftrightarrow \lambda > 0$.

2. Если $A^T = A > 0$, то $\exists\, A^{-1} > 0$.

3. Если $A^T = A, \rho > 0$, то $-\rho E < A < \rho E \Leftrightarrow A^2 < \rho^2 E$.

4. Если $A^T = A > 0$, то $\exists\, B \;:\; B^2 = A, B^T = B > 0$.

Определение. Матрица B называется *квадратным корнем* матрицы A и обозначается $A^{1/2}$ $(A^T = A \geqslant 0)$.

5. Если $A^T = A > 0, B^T = B > 0$, то $\alpha A > \beta B \Leftrightarrow \alpha B^{-1} > \beta A^{-1}$, где α, β — вещественные числа.

Замечание. В утверждениях 1,3,4,5 после слова «то» неравенства могут быть нестрогими.

Замечание. Аналогичные утверждения справедливы для линейных операторов в евклидовом пространстве H. При этом под C^T следует понимать оператор C^*.

Определение. *Матричной (операторной, энергетической) нормой* вектора v, порожденной симметричной положительно определенной матрицей A называется функционал $\|v\|_A = \sqrt{(Av, v)}$.

Замечание. $\|v\|_A = \sqrt{(A^{1/2}v, A^{1/2}v)} = \|A^{1/2}v\|$.

Ранее было показано, что погрешность z^k одношагового стационарного метода (1.11) удовлетворяет соотношению

$$z^{k+1} = Sz^k,\ S = E - \tau B^{-1}A.$$

Докажем следующие две леммы.

Лемма 1.1. *Пусть* $A^T = A > 0$, $B^T = B > 0$, $\rho > 0$ — *вещественное число, тогда неравенства*

$$\frac{1-\rho}{\tau}B \leqslant A \leqslant \frac{1+\rho}{\tau}B$$

необходимы и достаточны для того, чтобы при любых z^0 *для погрешности выполнялась оценка*

$$\|z^{k+1}\|_A \leqslant \rho\|z^k\|_A,\ k = 0, 1, \dots\ .$$

▼ Доказательство. Обозначим $v^k = A^{1/2}z^k$, тогда

$$v^{k+1} = A^{1/2}z^{k+1} = A^{1/2}Sz^k = A^{1/2}SA^{-1/2}v^k = \tilde{S}v^k,$$

где $\tilde{S} = E - \tau C$, $C = A^{1/2}B^{-1}A^{1/2}$, $C^T = C > 0$. Имеем

$$\|z^{k+1}\|_A \leqslant \rho\|z^k\|_A \ \Leftrightarrow\ \|v^{k+1}\| \leqslant \rho\|v^k\| \ \Leftrightarrow\ (\tilde{S}^2v^k, v^k) \leqslant \rho^2(v^k, v^k) \ \Leftrightarrow$$
$$\Leftrightarrow\ \tilde{S}^2 \leqslant \rho^2 E \ \Leftrightarrow\ -\rho E \leqslant \tilde{S} \leqslant \rho E \ \Leftrightarrow\ \frac{1-\rho}{\tau}E \leqslant C \leqslant \frac{1+\rho}{\tau}E \ \Leftrightarrow$$
$$\Leftrightarrow\ \frac{1-\rho}{\tau}C^{-1} \leqslant E \leqslant \frac{1+\rho}{\tau}C^{-1} \ \Leftrightarrow$$
$$\Leftrightarrow\ \frac{1-\rho}{\tau}A^{-1/2}BA^{-1/2} \leqslant E \leqslant \frac{1+\rho}{\tau}A^{-1/2}BA^{-1/2} \ \Leftrightarrow$$

$$\Leftrightarrow \frac{1-\rho}{\tau}B \leqslant A^{1/2}EA^{1/2} \leqslant \frac{1+\rho}{\tau}B.$$

▲ Утверждение доказано.

Лемма 1.2. *При условиях леммы 1.1 справедлива оценка*

$$\|z^{k+1}\|_B \leqslant \rho\|z^k\|_B,\ k = 0, 1, \dots\ .$$

▼ Доказательство. Обозначив $v^k = B^{1/2}z^k$, получим, что $v^{k+1} = \tilde{S}v^k$, где $\tilde{S} = E - \tau C$, $C = B^{-1/2}AB^{-1/2}$. Далее аналогично предыдущему доказательству

$$\|z^{k+1}\|_B \leqslant \rho\|z^k\|_B \ \Leftrightarrow\ \frac{1-\rho}{\tau}E \leqslant C \leqslant \frac{1+\rho}{\tau}E \ \Leftrightarrow$$
$$\Leftrightarrow \frac{1-\rho}{\tau}B^{1/2}EB^{1/2} \leqslant A \leqslant \frac{1+\rho}{\tau}B^{1/2}EB^{1/2}.$$

▲ Утверждение доказано.

Теорема 1.5. *Пусть* $A^T = A > 0$, $B^T = B > 0$, $\gamma_1 B \leqslant A \leqslant \gamma_2 B$, *где* γ_1, γ_2 — *вещественные числа такие, что* $\gamma_2 > \gamma_1 > 0$. *Тогда при* $\tau = 2/(\gamma_1 + \gamma_2)$ *итерационный метод* (1.11) *сходится и для погрешности справедливы оценки*

$$\|z^k\|_A \leqslant \rho^k\|z^0\|_A,\ \|z^k\|_B \leqslant \rho^k\|z^0\|_B,\ k = 1, 2, \dots\ ,$$

где $\rho = \dfrac{1-\xi}{1+\xi}$, $\xi = \dfrac{\gamma_1}{\gamma_2}$.

▼ Доказательство.

В неравенстве $\gamma_1 B \leqslant A \leqslant \gamma_2 B$ константы γ_1 и γ_2 выразим через τ и ρ

$$\gamma_1 = \frac{1-\rho}{\tau},\ \gamma_2 = \frac{1+\rho}{\tau}.$$

Тогда неравенство примет вид

$$\frac{1-\rho}{\tau} B \leqslant A \leqslant \frac{1+\rho}{\tau} B.$$

Согласно леммам 1.1, 1.2 неравенство равносильно оценкам погрешности

$$\|z^{k+1}\|_A \leqslant \rho \|z^k\|_A,\ \|z^{k+1}\|_B \leqslant \rho \|z^k\|_B,\ k = 0, 1, \ldots\ .$$

То есть

$$\|z^k\|_A \leqslant \rho \|z^{k-1}\|_A \leqslant \ldots \leqslant \rho^k \|z^0\|_A$$

и, аналогично,

$$\|z^k\|_B \leqslant \rho \|z^{k-1}\|_B \leqslant \ldots \leqslant \rho^k \|z^0\|_B.$$

▲ Утверждение доказано.

Замечание. Скорость сходимости в случае, когда величина ξ мала, равна

$$\ln\frac{1}{\rho} = \ln\frac{1+\xi}{1-\xi} = \ln\left(1+\frac{2\xi}{1-\xi}\right) \approx 2\xi = 2\frac{\gamma_1}{\gamma_2}$$

и, следовательно, число итераций, необходимое для достижения заданной точности ε равно

$$k_0(\varepsilon) = \frac{\ln(1/\varepsilon)}{\ln(1/\rho)} \approx \frac{\ln(1/\varepsilon)}{2\xi}.$$

Ускорить сходимость можно за счет увеличения константы γ_1 и уменьшения γ_2.

Замечание. Предполагая, что $A^T = A > 0$, $B^T = B > 0$, рассмотрим обобщенную задачу на собственные значения $A\mu = \lambda B\mu$, равносильную задаче поиска собственных значений и собственных векторов матрицы $B^{-1}A$. Уравнение задачи можно переписать в виде

$$(B^{-1/2}AB^{-1/2})(B^{1/2}\mu) = \lambda(B^{1/2}\mu) \;\Leftrightarrow\; C\tilde{\mu} = \lambda\tilde{\mu}.$$

Здесь $C = B^{-1/2}AB^{-1/2}$, $\tilde{\mu} = B^{1/2}\mu$. Поскольку $C^T = C > 0$, приходим к выводу, что все собственные значения λ матрицы C, они же собственные значения матрицы $B^{-1}A$, действительны и положительны. Тогда, неравенства

$$\gamma_1 B \leqslant A \leqslant \gamma_2 B$$

слева и справа умножая на $B^{-1/2}$, получаем

$$\gamma_1 E \leqslant B^{-1/2} A B^{-1/2} \leqslant \gamma_2 E \;\Leftrightarrow$$
$$\Leftrightarrow\; \gamma_1 \leqslant \lambda \leqslant \gamma_2,$$

где λ - любое собственное значение матрицы $B^{-1}A$. Отсюда вытекает, что $\gamma_1 = \lambda_{\min}(B^{-1}A)$, $\gamma_2 = \lambda_{\max}(B^{-1}A)$ — наиболее точные постоянные, с которыми выполняются неравенства $\gamma_1 B \leqslant A \leqslant \gamma_2 B$.

Определение. *Оптимальным итерационным параметром* метода (1.11) называется число

$$\tau = \frac{2}{\lambda_{\min}(B^{-1}A) + \lambda_{\max}(B^{-1}A)}.$$

Замечание. Оптимальный итерационный параметр минимизирует величину ρ на множестве всех положительных γ_1,γ_2, удовлетворяющих условиям $\gamma_1 B \leqslant A \leqslant \gamma_2 B$.

Замечание. Скорость сходимости максимальна, если выбрать $B = A$. Тогда $\rho = 0$ при $\gamma_1 = \lambda_{\min}(B^{-1}A) = \lambda_{\max}(B^{-1}A) = \gamma_2 = 1$, $\tau = 1$ и метод (1.11) дает точное решение уравнения $Ay = f$ на первой же итерации, поскольку $A(y^1 - y^0) + Ay^0 = f$. Однако для вычисления y^1 необходимо обратить матрицу A, что равносильно нахождению точного решения $y = A^{-1}f$.

Воспользуемся доказанной в этом пункте теоремой для сравнения скорости сходимости различных стационарных одношаговых итерационных методов. Тестировать итерационные методы будем на одной и тойже системе линейных алгебраических уравнений, которую назовем модельной задачей.

1.3.3 Модельная задача

Рассмотрим краевую задачу для ОДУ 2-го порядка:

$$-u''(x) = f(x),\ 0 < x < 1; u(0) = u(1) = 0.$$

Введем на отрезке $[0;1]$ равномерную разностную сетку с постоянным шагом h и узлами x_i

$$\Omega_h = \{x_i = ih;\ i = 0, 1, \ldots, N;\ hN = 1\}$$

и сопоставим дифференциальной задаче разностную схему

$$-y_{\bar{x}x,i} = f_i;\ i = 1, 2, \ldots, N-1;\ y_0 = y_N = 0. \tag{1.14}$$

Здесь $y_i = y(x_i)$, $f_i = f(x_i)$ — сеточные функции, определенные в узлах сетки x_i, а $y_{\bar{x}x,i}$ — сокращенная запись разностного отношения

$$y_{\bar{x}x,i} = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}.$$

Пусть функция $u(x)$ — достаточно гладкое решение дифференциальной задачи. Подставляя ее значения в узлах сетки в разностную схему, получим,

используя разложения по формуле Тейлора с центром в узле x_i, что сеточная функция

$$\psi_i = \frac{u(x_{i-1}) - 2u(x_i) + u(x_{i+1})}{h^2} + f(x_i) =$$
$$= u''(x_i) + \frac{h^2}{12}u^{(4)}(x_i) + O(h^4) + f(x_i) = \frac{h^2}{12}u^{(4)}(x_i) + O(h^4) = O(h^2).$$

Функция ψ_i называется погрешностью аппроксимации разностной схемы на решении дифференциальной задачи, а равенство $\psi_i = O(h^2)$ означает, что разностная схема аппроксимирует исходную задачу со вторым порядком по параметру h.

Введем векторы

$$y = (y_1,\, y_2,\, \ldots,\, y_{N-1})^T, \quad f = (f_1,\, f_2,\, \ldots,\, f_{N-1})^T;$$

и матрицу

$$A = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & 0 & \ldots & 0 & 0 & 0 \\ -1 & 2 & -1 & \ldots & 0 & 0 & 0 \\ 0 & -1 & 2 & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & -1 & 2 & -1 \\ 0 & 0 & 0 & \ldots & 0 & -1 & 2 \end{pmatrix}.$$

Тогда в матричной форме разностная схема (1.14) может быть записана в виде системы линейных алгебраических уравнений $Ay = f$, где матрица A является симметричной матрицей ($A^T = A$).

Именно эту систему линейных алгебраических уравнений в дальнейшем будем использовать как модельную задачу для тестирования и сопоставления между собой различных итерационных методов.

Для доказательства положительной определенности матрицы A, найдем ее собственные значения и убедимся в их положительности. Для этого рассмотрим разностную задачу на собственные значения

$$\mu_{\bar{x}x,i} + \lambda\mu_i = 0,\ \mu_0 = \mu_N = 0,\ i = 1, 2, \ldots, N-1,\ hN = 1. \tag{1.15}$$

Нахождение чисел λ_l и соответствующих им сеточных функций μ_i^l ($l = 1, 2, \ldots, N-1,\ i = 0, 1, \ldots, N$), являющихся решением этой задачи, эквивалентно нахождению собственных значений и собственных векторов матрицы A.

