Заметим, что матрицы типа тёплицевых возникают не только в связи со стационарными случайными процессами, но также и в связи с процессами, в определенном смысле близкими к стационарным (см. / 98 /). Роль быстрых алгоритмов особенно велика при конструировании приборов, выполняющих статистическую обработку, и, конечно, при моделировании.

§ 3. Итерационные методы и способы предобусловливания

3.I. Общие замечания

Пусть требуется решить систему линейных алгебраических уравнений

$$A_n \varphi_n = f_n, \tag{3.I.1}$$

где A_n - невырожденная плотная матрица порядка n. Предположим, что использование прямых методов нежелательно или даже невозможно из-за чрезмерно больших запросов к машинным ресурсам. В этой ситуации естественно обратиться к итерационным методам. Обычно быстрая сходимость бывает связана с выделением хорошей главной части для A_n, т.е. с выбором расщепления

$$A_n = T_n - K_n, \tag{3.I.2}$$

где T_n - невырожденная "легко обратимая" матрица, рассматриваемая как главная часть для A_n.

Пусть имеется одномерное сингулярное интегральное уравнение

$$\int_0^T \frac{1}{t_0 - t}\,\varphi(t)\,dt - \int_0^T k(t_0,t)\,\varphi(t)\,dt = f(t_0). \tag{3.I.3}$$

Тогда при формировании соответствующей ему алгебраической задачи естественно T_n и K_n связать с главной и регулярной частями для (3.I.3). Если используется метод дискретных особенностей, то матрица T_n будет тёплицевой. Вообще говоря, тёплицевость связана с использованием равномерных сеток. Однако заметим, что расщепление (3.I.3) может записываться для различных способов параметризации. Очевидно, для интегральных уравнений с логарифмической или степенной

особенностью возможны аналогичные построения.

Таким образом, целесообразно рассмотреть класс расщеплений (3.I.2) с тёплицевой матрицей T_n. В случае двумерных уравнений в качестве T_n будет выбираться дважды тёплицева матрица.

На базе расщепления (3.I.2) можно строить различные итерационные процессы, в том числе с использованием известных приемов ускорения за счет выбора итерационных параметров. Не касаясь последнего, рассмотрим здесь два основных метода. Первый - метод простой итерации:

$$T x_k = K x_{k-1} + f, \quad k = 1, 2, \dots, \tag{3.I.4}$$

или

$$x_k = x_{k-1} + T^{-1}(f - A x_{k-1}), \quad k = 1, 2, \dots \tag{3.I.5}$$

Второй - метод сопряженных градиентов с предобусловливанием (метод неполного эрмитова разложения):

$$\begin{aligned}
&z_{k-1} = T^{-1} r_{k-1},\\
&\beta_k = z^T_{k-1} r_{k-1} / z^T_{k-2} r_{k-2} \quad (\text{при } k \geq 2),\\
&d_k = z_{k-1} + \beta_k d_{k-1} \quad (\text{при } k \geq 2), \quad d_1 = z_0,\\
&\alpha_k = z^T_{k-1} r_{k-1} / z^T_{k-1} A d_k,\\
&x_k = x_{k-1} + \alpha_k d_k,\\
&r_k = r_{k-1} - \alpha_k A d_k.
\end{aligned} \tag{3.I.6}$$

Здесь $r_0 = f - A x_0$, где x_0 - начальное приближение.

Метод (3.I.5) сходится при любом начальном приближении x_0 тогда и только тогда, когда все собственные значения матрицы $I - T^{-1}A$ лежат внутри единичного круга. При выборе T из некоторого множества матриц $\mathcal{T}$ полезно рассмотреть следующую задачу оптимизации:

$$\| I - T^{-1} A \| \to \min, \quad T \in \mathcal{T}, \tag{3.I.7}$$

или, в упрощенной форме, такую задачу:

$$\|T-A\| \longrightarrow \min, \quad T\in\mathcal{T}. \qquad (3.I.8)$$

Если в (3.I.8) используется евклидова норма и $\mathcal{T}$ - множество тёплицевых матриц порядка n, то решением задачи (3.I.8) является матрица $T\equiv[t_{i-j}]$, где

$$t_{i-j}=\frac{\sum\limits_{p-q=i-j} a_{pq}}{n-|i-j|}. \qquad (3.I.9)$$

Этот факт проверяется непосредственно.

Метод (3.I.6) используется в случае симметричных A и T, причем T предполагается положительно определенной (в противном случае требуется некоторая модификация процесса). Если исходная система (3.I.4) несимметрична, то следует каким-либо способом перейти к симметричной системе, т.е. выполнить симметризацию.

В условиях точной арифметики метод (3.I.6) за конечное число шагов (не более n) должен дать точное решение (на практике же он ведет себя как итерационный метод). Обозначим через e_k отклонение k-го приближения от точного решения. Тогда известно, что в методе сопряженных градиентов на k-м шаге отклонение e_k минимизируется по всем векторам из подпространства Крылова, натянутого на векторы

$$T^{-1}f, \quad (T^{-1}A)T^{-1}f, \quad (T^{-1}A)^2T^{-1}f, \ldots \qquad (3.I.10)$$

Вследствие этого

$$\|e_k\|_2 \le \min_{P_k}\ \max_{\lambda}|P_k(\lambda)|\,\|e_0\|_2, \qquad (3.I.11)$$

где минимум берется по всем многочленам P_k степени не выше k со свободным членом I. Таким образом, проблема оценки скорости сходимости сводится к оценке минимума в (3.I.11).

Типичная ситуация - когда собственные значения

$$|\lambda_1^{(n)}|\le|\lambda_2^{(n)}|\le\ldots\le|\lambda_n^{(n)}|$$

матриц $T_n^{-1}K_n$ обладают следующим свойством: для любого $\varepsilon>0$ существует номер N такой, что

$$|\lambda_j^{(n)}|\le\varepsilon \quad \text{при} \quad j\le n-N.$$

В этом случае положим (см. / 77 /)

$$P_k(x)=\prod_{j=1}^{N}\left(1-\frac{x}{1-\lambda_{n+1-j}^{(n)}}\right)Q_{n-N}(x),$$

где $Q_{n-N}(x)$ - многочлен Чебышева первого рода, построенный для отрезка $[1-\varepsilon,\ 1+\varepsilon]$. Для достаточно малых ε имеем

$$\max_{|x|\le\varepsilon}\left|1-\frac{x}{1-\lambda_{n+1-j}^{(n)}}\right|\le K, \quad j=1,\ldots,N,$$

где константа K не зависит от ε. Кроме того,

$$\max_{1-\varepsilon\le x\le 1+\varepsilon}\left|Q_{n-N}(x)\right|\le c\,\varepsilon^{k-N}$$

Таким образом,

$$\|e_k\|\le c\,\varepsilon^{k-N}K^N\|e_0\|. \qquad (3.I.12)$$

Обратим внимание на то, что на каждой итерации основные вычислительные затраты определяют следующие два действия:

- решение системы с матрицей T ;
- умножение на вектор матрицы A .

В случае тёплицевой матрицы T первое действие (за исключением некоторого подготовительного этапа) требует $O(n\log_2 n)$ операций. Второе действие в общем случае занимает $O(n^2)$ операций. Оценка (3.I.12), следовательно, показывает, что при больших n вычислительные затраты на решение системы (3.I.I) не превышают $c_\delta n^2$, где c_δ определяется требуемой точностью δ. Учитывая специфику матрицы A, иногда удается предложить более быструю реализацию умножения ее на вектор; это влечет снижение общих вычислительных затрат.

3.2. Главные части для матриц метода дискретных особенностей

Пусть матрица A_n получена в результате применения метода дискретных особенностей при решении сингулярного интегрального уравнения (3.I.3). Тогда, как уже отмечалось, в качестве главной части для нее естественным образом выделяется тёплицева матрица

$$T_n = \left[\frac{1}{i-j+\frac{1}{2}}\right]_{ij=1}^{n} . \qquad (3.2.1)$$

Поэтому целесообразно изучить сначала именно её свойства.

Матрица T_n является одновременно тёплицевой и матрицей Коши (см. § I.I). Чтобы использовать последнее, нам кажется полезным вначале установить некоторые общие свойства матриц Коши.

Предположим, что $A_n = \left[\frac{1}{x_i - y_j}\right]$ - произвольная матрица Коши; $x_i \neq y_j$ для всех i, j. Вычтем последнюю строку в A_n из всех предыдущих, а затем то же проделаем со столбцами; после этого ведущая подматрица порядка $n-1$ с точностью до умножения слева и справа на диагональные матрицы оказывается равной A_{n-1}. Для определителей получаем рекуррентное соотношение

$$\det A_n = \det A_{n-1} \frac{1}{x_n - y_n} \prod_{i=1}^{n-1} \frac{(x_i - x_n)(y_n - y_i)}{(x_i - y_n)(x_n - y_i)};$$

следовательно,

$$\det A_n = \frac{\prod\limits_{1\le i<j\le n} (x_i - x_j) \prod\limits_{1\le i<j\le n} (y_j - y_i)}{\prod\limits_{1\le i,j\le n} (x_i - y_j)} . \qquad (3.2.2)$$

Таким образом, если $x_i \neq x_j$, $y_i \neq y_j$, то матрица A_n невырожденная.