Лемма 1.3. *Решения задачи* (1.15) *имеют вид:*

$$\lambda_l = \frac{4}{h^2}\sin^2\frac{\pi l}{2N},\ \mu_i^l = \sin\frac{\pi l i}{N},\ l = 1, 2, \ldots, N-1,\ i = 0, 1, \ldots, N.$$

▼ Доказательство. По аналогии с соответствующей дифференциальной задачей на собственные значения для дифференциального оператора второй

производной будем искать решение в виде

$$\mu_i = \sin(\alpha i),\ i = 0, 1, \ldots, N.$$

Подставляя $\mu_i = \sin(\alpha i)$ в уравнение (1.15)

$$\mu_{i-1} - 2(1 - \lambda h^2/2)\mu_i + \mu_{i+1} = 0,\ i = 1, 2, \ldots, N-1,$$

получим $2\sin(\alpha i)\cos\alpha - 2(1 - \lambda h^2/2)\sin(\alpha i) = 0$. Из этого следует

$$\cos\alpha = 1 - \frac{\lambda h^2}{2},\ \lambda = \frac{4}{h^2}\sin^2\frac{\alpha}{2}.$$

Поскольку $\mu_0 = \sin(\alpha 0) = 0$, осталось учесть граничное условие $\mu_N = 0$:

$$\sin(\alpha N) = 0 \Rightarrow \alpha = \pi l/N б\ l = 1, 2, \ldots, N-1.$$

▲ Утверждение доказано.

Тем самым показано, что собственные значения λ_l, $l = 1, 2, \ldots, N-1$ матрицы A различны и положительны. Таким образом, $A^T = A > 0$. При этом с учетом равенства $hN = 1$

$$\lambda_{\min} = \lambda_1 = \frac{4}{h^2}\sin^2\frac{\pi h}{2}, \quad \lambda_{\max} = \lambda_{N-1} = \frac{4}{h^2}\cos^2\frac{\pi h}{2}.$$

Используем построенную модельную задачу для тестирования итерационных методов. Применим для решения системы линейных алгебраических уравнений (1.14) с симметричной и положительно определенной матрицей итерационный метод простой итерации с оптимальным итерационным параметром.

Определение. Методом *простой итерации* для решения системы $Ay = f$ называется итерационный метод (1.11) при $B = E$, то есть

$$\frac{y^{k+1} - y^k}{\tau} + Ay^k = f;\ k = 0, 1, \dots\ .$$

Тогда справедливо следствие.

Следствие. *Пусть* $A^T = A > 0$, $\lambda_{\min}(A)$ *и* $\lambda_{\max}(A)$ *— соответственно минимальное и максимальное собственные значения матрицы* A. *Тогда для метода простой итерации при*

$$\tau = \frac{2}{\lambda_{\min}(A) + \lambda_{\max}(A)}$$

справедлива оценка погрешности

$$\|z^k\| \leqslant \rho^k \|z^0\|,\ k = 1, 2, \dots,\ \text{где } \rho = \frac{1 - \xi}{1 + \xi},\ \xi = \frac{\lambda_{\min}(A)}{\lambda_{\max}(A)}.$$

Замечание. Можно показать, что метод Якоби для решения уравнений разностной схемы (1.14) совпадает с методом простой итерации с оптимальным итерационным параметром (см. пункт 1.3.2) для модельной задачи.

Получим асимптотические оценки при $N \to \infty$ для скорости сходимости $\ln(1/\rho)$ и минимального числа итераций $k_0(\varepsilon)$, необходимых для достижения заданной точности ε:

$$\ln\frac{1}{\rho} = \ln\left(1 + \frac{2\xi}{1-\xi}\right) \approx 2\xi \ = 2\,\mathrm{tg}^2\,\frac{\pi h}{2} \approx \frac{\pi^2 h^2}{2} = \frac{\pi^2}{2N^2},$$

$$k_0(\varepsilon) = \frac{\ln(1/\varepsilon)}{\ln(1/\rho)} \approx \frac{2N^2}{\pi^2}\ln\frac{1}{\varepsilon}.$$

Пусть, например, $\varepsilon = 0,5 \cdot 10^{-4}$, тогда $\ln(1/\varepsilon) \approx 9,9$ и $k_0(\varepsilon) \approx 2N^2$. Это означает, что для нахождения приближенного решения разностных уравнений (1.14) с заданной точностью ε на сетке с $N = 10$ узлами требуется выполнить порядка 200 итераций, а для $N = 100$ — уже 20000 итераций. Такой быстрый рост числа итераций при увеличении размерности модельной задачи является характерной особенностью метода простой итерации с оптимальным итерационным параметром и метода Якоби.

1.4 Попеременно–треугольный итерационный метод

1.4.1 Алгебраическая теория

В пункте 1.2.2 некоторые стандартные итерационные методы приводились к каноническому виду

$$B\frac{y^{k+1}-y^k}{\tau}+Ay^k=f.$$

Здесь же проиллюстрируем возможность построения итерационного метода путем специального выбора матрицы B в канонической форме записи итерационного процесса.

Далее будем предполагать, что матрица системы уравнений $Ay=f$ симметрична и положительно определена. Введем матрицу

$$R=(r_{ij}),\quad r_{ij}=\begin{cases} a_{ij}, & i>j;\\ 0{,}5\,a_{ij}, & i=j;\\ 0, & i<j; \end{cases}\quad i,j=1,2,\ldots,n.$$

Матрица R является нижней треугольной матрицей, а транспонированная по отношению к ней матрица R^T — верхней треугольной. Матрица A пред-

ставима в виде $A = R + R^T$, причем

$$0 < (Av, v) = ((R + R^T)v, v) = 2(Rv, v),\ \forall\, v \neq 0 \Rightarrow R, R^T > 0.$$

Для попеременно–треугольного итерационного метода матрица B определяется как произведение

$$B = (E + \omega R^T)(E + \omega R),$$

где E — единичная матрица, а $\omega > 0$ — числовой параметр. Такой выбор матрицы B обусловлен следующими обстоятельствами.

1) Используя вспомогательное промежуточное значение $y^{k+1/2}$, где

$$(E + \omega R^T)\underbrace{(E + \omega R)y^{k+1}}_{y^{k+1/2}} = \underbrace{(B - \tau A)y^k + \tau f}_{\varphi_k},$$

решение на новой итерации легко находится в два этапа:

$(E + \omega R^T)y^{k+1/2} = \varphi_k$ — система с верхней треугольной матрицей;
$(E + \omega R)y^{k+1} = y^{k+1/2}$ — система с нижней треугольной матрицей.

Замечание. Отсюда название метода.

2) Поскольку $B^T = B > 0$, так как

$$B = E + \omega A + \omega^2 R^T R \;\Rightarrow B^T = B,$$
$$(Bv, v) = ((E + \omega R)v, (E + \omega R)v) > 0 \;\Rightarrow B > 0,$$

то для попеременно-треугольного итерационного метода можно использовать полученные ранее оценки сходимости.

Лемма 1.4. *Пусть существуют положительные постоянные δ и Δ такие, что выполнены матричные неравенства $A \geqslant \delta E$, $4R^T R \leqslant \Delta A$. Тогда для матриц $A = R + R^T$ и $B(\omega) = (E + \omega R^T)(E + \omega R)$ справедливы неравенства*

$$\gamma_1 B \leqslant A \leqslant \gamma_2 B, \text{ где } \gamma_1 = \left(\frac{1}{\delta} + \omega + \frac{\omega^2 \Delta}{4}\right)^{-1}, \; \gamma_2 = \frac{1}{2\omega}.$$

▼ Доказательство.

$$B(\omega) = E + \omega A + \omega^2 R^T R \leqslant \left(\frac{1}{\delta} + \omega + \frac{\omega^2 \Delta}{4}\right) A;$$
$$B(\omega) = E + \omega A + \omega^2 R^T R = E - \omega A + \omega^2 R^T R + 2\omega A =$$
$$= (E - \omega R^T)(E - \omega R) + 2\omega A \Rightarrow B(\omega) \geqslant 2\omega A.$$

▲ Утверждение доказано.

Замечание. Тем самым показано, что нахождение постоянных γ_1 и γ_2 сводится к нахождению постоянных δ и Δ. При выполнении неравенств $A \geqslant \delta E$, $4R^TR \leqslant \Delta A$ для произвольного вектора $v \neq 0$ имеем

$$\delta\|v\|^2 \leqslant (Av,v) = \frac{(Av,v)^2}{(Av,v)} = \frac{4(Rv,v)^2}{(Av,v)} \leqslant \frac{4\|Rv\|^2\|v\|^2}{(Av,v)} =$$
$$= \frac{4(R^TRv,v)\|v\|^2}{(Av,v)} \leqslant \frac{\Delta(Av,v)\|v\|^2}{(Av,v)} = \Delta\|v\|^2.$$

Отсюда следует, что $\delta \leqslant \Delta$. В качестве константы δ можно взять минимальное собственное значение $\lambda_{\min}(A)$ матрицы A. Также отметим, что, поскольку $(Av,v) \leqslant \Delta\|v\|^2$, то выполняется неравенство $\Delta \geqslant \lambda_{\max}(A)$, где $\lambda_{\max}(A)$ — максимальное собственное значение матрицы A.

Теорема 1.6. *Предположим, что для симметричной и положительно определенной матрицы $A = R + R^T$ известны положительные постоянные δ и Δ, при которых выполнены неравенства $A \geqslant \delta E$, $4R^TR \leqslant \Delta A$. Пусть*

$$\omega = \frac{2}{\sqrt{\delta\Delta}},\ \tau = \frac{2}{\gamma_1+\gamma_2},\ \textit{где}\ \gamma_1 = \frac{\delta}{2(1+\sqrt{\eta})},\ \gamma_2 = \frac{\delta}{4\sqrt{\eta}},\ \eta = \frac{\delta}{\Delta}.$$

Тогда попеременно–треугольный итерационный метод сходится и для его погрешности справедлива оценка

$$\|y^k - y\|_A \leqslant \rho^k \|y^0 - y\|_A, \text{ где } \rho = \frac{1-\sqrt{\eta}}{1+3\sqrt{\eta}}.$$

▼ Доказательство. Согласно теореме 1.5 для выполнения требуемой оценки погрешности с константой

$$\rho(\omega) = \frac{1-\xi}{1+\xi} = 1 - \frac{2\xi}{1+\xi} = 1 - \frac{2}{1+\xi^{-1}},\ \xi(\omega) = \frac{\gamma_1(\omega)}{\gamma_2(\omega)}$$

достаточно положить $\tau = 2/(\gamma_1+\gamma_2)$. Выберем параметр $\omega > 0$ так, чтобы минимизировать $\rho(\omega)$. Для этого достаточно найти значение $\omega = \omega_0$, при котором функция $\xi^{-1}(\omega)$ достигает минимума. Согласно лемме 1.4

$$\xi^{-1}(\omega) = \frac{\gamma_2(\omega)}{\gamma_1(\omega)} = \frac{1}{2\omega}\left(\frac{1}{\delta} + \omega + \frac{\omega^2\Delta}{4}\right) = \frac{1}{2} + \frac{1}{2}\left(\frac{1}{\omega\delta} + \frac{\omega\Delta}{4}\right) =$$
$$= \frac{1}{2} + \frac{1}{2}\left(\frac{1}{\sqrt{\omega\delta}} - \frac{\sqrt{\omega\Delta}}{2}\right)^2 + \frac{1}{2}\sqrt{\frac{\Delta}{\delta}}.$$

Отсюда находим точку минимума $\omega_0 = 2/\sqrt{\delta\Delta}$. Подставляя это значение ω_0 в выражения для γ_1 и γ_2, получим

$$\gamma_1(\omega_0) = \left(\frac{1}{\delta} + \frac{2}{\sqrt{\delta\Delta}} + \frac{1}{\delta}\right)^{-1} = \frac{1}{2}\left(\frac{\sqrt{\delta}+\sqrt{\Delta}}{\delta\sqrt{\Delta}}\right)^{-1} = \frac{\delta}{2(1+\sqrt{\eta})},$$

$$\gamma_2(\omega_0) = \frac{\sqrt{\delta\Delta}}{4} = \frac{\delta}{4\sqrt{\eta}}, \text{ где } \eta = \frac{\delta}{\Delta} \in (0;1].$$

$$\text{Тогда} \xi(\omega_0) = \frac{\gamma_1}{\gamma_2} = \frac{2\sqrt{\eta}}{1+\sqrt{\eta}}, \ \rho(\omega_0) = \frac{1-\xi}{1+\xi} = \frac{1-\sqrt{\eta}}{1+3\sqrt{\eta}} \in [0;1).$$

▲ Утверждение доказано.

Применение попеременно–треугольного метода к модельной задаче.

Обсудим применение попеременно–треугольного итерационного метода к модельной задаче (1.14). Напомним, что для модельной задачи $A^T = A > 0$.

Для того, чтобы применить попеременно–треугольный итерационный метод необходимо знать постоянные δ и Δ, определяющие параметры метода. Как уже отмечалось, в качестве δ и Δ можно взять минимальное собственное

значение матрицы A и, соответственно, максимальное собственное значение матрицы A, то есть

$$\delta = \lambda_{\min}(A) = \frac{4}{h^2}\sin^2\frac{\pi h}{2},\ \Delta = \lambda_{\max}(A) = \frac{4}{h^2}\cos^2\frac{\pi h}{2}.$$

Согласно теореме 1.6 при выборе параметров метода

$$\omega = \frac{2}{\sqrt{\delta\Delta}},\ \tau = \frac{2}{\gamma_1+\gamma_2},\ \text{где}$$
$$\gamma_1 = \frac{\delta}{2(1+\sqrt{\eta})},\ \gamma_2 = \frac{\delta}{4\sqrt{\eta}},\ \eta = \frac{\delta}{\Delta} = \operatorname{tg}^2\frac{\pi h}{2},$$

справедлива оценка погрешности с постоянной

$$\rho = \frac{1-\sqrt{\eta}}{1+3\sqrt{\eta}} = \frac{1-\operatorname{tg}\frac{\pi h}{2}}{1+3\operatorname{tg}\frac{\pi h}{2}} \approx \left(1-\frac{\pi h}{2}\right)\left(1-3\frac{\pi h}{2}\right) \approx 1-2\pi h.$$

Отсюда минимальное число итераций, необходимое для достижения заданной точности ε, равно

$$k_0(\varepsilon) = \frac{\ln(1/\varepsilon)}{\ln(1/\rho)} \approx \frac{\ln(1/\varepsilon)}{-\ln(1-2\pi h)} \approx \frac{\ln(1/\varepsilon)}{2\pi h} = \frac{N}{2\pi}\ln\frac{1}{\varepsilon}.$$

При $\varepsilon = 0,5 \cdot 10^{-4}$ получаем, что $k_0(\varepsilon) \approx 1,6N$.