Рассмотрим линейную алгебраическую систему

$$A_n z = b,$$

$$z = [z_1 \dots z_n]^T, \quad b = [b_1 \dots b_n]^T .$$

Учитывая и необходимым образом модифицируя (3.2.2), согласно правилу Крамера находим

$$z_l = \frac{(-1)^l}{\det A_n} \sum_{k=1}^{n} b_k (-1)^k \frac{\prod\limits_{\substack{1\le i<j\le n\\ i,j\neq k}} (x_i - x_j) \prod\limits_{\substack{1\le i<j\le n\\ i,j\neq l}} (y_j - y_i)}{\prod\limits_{\substack{1\le i,j\le n\\ i\neq k,\ j\neq l}} (x_i - y_j)} ,$$

или, после подстановки выражения для $\det A_n$ и некоторых преобразований,

$$z_l = u_l \sum_{k=1}^{n} v_k \frac{b_k}{x_k - y_l} , \quad l = 1, \dots, n, \qquad (3.2.3)$$

где

$$u_l = \prod_{i=1}^{n} (x_i - y_l) \Big/ \prod_{\substack{i=1\\ i\neq l}}^{n} (y_i - y_l), \quad l = 1, \dots, n, \qquad (3.2.4)$$

$$v_k = \prod_{j=1}^{n} (x_k - y_j) \Big/ \prod_{\substack{j=1\\ j\neq k}}^{n} (x_k - x_j), \quad k = 1, \dots, n. \qquad (3.2.5)$$

Полученным соотношениям полезно дать следующую матричную трактовку.

Лемма 3.2.I. Пусть A_n - невырожденная матрица Коши порядка n. Тогда для обратной матрицы имеет место представление

$$A_n^{-1} = U_n A_n^T V_n, \tag{3.2.6}$$

где

$$U_n = diag(u_1, \dots, u_n), \quad V_n = diag(v_1, \dots, v_n), \tag{3.2.7}$$

а компоненты $u_1, \dots, u_n$ и $v_1, \dots, v_n$ составляют решения следующих линейных алгебраических систем:

$$A_n [u_1 \dots u_n]^T = [1 \dots 1]^T, \tag{3.2.8}$$

$$[v_1 \dots v_n] A_n = [1 \dots 1]. \tag{3.2.9}$$

При этом справедливы формулы (3.2.4), (3.2.5).

Доказательство. Формула (3.2.6) есть не что иное, как матричная запись для (3.2.3). Пусть $v_1, \dots, v_n$ определяются согласно (3.2.5). Тогда (3.2.9) равносильно следующим соотношениям:

$$\sum_{k=1}^{n} \frac{f_l(x_k)}{\prod\limits_{\substack{j=1 \\ j\neq k}}^{n} (x_k - x_j)} = 1, \quad l = 1, \dots, n, \tag{3.2.10}$$

где

$$f_l(x) = \prod_{\substack{j=1 \\ j\neq l}}^{n} (x - y_j), \quad l = 1, \dots, n.$$

Левая часть (3.2.10) представляет собой разделенную разность $f_l(x_1; \dots; x_n)$ порядка $n-1$ для функции $f_l(x)$ (см. / 6 /). Поскольку $f_l(x)$ есть нормированный многочлен степени $n-1$, эта разделенная разность равна 1, т.е. соотношение (3.2.9) установлено. Далее, компоненты $z_1, \dots, z_n$ решения системы (3.2.8) имеют вид (3.2.3), где $b_k = 1$ для всех k. В силу (3.2.9) находим $z_l = u_l$

для всех l, где u_l определяется согласно (3.I.4). Лемма доказана.

При $x_i = i + \frac{1}{2}$, $y_j = j$ матрица A_n превращается в интересующую нас матрицу T_n вида (3.2.I). В этом случае

$$\prod_{i=1}^{n} (x_i - y_l) = \prod_{i=1}^{n} \left(i + \frac{1}{2} - l\right) = \frac{(-1)^{l-1}}{2} \prod_{i=1}^{l-1} \left(i - \frac{1}{2}\right) \prod_{i=1}^{n-l} \left(i + \frac{1}{2}\right),$$

$$\prod_{i=1}^{n} (x_k - y_i) = \prod_{i=1}^{n} \left(k + \frac{1}{2} - i\right) = \frac{(-1)^{n-k}}{2} \prod_{i=1}^{k-1} \left(i + \frac{1}{2}\right) \prod_{i=1}^{n-k} \left(i - \frac{1}{2}\right),$$

$$\prod_{\substack{i=1 \\ i\neq l}}^{n} (y_i - y_l) = (-1)^{l-1} (l-1)! \, (n-l)!,$$

$$\prod_{\substack{i=1 \\ i\neq k}}^{n} (x_k - x_i) = (-1)^{n-k} (k-1)! \, (n-k)!$$

Следовательно, согласно (3.2.4), (3.2.5) находим

$$u_l = \frac{1}{2} \prod_{i=1}^{l-1} \left(1 - \frac{1/2}{i}\right) \prod_{i=1}^{n-l} \left(1 + \frac{1/2}{i}\right), \quad l = 1, \dots, n, \tag{3.2.11}$$

$$v_k = \frac{1}{2} \prod_{i=1}^{k-1} \left(1 + \frac{1/2}{i}\right) \prod_{i=1}^{n-k} \left(1 - \frac{1/2}{i}\right), \quad k = 1, \dots, n. \tag{3.2.12}$$

Отсюда, в частности, вытекает, что компоненты диагональных матриц U_n, V_n в случае матрицы T_n могут быть вычислены с затратой лишь $O(n)$ арифметических операций. Вследствие (3.2.6) и тёплицевости T_n^T решение произвольной линейной алгебраической системы с матрицей T_n может быть найдено ценой $O(n\log_2 n)$ арифметических операций.

Лемма 3.2.2. Для любого n

$$\pi - \alpha_n \le \|T_n\|_2 \le \pi, \qquad (3.2.13)$$

где $\alpha_n > 0$, $\alpha_n = O\left(\frac{1}{n}\right)$. Более того, для любого фиксированного $\varepsilon > 0$ количество сингулярных чисел матрицы T_n вне отрезка $[\pi-\varepsilon, \pi]$ есть $o(n)$.

Доказательство. Обозначим через $\rho_{1n} \ge \ldots \ge \rho_{nn} > 0$ сингулярные числа матрицы T_n. Тогда собственные значения симметричной матрицы

$$\tilde{T}_n = \begin{bmatrix} 0 & T_n^T \\ T_n & 0 \end{bmatrix}$$

суть $\pm\rho_{1n}, \pm\rho_{2n}, \ldots, \pm\rho_{nn}$. Согласно лемме о перестановках уровней (см. § I.I) матрица $\tilde{T}_n$ перестановочно подобна блочно-тёплицевой матрице $\hat{T}_n$ с блоками порядка 2. Матрица перестановок отвечает такой последовательности индексов: I, n + I, 2, n + 2, ..., n, $2n$. Матрица $\hat{T}_n$ блочно-тёплицева, но не тёплицева:

$$\hat{T}_n = \begin{bmatrix} 0 & 2 & 0 & 2/3 & 0 & 2/5 & \ldots \\ 2 & 0 & -2 & 0 & -2/3 & 0 & \ldots \\ 0 & -2 & 0 & 2 & 0 & 2/3 & \ldots \\ 2/3 & 0 & 2 & 0 & -2 & 0 & \ldots \\ 0 & -2/3 & 0 & -2 & 0 & 2 & \ldots \\ 2/5 & 0 & 2/3 & 0 & 2 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix}.$$

После умножения всех строк с нечетными номерами на мнимую единицу i, а строк с четными номерами на $-i$ получаем, как замечено в / I05 /, тёплицеву эрмитову матрицу $A = [a_{k-l}]_{kl=0}^{2n-1}$, где

$$a_m = \begin{cases} 0, & m - \text{четно}, \\ i\,\frac{2}{m}, & m - \text{нечетно}. \end{cases}$$

Непосредственно проверяется, что

$$a_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-imx} f(x)\,dx,$$

где

$$f(x) = \begin{cases} -\pi, & -\pi \le x < 0, \\ \pi, & 0 \le x \le \pi. \end{cases}$$

Далее, при $w = [w_0 \ldots w_{2n-1}]^T$ имеем

$$(Aw, w) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |w_0 + w_1 e^{ix} + \ldots + w_{2n-1} e^{i(2n-1)x}|^2 f(x)\,dx,$$

$$(w, w) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |w_0 + w_1 e^{ix} + \ldots + w_{2n-1} e^{i(2n-1)x}|^2\,dx.$$

Учитывая вид функции $f(x)$, находим

$$\|T_n\|_2 = \|A\|_2 = \max_{w \ne 0}\left|\frac{(Aw, w)}{(w, w)}\right| \le \pi.$$

Для того чтобы оценить $\|T_n\|_2$ снизу, воспользуемся очевидным неравенством

$$\|T_n e_j\|_2 \le \|T_n\|_2,$$

где e_j есть j-й столбец единичной матрицы. Возьмем $j = \left[\frac{n}{2}\right]$ и запишем

$$\|T_n e_j\|_2^2 = \sum_{k=1}^{n} \frac{1}{(k-j+\frac{1}{2})^2} =$$

$$= 4\left(\sum_{k=0}^{j-1} \frac{1}{(2k+1)^2} + \sum_{k=0}^{n-j} \frac{1}{(2k+1)^2}\right).$$

Поскольку

$$\sum_{k=0}^{\infty}\frac{1}{(2k+1)^2}=\frac{\pi^2}{8},$$

получаем, что $\|T_n e_{[n/2]}\|_2^2 \to \pi^2$ при $n\to\infty$. Положим

$$d_k = 4\left(\sum_{k=[\frac{n}{2}]-1}^{\infty}\frac{1}{(2k+1)^2}+\sum_{k=n-[\frac{n}{2}]+1}^{\infty}\frac{1}{(2k+1)^2}\right);$$

тогда $d_n = O(\frac{1}{n})$, и при $d_n\le\pi$ находим

$$\pi - d_n \le \sqrt{\pi^2-d_n}\le\|T_n\|_2.$$

Теперь фиксируем произвольное $\delta>0$ и обозначим через m_δ такой номер, что при $j\ge m_\delta$

$$\sum_{k=0}^{j}\frac{1}{(2k+1)^2}\ge\frac{\pi^2}{8}-\delta.$$

Тогда при $m_\delta+2\le j\le n-m_\delta$ имеем

$$\|T_n e_j\|_2^2\ge\pi^2-8\delta.$$

Следовательно,

$$\frac{(\pi^2-8\delta)(n-2m_\delta-1)}{n}\le\frac{\sum_{j=1}^{n}\|T_n e_j\|_2^2}{n}\le\pi^2.$$

Выражение в левой части при $n\to\infty$ стремится к $\pi^2-8\delta$. Принимая во внимание равенство

$$\sum_{j=1}^{n} \rho_{jn}^2=\sum_{j=1}^{n}\|T_n e_j\|_2^2\quad(=\mathrm{tr}(T_n^T T_n)),$$

в силу произвольности δ находим

$$\lim_{n\to\infty}\frac{\rho_{1n}^2+\dots+\rho_{nn}^2}{n}=\pi^2. \qquad (3.2.14)$$

Пусть $0<\varepsilon<\pi$ и γ_n обозначает количество сингулярных чисел ρ_{jn}, удовлетворяющих неравенству $\rho_{jn}<\pi-\varepsilon$. Тогда

$$\frac{\rho_{1n}^2+\dots+\rho_{nn}^2}{n}<\frac{\gamma_n(\pi-\varepsilon)^2}{n}+\frac{n-\gamma_n}{n}\pi^2$$

и вследствие (3.2.14) $\gamma_n=o(n)$. Лемма доказана.

Замечание. То, что сингулярные числа матрицы T_n почти все сосредоточены вблизи π, вначале было обнаружено экспериментально. Если взять $\varepsilon=10^{-4}$, то при n = 40, 60, I00 получаем γ_n = 6, 7, 8. Вот соответствующие минимальные сингулярные числа:

n	40	60	I00	400	I000	I500
ρ_{nn}	I,063I	I,0086	0,94735	0,8I267	0,74258	0,7I523

Как видим, с ростом n происходит уменьшение минимального сингулярного числа матрицы T_n. Остается ли оно ограниченным снизу? Ответ на этот вопрос дает следующая

Лемма 3.2.3. При $n\to\infty$ $\|T_n^{-1}\|_2\to\infty$ и выполняется соотношение

$$\|T_n^{-1}\|_2=O(\sqrt{n}). \qquad (3.2.15)$$

Доказательство. Возьмем $x^{(n)}=[1\dots1]^T$ и покажем, что $\|T_n^{-1}\|_2/\|x^{(n)}\|_2\to\infty$ при $n\to\infty$. Согласно лемме (3.2.I) $T_n^{-1}x^{(n)}=[u_1\dots u_n]^T$, где u_l определяются формулой (3.2.II). Воспользуемся следующими соотношениями (из теории гамма-функции):

$$\prod_{k=1}^{m}\left(1+\frac{1/2}{k}\right)=\frac{2m^{1/2}}{\sqrt{\pi}}+O\left(m^{-1/2}\right), \tag{3.9.I6}$$

$$\prod_{k=1}^{m}\left(1-\frac{1/2}{k}\right)=\frac{m^{-1/2}}{\sqrt{\pi}}+O\left(m^{-3/2}\right).$$

Вследствие (3.2.II) и (3.2.I6) получаем (см. также / 7 /):

$$u_l=\frac{1}{\pi}\sqrt{\frac{n-l+1}{l}}+O\left(\frac{1}{l}\sqrt{\frac{n-l+1}{l}}\right)+$$

$$+O\left(\frac{1}{\sqrt{l(n-l+1)}}\right); \tag{3.2.7}$$

отсюда

$$u_l^2=\frac{1}{\pi^2}\,\frac{n-l+1}{l}+O\left(\frac{n-l+1}{l^2}\right)+O\left(\frac{1}{l}\right),$$

т.е. для некоторых положительных констант c_1, c_2

$$u_l^2\ge\frac{1}{\pi^2}\,\frac{n-l+1}{l}-c_1\frac{n-l+1}{l^2}-c_2\frac{1}{l}.$$

Суммируя эти неравенства, находим

$$\sum_{l=1}^{n}u_l^2\ge\frac{1}{\pi^2}(n+1)\sum_{l=1}^{n}\frac{1}{l}-c_1(n+1)\sum_{l=1}^{n}\frac{1}{l^2}+(c_1-c_2)\sum_{l=1}^{n}\frac{1}{l}-\frac{n}{\pi^2}.$$

Принимая во внимание сходимость ряда $\sum_{l=1}^{\infty}\frac{1}{l^2}=\frac{\pi^2}{6}$ и существование конечного предела

$$\lim_{n\to\infty}\left(\sum_{l=1}^{n}\frac{1}{l}-\ln n\right)\equiv\gamma=0{,}577\ldots,$$

получаем, что

$$\sum_{l=1}^{n}u_l^2\ge\lambda n\ln n-\beta n,$$

где λ, β - положительные константы.

Таким образом,

$$\|T_n^{-1}\|_2\ge\frac{\|T_n^{-1}x^{(n)}\|_2}{\|x^{(n)}\|_2}\ge\sqrt{\lambda\ln n-\beta}, \tag{3.2.I8}$$

т.е. при $n\to\infty$ $\|T_n^{-1}\|_2\to\infty$.

Далее, согласно (3.2.II), (3.2.I2), $u_l>0$, $v_k>0$ для всех l, k. Значит, диагональные элементы матриц U_n, V_n из (3.2.6) положительны, и мы имеем право рассмотреть матрицу $Q_n=V_n^{1/2}T_nU_n^{1/2}$. В силу (3.2.6) $Q_n^{-1}=U_n^{-1/2}(U_nT_n^TV_n)V_n^{-1/2}=U_n^{1/2}T_n^TV_n^{1/2}=Q_n^T$, т.е. матрица Q_n ортогональная. Итак, $T_n^{-1}=U_n^{1/2}Q_n^TV_n^{1/2}$ и, следовательно,

$$\|T_n^{-1}\|_2\le\|U_n^{1/2}\|_2\,\|V_n^{1/2}\|_2, \tag{3.2.I9}$$

так как спектральная норма любой ортогональной матрицы равна I. Вследствие (3.2.I7) и аналогичных соотношений, получаемых с учетом (3.2.I6) для v_k, находим

$$\|U_n\|_2=O(\sqrt{n}),\quad \|V_n\|_2=O(\sqrt{n}). \tag{3.2.20}$$

Поэтому согласно (3.2.I9) $\|T_n^{-1}\|_2=O(\sqrt{n})$. Лемма доказана.

В соответствии с методом дискретных особенностей для расщепления (3.I.2), связанного с (3.I.3), имеем

$$K_n=\left[k(t_{0i},t_j)h\right],\quad h=\tau/n+1. \tag{3.2.21}$$

Предположим, что при всех t_0, t

$$|k(t_0,t)|\le R. \tag{3.2.22}$$

Тогда получаем

$$\|T_n^{-1} K_n\|_E \le \|T_n^{-1}\|_2 \, \|K_n\|_E < \|T_n^{-1}\|_2 \cdot R\tau. \qquad (3.2.23)$$

Возьмем, например, $\tau = 1$ и $n = 1000$. Учитывая замечание к лемме 3.1.2, заключаем, что метод простой итерации будет сходиться при $R \le 0,7425$.