Замечание. Напомним, что для метода простой итерации число итераций при больших N оценивалось как $O(N^2)$. То есть попеременно–треугольный итерационный метод обеспечивает на порядок более быструю сходимость.

1.4.2 Чебышевский набор итерационных параметров

Используем для решения уравнения $Ay = f$ следующую нестационарную итерационную схему

$$B\frac{y^l - y^{l-1}}{\tau_l} + Ay^{l-1} = f, l = 1, 2, \ldots, k.$$

Повысить скорость сходимости итерационного метода можно за счет использования переменного итерационного параметра τ_l, зависящего от номера итерации. При фиксированном числе итераций k можно указать набор итерационных параметров τ_1, τ_2, ..., τ_k , обеспечивающий наилучшую скорость сходимости вне зависимости от выбора начального приближения.

Перейдем от y^l-го итерационного приближения к погрешности $z^l = y^l - y$. Тогда получим следующее равенство

$$B\frac{z^l - z^{l-1}}{\tau_l} + Az^{l-1} = 0.$$

Отсюда выразим погрешность на l-й итерации z^l

$$z^l = (E - \tau_l B^{-1}A)z^{l-1}, \quad l = \overline{1,\, k}.$$

Рекурсивно применяя эту формулу для z^k, получим выражение для z^k

$$z^k = (E - \tau_k B^{-1}A)(E - \tau_{k-1}B^{-1}A)\dots(E - \tau_1 B^{-1}A)z^0.$$

Фиксируем число итераций (k) и постараемся выбрать τ_l так, чтобы погрешность на k-й итерации была бы минимально возможной:

$$||z^k|| \longrightarrow \inf_{\tau_l}.$$

Эта задача имеет точное решение (см. [2]). Соответствующий набор итерационных параметров τ_l принято называть чебышевским набором итерационных параметров.

Справедлива следующая теорема (доказательство см. в [2]).

Теорема 1.7. *Пусть* $A^T = A > 0$, $B^T = B > 0$, *а* τ_l *вычисляются по формуле:*

$$\tau_l = \frac{\tau_0}{1 + \rho_0 t_l}, \text{ где } \tau_0 = \frac{2}{\lambda_{min}(B^{-1}A) + \lambda_{max}(B^{-1}A)}, \quad \rho_0 = \frac{1-\xi}{1+\xi},$$

$$\xi = \frac{\lambda_{min}(B^{-1}A)}{\lambda_{max}(B^{-1}A)}, \quad t_l = \cos\frac{(2l-1)\pi}{2k}, \; l = \overline{1,\, k}.$$

Тогда погрешность $||y^k - y||_A$ *будет минимально возможной, и для нее справедлива оценка*

$$||y^k - y||_A \leqslant q_k ||y^0 - y||_A,$$

где $q_k = \dfrac{\rho_1^k}{1 + \rho_1^{2k}}$, $\rho_1 = \dfrac{1 - \sqrt{\xi}}{1 + \sqrt{\xi}}$.

Замечание. Выясним, при каких значениях k выполняется условие выхода из итерационного процесса $||y^k - y||_A \leqslant \varepsilon ||y^0 - y||_A$.

Из предыдущей теоремы следует, что это условие будет выполнено, если $q_k \leqslant \varepsilon$. То есть,

$$\frac{\rho_1^k}{1 + \rho_1^{2k}} \leqslant \varepsilon \iff \varepsilon(\rho_1^k)^2 - \rho_1^k + \varepsilon \geqslant 0.$$

Корнями этого квадратного неравенства будут

$$\rho_1^k = \frac{1 \pm \sqrt{1 - 4\varepsilon^2}}{2\varepsilon} \approx \frac{1 \pm (1 - 2\varepsilon^2)}{2\varepsilon} = \left[\begin{array}{l} \rho_1^k = \varepsilon; \\ \rho_1^k = \dfrac{1}{\varepsilon} - \varepsilon. \end{array} \right.$$

Отсюда следует, что $q_k \leqslant \varepsilon$ при

$$\left[\begin{array}{l} \rho_1^k \leqslant \varepsilon; \\ \rho_1^k \geqslant \dfrac{1}{\varepsilon} - \varepsilon. \end{array}\right.$$

Из условий предыдущей теоремы следует, что ρ_1^k меньше единицы. Поэтому

$$q_k \leqslant \varepsilon \iff \rho_1^k \leqslant \varepsilon.$$

Таким образом, условие выхода из итерационного процесса выполнено при

$$k \geqslant k_0(\varepsilon) = \left[\frac{\ln \frac{1}{\varepsilon}}{\ln \frac{1}{\rho_1}}\right].$$

Для скорости сходимости $(\ln \frac{1}{\rho_1})$ верна оценка

$$\ln \frac{1}{\rho_1} = \ln \frac{1+\sqrt{\xi}}{1-\sqrt{\xi}} \approx \ln(1 + 2\sqrt{(}\xi)) \approx 2\sqrt{\xi}.$$

Применение метода Ричардсона к модельной задаче.

Воспользуемся сформулированной теоремой для оценки числа итераций $k_0(\varepsilon)$ в случае применения метода Ричардсона ($B = E$) с чебышевскими итерационными параметрами для решения модельной задачи.

В рассматриваемом случае собственные значения $\lambda_{min}(B^{-1}A)$ и $\lambda_{max}(B^{-1}A)$ равны

$$\lambda_{min}(B^{-1}A) = \lambda_{min}(A) = \frac{4}{h^2}\sin^2(\frac{\pi h}{2}), \quad \lambda_{max}(B^{-1}A) = \lambda_{max}(A) = \frac{4}{h^2}\cos^2(\frac{\pi h}{2}),$$

поэтому $\xi = \frac{\lambda_{min}}{\lambda_{max}} = \mathrm{tg}^2(\frac{\pi h}{2})$.

Пусть, как и прежде, $\varepsilon = 0,5 \cdot 10^{-4} \approx e^{-10}$. В этом случае, в соответствии с замечанием к теореме, получаем, что

$$k_0(\varepsilon) = \frac{\ln\frac{1}{\varepsilon}}{\ln\frac{1}{\rho_1}} \approx \frac{10}{2\,\mathrm{tg}(\frac{\pi h}{2})} \approx \frac{10}{\pi h} = \ \{\mathsf{h} = \frac{1}{\mathsf{N}}\} \ = \frac{10}{\pi}N \approx 3,2N.$$

Тогда при $N = 10$ нам понадобится 32 итерации, а в случае $N = 100 - 320$ итераций, что сопоставимо с соответствующими величинами для попеременно–треугольного итерационного метода.

Упорядоченный набор чебышевских параметров.

Замечание. Оценка числа итераций $k_0(\varepsilon)$ не зависит от порядка, в котором применяются итерационные параметры τ_l, однако этот порядок существенно влияет на вычислительную устойчивость алгоритма. При практическом применении данного метода используется специальный алгоритм построения упорядоченного набора итерационных параметров, обеспечивающий устойчивость вычислений.

Описанный в предыдущем разделе итерационный процесс гарантирует минимальное значение нормы погрешности итерационного приближения на k-й итерации. Расчет k-ого итерационного приближения осуществляется последовательно от y^0 до y^k в соответствии с используемой расчетной схемой итерационного процесса. Данная расчетная схема не гарантирует монотонного по итерациям убывания нормы погрешности итерационного приближения. Поэтому, при реализации итерационного метода с чебышевским набором параметров, возможен рост нормы погрешности на нескольких соседних итерациях, что может приводить к возникновению неприятных ситуаций (переполнению арифметических устройств). Для предотвращения таких неприятностей рекомендуется использовать, так называемый, упорядоченный чебышевский набор итерационных параметров.

В рассматриваемом методе формула для итерационного параметра τ_l содержит параметр $t_l = \cos(\frac{(2l-1)\pi}{2k})$. Запишем это выражение следующим образом

$$t_l = \cos\left(\frac{(2l-1)\pi}{2k}\right) = \cos\left(\frac{\pi}{2k}\,\theta_l^k\right), \quad l = \overline{1,\,k},$$

где θ_l^k нечетное число из множества θ^k, состоящего из нечетных чисел от 1 до $2k-1$.

Сформулируем, как пример, правило упорядочевания элементов множества θ^k в одном частном случае. Для произвольного k правило упорядочевания элементов множества θ^k можно найти в [2].

Пусть $k = 2^p$. Тогда элементы θ_l^k будем упорядочивать используя следующие рекуррентные формулы

$$\theta_1^1 = 1, \quad \theta_{2i-1}^{2m} = \theta_i^m, \quad \theta_{2i}^{2m} = 4m - \theta_{2i-1}^{2m}, \quad i = \overline{1,\,m}, \quad m = 1, 2, 4, \ldots, 2^{p-1}.$$

Проиллюстрируем эти формулы на примере $k = 2^3 = 8$. Так как $p = 3$, то $m = 1, 2, 4$. Итак

$$\begin{aligned} & \theta^1 = \left\{\theta_1^1\right\} = \{1\} \\ m = 1: \quad & \theta^2 = \left\{\theta_1^2, \theta_2^2\right\} = \left\{\theta_1^2 = \theta_1^1 = 1, \theta_2^2 = 4 - \theta_1^1 = 3\right\} \end{aligned}$$

$$
\begin{aligned}
m=2: \quad \theta^4 &= \left\{\theta_1^4, \theta_2^4, \theta_3^4, \theta_4^4\right\} = \left\{\theta_1^4 = \theta_1^2 = 1, \theta_2^4 = 8 - \theta_1^4 = 7,\right. \\
&\left. \theta_3^4 = \theta_2^2 = 3, \theta_4^4 = 8 - \theta_3^4 = 5\right\} \\
m=4: \quad \theta^8 &= \left\{\theta_1^8, \theta_2^8, \ldots, \theta_8^8\right\} = \left\{\theta_1^8 = \theta_1^4 = 1, \theta_2^8 = 16 - \theta_1^8 = 15,\right. \\
&\theta_3^8 = \theta_2^4 = 7, \theta_4^8 = 16 - \theta_3^8 = 9, \theta_5^8 = \theta_3^4 = 3, \\
&\left. \theta_6^8 = 16 - \theta_5^8 = 13, \theta_7^8 = \theta_4^4 = 5, \theta_8^8 = 16 - \theta_7^8 = 11\right\}.
\end{aligned}
$$

1.4.3 Попеременно–треугольный итерационный метод с упорядоченным набором чебышевских параметров

Рассмотрим итерационный метод соединяющий в себе достоинства как попеременно–треугольного метода, так и метода с чебышевским набором итерационных параметров. Пусть система линейных алгебраических уравнений $Ay = f$ имеет симметричную и положительно определенную матрицу A. Используем для ее решения линейный одношаговый нестационарный неявный итерационный метод, для которого справедливо следующее утверждение.

Теорема 1.8. *Пусть для решения уравнения $Ay = f$ используется итерационная схема*

$$
B\frac{y^l - y^{l-1}}{\tau_l} + Ay^{l-1} = f, \quad l = \overline{1,\, k},
$$

где $B = (E+\omega R^T)(E+\omega R), R+R^T = A$. *Пусть известны положительные постоянные* δ *и* Δ *такие, что выполнены матричные неравенства*

$$A \geqslant \delta E,\ \Delta A \geqslant 4R^T R.$$

Пусть

$$\tau_l = \frac{\tau_0}{1+\rho_0 t_l}, \quad \tau_0 = \frac{2}{\gamma_1+\gamma_2}, \quad \rho_0 = \frac{\gamma_2-\gamma_1}{\gamma_2+\gamma_1},$$

$$\gamma_1 = \frac{\delta}{2(1+\sqrt{\eta})}, \quad \gamma_2 = \frac{\delta}{4\sqrt{\eta}}, \quad \eta = \frac{\delta}{\Delta},$$

$$t_l = \cos\left(\frac{\pi}{2k}\,\theta_l^k\right), \quad \omega = \frac{2}{\sqrt{\delta\Delta}}.$$

Тогда, для выполнения на k*-ой итерации неравенства*

$$||A(y^k - y)||_{B^{-1}} \leqslant \varepsilon ||A(y^0 - y)||_{B^{-1}}$$

достаточно $k > k_0(\varepsilon) = \dfrac{\ln\frac{2}{\varepsilon}}{2\sqrt{2}\,\sqrt[4]{\eta}}$ *итераций.*

Замечание. Доказательство теоремы приведено в [2].

Применение к модельной задаче.

Воспользуемся этой теоремой для оценки числа итераций $k_0(\varepsilon)$ в случае применения попеременно-треугольного метода с чебышевскими итерационными параметрами для решения модельной задачи. Для модельной задачи $\delta = \frac{4}{h^2}\sin^2\frac{\pi h}{2}$ и $\Delta = \frac{4}{h^2}\cos^2\frac{\pi h}{2}$.

Тогда, при $\varepsilon = 0,5 \cdot 10^{-4} \approx e^{-10}$ для числа итераций $k_0(\varepsilon)$ справедлива приближенная оценка

$$k_0(\varepsilon) = \frac{\ln 2 + 10}{2\sqrt{2}\sqrt[4]{\mathrm{tg}^2(\frac{\pi h}{2})}} \approx \frac{\ln 2 + 10}{2\sqrt{2}\sqrt{\frac{\pi h}{2}}} = \{\mathsf{h} = \frac{1}{\mathsf{N}}\} = \frac{(\ln 2 + 10)\sqrt{N}}{2\sqrt{\pi}} \approx 3\sqrt{N}.$$

Таким образом, при решении системы с числом уравнений $N = 10$ понадобится 10 итераций, а в случае $N = 100 - 30$ итераций.