Следующая теорема представляет более тонкий результат о сходимости метода простой итерации.

Теорема 3.2.1. Пусть для матрицы A_n имеет место расщепление (3.1.2), где матрицы T_n и K_n соответственно вида (3.2.1) и (3.2.21). Предположим, что выполняется неравенство (3.2.22). Тогда если

$$R\tau < 4, \qquad (3.2.24)$$

то метод простой итерации (3.1.5) сходится при достаточно больших n для любого начального приближения x_0.

Доказательство. Рассмотрим представление (3.2.6), где диагональные матрицы U_n, V_n определяются согласно (3.2.11), (3.2.12). Как уже отмечалось при доказательстве леммы 3.2.3, матрица $Q_n = V_n^{1/2} T_n U_n^{1/2}$ является ортогональной.

Следовательно, $T_n^{-1} = U_n^{1/2} Q_n^T V_n^{1/2}$.

Очевидно, спектральный радиус матрицы $T_n^{-1} K_n$ совпадает со спектральным радиусом матрицы $U_n^{-1/2} T_n^{-1} K_n U_n^{1/2} = Q_n^T V_n^{1/2} K_n U_n^{1/2}$.

Далее,

$$\|V_n^{1/2} K_n U_n^{1/2}\|_E^2 \le (R\tau)^2 \sum_{l=1}^{n} \frac{u_l}{n+1} \sum_{k=1}^{n} \frac{v_k}{n+1}.$$

Учитывая (3.2.17), находим

$$\sum_{l=1}^{n} \frac{u_l}{n+1} = \frac{1}{\pi} \sum_{l=1}^{n} \sqrt{\frac{n+l-1}{l}} \, \frac{1}{n+1} + O(1) \le$$

$$\le \frac{1}{\pi} \int_0^1 \sqrt{\frac{1-t}{t}}\, dt + o(n) = \frac{1}{4} + o(1).$$

Аналогичная оценка имеет место для суммы $\sum_{k=1}^{n} \frac{v_k}{n+1}$. Таким образом, при достаточно больших n

$$(R\tau)^2 \sum_{l=1}^{n} \frac{u_l}{n+1} \sum_{k=1}^{n} \frac{v_k}{n+1} \le \left(\frac{R\tau}{4}\right)^2 + o(1) < 1.$$

Поэтому

$$\|U_n^{-1/2} T_n^{-1} K_n U_n^{1/2}\|_E < 1,$$

и, значит, все собственные значения матрицы $T_n^{-1} K_n$ лежат внутри единичного круга. Теорема доказана.

Следствие. Пусть матрица A_n отвечает незамкнутому контуру с параметризацией (2.3.1), где функции $x(t)$, $y(t)$ дважды непрерывно дифференцируемы и удовлетворяют условию

$$\sup \frac{(x''(\eta_1))^2 + (y''(\eta_2))^2}{(x'(\xi_1))^2 + (y'(\xi_2))^2} < \frac{4}{\tau}, \qquad (3.2.25)$$

$$0 \le \xi_1, \xi_2, \eta_1, \eta_2 \le \tau.$$

Тогда метод простой итерации (3.1.5) сходится при достаточно больших n для любого начального приближения x_0.

Доказательство. Любой элемент матрицы $K_n = [\Delta_{ij}]$ вычисляется по формуле (при соответствующем выборе t_0, t)

$$\Delta_{ij} = \frac{\dfrac{\Delta x}{\Delta t}\,\dfrac{\frac{\Delta x}{\Delta t} - x'(t_0)}{\Delta t} + \dfrac{\Delta y}{\Delta t}\,\dfrac{\frac{\Delta y}{\Delta t} - y'(t_0)}{\Delta t}}{\left(\dfrac{\Delta x}{\Delta t}\right)^2 + \left(\dfrac{\Delta y}{\Delta t}\right)^2}\, h = k_{ij} h,$$

где

$$\Delta x = x(t_0) - x(t), \quad \Delta y = y(t_0) - y(t), \quad \Delta t = t_0 - t.$$

Отсюда находим

$$|k_{ij}\tau| = \left| \frac{x'(\xi_1)\,x''(\eta_1) + y'(\xi_2)\,y''(\eta_2)}{(x'(\xi_1))^2 + (y'(\xi_2))^2}\,\tau \right| < 4,$$

т.е. в силу теоремы 3.2.I метод простой итерации сходится при достаточно больших n. Следствие доказано.

В действительности выбор в качестве T_n матрицы вида (3.2.I) может оказаться не самым удачным. Расчеты показывают, что практически во всех случаях лучше брать в качестве T_n тёплицеву матрицу, в которой первые строка и столбец те же, что и в матрице A_n. Еще один тип матрицы T_n может быть определен согласно формулам (3.I.9). Чтобы различать эти три типа главных частей, будем обозначать их через $T^{(1)}$, $T^{(2)}$ и $T^{(3)}$.

В качестве примера рассмотрим обтекание параболического контура:

$$x(t) = t, \quad y(t) = t^2, \quad 0 < t < \tau. \qquad (3.2.26)$$

Пусть $\tau = 1$; $n = 80$. В табл. 3.2.I приведены результаты применения метода простой итерации для указанных выше трех типов главных частей T_n. В каждой графе число сверху - это норма невязки $\|f - A_n x_N\|_\infty$, число снизу - это условная относительная погрешность $\|x_N - x_{N-1}\|_\infty / \|x_N\|_\infty$. При уменьшении нормы невязки в 10^6 раз итерации прекращались.

Таблица 3.2.I

Метод простой итерации для параболического контура

Номер итерации	0	I	2	3	4	5	6	7
$T^{(1)}$	0,99 -	0,I2 I,0	2,8 10^{-2} I,I3	6,4 10^{-3} 0,3I	I,5 10^{-3} 6,7 10^{-2}	3,4 10^{-4} I,6 10^{-2}	8,0 10^{-5} 3,6 10^{-3}	I,8 10^{-5} 8,2 10^{-4}
$T^{(2)}$	0,99 -	6,I 10^{-2} I,0	3,3 10^{-3} 0,36	8,6 10^{-5} I,6 10^{-2}	2,5 10^{-6} 2,6 10^{-4}	6,9 10^{-7} 9,5 10^{-6}	- -	- -
$T^{(3)}$	0,99 -	I,5 10^{-2} I,0	3,4 10^{-4} I,I 10^{-2}	2,5 10^{-6} 9,I 10^{-2}	6,3 10^{-7} I,2 10^{-5}	- -	- -	- -

Как видим, третий способ выбора T_n оказался наилучшим. Для него в случае τ = 2, 3, 4, 5 число итераций, уменьшающих норму невязки в 10^6 раз, оказалось соответственно равным 5, 6, 6, 7.

Заметим, что первый способ выбора T_n может давать очень медленную сходимость. Рассмотрим, например, дугу окружности:

$$x(t) = 1 - \cos t, \quad y(t) = \sin t, \quad 0 < t < \tau. \qquad (3.2.27)$$

В этом случае матрица A_n оказывается тёплицевой, поэтому во втором и в третьем способах имеем $T_n = A_n$, и первая итерация должна давать решение системы. При $\tau = 7\pi/4$ в первом способе на 9-й итерации достигается условная относительная погрешность $2{,}5 \cdot 10^{-3}$. Если же взять $\tau = 2\pi$ (т.е. контур почти замкнут), то на 9-й итерации норма невязки уменьшается лишь в 7 раз, а условная относительная погрешность приближения к решению системы равна 0,3I.

Второй и третий способы выбора T_n также могут давать медленную сходимость. Рассмотрим часть эллипса:

$$x(t) = 1 - \cos t, \quad y(t) = 2\sin t, \quad 0 < t < \tau. \qquad (3.2.28)$$

При $\tau = \pi$ на 9-й итерации норма невязки составила 0,4I от ее начального уровня, относительная погрешность приближения равна 0,54.

Таким образом, в ряде случаев желательно иметь более быстрый итерационный процесс. Расчеты показали, что практически для любых контуров быстрая сходимость обеспечивается методом сопряженных градиентов с тёплицевым предобусловливанием. Он применялся к симметризованной системе $A_n^* A_n \varphi = A^* f$; в качестве предобусловливателя выбиралась матрица $T = G^* G$, где G - тёплицева матрица с первыми строкой и столбцом такими же, как в A_n. В таблице 3.2.2 приведены результаты для различных τ в случае контура (3.2.28). Здесь n = 80; число сверху есть отношение $\|r_k\|_\infty / \|r_0\|_\infty$, число снизу - условная относительная погрешность.