Приведем асимптотические оценки при $N \to \infty$ минимального числа итераций $k_0(\varepsilon)$, необходимого для достижения заданной точности ε, для рассмотренных итерационных методов решения модельной задачи:

$$k_0(\varepsilon) \approx \frac{2N^2}{\pi^2}\ln\frac{1}{\varepsilon} \qquad - \text{для метода простой итерации;}$$

$$k_0(\varepsilon) \approx \frac{N}{2\pi}\ln\frac{1}{\varepsilon} \qquad - \text{для ПТИМ;}$$

$$k_0(\varepsilon) \approx \frac{\sqrt{N}}{2\sqrt{\pi}} \ln \frac{2}{\varepsilon} \qquad - \text{для ПТИМ с чебышевскими параметрами.}$$

Здесь аббревиатура ПТИМ является сокращенным обозначением попеременно–треугольного итерационного метода. Превосходство в скорости сходимости ПТИМ с чебышевскими параметрами над другими рассмотренными итерационными методами очевидно.

1.5 Итерационные методы вариационного типа

1.5.1 Одношаговые итерационные методы вариационного типа

Рассмотрим нестационарный одношаговый итерационный метод решения системы $Ay = f$ вида

$$B\frac{y^{k+1} - y^k}{\tau_{k+1}} + Ay^k = f,\ k = 0, 1, \ldots. \tag{1.16}$$

Здесь невырожденная матрица B не зависит от номера итерации k.

Пусть, как и ранее, $z^k = y^k - y$ есть погрешность на $k-$ой итерации, удовлетворяющая соотношению

$$z^{k+1} = (E - \tau_{k+1}B^{-1}A)z^k.$$

Пусть D — некоторая матрица, удовлетворяющая условиям $D^T = D > 0$. Будем выбирать итерационные параметры τ_{k+1}, минимизирующие энергетическую норму (см. пункт 1.3.2) погрешности $\|z^{k+1}\|_D$. Такой способ построения итерационного процесса называется *локальной минимизацией*.

Обозначим $w^{k+1} = D^{1/2} z^{k+1}$, тогда

$$\|z^{k+1}\|_D = \sqrt{(Dz^{k+1}, z^{k+1})} = \sqrt{(D^{1/2} z^{k+1}, D^{1/2} z^{k+1})} = \|w^{k+1}\|,$$

и минимизация энергетической нормы $\|z^{k+1}\|_D$ эквивалентна минимизации нормы $\|w^{k+1}\|$. Перепишем уравнение для погрешности в следующей форме

$$D^{1/2} z^{k+1} = D^{1/2}(E - \tau_{k+1} B^{-1} A) D^{-1/2} D^{1/2} z^k \;\Leftrightarrow$$
$$\Leftrightarrow\; w^{k+1} = (E - \tau_{k+1} D^{1/2} B^{-1} A D^{-1/2}) w^k \;\Leftrightarrow\; w^{k+1} = (E - \tau_{k+1} C) w^k.$$

Здесь $C = D^{1/2} B^{-1} A D^{-1/2}$ — невырожденная матрица. Считая что $w^k \neq 0$ (иначе на $k-$ой итерации найдено точное решение), получим

$$\|w^{k+1}\|^2 = ((E - \tau_{k+1} C) w^k, (E - \tau_{k+1} C) w^k) =$$
$$= \|w^k\|^2 + \tau_{k+1}^2 (Cw^k, Cw^k) - 2\tau_{k+1} (Cw^k, w^k) =$$
$$= \|w^k\|^2 + (Cw^k, Cw^k) \left(\tau_{k+1}^2 - 2\tau_{k+1} \frac{(Cw^k, w^k)}{(Cw^k, Cw^k)} \right) =$$
$$= \|w^k\|^2 + (Cw^k, Cw^k) \left(\tau_{k+1} - \frac{(Cw^k, w^k)}{(Cw^k, Cw^k)} \right)^2 - \frac{(Cw^k, w^k)^2}{(Cw^k, Cw^k)}.$$

Отсюда следует, что минимальное значение $\|w^{k+1}\|$ достигается при

$$\tau_{k+1} = \frac{(Cw^k, w^k)}{(Cw^k, Cw^k)}, \text{ где } \tau_{k+1} > 0, \quad \text{если } C > 0.$$

В дальнейшем будем предполагать условие $C > 0$ выполненным. Отметим, что при таком выборе τ_{k+1} верно равенство

$$\|w^{k+1}\| = \rho_{k+1}\|w^k\|, \text{ где } \rho_{k+1}^2 = 1 - \frac{(Cw^k, w^k)^2}{(Cw^k, Cw^k)(w^k, w^k)} < 1.$$

Учитывая, что $Cw^k = D^{1/2}B^{-1}Az^k$ и $w^k = D^{1/2}z^k$, получим

$$\tau_{k+1} = \frac{(DB^{-1}Az^k, z^k)}{(DB^{-1}Az^k, B^{-1}Az^k)}.$$

Перепишем уравнение итерационного метода (1.16) в виде

$$y^{k+1} = y^k - \tau_{k+1}B^{-1}(Ay^k - f) = y^k - \tau_{k+1}v^k.$$

Здесь вектор $v^k = B^{-1}(Ay^k - f) = B^{-1}r^k$ называется *поправкой*, а вектор $r^k = Ay^k - f = Az^k$ — *невязкой* на $k-$ой итерации. Поскольку $v^k = B^{-1}Az^k$,

приходим к равенству

$$\tau_{k+1} = \frac{(Dv^k, z^k)}{(Dv^k, v^k)}. \tag{1.17}$$

Это выражение содержит погрешность z^k, которую нельзя вычислить, поскольку неизвестно точное решение y задачи $Ay = f$. Однако за счет выбора матрицы D можно выразить τ_{k+1} через значения v^k и r^k, которые могут быть вычислены на каждой итерации. Например, если $A^T = A > 0$, то можно выбрать матрицу $D = A$. В этом случае

$$\tau_{k+1} = \frac{(Av^k, z^k)}{(Av^k, v^k)} = \frac{(v^k, Az^k)}{(Av^k, v^k)} = \frac{(v^k, r^k)}{(Av^k, v^k)}.$$

Таким образом, путем выбора матриц B и D можно строить различные одношаговые итерационные методы вариационного типа. Далее будем называть такие методы *градиентными*.

Замечание. Рассмотрим выражение для итерационного параметра τ_1. Согласно (1.17) в градиентных методах

$$\tau_1 = \frac{(Dv^0, z^0)}{(Dv^0, v^0)} = \frac{(DB^{-1}r^0, z^0)}{(DB^{-1}r^0, B^{-1}r^0)} = \frac{(DB^{-1}Az^0, z^0)}{(DB^{-1}Az^0, B^{-1}Az^0)}.$$

Отметим, что если в градиентных методах в качестве матрицы B взять матрицу A, определяющую решаемую систему линейных алгебраических уравнений, то получим, что

$$\tau_1 = \frac{(Dz^0, z^0)}{(Dz^0, z^0)} = 1.$$

Тогда для итерационного приближения y^1 верно уравнение

$$A\frac{y^1 - y^0}{1} + Ay^0 = f \;\Leftrightarrow\; Ay^1 = f.$$

То есть, y^1 будет совпадать с точным решением исходной системы линейных уравнений. Это означает сходимость градиентных методов с матрицей $B = A$ за одну итерацию к точному решению задачи.

Примеры градиентных методов будут приведены в следующих пунктах. Здесь же будем предполагать, что выбор матриц D и B позволяет вычислять параметры τ_{k+1} на каждой итерации.

Далее обсудим сходимость градиентных методов. Ограничимся так называемым самосопряженным случаем, когда матрица $C^T = C > 0$. Тогда существуют такие положительные постоянные γ_1 и γ_2, что

$$\gamma_1 E \leqslant C \leqslant \gamma_2 E, \text{ где } C = D^{1/2}B^{-1}AD^{-1/2} = D^{-1/2}(DB^{-1}A)D^{-1/2}.$$

Отсюда получим, что требование $C^T = C > 0$ равносильно условиям

$$(DB^{-1}A)^T = DB^{-1}A > 0,\ \gamma_1 D \leqslant DB^{-1}A \leqslant \gamma_2 D,\ \gamma_1 > 0. \tag{1.18}$$

Будем считать, что γ_1 и γ_2 — минимальное и максимальное собственное значение матрицы C, соответственно. Это наиболее точные значения постоянных γ_1 и γ_2, при которых выполнены неравенства в (1.18).

Теорема 1.9. *Пусть выполнены условия* (1.18). *Тогда итерационный метод* (1.16), (1.17) *сходится и для его погрешности справедлива оценка*

$$\|z^k\|_D \leqslant \rho^k \|z^0\|_D,\ \textit{где}\ \rho = \frac{1-\xi}{1+\xi},\ \xi = \frac{\gamma_1}{\gamma_2}.$$

▼ Доказательство. Напомним, что векторы $w^k = D^{1/2}z^k$ при каждом $k = 0, 1, \ldots$ удовлетворяют соотношению $w^{k+1} = (E - \tau_{k+1}C)w^k$. Сопоставим этому равенству уравнение

$$\frac{w^{k+1} - w^k}{\tau_0} + Cw^k = 0,\ \text{где}\ \tau_0 = \frac{2}{\gamma_1 + \gamma_2},$$

в котором параметр τ_{k+1} заменен значением τ_0. Последнее равенство совпадает с уравнением для погрешности метода простой итерации, примененного

для решения линейной системы с матрицей C, где $C^T = C > 0$, и оптимальным параметром τ. Поэтому в силу следствия из теоремы 1.5 справедлива оценка

$$\|w^{k+1}\|\big|_{\tau=\tau_0} \leqslant \rho\|w^k\|.$$

Поскольку в итерационном методе (1.16) параметр τ_{k+1} выбирается исходя из минимизации $\|w^{k+1}\|$, также верно неравенство

$$\|w^{k+1}\|\big|_{\tau=\tau_{k+1}} = \min_{\tau>0}\|w^{k+1}\| \leqslant \|w^{k+1}\|\big|_{\tau=\tau_0}.$$

Приходим к тому, что при каждом $k = 0, 1, \ldots$ справедлива оценка

$$\|w^{k+1}\|\big|_{\tau=\tau_{k+1}} \leqslant \rho\|w^k\| \;\Rightarrow \|z^k\|_D \leqslant \rho^k\|z^0\|_D.$$

▲ Утверждение доказано.

Таким образом показано, что в самосопряженном случае любой градиентный метод сходится не хуже, чем соответствующий ему метод простой итерации с симметричной матрицей C. Это не означает, что скорость градиентного метода не может оказаться выше. Например, если в качестве начального приближения выбрать вектор y^0 такой, что соответствующее ему значение $w^0 = D^{1/2}z^0 = D^{1/2}(y^0 - y) = \mu$, где μ — произвольный собственный вектор

матрицы C, отвечающий собственному значению λ, то получим

$$\tau_1 = \frac{(C\mu, \mu)}{(C\mu, C\mu)} = \frac{1}{\lambda},\ \rho_1^2 = 1 - \frac{(C\mu, \mu)^2}{(C\mu, C\mu)(\mu, \mu)} = 0.$$

Отсюда $\|w^1\| = \rho_1 \|w^0\| = 0$, что равносильно равенству $y^1 = y$, которое означает сходимость метода за одну итерацию. Таким образом, при «удачном» выборе начального приближения градиентные методы могут иметь существенно более высокую скорость сходимости по сравнению с соответствующими им методами простой итерации.

Замечание. Аналогично можно показать (см. [8]), что при «неудачном» выборе w^0 для всех $k = 0, 1, \ldots$ будет верным равенство $\|w^{k+1}\| = \rho \|w^k\|$. Это означает сходимость со скоростью $\ln(1/\rho)$ и неулучшаемость оценки, доказанной в предыдущей теореме.

Рассмотрим случай, когда $A^T = A > 0$, $B^T = B > 0$, $D = A$. Тогда матрица $DB^{-1}A$ симметрична и неравенства (1.18) принимают вид

$$\gamma_1 A \leqslant AB^{-1}A \leqslant \gamma_2 A \ \Leftrightarrow\ \gamma_1 E \leqslant A^{1/2}B^{-1}A^{1/2} \leqslant \gamma_2 E \ \Leftrightarrow$$
$$\Leftrightarrow\ \gamma_1 A^{-1} \leqslant B^{-1} \leqslant \gamma_2 A^{-1} \ \Leftrightarrow\ \gamma_1 B \leqslant A \leqslant \gamma_2 B.$$

Отсюда вытекает (см. теорему 1.5), что градиентный метод будет сходиться не хуже стационарного метода (1.11) с теми же матрицами A, B и опти-

мальным итерационным параметром. При этом для задания оптимального параметра необходимо находить минимальное и максимальное собственные значения γ_1 и γ_2 матрицы $B^{-1}A$. Для соответствующего градиентного метода эти данные не требуются, хотя и могут быть использованы для априорных оценок скорости сходимости.

Замечание. Например, если матрица B выбирается так же, как в попеременно–треугольном итерационном методе, для сходимости соответствующего градиентного метода можно получить априорную оценку сходимости на основе теоремы 1.6.

В заключение отметим, что минимальными требованиями, необходимыми для построения градиентных методов, являются условия $D^T = D > 0$, $C = D^{1/2}B^{-1}AD^{-1/2} > 0$, а также возможность нахождения параметров τ_{k+1} по формуле (1.17). Эти условия не предполагают, что матрица A исходной системы $Ay = f$ непременно должна быть симметричной и(или) положительно определенной.