Таблица 3.2.2

Метод сопряженных градиентов для симметризованной системы (эллиптический контур)

Номер итерации	I	2	3	4	5	6	7	8
$\tau = \pi$	0,7 I,0	3,7 IO^{-2} 0,82	I,9 IO^{-2} 4,7 IO^{-2}	6,I IO^{-4} I,3 IO^{-2}	9,0 IO^{-5} 2,5 IO^{-4}	2,8 IO^{-5} 3,0 IO^{-4}	2,I IO^{-7} 7,3 IO^{-6}	- -
$\tau = \frac{3}{2}\pi$	0,89 I,0	0,I2 I,3	6,5 IO^{-2} 0,I6	8,3 IO^{-3} 5,2 IO^{-2}	3,7 IO^{-3} I,9 IO^{-2}	9,I IO^{-4} 7,6 IO^{-3}	I,9 IO^{-5} 8,0 IO^{-4}	2,I IO^{-7} 7,7 IO^{-6}

Ниже мы рассмотрим некоторое другое преобразование системы, также ориентированное на применение метода сопряженных градиентов.

3.3. Модифицированное предобусловливание

Преобразование исходной системы $A\varphi = f$ к виду $BAQ\hat{\varphi} = Bf$, $\varphi = Q\hat{\varphi}$, - именно это обычно и называется предобусловливанием. Во многих случаях, однако, более полезен переход к системе $(BAQ + R)\tilde{\varphi} = Bf$, где R - матрица малого ранга; обычно $\operatorname{rg} R = $ I. Такой переход будем называть модифицированным предобусловливанием.

Положим $S \equiv BAQ + R$ и будем считать, что матрицы A, B, Q, S невырожденные. Для определенности пусть $\operatorname{rg} R = $ I; следовательно, $R = uv^T$, где u, v - вектор-столбцы размерности n. Тогда

$$\varphi = Q(BAQ)^{-1}Bf =$$

$$= Q\left(S^{-1} + \frac{1}{1 - v^T S^{-1} u} S^{-1} u v^T S^{-1}\right) Bf.$$

Поэтому, чтобы вычислить вектор φ, достаточно выполнить следующие действия:

I) решить систему $S\hat{u} = u$;

2) $h = \hat{u} / (1 - v^T \hat{u})$;

3) решить систему $Sz = \hat{f}$, $\hat{f} = Bf$;

4) $\varphi = Q(z + h(v^T z))$,

Таким образом, дважды решается система с матрицей S (при этом могут быть использованы те или иные способы предобусловливания); действия I), 2) не используют никакой информации о правой части f.

Далее посмотрим, что дает модифицированное предобусловливание в случае, когда $B = I$, а Q имеет вид

$$Q = \begin{bmatrix} 1 & & & & 0 \\ -1 & 1 & & & \\ & & \ddots & & \\ & & -1 & 1 & \\ 0 & & & -1 & 1 \end{bmatrix}. \qquad (3.3.I)$$

Заметим, что если матрица A тёплицева, то матрица AQ будет отличаться от тёплицевой лишь последним столбцом, т.е. можно выбрать R таким образом, что матрица S будет тёплицевой. Следующая лемма обобщает результат И.К. Лифанова (см. /45/).

Лемма 3.3.I. Пусть вещественная матрица $A = [a_{ij}]$ порядка n такова, что в любой ее строке хотя бы один элемент отличен от нуля, и предположим, что: $a_{ij} \geqslant 0$ при $i \geqslant j$; $a_{ij} \leqslant 0$ при $i < j$; $|a_{ik}| \geqslant |a_{il}|$ при $k < l$, если $i \leqslant \min(k,l)$, или при $k > l$, если $i \leqslant \min(k,l)$. Тогда матрица A невырожденная.

Если Q имеет вид (3.3.I), то матрица $\hat{A} \equiv AQ$ лишь последним столбцом отличается от M-матрицы.

Доказательство. Легко видеть, что матрица $\hat{A} = [\hat{a}_{ij}]$ имеет неположительные внедиагональные элементы. Кроме того, $\hat{a}_{ii} > 0$ для всех i; в противном случае для какого-то i $a_{ii} = a_{i\,i+1}$ и, следовательно, $a_{ii} = a_{i\,i+1} = 0$, а это означает, что вся i-я строка матрицы A нулевая, что противоречит условиям леммы. Фиксируем $\varepsilon > 0$, положим $a_{i\,n+1} = -\varepsilon$ $(i = 1, \ldots, n)$ и определим $S = [s_{ij}]$ следующим образом:

$$S = AQ + uv^T,$$

$$u = -\varepsilon[1 \ldots 1]^T, \quad v = [0 \ldots 0\, 1]^T.$$

8957

Заметим, что $s_{ii} = \hat{a}_{ii} > 0$ и $s_{ij} \le 0$ при $i \ne j$. Далее,

$$\sum_{j=1}^{n} s_{ij} = \sum_{j=1}^{k} (a_{ij} - a_{ij+1}) = a_{ii} + \varepsilon > 0;$$

следовательно, для всех

$$s_{ii} > \sum_{\substack{j=1 \\ j \ne i}}^{n} |s_{ij}|,$$

т.е. матрица S имеет доминирующую диагональ. Поэтому матрица S невырожденная и в действительности является M-матрицей при любом $\varepsilon > 0$. Значит, все элементы матрицы S^{-1} неотрицательны и, следовательно,

$$b \equiv [0 \dots 0\ 1]\, S^{-1} [1 \dots 1]^T > 0.$$

Таким образом, $\gamma = 1 - v^T S^{-1} u = 1 + \varepsilon b > 0$, что влечет невырожденность матрицы $I - S^{-1} u v^T$, поскольку

$$(I - S^{-1} u v^T)\left(I + \frac{1}{\gamma} S^{-1} u v^T\right) =$$

$$= I - S^{-1} u v^T + \frac{1}{\gamma} S^{-1} u v^T - \frac{1}{\gamma} S^{-1} u (v^T S^{-1} u) v^T =$$

$$= I - S^{-1} u v^T + \frac{1}{\gamma} S^{-1} u (1 - v^T S^{-1} u) v^T = I.$$

Значит, матрица $S = uv^T = AQ$ невырожденная. Отсюда вытекает невырожденность матрицы A. Лемма доказана.

Теперь предположим, что матрица A имеет вид (3.2.I). Легко видеть, что для нее выполнены все условия леммы 3.3.I. В данном случае положим

$$a_{in+1} = \frac{1}{i-(n+1)+1/2}, \quad s_{ij} = a_{ij} - a_{ij+1}.$$

Матрица $S_n = [s_{ij}]_{ij=1}^{n}$ является симметричной M-матрицей (такие матрицы называются также матрицами Стилтьеса):

$$S_n = \left[-\frac{1}{(i-j)^2 - 1/4}\right]. \qquad (3.3.2)$$

Лемма 3.3.2. Для матриц S_n имеют место следующие оценки:

$$\|S_n\|_\infty \le 8 - 4/n, \qquad (3.3.3)$$

$$\|S_n^{-1}\|_\infty \le n/4. \qquad (3.3.4)$$

Кроме того,

$$\|S_n\|_2 \le 2\pi. \qquad (3.3.5)$$

Доказательство. Для i-й строчной суммы матрицы S_n находим

$$\sum_{j=1}^{n} |s_{ij}| = 4 + \sum_{j=1}^{i-1} (a_{ij+1} - a_{ij}) + \sum_{j=i+1}^{n} (a_{ij+1} - a_{ij}) =$$

$$= 4 + (a_{ii} - a_{i1}) + (a_{in+1} - a_{ii+1}) = 8 - \frac{n}{(i-\frac{1}{2})(n-(i-\frac{1}{2}))} \le$$

$$\le 8 - \frac{n}{n^2/4} = 8 - \frac{4}{n}.$$

Далее, $\|S_n^{-1}\|_\infty = \max_{x \ne 0} \frac{\|x\|_\infty}{\|S_n x\|_\infty}$. Возьмем произвольный вектор $x \ne 0$ и предположим, что $|x_i| = \|x\|_\infty$. Тогда

$$\|S_n x\|_\infty \ge \left(s_{ii} - \sum_{\substack{j=1 \\ j \ne i}}^{n} |s_{ij}|\right) \|x\|_\infty =$$

$$= (a_{i1} - a_{in+1}) \|x\|_\infty = \frac{n}{(i-\frac{1}{2})(n-(i-\frac{1}{2}))} \|x\|_\infty.$$

Следовательно,

$$\| S_n^{-1} \|_\infty \le \max_i \frac{\left(i-\frac{1}{2}\right)\left(n-\left(i-\frac{1}{2}\right)\right)}{n} \le \frac{n}{4}.$$

Чтобы получить оценку (3.3.5), заметим, что матрицы S_n вложены в бесконечную теплицеву матрицу, отвечающую ряду Фурье

$$f(x)=\sum_{k=-\infty}^{\infty} \frac{-1}{k^2-1/4}\, e^{ikx}. \qquad (3.3.6)$$

Нетрудно проверить, что

$$f(x) = 2\pi \left| \sin \frac{x}{2} \right|. \qquad (3.3.7)$$

Если $w=[w_0 \ \dots \ w_{n-1}]^T$, то

$$(S_n w,\ w)=\frac{1}{2\pi}\int_{-\pi}^{\pi} \left| w_0 + w_1 e^{ix} + \dots + w_{n-1} e^{i(n-1)x} \right|^2 f(x)\,dx$$

$$(w,w) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left| w_0 + w_1 e^{ix} + \dots + w_{n-1} e^{i(n-1)x} \right|^2 dx.$$

Поэтому

$$\| S_n \|_2 \le \max_{-\pi \le x \le \pi} |f(x)| \le 2\pi.$$

Лемма доказана.

Замечание I. Неравенство (3.3.3) можно уточнить, поскольку из проведенных выкладок видно, что

$$\| S_n \|_\infty = 8 - \frac{n}{\left[\frac{n}{2}\right]\left(n-\left[\frac{n}{2}\right]\right)}. \qquad (3.3.8)$$

Отсюда следует, что

$$\lim_{n\to\infty} \| S_n \|_\infty = 8. \qquad (3.3.9)$$

Замечание 2. Согласно (3.3.6), (3.3.7) из теоремы Г. Сегё вытекает, что собственные значения

$$\lambda_1^{(n)} \le \lambda_2^{(n)} \le \dots \le \lambda_n^{(n)}$$

матриц S_n распределены на отрезке $[0,\ 2\pi]$ так же, как последовательность чисел

$$f_k^{(n)} = 2\pi \left| \sin\left(-\frac{\pi}{2}+\frac{k\pi}{n+1}\right) \right|, \quad k=1, \dots, n. \qquad (3.3.10)$$

В частности, для любого отрезка $\Delta \in [0, 2\pi]$ обозначим через $\lambda^{(n)}(\Delta)$ и $f^{(n)}(\Delta)$ количество чисел $\lambda_k^{(n)}$ и $f_k^{(n)}$, принадлежащих Δ; тогда при $n \to \infty$

$$\frac{\lambda^{(n)}(\Delta) - f^{(n)}(\Delta)}{n} \to 0. \qquad (3.3.11)$$

Отсюда следует, что

$$\lim_{n\to\infty} \| S_n \|_2 = 2\pi, \qquad (3.3.12)$$

$$\lim_{n\to\infty} \| S_n^{-1} \|_2 = \infty. \qquad (3.3.13)$$

Далее рассмотрим матрицу A, полученную в согласии с методом дискретных особенностей для сингулярного итерационного уравнения (3.I.3). Доопределим $a_{i\,n+1}$ естественным образом:

$$a_{i\,n+1} = \frac{1}{t_{0i}-t_{n+1}} - k(t_{0i},\ t_{n+1})\,h, \qquad (3.3.14)$$

$$h = \tau/(n+1).$$

Матрица AQ лишь последним столбцом отличается от матрицы

$$\tilde{A}_n = \left[a_{ij} - a_{i\,j+1} \right]_{ij=1}^{n} \qquad (3.3.15)$$

Естественно ожидать, что матрица S_n вида (3.3.2) может представлять главную часть для $\tilde{A}_n$.

Теорема 3.3.I. Предположим, что функция $k(t_0,t)$ удовлетворяет условию

$$\sup_{0\le t_0,t\le\tau}\left|\frac{\partial k(t_0,t)}{\partial t}\right| < \frac{4}{\tau^2} \qquad (3.3.16)$$

Тогда при достаточно больших n все собственные значения матрицы $I - S_n^{-1}\tilde{A}_n$ находятся внутри единичного круга.

Доказательство. Согласно (3.3.I6) при достаточно больших n

$$\left|\frac{k(t_{0i},t_j) - k(t_{0i},t_{j+1})}{h}\right| < \frac{4}{\tau^2}.$$

Следовательно, $\| S_n - \tilde{A}_n \|_E < \frac{4h^2}{\tau^2} n = \frac{4n}{(n+1)^2}$. Учитывая симметричность матрицы S_n^{-1}, находим $\|S_n^{-1}\|_2 \le \|S_n^{-1}\|_\infty$. В силу (3.3.4)

$$\| I - S_n^{-1}\tilde{A}_n \|_E \le \|S_n^{-1}\|_2 \; \|S_n - \tilde{A}_n\|_E < \left(\frac{n}{n+1}\right)^2 < 1.$$

Теорема доказана.

Примечательно, что в практически интересных случаях матрица $\tilde{A}_n$ оказывается почти симметричной.

Лемма 3.3.3. Пусть $\Phi(t_0,t) \equiv \frac{\partial}{\partial t} k(t_0,t)$, - и предположим, что для всех t_0, t

$$\Phi(t_0,t) = \Phi(t,t_0). \qquad (3.3.17)$$

Тогда матрицы $\tilde{A}_n$ почти симметричны в том смысле, что

$$\|\tilde{A}_n - \tilde{A}_n^T\|_E = O\left(\frac{1}{n^2}\right). \qquad (3.3.18)$$

Доказательство. Заметим, что при $i \ne j$

$$k(t_{0i},t_{j+1}) - k(t_{0i},t_j) = \int_{t_j}^{t_{j+1}} \Phi(t_{0i},t)\,dt = \Phi(t_{0i},t_{0j})h + O(h^2).$$

Учитывая определение матрицы $\tilde{A}_n$, приходим к (3.3.I8). Лемма доказана.

Обратим внимание на то, что в случае сингулярного интегрального уравнения (2.3.2) мы находимся именно в условиях леммы 3.3.3. Пренебрегая малой кососимметричной частью матрицы $\tilde{A}_n$, мы можем использовать методы, ориентированные на симметричные матрицы, например метод сопряженных градиентов. В качестве предобусловливателя можно взять симметричную тёплицеву матрицу S_n вида (3.3.2). Целесообразно также рассмотреть задачи оптимизации (3.I.7) и (3.I.8), считая $\mathcal{T}$ множеством симметричных тёплицевых матриц.

Обычно быстрая сходимость достигается, если в качестве предобусловливателя взять тёплицеву матрицу с такими же первыми строкой и столбцом, как и в $\tilde{A}_n$. По крайней мере, именно это наблюдалось в расчетах для различных типов профилей, в том числе для почти замкнутых и сильно изогнутых.

3.4. Циркулянтное предобусловливание и его оптимизация

Тестовые расчеты показывают, что во многих случаях целесообразна замена тёплицевых предобусловливателей на циркулянтные, дважды тёплицевых - на дважды циркулянтные. Число итераций при этом несколько возрастает, однако, согласно теореме I.8.2, многоуровневые циркулянты, в том числе дважды циркулянтные матрицы, являются столь же "легко обратимыми", как и обычные циркулянты. В отношении дважды тёплицевых матриц этого сказать уже нельзя. Поэтому при решении алгебраических систем, связанных с двумерными уравнениями, в том числе при решении дважды тёплицевых систем, наиболее подходящий тип предобусловливателей представляют дважды циркулянтные матрицы. Многие свойства дважды циркулянтного предобусловливания можно получить по аналогии со свойствами циркулянтного предобусловливания. Поскольку последнее изучается несколько проще (без громоздких выкладок), то начнем именно с него.

Итак, пусть $A = [a_{ij}]$ - вещественная невырожденная матрица порядка n и $C = [c_{i-j}]$ - вещественная циркулянтная матрица того же порядка. Предположим, что C есть решение задачи оптимизации типа (3.I.8):

$$\|C-A\|_E \to \min, \quad C \in \mathcal{T}_n, \tag{3.4.1}$$

где $\mathcal{T}_n$ - множество вещественных циркулянтов порядка n. Тогда несложно убедиться в том, что

$$c_k = \frac{1}{n} \sum_{p-q=k(\mathrm{mod}\, n)} a_{pq}, \quad k = 0, 1, \ldots, n-1. \tag{3.4.2}$$

Лемма 3.4.1. Пусть A - произвольная вещественная невырожденная матрица порядка n, такая, что $(Ax, x) \neq 0$ при $x \neq 0$, и циркулянтная матрица C' определяется согласно (3.4.2). Тогда матрица C невырожденная.

Если A симметричная, то и C симметричная; при этом если A положительно определенная, то и C положительно определенная. Для упорядоченных по невозрастанию собственных значений матриц A и C справедливы следующие соотношения:

$$\lambda_1(A) + \ldots + \lambda_j(A) \geq \lambda_1(C) + \ldots + \lambda_j(C), \tag{3.4.3}$$

$$\lambda_n(A) + \ldots + \lambda_{n-j+1}(A) \leq \lambda_n(C) + \ldots + \lambda_{n-j+1}(C), \tag{3.4.4}$$
$$j = 1, \ldots, n.$$

Доказательство. Рассмотрим ортонормированную систему векторов

$$u_k = \frac{1}{\sqrt{n}} \left[\varepsilon^{k\cdot 0}\ \varepsilon^{k\cdot 1}\ \ldots\ \varepsilon^{k(n-1)} \right]^T, \quad k = 0, 1, \ldots, n-1,$$

где $\varepsilon = \exp(2\pi i/n)$, i - мнимая единица. Согласно теореме 1.8.1 u_k суть собственные векторы циркулянтной матрицы C, отвечающие собственным значениям (Cu_k, u_k). В соответствии с (3.4.2) находим

$$(Cu_k, u_k) = \frac{1}{n} \sum_{p=0}^{n-1} \varepsilon^{-kp} \sum_{q=0}^{n-1} c_{p-q}\, \varepsilon^{kq} =$$

$$= \sum_{l=0}^{n-1} c_l\, \varepsilon^{-kl} = \sum_{l=0}^{n-1} \varepsilon^{-kl} \frac{1}{n} \sum_{p-q=l(\mathrm{mod}\, n)} a_{pq} = \tag{3.4.5}$$

$$= \frac{1}{n} \sum_{p=0}^{n-1} \varepsilon^{-kp} \sum_{q=0}^{n-1} a_{pq}\, \varepsilon^{kq} = (Au_k, u_k).$$

По условию леммы $(Au_k, u_k) \neq 0$ для всех k; поэтому все собственные значения матрицы C отличны от нуля, вследствие чего матрица C невырожденная.