1.5.2 Примеры одношаговых итерационных методов вариационного типа

Метод скорейшего спуска.

Метод скорейшего спуска применим для случая симметричной и положительно определенной матрицы системы $Ay = f$, то есть $A^T = A > 0$. Тогда можно выбрать матрицу $D = A$ и, как уже отмечалось,

$$\tau_{k+1} = \frac{(Av^k, z^k)}{(Av^k, v^k)} = \frac{(v^k, Az^k)}{(Av^k, v^k)} = \frac{(v^k, r^k)}{(Av^k, v^k)}.$$

Условие $C > 0$ приводит к ограничению $B > 0$, поскольку

$$(Cy, y) = (A^{1/2}B^{-1}A^{1/2}y, y) = (B^{-1}A^{1/2}y, A^{1/2}y) > 0 \ \Rightarrow B > 0.$$

Пусть $B = E$, тогда поправка $v^k = B^{-1}r^k$ совпадает с невязкой $r^k = Ay^k - f$, метод является явным и расчетные формулы принимают вид

$$y^{k+1} = y^k - \tau_{k+1}r^k, \ \tau_{k+1} = \frac{(r^k, r^k)}{(Ar^k, r^k)}, \ k = 0, 1, \ldots.$$

Метод минимальных невязок.

В *методе минимальных невязок* выбирается $D = A^T A$, тем самым $D^T = D > 0$. Итерационные параметры вычисляются следующим образом:

$$\tau_{k+1} = \frac{(A^T A v^k, z^k)}{(A^T A v^k, v^k)} = \frac{(A v^k, A z^k)}{(A v^k, A v^k)} = \frac{(A v^k, r^k)}{(A v^k, A v^k)}.$$

Условие $C > 0$ приводит к следующему ограничению применимости метода:

$$\begin{aligned}(Cy, y) &= (D^{1/2} B^{-1} A D^{-1/2} y, y) = \{\bar{y} = D^{-1/2} y\} = (D^{1/2} B^{-1} A \bar{y}, D^{1/2} \bar{y}) = \\ &= (D B^{-1} A \bar{y}, \bar{y}) = (A^T A B^{-1} A \bar{y}, \bar{y}) = (A B^{-1} A \bar{y}, A \bar{y}) = \{\tilde{y} = B^{-1} A \bar{y}\} = \\ &= (A\tilde{y}, B\tilde{y}) = (B^T A \tilde{y}, \tilde{y}) > 0 \ \Rightarrow B^T A > 0.\end{aligned}$$

Если матрица $A > 0$, можно использовать явный метод, выбирая $B = E$. В противном случае необходимо подбирать легко обратимую матрицу B, для которой условие $B^T A > 0$ выполняется.

Название данного метода объясняется тем, что энергетическая норма погрешности

$$\|z^{k+1}\|_D = \sqrt{(A^T A z^{k+1}, z^{k+1})} = \sqrt{(A z^{k+1}, A z^{k+1})} = \|r^{k+1}\|.$$

Поэтому минимизация нормы $\|z^{k+1}\|_D$ в рассматриваемом методе эквивалентна минимизации нормы невязки $\|r^{k+1}\|$.

Метод минимальных поправок.

В *методе минимальных поправок* выбирается $D = A^T B^{-1} A$. Условие $D^T = D > 0$ приводит к ограничениям на выбор матрицы B, а именно $B^T = B > 0$. Итерационные параметры вычисляются следующим образом:

$$\tau_{k+1} = \frac{(A^T B^{-1} A v^k, z^k)}{(A^T B^{-1} A v^k, v^k)} = \frac{(A v^k, B^{-1} A z^k)}{(B^{-1} A v^k, A v^k)} = \frac{(A v^k, v^k)}{(B^{-1} A v^k, A v^k)}.$$

Условие $C > 0$ приводит к дополнительному ограничению:

$$(Cy, y) = (D^{1/2} B^{-1} A D^{-1/2} y, y) = \{\bar{y} = D^{-1/2} y\} = (D^{1/2} B^{-1} A \bar{y}, D^{1/2} \bar{y}) =$$
$$= (D B^{-1} A \bar{y}, \bar{y}) = (A^T B^{-1} A B^{-1} A \bar{y}, \bar{y}) = (A B^{-1} A \bar{y}, B^{-1} A \bar{y}) =$$
$$= \{\tilde{y} = B^{-1} A \bar{y}\} = (A\tilde{y}, \tilde{y}) > 0 \ \Rightarrow A > 0.$$

Итак, метод применим для случая положительно определенной матрицы A.

Для энергетической нормы погрешности справедливы равенства

$$\|z^{k+1}\|_D^2 = (A^T B^{-1} A z^{k+1}, z^{k+1}) = (B^{-1} r^{k+1}, r^{k+1}) = (v^{k+1}, B v^{k+1}) = \|v^{k+1}\|_B^2.$$

Поэтому минимизация нормы $\|z^{k+1}\|_D$ в рассматриваемом методе эквивалентна минимизации нормы поправки $\|v^{k+1}\|_B$. Отсюда и название метода.

Метод минимальных погрешностей.

Метод минимальных погрешностей определяется следующим выбором матриц D и B:

$$D = B_0,\ B = (A^T)^{-1}B_0,\ \text{где } B_0^T = B_0 > 0.$$

Здесь в качестве B_0 можно выбрать произвольную симметричную положительно определенную матрицу, которая легко обратима, например, диагональную. Итерационные параметры вычисляются следующим образом

$$\tau_{k+1} = \frac{(B_0 v^k, z^k)}{(B_0 v^k, v^k)} = \frac{(B_0 B^{-1} r^k, z^k)}{(B_0 B^{-1} r^k, v^k)} = \frac{(B_0 B_0^{-1} A^T r^k, z^k)}{(B_0 B_0^{-1} A^T r^k, v^k)} =$$
$$= \frac{(r^k, Az^k)}{(r^k, Av^k)} = \frac{(r^k, r^k)}{(r^k, Av^k)}.$$

Проверим выполнение условия $C > 0$:

$$(Cy, y) = (D^{1/2} B^{-1} A D^{-1/2} y, y) = \{\bar{y} = D^{-1/2} y\} = (D^{1/2} B^{-1} A \bar{y}, D^{1/2} \bar{y}) =$$

$$= (DB^{-1}A\bar{y}, \bar{y}) = (B_0 B_0^{-1} A^T A\bar{y}, \bar{y}) = (A\bar{y}, A\bar{y}) > 0.$$

Метод применим для произвольной невырожденной матрицы A.

Название метода объясняется тем, что на каждой итерации минимизируется норма погрешности $\|z^{k+1}\|_D = \|z^{k+1}\|_{B_0}$.

1.6 Методы сопряженных направлений

Методы решения систем линейных алгебраических уравнений, в которых вектор неизвестных представляется в виде линейной комбинации векторов, сопряженных (ортогональных) в какой-либо метрике, связанной с матрицей решаемой системы уравнений, называют методами сопряженных направлений [7]. Одним из таких методов является метод сопряженных градиентов.

1.6.1 Метод сопряженных градиентов

Пусть невырожденная квадратная матрица $A = (a_{ij})$ $(i, j = 1, 2, \ldots, n)$ является симметричной и положительно определенной $\left(A = A^T > 0\right)$. По-прежнему рассматривается решение системы линейных алгебраических уравнений

$$Ay = f, \tag{1.19}$$

где $f = \begin{pmatrix} f_1 & f_2 & \ldots & f_n \end{pmatrix}^T$ заданный вектор $(f \neq 0)$, а вектор

$y = \begin{pmatrix} y_1 & y_2 & \ldots & y_n \end{pmatrix}^T$ является искомым решением системы уравнений (1.19).

Как уже отмечалось, при $A = A^T > 0$ билинейный функционал

$(Au, v) = (u, Av)$ удовлетворяет всем аксиомам скалярного произведения, что позволяет использовать обозначения $(u, v)_A := (Au, v)$ и $\|u\|_A := \sqrt{(u, u)_A}$ для подчиненной этому скалярному произведению нормы.

Пусть некоторые n-мерные векторы $b^0, b^1, \ldots, b^{n-1}$ являются линейно независимыми. Используя эти векторы, по аналогии с процедурой ортогонализации Грамма-Шмидта (см. [5]), построим векторы $e^1, e^2, \ldots, e^n$ следующим образом:

$$
\begin{aligned}
e^1 &= \frac{s^1}{\|s^1\|_A}, \text{ где } s^1 = b^0; \\
e^2 &= \frac{s^2}{\|s^2\|_A}, \text{ где } s^2 = b^1 - \left(b^1, e^1\right)_A e^1; \\
e^k &= \frac{s^k}{\|s^k\|_A}, \text{ где } s^k = b^{k-1} - \sum_{m=1}^{k-1} \left(b^{k-1}, e^m\right)_A e^m; \ k = 3, 4, \ldots, n.
\end{aligned} \tag{1.20}
$$

Покажем по индукции, что система векторов $e^1, e^2, \ldots, e^n$ является *A-ортонормированной*, то есть $\left(e^i, e^j\right)_A = \delta_{ij}$, где δ_{ij} – символ Кронекера ($i, j = 1, 2, \ldots, n$), а, следовательно, система векторов $s^1, s^2, \ldots, s^n$ является

A-ортогональной, то есть $\left(s^i, s^j\right)_A = 0$ при $i \neq j$.

По построению $\left\|e^i\right\|_A = 1, (i = 1, 2, \ldots, n)$. Кроме того (базис индукции)

$$\left(e^2, e^1\right)_A = \frac{1}{\left\|s^2\right\|_A}\left(b^1 - \left(b^1, e^1\right)_A e^1, e^1\right)_A =$$

$$= \frac{1}{\left\|s^2\right\|_A}\left[\left(b^1, e^1\right)_A - \left(b^1, e^1\right)_A \underbrace{\left\|e^1\right\|_A^2}_{=1}\right] = 0.$$

Пусть система векторов $e^1, e^2, \ldots, e^k$ является A-ортонормированной (предположение индукции). Покажем, что тогда вектор e^{k+1} A-ортогонален каждому вектору этой системы (шаг индукции). Для произвольного $i \in \{1, 2, \ldots, k\}$

$$\left(e^{k+1}, e^i\right)_A = \frac{1}{\left\|s^{k+1}\right\|_A}\left(b^k - \sum_{m=1}^{k}\left(b^k, e^m\right)_A e^m, e^i\right)_A =$$

$$= \frac{1}{\left\|s^{k+1}\right\|_A}\left[\left(b^k, e^i\right)_A - \left(b^k, e^i\right)_A \underbrace{\left\|e^i\right\|_A^2}_{=1} - \sum_{\substack{m=1 \\ m\neq i}}^{k}\left(b^k, e^m\right)_A \underbrace{\left(e^m, e^i\right)_A}_{=0}\right] = 0.$$

Таким образом, доказано, что система векторов $e^1, e^2, \ldots, e^n$ является A-ортонормированной, откуда вытекает линейная независимость этих векторов.

Действительно, если их линейная комбинация $\alpha_1 e^1 + \alpha_2 e^2 + \cdots + \alpha_n e^n = 0$, то после скалярного умножения этого равенства на вектор Ae^i получим, что $\alpha_i = 0$ для любого $i = 1, 2, \ldots, n$.

Линейная независимость системы векторов $e^1, e^2, \ldots, e^n$ позволяет искать решение задачи (1.19) в виде $y = \sum\limits_{i=1}^{n} c_i e^i$, где c_i числовые коэффициенты. Подставляя указанное представление в уравнение (1.19) и умножая полученное равенство скалярно на векторы e^j $(j = 1, 2, \ldots, n)$, получим

$$\sum_{i=1}^{n} c_i \left(e^j, Ae^i\right) = \sum_{i=1}^{n} c_i \left(e^j, e^i\right)_A = \sum_{i=1}^{n} c_i \delta_{ij} = c_j = \left(e^j, f\right).$$

То есть, решением уравнения (1.19) является вектор $y = \sum\limits_{i=1}^{n} \left(e^i, f\right) e^i$.

Введём в рассмотрение вспомогательные векторы

$$y^k = \sum_{i=1}^{k} \left(e^i, f\right) e^i, \ k = 1, 2, \ldots, n,$$

которые также можно определить рекуррентным соотношением

$$y^k = y^{k-1} + \left(e^k, f\right) e^k, \ k = 1, 2, \ldots, n, \ y^0 = 0. \tag{1.21}$$

Тогда решение уравнения (1.19) $y = y^n$. Умножая соотношение (1.21) на матрицу A слева и вычитая из обеих частей полученного равенства вектор f, получим

$$Ay^k - f = Ay^{k-1} - f + \left(e^k, f\right) Ae^k.$$

Вектор $r^k = Ay^k - f$ представляет собой невязку приближенного решения y^k системы уравнений (1.19). При этом $r^0 = Ay^0 - f = -f$. Таким образом, для невязок r^k верно рекуррентное соотношение

$$r^k = r^{k-1} - \left(e^k, r^0\right) Ae^k,\ k = 1, 2, \ldots, n-1,\ r^n = 0,\ r^0 = -f. \qquad (1.22)$$

Предполагая линейную независимость системы векторов $r^0, r^1, \ldots, r^{n-1}$ (доказана далее), выберем эти векторы в качестве $b^0, b^1, \ldots, b^{n-1}$ в формулах (1.20). Тогда

$$s^k = r^{k-1} - \sum_{m=1}^{k-1} \left(r^{k-1}, e^m\right)_A e^m,\ k = 2, 3, \ldots, n,\ s^1 = r^0,\ e^k = \frac{s^k}{\left\|s^k\right\|_A}. \qquad (1.23)$$

Докажем, что векторы r^k и e^k, задаваемые рекуррентными соотношениями (1.22), (1.23), обладают следующими свойствами:

1. Верны равенства

$$\left(e^k, r^j\right) = \left(e^k, r^0\right),\ k = 1, 2, \ldots, n,\ j = 0, 1, \ldots, k-1 \tag{1.24}$$

и условия ортогональности

$$(e^k, r^k) = 0,\ k = 1, 2, \ldots, n; \tag{1.25}$$

2. Выполнены условия ортогональности

$$\left(r^k, r^j\right) = 0,\ k = 1, 2, \ldots, n-1,\ j = 0, 1, \ldots, k-1 \tag{1.26}$$

и

$$\left(r^k, e^j\right)_A = 0,\ k = 2, 3, \ldots, n-1,\ j = 1, 2, \ldots, k-1. \tag{1.27}$$

Замечание. 1. Выполнение (1.26) означает, что векторы $r^0,\ r^1,\ \ldots, r^{n-1}$ попарно ортогональны и, следовательно, линейно независимы. 2. Из (1.27) следует, что вектор r^k $(k = 1, 2, \ldots, n-1)$ A-ортогонален любой линейной комбинации векторов e^j $(j = 1, 2, \ldots, k-1)$ и соотношение (1.23) для векторов s^k $(k = 1, 2, \ldots, n)$ упрощается и принимает вид

$$s^k = r^{k-1} - \left(r^{k-1}, e^{k-1}\right)_A e^{k-1},\ k = 2, 3, \ldots, n,\ s^1 = r^0,\ e^k = \frac{s^k}{\left\|s^k\right\|_A}. \tag{1.28}$$

Установим справедливость первой группы свойств. При $j < k$

$$\left(e^k, r^j\right) = \left(e^k, r^{j-1} - \left(e^j, r^0\right) Ae^j\right) = \left(e^k, r^{j-1}\right) = \cdots = \left(e^k, r^0\right).$$

Учитывая данное равенство,

$$\left(e^k, r^k\right) = \left(e^k, r^{k-1} - \left(e^k, r^0\right) Ae^k\right) = \left(e^k, r^{k-1}\right) - \left(e^k, r^0\right) =$$
$$= \left(e^k, r^0\right) - \left(e^k, r^0\right) = 0.$$

Вторую группу свойств докажем методом математической индукции (используя соотношения (1.22), (1.23) и (1.25)). Проверим их выполнение для $k = 2$ (базис индукции):

$$\left(r^1, r^0\right) = \left(r^0 - \left(e^1, r^0\right) Ae^1, r^0\right) = \left(r^0, r^0\right) - \left(r^0, r^0\right) = 0;$$
$$\left(r^2, r^0\right) = \left(r^1 - \left(e^2, r^0\right) Ae^2, r^0\right) = -\left(e^2, r^0\right) \left(Ae^2, \left\|s^1\right\|_A e^1\right) = 0;$$
$$\left(r^2, r^1\right) = \left(r^2, \left\|s^2\right\|_A e^2 + \left(r^1, e^1\right)_A e^1\right) = \left(r^1, e^1\right)_A \left(r^2, e^1\right) =$$
$$= \left(r^1, e^1\right)_A \left(r^1 - \left(e^2, r^0\right) Ae^2, e^1\right) = 0;$$
$$\left(r^2, e^1\right)_A = \left(r^2, Ae^1\right) = \left(r^2, \frac{r^0 - r^1}{\left(e^1, r^0\right)}\right) = 0.$$

Знаменатель $\left(e^1, r^0\right) = \left(\dfrac{r^0}{\left\|r^0\right\|_A}, r^0\right) = \left\{r^0 = -f,\ f \neq 0\right\} \neq 0.$

Пусть вторая группа свойств справедлива для некоторого $k \geq 2$ (предположение индукции).

Достаточно доказать (шаг индукции), что

$$\left(r^{k+1}, r^j\right) = 0,\ j = 0, 1, \ldots, k;$$
$$\left(r^{k+1}, e^j\right)_A = 0,\ j = 1, 2, \ldots, k.$$

Покажем справедливость этих равенств:

$$\left(r^{k+1}, r^j\right) = \left(r^k - \left(e^{k+1}, r^0\right) Ae^{k+1}, r^j\right) = -\left(e^{k+1}, r^0\right)\left(Ae^{k+1}, r^j\right) =$$
$$= -\left(e^{k+1}, r^0\right)\left(Ae^{k+1}, \left\|s^{j+1}\right\|_A e^{j+1} + \left(r^j, e^j\right)_A e^j\right) = 0, \text{ для } j < k;$$
$$\left(r^{k+1}, r^k\right) = \left(r^k - \left(e^{k+1}, r^0\right) Ae^{k+1}, \left\|s^{k+1}\right\|_A e^{k+1} + \left(r^k, e^k\right)_A e^k\right) =$$
$$= \left\|s^{k+1}\right\|_A \left(r^k, e^{k+1}\right) - \left\|s^{k+1}\right\|_A \left(e^{k+1}, r^0\right) = 0, \text{ для } j = k;$$
$$\left(r^{k+1}, e^j\right)_A = \left(r^{k+1}, Ae^j\right) = \left(r^{k+1}, \frac{r^{j-1} - r^j}{\left(e^j, r^0\right)}\right) = 0.$$

Замечание. Если $\left(e^j, r^0\right) = 0$, то $r^j = r^{j-1}$. Тогда, условие ортогональнальности $\left(r^j, r^{j-1}\right) = 0$ выполнено только при $r^{j-1} = 0$. Это означает, что решение задачи (1.19) $y = y^{j-1}$ найдено на $(j-1)$-ом шаге применения рекуррентной формулы (1.21) (в совокупности с (1.22), (1.23)). В этом случае выполнения свойства (1.27) для $k + 1 > j$ не требуется.

Из соотношений (1.21), (1.22), (1.28) следует, что вектор y, являющийся решением системы линейных алгебраических уравнений (1.19), может быть вычислен по следующему алгоритму.

Пусть $y^0 = 0$. Тогда $r^0 = Ay^0 - f = -f$. Пусть $s^1 = r^0$.

Далее, последовательно, вычисляются

$$
\begin{aligned}
y^k &= y^{k-1} - \frac{\left(s^k,\ r^0\right)}{\left(s^k,\ As^k\right)} s^k, \\
r^k &= r^{k-1} - \frac{\left(s^k,\ r^0\right)}{\left(s^k,\ As^k\right)} As^k \ \text{ или } \ r^k = Ay^k - f \ \text{ для } \ k = 1, 2, \ldots, n \ , \\
s^{k+1} &= r^k - \frac{\left(r^k,\ As^k\right)}{\left(s^k,\ As^k\right)} s^k \ \text{ для } \ k = 1, 2, \ldots, (n-1).
\end{aligned}
\tag{1.29}
$$

Вектор $y^n = y$ является искомым решением системы (1.19).

Преобразуем формулы (1.29) к виду, который, как показала практика вычислений, менее чувствителен к ошибкам машинного округления чисел.

Из (1.24) и (1.28) следует, что при $k = 1, 2, \ldots, n$

$$\begin{aligned}\left(s^k,\ r^0\right) = \left(s^k,\ r^{k-1}\right) &= \left(r^{k-1} - \frac{\left(r^{k-1},\ As^{k-1}\right)}{\left(s^{k-1},\ As^{k-1}\right)} s^{k-1},\ r^{k-1}\right) = \\ &= \left(r^{k-1},\ r^{k-1}\right) - \frac{\left(r^{k-1},\ As^{k-1}\right)}{\left(s^{k-1},\ As^{k-1}\right)} \underbrace{\left(s^{k-1},\ r^{k-1}\right)}_{=0\ (1.25)} = \left(r^{k-1},\ r^{k-1}\right).\end{aligned} \tag{1.30}$$

Рассмотрим скалярное произведение $\left(r^k,\ r^k\right)$. В силу (1.22), (1.26) и (1.30), имеем

$$\begin{aligned}\left(r^k,\ r^k\right) = \left(r^{k-1} - \frac{\left(s^k,\ r^0\right)}{\left(s^k,\ As^k\right)} As^k,\ r^k\right) &= \underbrace{\left(r^{k-1},\ r^k\right)}_{=0\ (1.26)} - \\ - \frac{\left(s^k,\ r^0\right)}{\left(s^k,\ As^k\right)} \left(As^k,\ r^k\right) &= -\frac{\left(r^{k-1},\ r^{k-1}\right)}{\left(s^k,\ As^k\right)} \left(As^k,\ r^k\right),\end{aligned}$$

где $k = 1, 2, \ldots, n$.

Следовательно

$$\frac{\left(r^k,\ As^k\right)}{\left(s^k,\ As^k\right)} = -\frac{\left(r^k,\ r^k\right)}{\left(r^{k-1},\ r^{k-1}\right)} \tag{1.31}$$

при $k = 1, 2, \ldots, n$.

Введём в рассмотрение вектор p^k, связанный с вектором s^{k+1} соотношением (см. [11])

$$s^{k+1} = p^k \left(r^k,\ r^k\right) \tag{1.32}$$

при $k = 0, 1, \ldots, (n-1)$.

Вектора $p^0,\ p^1,\ \ldots, p^{n-1}$ являются A-ортогональными векторами.

В результате подстановки (1.30), (1.31) и (1.32) в (1.29) получим следующие рассчётные формулы метода сопряжённых градиентов:

$$\begin{aligned}
&y^0 = 0,\ \ r^0 = -f,\ \ p^0 = \frac{r^0}{(r^0,\ r^0)},\ \ \text{далее} \\
&y^k = y^{k-1} - \frac{p^{k-1}}{(p^{k-1},\ Ap^{k-1})}, \\
&r^k = r^{k-1} - \frac{Ap^{k-1}}{(p^{k-1},\ Ap^{k-1})} \ \ \text{или} \ \ r^k = Ay^k - f \\
&\qquad \text{для} \ \ k = 1, 2, \ldots, n \ \ , \\
&p^k = p^{k-1} + \frac{r^k}{(r^k,\ r^k)} \ \ \text{для} \ \ k = 1, 2, \ldots, (n-1).
\end{aligned} \tag{1.33}$$

При реализации алгоритма (1.33) требуется выполнить $O(n^3)$ операций умножения.

Минимизационные свойства метода сопряженных градиентов

Пусть вектор $y = (y_1, \ \dots \ , \ y_n)^T$ является искомым решением системы линейных алгебраических уравнений $Ay = f$ (1.19), где матрица $A = A^T > 0$, а $x = (x_1, \ \dots \ , \ x_n)^T$ - произвольный вектор. Вектор $z = x - y$ является ошибкой (погрешностью) определения вектора y. При наличии положительной определенности у матрицы A, удобной количественной мерой точности решения системы (1.19) является функция многих переменных $f\,(z) = (Az, \ z) \geq 0$, которую называют функцией ошибок.

Функция ошибок принимает положительные значения при любых $x \neq y$ и имеет минимальное значение равное нулю при $x = y$. Следовательно, решение системы (1.19) эквивалентно поиску вектора x, при котором функция ошибок принимает минимальное значение равное нулю.

Преобразуем выражение для функции ошибок к виду

$$
\begin{aligned}
&f\,(z) = (Az, \ z) = (Ax - Ay, \ x - y) = \{Ay = f\} = \\
&= (Ax - f, \ x - y) = (Ax, \ x) - (Ax, \ y) - (f, \ x) + (f, \ y) = \\
&= \{(Ax, \ y) = (x, \ Ay) = (x, \ f)\} = (Ax, \ x) - 2\,(f, \ x) + (f, \ y) = \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = J(x) + (f, \ y)\,,
\end{aligned}
$$

где функция $J(x) = (Ax, \ x) - 2\,(f, \ x)\,.$

Итак, решение системы линейных алгебраических уравнений (1.19) эквивалентно поиску вектора x, доставляющего минимум функции

$J(x) = (Ax,\ x) - 2\,(f,\ x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j - 2 \sum_{i=1}^{n} f_i x_i.$

Рассмотрим вектор $\operatorname{grad} J(x) = \left(\frac{\partial J}{\partial x_1},\ \frac{\partial J}{\partial x_2},\ \ldots,\ \frac{\partial J}{\partial x_n} \right)^T$, определяющий направление самого быстрого роста значений функции $J(x)$. Компонента $\frac{\partial J(x)}{\partial x_l}$ $(l = 1, 2, \ldots, n)$ этого вектора равна

$$\frac{\partial J(x_1,\ \ldots\ ,\ x_n)}{\partial x_l} = \frac{\partial}{\partial x_l} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j - 2 \sum_{i=1}^{n} f_i x_i \right) =$$

$$= \frac{\partial}{\partial x_l} \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{l-1} a_{ij} x_j + a_{il} x_l + \sum_{j=l+1}^{n} a_{ij} x_j - 2 f_i \right) =$$

$$= \frac{\partial}{\partial x_l} \left(\sum_{i=1}^{l-1} x_i \left(\sum_{j=1}^{l-1} a_{ij} x_j + a_{il} x_l + \sum_{j=l+1}^{n} a_{ij} x_j - 2 f_i \right) + \right.$$

$$\left. + x_l \left(\sum_{j=1}^{l-1} a_{lj} x_j + a_{ll} x_l + \sum_{j=l+1}^{n} a_{lj} x_j - 2 f_l \right) + \sum_{i=l+1}^{n} x_i \left(\sum_{j=1}^{l-1} a_{ij} x_j + \right. \right.$$

$$+a_{il}x_l + \sum_{j=l+1}^{n} a_{ij}x_j - 2f_i\Bigg)\Bigg) = \frac{\partial}{\partial x_l}\left(x_l \sum_{i=1}^{l-1} a_{il}x_i + x_l \sum_{j=1}^{l-1} a_{lj}x_j + \right.$$

$$\left. + x_l a_{ll} x_l + x_l \sum_{j=l+1}^{n} a_{lj}x_j - x_l 2f_l + x_l \sum_{i=l+1}^{n} x_i a_{il}\right) = \sum_{i=1}^{l-1} a_{il}x_i +$$

$$+ \sum_{j=1}^{l-1} a_{lj}x_j + 2a_{ll}x_l + \sum_{j=l+1}^{n} a_{lj}x_j - 2f_l + \sum_{i=l+1}^{n} a_{il}x_i = \sum_{i=1}^{n} a_{il}x_i +$$

$$+ \sum_{j=1}^{n} a_{lj}x_j - 2f_l = \left\{A = A^T\right\} = 2\left(\sum_{j=1}^{n} a_{lj}x_j - f_l\right) =$$

$$= 2\left((Ax)_l - f_l\right) = 2\left(Ax - f\right)_l = 2r_l,$$

где $r = Ax - f$ вектор невязки в системе линейных алгебраических уравнений (1.19).