Предположим, что матрица A симметричная. Тогда

$$\sum_{p-q=k(\mathrm{mod}\, n)} a_{pq} = \sum_{p-q=n-k\,(\mathrm{mod}\, n)} a_{pq}, \quad k = 1, \ldots, n-1;$$

согласно (3.4.2) получаем

$$c_k = c_{n-k}, \quad k = 1, \ldots, n-1,$$

что равносильно симметричности матрицы C. Если матрица A положительно определенная, то $(Au_k, u_k) > 0$ для всех k и вследствие (3.4.5) все собственные значения для C положительны. Значит, матрица C положительно определенная.

Далее, обозначим через $v_1, \ldots, v_n$ те же векторы $u_0, \ldots, u_{n-1}$, взятые в таком порядке, что

$$(Cv_1, v_1) \geq (Cv_2, v_2) \geq \ldots \geq (Cv_n, v_n).$$

Тогда $\lambda_j(C) = (Cv_j, v_j)$. Для любой ортонормированной системы векторов $z_1, \ldots, z_j$ имеют место неравенства (см. / 17 /)

$$\sum_{l=1}^{j} \lambda_{n-l+1}(A) \leq \sum_{l=1}^{j} (Az_j, z_j) \leq \sum_{l=1}^{j} \lambda_l(A),$$

откуда и вытекают соотношения (3.4.3), (3.4.4). Лемма доказана.

Теперь предположим, что C есть решение более трудной задачи оптимизации типа (3.I.7):

$$\|I - C^{-1}A\|_E \to \min, \quad C \in \mathring{\mathfrak{T}}_n, \tag{3.4.6}$$

где $\mathring{\mathfrak{T}}_n$ - множество вещественных невырожденных циркулянтов порядка n. Запишем

$$\hat{C} = C^{-1} = \sum_{j=0}^{n-1} \alpha_j Q^j, \tag{3.4.7}$$

где

$$Q = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & I \\ I & 0 & 0 & \dots & 0 & 0 \\ 0 & I & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & I & 0 \end{bmatrix} \tag{3.4.8}$$

Вместо задачи (3.4.6) рассмотрим следующую задачу:

$$\|I - \hat{C}A\|_E \to \min, \quad \hat{C} \in \mathfrak{T}_n. \tag{3.4.9}$$

Если минимум в (3.4.6) достигается, то для соответствующей циркулянтной матрицы C имеем $C^{-1} = \hat{C}$.

Чтобы решить задачу (3.4.9), исследуем функционал

$$\mathfrak{F}(\alpha_0, \alpha_1, \dots, \alpha_{n-1}) \equiv \|I - \sum_{j=0}^{n-1} \alpha_j Q^j A\|_E^2 =$$

$$= tr\left(\left(I - \sum_{i=0}^{n-1} \alpha_i A^T Q^{-i}\right)\left(I - \sum_{j=0}^{n-1} \alpha_j Q^j A\right)\right) = \tag{3.4.10}$$

$$= n - \sum_{i=0}^{n-1} \alpha_i f_i + \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \alpha_i \alpha_j u_{ij},$$

где

$$f_i = tr(A^T Q^i + Q^i A), \tag{3.4.11}$$

$$u_{ij} \equiv tr(A^T Q^{j-i} A). \tag{3.4.12}$$

Вычислим для $\mathfrak{F}$ частные производные и приравняем их к нулю:

$$\frac{\partial \mathfrak{F}}{\partial \alpha_i} = -f_i + \sum_{\substack{j=0 \\ j \neq i}}^{n-1} \alpha_j (u_{ij} + u_{ji}), \tag{3.4.13}$$

или, в матрично-векторной записи,

$$(U + U^T)\begin{bmatrix} \alpha_0 \\ \dots \\ \alpha_{n-1} \end{bmatrix} = \begin{bmatrix} f_0 \\ \dots \\ f_{n-1} \end{bmatrix}, \tag{3.4.14}$$

$$U = [u_{ij}]_{ij=0}^{n-1}. \tag{3.4.15}$$

Очевидно, решение линейной алгебраической системы (3.4.14) будет определять именно точку минимума для $\mathfrak{F}$.

Лемма 3.4.2. Пусть A - произвольная вещественная невырожденная матрица порядка n. Тогда линейная алгебраическая система (3.4.14) имеет единственное решение. Если $(Ax, x) \neq 0$ при $x \neq 0$, то циркулянтная матрица $\hat{C}$ с первым столбцом $[\alpha_0 \dots \alpha_{n-1}]^T$ невырожденная.

Если A симметричная, то и $\hat{C}$ симметричная; при этом если A положительно определенная, то и $\hat{C}$ положительно определенная.

Доказательство. Матрица U, определяемая согласно (3.4.15), (3.4.12), очевидно циркулянтная. В силу свойств следа матрицы

$$u_{ij} = tr(A^T Q^{j-i} A) = tr(Q^{j-i} A A^T). \tag{3.4.16}$$

Учитывая (3.4.2), приходим к выводу, что матрица $\tilde{U} = \frac{1}{n} U$ является решением следующей задачи оптимизации:

$$\|\tilde U - AA^T\|_E \to \min, \quad \tilde U \in \mathcal{T}_n. \qquad (3.4.17)$$

Согласно лемме (3.4.I) матрица $\tilde U$ симметричная положительно определенная, поскольку таковой является матрица AA^T. Таким образом, матрица $U + U^T = 2U$ симметричная положительно определенная; следовательно, система (3.4.I4) имеет единственное решение.

Обозначим через F циркулянтную матрицу с первым столбцом $[f_0 \dots f_{n-1}]^T$. Поскольку произведение циркулянтов остается циркулянтом, вследствие (3.4.I4) имеем

$$2U\hat C = F. \qquad (3.4.18)$$

Легко видеть, что $F = [2\,tr(Q^{i-j}A)]_{ij=0}^{n-1}$, т.е. циркулянтная матрица $\tilde F = \frac{1}{n} F$ есть решение задачи оптимизации

$$\|\tilde F - A\|_E \to \min, \quad \tilde F \in \mathcal{T}_n, \qquad (3.4.19)$$

и, согласно лемме 3.4.I, если $(Ax, x) \neq 0$ при $x \neq 0$, то матрица $\tilde F$ невырожденная. В силу (3.4.I8) приходим к выводу о невырожденности матрицы $\hat C$.

Если матрица A симметричная, то по лемме 3.4.I матрица F также симметричная. Поскольку произведение симметричных циркулянтов остается симметричным циркулянтом, вследствие (3.4.I8) матрица $\hat C$ симметричная. Если A положительно определечная, то такой же будет и F (по лемме 3.4.I). В силу (3.4.I8) положительная определенность симметричных циркулянтных матриц U, F влечет положительную определенность симметричной циркулянтной матрицы $\hat C$. Лемма доказана.

При практическом использовании циркулянтного предобусловливателя $\hat C$, очевидно, необходимо научиться получать его достаточно быстро. Мы уже знаем, что его компоненты $\alpha_0, \dots, \alpha_{n-1}$ определяются как решение системы (3.4.I4) с симметричной положительно определенной циркулянтной матрицей. Чтобы найти решение такой системы, требуется $O(n \log_2 n)$ арифметических операций. Несколько бо́льшие затраты связаны с вычислением компонент u_{ij}, f_i. Вектор $[f_0 \dots f_{n-1}]^T$ - правая часть системы (3.4.I4)-вычисляется, как легко

видеть, с затратой $O(n^2)$ операций. Далее, матрица U определяется своей первой строкой, компоненты которой имеют вид

$$u_{0j} = tr(A^T Q^j A) = \sum_{\nu=0}^{n-1} a_\nu^T Q^j a_\nu, \qquad (3.4.20)$$

где $a_0, \dots, a_{n-1}$ - столбцы матрицы A. Фиксируем ν и найдем сначала все величины

$$\alpha_{\nu j} \equiv a_\nu^T Q^j a_\nu. \qquad (3.4.21)$$

Для этого заметим, что имеет место соотношение

$$[\alpha_{\nu 0} \dots \alpha_{\nu n-1}] = a_\nu^T C_\nu,$$

$$C_\nu = [Q^0 a_\nu, \dots, Q^{n-1} a_\nu], \qquad (3.4.22)$$

причем матрица C_ν является циркулянтной. Для каждого ν операция умножения вектор-строки на циркулянт выполняется с затратой $O(n \log_2 n)$ операций; следовательно, все величины $\alpha_{\nu j}$ могут быть получены с затратой $O(n^2 \log_2 n)$ операций. Еще $O(n^2)$ операций сложения потребуется для того, чтобы по найденным $\alpha_{\nu j}$ вычислить u_{0j}. Таким образом, для вычисления циркулянтной матрицы $\hat C$ достаточно $O(n^2 \log_2 n)$ операций в случае матрицы A общего вида. Если матрица A теплицева, то затраты, идущие на построение $\hat C$, могут быть снижены до $O(n \log_2 n)$ операций (см. § 4.5).