То есть, вектор $\operatorname{grad} J(x) = 2r$. Следовательно, вектор $(-r)$ определяет направление самого быстрого убывания значений функции J в точке x.

Рассмотрим функцию J от аргумента $x - \tau r$, принадлежащего направлению антиградиента функции $J(x)$ при значениях числового параметра $\tau > 0$:

$$J(x-\tau r) = (A(x-\tau r),\ x-\tau r) - 2(f,\ x-\tau r) = (Ax,\ x) -$$
$$- 2\tau(Ax,\ r) + \tau^2(Ar,\ r) - 2(f,\ x) + 2\tau(f,\ r) = (Ax,\ x) - 2(f,\ x) -$$
$$- 2\tau(r,\ r) + \tau^2(Ar,\ r) = J(x) + (Ar,\ r)\left(\tau^2 - 2\tau\frac{(r,\ r)}{(Ar,\ r)}\right) =$$
$$= J(x) + (Ar,\ r)\left(\tau - \frac{(r,\ r)}{(Ar,\ r)}\right)^2 - \frac{(r,\ r)^2}{(Ar,\ r)}.$$

Пусть для параметра τ выполнено условие

$$\tau = \frac{(r,\ r)}{(Ar,\ r)}. \tag{1.34}$$

Тогда функция J в точке $x - \frac{(r,\ r)}{(Ar,\ r)}r$ будет иметь минимальное значение вдоль направления $(-r)$.

Выражение (1.34) совпадает с формулой для итерационного параметра τ_{k+1} в методе скорейшего спуска (см. п.1.5.2).

Рассмотрим частный случай $x = (x_1,\ x_2)^T$ $(n = 2)$. При $n = 2$ поверхность $J(x)$ будет иметь вид эллипсоида. Линия $J(x) = const$ представляет собой эллипс, который может быть сильно вытянут вдоль обной из полуосей. Если точка x, являющаяся аргументом функции $J(x)$, расположена близко к концу большей полуоси эллипса, то и следующее приближение, точка $x - \frac{(r,\ r)}{(Ar,\ r)}r$,

будет расположено близко к концу большей полуоси эллипса. То есть, выбор направления антиградиента функции $J(x)$ не является оптимальным для приближения к искомому решению y.

В методе сопряжённых градиентов очередное приближение

$y^k = y^{k-1} - \frac{p^{k-1}}{(p^{k-1},\ Ap^{k-1})}$ (1.33) к искомому решению расположено на направлении $\left(-p^{k-1}\right)$, не совпадающем с направлением антиградиента функции J.

Для значения функции $J(y^k)$ верна следующая оценка:

$$
\begin{aligned}
&J\left(y^k\right) = \left(Ay^k,\ y^k\right) - 2\left(f,\ y^k\right) = \left\{ y^k = y^{k-1} - \frac{p^{k-1}}{(p^{k-1},\ Ap^{k-1})}\ (1.33) \right\} = \\
&= \left(A\left(y^{k-1} - \frac{p^{k-1}}{(p^{k-1},\ Ap^{k-1})} \right),\ y^{k-1} - \frac{p^{k-1}}{(p^{k-1},\ Ap^{k-1})} \right) - \\
&- 2\left(f,\ y^{k-1} - \frac{p^{k-1}}{(p^{k-1},\ Ap^{k-1})} \right) = \left(Ay^{k-1},\ y^{k-1}\right) - \\
&- \frac{1}{(p^{k-1},\ Ap^{k-1})}\left(Ay^{k-1},\ p^{k-1}\right) - \frac{1}{(p^{k-1},\ Ap^{k-1})}\left(Ap^{k-1},\ y^{k-1}\right) + \\
&+ \frac{1}{(p^{k-1},\ Ap^{k-1})^2}\left(Ap^{k-1},\ p^{k-1}\right) - 2\left(f,\ y^{k-1}\right) + \frac{2}{(p^{k-1},\ Ap^{k-1})}\left(f,\ p^{k-1}\right) =
\end{aligned}
$$

$$= J\left(y^{k-1}\right) - \frac{2}{\left(p^{k-1},\ Ap^{k-1}\right)}\left(Ay^{k-1} - f,\ p^{k-1}\right) + \frac{1}{\left(p^{k-1},\ Ap^{k-1}\right)} =$$

$$= J\left(y^{k-1}\right) - \frac{2}{\left(p^{k-1},\ Ap^{k-1}\right)}\left(r^{k-1},\ p^{k-1}\right) + \frac{1}{\left(p^{k-1},\ Ap^{k-1}\right)} =$$

$$= J\left(y^{k-1}\right) - \frac{1}{\left(p^{k-1},\ Ap^{k-1}\right)}\left(2\left(r^{k-1},\ p^{k-1}\right) - 1\right) =$$

$$= \left\{\left(r^{k-1},\ p^{k-1}\right) = \left(p^{k-1},\ r^0\right)\ \ (1.24)\right\} = J\left(y^{k-1}\right) -$$

$$- \frac{1}{\left(p^{k-1},\ Ap^{k-1}\right)}\left(2\left(p^{k-1},\ r^0\right) - 1\right) = \left\{2\left(p^{k-1},\ r^0\right) - 1 =\right.$$

$$= 2\left(\underbrace{p^{k-2} + \frac{r^{k-1}}{\left(r^{k-1},\ r^{k-1}\right)}}_{(1.33)},\ r^0\right) - 1 = 2\left(p^{k-2},\ r^0\right) +$$

$$+ \frac{2}{\left(r^{k-1},\ r^{k-1}\right)}\underbrace{\left(r^{k-1},\ r^0\right)}_{=0\ \ (1.26)} - 1 = 2\left(p^{k-2},\ r^0\right) - 1 = \ldots = 2\left(p^0,\ r^0\right) - 1 =$$

$$\left. = 2\left(\frac{r^0}{\left(r^0,\ r^0\right)},\ r^0\right) - 1 = 1\right\} = J\left(y^{k-1}\right) - \frac{1}{\left(p^{k-1},\ Ap^{k-1}\right)} < J\left(y^{k-1}\right).$$

То есть, алгоритм метода сопряжённых градиентов гарантирует монотонное убывание значений функции J.

1.6.2 Метод Крейга

Метод Крейга применим для решения систем линейных алгебраических уравнений

$$Ay = f \tag{1.35}$$

с невырожденной матрицей A.

Будем искать вектор y, используя вспомогательный вектор x такой, что $y = A^T x$. Тогда исходная система (1.35) примет вид $AA^T x = f$. Матрица AA^T преобразованной системы является симметричной и положительно определенной. Поэтому, для решения преобразованной системы можно использовать метод сопряженных градиентов.

Алгоритм метода сопряжённых градиентов (1.33) для преобразованной системы имеет вид:

$$x^0 = 0, \;\; r^0 = -f, \;\; p^0 = \frac{r^0}{(r^0,\; r^0)}, \;\; \text{далее}$$
$$x^k = x^{k-1} - \frac{p^{k-1}}{(p^{k-1},\; AA^T p^{k-1})},$$
$$r^k = AA^T x^k - f \;\; \text{для} \;\; k = 1, 2, \ldots, n \;\; ,$$
$$p^k = p^{k-1} + \frac{r^k}{(r^k,\; r^k)} \;\; \text{для} \;\; k = 1, 2, \ldots, (n-1).$$

Умножая соотношение для очередного приближения x^k слева на матрицу A^T и учитывая, что $y = A^T x$, получим следующий алгоритм решения систе-

мы (1.35) :

$$
\begin{aligned}
& y^0 = 0, \;\; r^0 = -f, \;\; p^0 = \frac{r^0}{(r^0,\; r^0)}, \;\; \text{далее} \\
& y^k = y^{k-1} - \frac{A^T p^{k-1}}{(A^T p^{k-1},\; A^T p^{k-1})}, \\
& r^k = Ay^k - f \;\; \text{для} \;\; k = 1, 2, \ldots, n \;\; , \\
& p^k = p^{k-1} + \frac{r^k}{(r^k,\; r^k)} \;\; \text{для} \;\; k = 1, 2, \ldots, (n-1).
\end{aligned}
\tag{1.36}
$$

Алгоритм (1.36) можно использовать и для решения недоопределенных систем линейных алгебраических уравнений.

1.6.3 Симметризованные сопряжённые градиенты

Метод применим для решения систем линейных алгебраических уравнений $Ay = f$ с невырожденной матрицей A.

Умножим исходную систему $Ay = f$ слева на матрицу A^T. Для решения преобразованной системы $A^T Ay = A^T f$ используем метод сопряжённых градиентов. Вычислительный алгоритм решения преобразованной системы имеет

вид:

$$
\begin{aligned}
&y^0 = 0, \;\; R^0 = -A^T f, \;\; p^0 = \frac{R^0}{(R^0,\; R^0)}, \;\; \textit{далее} \\
&y^k = y^{k-1} - \frac{p^{k-1}}{(Ap^{k-1},\; Ap^{k-1})}, \\
&R^k = A^T \left(Ay^k - f \right) \;\; \textit{для} \;\; k = 1, 2, \ldots, n \;\; \textit{и} \\
&p^k = p^{k-1} + \frac{R^k}{(R^k,\; R^k)} \;\; \textit{для} \;\; k = 1, 2, \ldots, (n-1).
\end{aligned}
\tag{1.37}
$$

Алгоритм (1.37) можно использовать и для решения переопределенных систем линейных алгебраических уравнений.

Глава 2

Задачи на собственные значения

Пусть $A = (a_{ij})$ — матрица размера $n \times n$, а $y = (y_1, y_2, \ldots, y_n)^T$ — вектор неизвестных. Тогда поиск таких констант λ и векторов $y \neq 0$, что

$$Ay = \lambda y,$$

называется задачей на собственные значения. Эта задача эквивалентна поиску таких y, для которых

$$(A - \lambda E)y = 0.$$

При этом λ называют собственными значениями матрицы A, а соответствующие им вектора y — собственными векторами.

Известно, что если $\det(A - \lambda E) = 0$, то решение задачи существует. Этот определитель является полиномом степени n от λ с коэффициентами составленными из элементов матрицы A. Корни полинома легко найти при $n \leq 3$ или если матрица A является диагональной либо треугольной.

Задачу нахождения всех собственных значений матрицы A называют полной проблемой собственных значений. Если нужно найти лишь некоторые λ, то такую задачу называют частичной проблемой собственных значений.

2.1 Поиск собственных значений методом вращений

Метод вращений, предложенный К. Якоби в 1846 году, позволяет найти все собственные значения (решает полную проблему собственных значений) вещественной симметричной матрицы $A = A^T$.

Для матрицы $A = A^T$ справедливо представление вида

$$A = Q^T \Lambda Q, \tag{2.1}$$

где Q — ортогональная матрица ($Q^T = Q^{-1}$), а Λ — диагональная матрица, элементами которой являются собственные значения матрицы A, $\Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$.

Если умножить равенство (2.1) на $(Q^T)^{-1} = Q$ слева и на матрицу $Q^{-1} = Q^T$ справа, то получим

$$QAQ^T = \Lambda.$$

Итак, для нахождения собственных значений матрицы A необходимо построить ортогональную матрицу Q и провести два матричных умножения.

Заметим, что произведение нескольких ортогональных матриц является ортогональной матрицей. Матрицу Q будем строить итерационно, проводя с

помощью специальных ортогональных матриц преобразования матрицы A с целью уменьшения абсолютных значений ее недиагональных элементов.

Рассмотрим последовательность матриц

$$\begin{aligned} A_1 &= V_{ij}^1\, A\, (V_{ij}^1)^T, \\ A_2 &= V_{ij}^2\, A_1\, (V_{ij}^2)^T, \ \cdots\ , \\ A_k &= V_{ij}^k\, A_{k-1}\, (V_{ij}^k)^T, \cdots\ . \end{aligned}$$

Здесь V_{ij}^k — ортогональные матрицы следующего вида

$$V_{ij}^k = \begin{pmatrix} 1 & & & & 0 & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & \cos\varphi & & -\sin\varphi & & \\ & & & & 1 & & 0 & \\ 0 & & & & & \ddots & & 0 \\ & & & 0 & & & 1 & \\ & & & \sin\varphi & & \cos\varphi & & \\ & & & & & & & 1 \\ & & & & & & & & \ddots \\ & & & & 0 & & & & 1 \end{pmatrix}.$$

Элементы матрицы $V_{ij}^k = (v_{lm})$ задаются следующим образом. Диагональные элементы $v_{ll} = 1$ при $l \neq i\,, j$ и $v_{ll} = \cos\varphi$ при $l = i, j$. Вне диагонали

$v_{ij} = -\sin\varphi$, $v_{ji} = \sin\varphi$, а все остальные элементы равны нулю. Здесь φ — пока свободный параметр. Матрицы V_{ij}^k являются ортогональными матрицами, так как $V_{ij}^k \, (V_{ij}^k)^T = E$. Индекс k — номер итерации. Индексы i и j выбираются на каждой итерации k равными соответствующим индексам максимального по модулю элемента матрицы $A_{k-1} = (a_{lm}^{k-1})$, являющейся $(k-1)$-ым итерационным приближением к матрице Λ. Итак

$$|a_{ij}^{k-1}| = \max_{\substack{l,m \\ l \neq m}} |a_{lm}^{k-1}| \, .$$

Если максимальных по модулю элементов матрицы A_{k-1} несколько, то используется любой из них. Если все недиагональные элементы матрицы A_{k-1} равны нулю, то итерационный процесс построения матриц A_k прекращается.

В качестве количественной характеристики диагональности матрицы A_k выберем число

$$t(A_k) = \sum_{m=1}^{n} \sum_{\substack{s=1 \\ s \neq m}}^{n} (a_{ms}^k)^2 \, .$$

Если числовая последовательность $t(A_k) \overset{k \to \infty}{\longrightarrow} 0$, то последовательность матриц A_k сходится к диагональной матрице.

Установим соотношения, связывающие элементы матриц A_{k+1} и A_k. Итак

$$A_{k+1} = V_{ij}^{k+1}\, A_k\, (V_{ij}^{k+1})^T.$$

Введем вспомогательные обозначения $B = A_k\,(V_{ij}^{k+1})^T = (b_{ms})$ и $(V_{ij}^{k+1})^T = (\overline{v}_{lm})$. Тогда, по определению произведения матриц,

$$b_{ms} = \sum_{p=1}^{n} a_{mp}^k\, \overline{v}_{ps} = \begin{cases} a_{ms}^k, & s \neq i, j; \\ a_{mi}^k \cos\varphi - a_{mj}^k \sin\varphi, & s = i; \\ a_{mi}^k \sin\varphi + a_{mj}^k \cos\varphi, & s = j. \end{cases} \tag{2.2}$$

То есть, в матрицах B и A_k не совпадают элементы только в столбце с номером i и в столбце с номером j.

Для элементов матрицы $A_{k+1} = V_{ij}^{k+1} B$ верны соотношения

$$a_{ms}^{k+1} = \sum_{p=1}^{n} v_{mp}\, b_{ps} = \begin{cases} b_{ms}, & m \neq i, j; \\ b_{is} \cos\varphi - b_{js} \sin\varphi, & m = i; \\ b_{is} \sin\varphi + b_{js} \cos\varphi, & m = j. \end{cases} \tag{2.3}$$

В матрицах A_{k+1} и B не совпадают элементы только в строке с номером i и в строке с номером j.

Используя (2.3), получаем, что

$$a_{ij}^{k+1} = b_{ij}\cos\varphi - b_{jj}\sin\varphi.$$

Подставляя в это соотношение b_{ij} и b_{jj}, взятые из (2.2), и проводя ряд преобразований приходим к следующему выражению

$$\begin{aligned} a_{ij}^{k+1} &= (a_{ii}^k \sin\varphi + a_{ij}^k \cos\varphi)\cos\varphi - (a_{ji}^k \sin\varphi + a_{jj}^k \cos\varphi)\sin\varphi = \\ &= \{\mathsf{A} = \mathsf{A}^\mathsf{T} \Rightarrow \mathsf{A_k} = \mathsf{A_k^T}\} = (a_{ii}^k - a_{jj}^k)\sin\varphi\cos\varphi + a_{ij}^k(\cos^2\varphi - \sin^2\varphi) = \\ &= \frac{(a_{ii}^k - a_{jj}^k)\sin 2\varphi}{2} + a_{ij}^k \cos 2\varphi. \end{aligned}$$

Элемент a_{ij}^k является максимальным по модулю внедиагональным элементом матрицы A_k. Потребуем выполнения равенства $a_{ij}^{k+1} = 0$. Тогда предыдущее выражение превращается в уравнение относительно φ. Решая его, находим значение параметра φ, используемого для вычисления элементов матрицы V_{ij}^{k+1}

$$\varphi = \frac{1}{2}\operatorname{arctg}\frac{2a_{ij}^k}{a_{jj}^k - a_{ii}^k}.$$

Вычислим количественную характеристику диагональности матрицы A_{k+1}

$$t(A_{k+1}) = \sum_{m=1}^{n} \sum_{\substack{s=1 \\ s \neq m}}^{n} (a_{ms}^{k+1})^2.$$

Согласно формулам (2.2) и (2.3), элементы матрицы A_{k+1} отличаются от элементов матрицы A_k только в i-х и j-х строках и столбцах.

Выделим в $t(A_{k+1})$ совпадающие элементы матриц A_{k+1} и A_k в отдельную сумму и проведем следующие преобразования

$$t(A_{k+1}) = \sum_{\substack{m=1\\m\neq i,j}}^{n}\sum_{\substack{s=1\\s\neq i,j,m}}^{n}(a_{ms}^k)^2 + \sum_{\substack{m=1\\m\neq i,j}}^{n}\left[b_{mi}^2 + b_{mj}^2\right] +$$
$$+ \sum_{\substack{s=1\\s\neq i,j}}^{n}\left[(a_{is}^{k+1})^2 + (a_{js}^{k+1})^2\right] = \sum_{\substack{m=1\\m\neq i,j}}^{n}\sum_{\substack{s=1\\s\neq i,j,m}}^{n}(a_{ms}^k)^2 +$$
$$+ \sum_{\substack{m=1\\m\neq i,j}}^{n}\left[(a_{mi}^k)^2\cos^2\varphi + (a_{mj}^k)^2\sin^2\varphi - 2a_{mi}^k a_{mj}^k\sin\varphi\cos\varphi +\right.$$
$$\left.+(a_{mi}^k)^2\sin^2\varphi + (a_{mj}^k)^2\cos^2\varphi + 2a_{mi}^k a_{mj}^k\sin\varphi\cos\varphi\right] +$$
$$+ \sum_{\substack{s=1\\s\neq i,j}}^{n}\left[b_{is}^2\cos^2\varphi + b_{js}^2\sin^2\varphi - 2b_{is}b_{js}\sin\varphi\cos\varphi +\right.$$
$$\left.+b_{is}^2\sin^2\varphi + b_{js}^2\cos^2\varphi + 2b_{is}b_{js}\sin\varphi\cos\varphi\right] =$$
$$= \sum_{\substack{m=1\\m\neq i,j}}^{n}\sum_{\substack{s=1\\s\neq i,j,m}}^{n}(a_{ms}^k)^2 + \sum_{\substack{m=1\\m\neq i,j}}^{n}\left[(a_{mi}^k)^2 + (a_{mj}^k)^2\right] + \sum_{\substack{s=1\\s\neq i,j}}^{n}\left[(a_{is}^k)^2 + (a_{js}^k)^2\right] +$$
$$+2(a_{ij}^k)^2 - 2(a_{ij}^k)^2 = t(A_k) - 2(a_{ij}^k)^2.$$

То есть $t(A_{k+1}) < t(A_k)$. Уменьшение количественной характеристики диагональности происходит монотонно с ростом номера итерации k на величину равную $2(a_{ij}^k)^2$ — удвоенный квадрат максимального внедиагонального элемента матрицы A_k. Следовательно, последовательность матриц A_k сходится к диагональной матрице.

Получим оценку на количественную характеристику диагональности матрицы A_k. Так как a_{ij}^k — максимальный по модулю внедиагональный элемент, то верно неравенство

$$t(A_k) \leqslant n(n-1)(a_{ij}^k)^2.$$

Отсюда следует, что $(a_{ij}^k)^2 \geqslant \dfrac{t(A_k)}{n(n-1)}$ для $n \geqslant 2$. Подставляя это неравенство в соотношение, связывающее $t(A_k)$ и $t(A_{k+1})$, имеем

$$t(A_{k+1}) = t(A_k) - 2(a_{ij}^k)^2 \leqslant t(A_k) - \frac{2}{n(n-1)} t(A_k) = \rho t(A_k),$$

где $\rho = 1 - \dfrac{2}{n(n-1)} < 1.$

Применив эту оценку k раз, получим

$$t(A_k) \leqslant \rho^k t(A).$$

Итак, последовательность матриц A_k сходится к диагональной матрице Λ со скоростью геометрической прогрессии со знаменателем ρ.

Замечание. Реализация правила выбора индексов i и j в матрице V_{ij}^k требует сравнения $n^2/2$ чисел. Возможна некоторая оптимизация вычислительных затрат. Например, сначала выбирается строка i с максимальной суммой квадратов значений недиагональных элементов. Затем в этой строке выбирается максимальный по модулю элемент a_{ij}^k.

2.2 Степенной метод поиска собственных значений

Рассмотрим задачу поиска максимального по модулю собственного значения симметричной матрицы $A = A^T$ ($\lambda(A)$ — вещественные числа).

Пусть все собственные числа $\lambda(A)$ различны и пронумерованы так, что $|\lambda_1| > |\lambda_2| > |\lambda_3| > \ldots > |\lambda_n|$.

Примечание. Так как нет совпадающих собственных значений, то все собственные вектора $\xi_i, i = 1, \ldots, n$ матрицы A ортогональны и образуют базис, который будем считать ортонормированным - $(\xi_i, \xi_j) = 1$ при $i = j$ и $(\xi_i, \xi_j) = 0$ если $i \neq j$.

Выберем произвольный вектор y^0, отличный от нуля, и построим последовательность векторов y^k

$$y^{k+1} = Ay^k, \quad k = 0, 1, \ldots. \tag{2.4}$$

Используем вектора y^k для вычисления элементов числовой последовательности $\{\Lambda_1^k\}$

$$\Lambda_1^k = \frac{(y^{k+1}, y^k)}{(y^k, y^k)}.$$

Покажем, что последовательность $\{\Lambda_1^k\}$ сходиться к λ_1.

Представим начальное приближение y^0 в виде разложения по базису из собственных векторов матрицы A. То есть $y^0 = \sum_{i=1}^{n} \alpha_i \xi_i$.

Из (2.4) следует, что $y^k = Ay^{k-1} = \ldots = A^k y^0$.

Тогда верно следующее

$$y^k = A^k \sum_{i=1}^{n} \alpha_i \xi_i = A^{k-1} \sum_{i=1}^{n} \alpha_i A\xi_i = A^{k-1} \sum_{i=1}^{n} \alpha_i \lambda_i \xi_i = \ldots = \sum_{i=1}^{n} \alpha_i \lambda_i^k \xi_i.$$

Вычислим два скалярных произведения :

$$(y^k, y^k) = (\sum_{i=1}^{n} \alpha_i \lambda_i^k \xi_i, \sum_{j=1}^{n} \alpha_j \lambda_j^k \xi_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \lambda_i^k \lambda_j^k (\xi_i, \xi_j) =$$

$$= \sum_{i=1}^{n} \alpha_i^2 \lambda_i^{2k} = \alpha_1^2 \lambda_1^{2k} + \sum_{i=2}^{n} \alpha_i^2 \lambda_i^{2k};$$

$$(y^{k+1}, y^k) = (\sum_{i=1}^{n} \alpha_i \lambda_i^{k+1} \xi_i, \sum_{j=1}^{n} \alpha_j \lambda_j^k \xi_j) = \ldots =$$

$$= \sum_{i=1}^{n} \alpha_i^2 \lambda_i^{2k+1} = \alpha_1^2 \lambda_1^{2k+1} + \sum_{i=2}^{n} \alpha_i^2 \lambda_i^{2k+1}.$$

Тогда, для элементов последовательности Λ_1^k, получаем

$$\Lambda_1^k = \frac{(y^{k+1}, y^k)}{(y^k, y^k)} = \frac{\alpha_1^2 \lambda_1^{2k+1} + \sum_{i=2}^{n} \alpha_i^2 \lambda_i^{2k+1}}{\alpha_1^2 \lambda_1^{2k} + \sum_{i=2}^{n} \alpha_i^2 \lambda_i^{2k}} =$$

$$= \frac{\alpha_1^2\lambda_1^{2k+1}\left(1+\sum\limits_{i=2}^{n}\left(\frac{\alpha_i}{\alpha_1}\right)^2\left(\frac{\lambda_i}{\lambda_1}\right)^{2k+1}\right)}{\alpha_1^2\lambda_1^{2k}\left(1+\sum\limits_{i=2}^{n}\left(\frac{\alpha_i}{\alpha_1}\right)^2\left(\frac{\lambda_i}{\lambda_1}\right)^{2k}\right)} =$$

$$= \lambda_1\frac{1+O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2k+1}\right)}{1+O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\right)} = \lambda_1\left(1+O\left(\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\right)\right) \longrightarrow \lambda_1$$

при $k \to \infty$.

Последовательность Λ_1^k сходится к λ_1 - искомому максимальному по модулю собственному значению матрицы A.

Рассмотрим последовательность векторов $\frac{y^k}{||y^k||}$. Верны следующие преобразования

$$\frac{y^k}{||y^k||} = \frac{\alpha_1\lambda_1^k\xi_1 + \sum\limits_{i=2}^{n}\alpha_i\lambda_i^k\xi_i}{\sqrt{\alpha_1^2\lambda_1^{2k} + \sum\limits_{i=2}^{n}\alpha_i^2\lambda_i^{2k}}} = \frac{\alpha_1\lambda_1^k\xi_1 + \sum\limits_{i=2}^{n}\alpha_i\lambda_i^k\xi_i}{|\alpha_1||\lambda_1|^k\left(1 + O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2k}\right)\right)} =$$

$$= \pm\xi_1 + \sum_{i=2}^{n}\frac{\alpha_i}{|\alpha_1|}O\left(\left(\frac{\lambda_i}{\lambda_1}\right)^k\right)\xi_i.$$

То есть, вектор $\frac{y^k}{||y^k||}$ с ростом итерационного индекса k приближается к направлению собственного вектора ξ_1 .