Полученные выше результаты в основном переносятся на случай дважды циркулянтного предобусловливания, а также и на более общий случай кратного (многоуровневого) циркулянтного предобусловливания. Ниже мы покажем, как выглядят эти обобщения.

Пусть матрица A порядка n рассматривается как многоуровневая. Это означает, что любой ее элемент a_{ij} $(0 \le i, j \le n-1)$ может быть указан с помощью s пар индексов i_l, j_l $(0 \le i_l, j_l \le n_l - 1;\ n = n_1 \dots n_s)$:

$$a_{ij} = a_{i_1 j_1; \dots; i_s j_s},$$

$$i = i_1 n_2 n_3 \dots n_s + i_2 n_3 \dots n_s + \dots + i_{s-1} n_s + i_s, \quad (3.4.23)$$

$$j = j_1 n_2 n_3 \dots n_s + j_2 n_3 \dots n_s + \dots + j_{s-1} n_s + j_s.$$

В соответствии с терминологией, введенной в § I.I, матрица A имеет тип $G_{n_1} \dots G_{n_s}$, где G_{n_ℓ} обозначает тип матрицы общего вида порядка n_ℓ ; s - число уровней матрицы A .

Рассмотрим задачу оптимизации (3.4.I) в предположении, что $\mathcal{T}_n$ есть множество вещественных s-уровневых циркулянтов, т.е. матриц, имеющих композиционный тип $C_{n_1} \dots C_{n_s}$ (см. § I8). Матрица C задается, очевидно, своим первым столбцом; положим

$$c_{i_1; \dots; i_s} \equiv c_{i_1 0; \dots; i_s 0}. \quad (3.4.24)$$

Несложно убедиться в том, что элементы оптимальной матрицы C определяются по следующим формулам:

$$c_{i_1; \dots; i_s} = \frac{1}{n} \sum_{\substack{p_1 - q_1 = i_1 (\mathrm{mod}\, n_1) \\ \cdots \\ p_s - q_s = i_s (\mathrm{mod}\, n_s)}} a_{p_1 q_1; \dots; p_s q_s}. \quad (3.4.25)$$

Лемма 3.4.3. Пусть A - произвольная вещественная невырожденная матрица порядка $n = n_1 \dots n_s$ и матрица C есть s-уровневый циркулянт, имеющий композиционный тип $C_{n_1} \dots C_{n_s}$ и определяемый согласно (3.4.25). Предположим, что $(Ax, x) \neq 0$ при $x \neq 0$. Тогда матрица C невырожденная.

Если A симметричная, то и C симметричная; при этом если A положительно определенная, то и C положительно определенная. Для упорядоченных по возрастанию собственных значений матриц A и C справедливы неравенства (3.4.3), (3.4.4).

Доказательство. Обозначим через ε_ℓ любой корень из единицы

степени n_ℓ , положим

$$u_\ell = \frac{1}{\sqrt{n_\ell}} \left[\varepsilon_\ell^0 \ \varepsilon_\ell^1 \ \dots \ \varepsilon_\ell^{n_\ell - 1} \right]^T, \quad \ell = 1, \dots, s,$$

и рассмотрим вектор $u = u_1 \times \dots \times u_s$. Вектор u является нормированным собственным вектором для C , отвечающим собственному значению (Cu, u) . Выбирая различные корни u_ℓ и составляя соответствующие векторы u , мы можем получить из них ортонормированную систему собственных векторов матрицы C . В силу (3.4.5) находим

$$(Cu, u) = (Au, u) \neq 0,$$

откуда и вытекает невырожденность матрицы C .

Если матрица A симметричная, то выполняются соотношения

$$\sum_{\substack{p_1 - q_1 = k_1 (\mathrm{mod}\, n_1) \\ \cdots \\ p_s - q_s = k_s (\mathrm{mod}\, n_s)}} a_{p_1 q_1; \dots; p_s q_s} = \sum_{\substack{p_1 - q_1 = n_1 - k_1 (\mathrm{mod}\, n_1) \\ \cdots \\ p_s - q_s = n_s - k_s (\mathrm{mod}\, n_s)}} a_{p_1 q_1; \dots; p_s q_s} \quad (3.4.26)$$

$$0 \le k_1 \le n_1 - 1, \ \dots, \ 0 \le k_s \le n_s - 1.$$

Отсюда заключаем, что матрица C симметричная. По поводу собственных значений рассуждения дословно совпадают с приведенными при доказательстве леммы 3.4.I. Лемма доказана.

Лемма 3.4.4. Пусть A - произвольная вещественная невырожденная матрица порядка $n = n_1 \dots n_s$. Рассмотрим задачу оптимизации (3.4.9) в предположении, что $\mathcal{T}_n$ есть множество вещественных s-уровневых циркулянтов, т.е. матриц типа $C_{n_1} \dots C_{n_s}$. Тогда эта задача имеет, и притом единственное, решение $\hat{C}$. Если $(Ax, x) \neq 0$ при $x \neq 0$, то соответствующая матрица $\hat{C}$ невырожденная. Если A симметричная, то и $\hat{C}$ симметричная; при этом если A положительно определенная, то и $\hat{C}$ положительно определенная.

Обозначим через Q_ℓ матрицу вида (3.4.8) порядка n_ℓ . Тогда имеет место представление

$$\hat{C} = \sum_{j_1 = 0}^{n_1 - 1} \dots \sum_{j_s = 0}^{n_s - 1} d_{j_1; \dots; j_s} Q_1^{j_1} \times \dots \times Q_s^{j_s}, \quad (3.4.27)$$

где $d_{j_1,\dots,j_s}$ суть компоненты первого столбца d матрицы $\hat{C}$. Рассмотрим функционал

$$\mathfrak{F} \equiv \| I - \hat{C} A \|_E^2 =$$

$$= n - \sum_{\substack{0 \le i_1 \le n_1 - 1 \\ \dots \\ 0 \le i_s \le n_s - 1}} d_{i_1,\dots,i_s} \, f_{i_1,\dots,i_s} + \qquad (3.4.28)$$

$$+ \sum_{\substack{0 \le i_1, j_1 \le n_1 - 1 \\ \dots \\ 0 \le i_s, j_s \le n_s - 1}} d_{i_1,\dots,i_s} \, d_{j_1,\dots,j_s} \, u_{i_1 j_1,\dots,i_s j_s},$$

где

$$f_{i_1,\dots,i_s} \equiv tr\left(A^T Q_1^{-i_1} \times \dots \times Q_s^{-i_s} + Q_1^{i_1} \times \dots \times Q_s^{i_s} A\right), \qquad (3.4.29)$$

$$u_{i_1 j_1,\dots,i_s j_s} \equiv tr\left(A^T Q_1^{j_1 - i_1} \times \dots \times Q_s^{j_s - i_s} A\right). \qquad (3.4.30)$$

Для определения d получаем систему линейных алгебраических уравнений

$$(U + U^T) d = f, \qquad (3.4.31)$$

где U есть s-уровневая матрица с компонентами $u_{i_1 j_1,\dots,i_s j_s}$; f - вектор с компонентами $f_{i_1,\dots,i_s}$, упорядоченный таким образом, что сначала возрастает индекс j_s, затем i_{s-1} и т.д. Система (3.4.31) исследуется таким же способом, как изученная ранее система (3.4.14).

Если матрица A не обладает какой-либо спецификой, то s-

уровневый циркулянт $\hat{C}$ можно получить ценой $O(n^2 \log_2 n)$ операций. Построения вполне аналогичны уже рассмотренным в случае обычного циркулянта $\hat{C}$.

Предположим теперь, что матрица A является дважды тёплицевой. В этом случае дважды циркулянт $\hat{C}$ строится с затратой $O(n \log_2 n)$ операций (см. §4.5).

В качестве примера рассмотрим линейную алгебраическую систему (2.4.8) с дважды тёплицевой и дважды симметричной матрицей, элементы которой имеют вид (2.4.9). Эта система соответствует сингулярному интегральному уравнению (2.4.4), где область интегрирования D есть прямоугольник (2.4.5). В табл. 3.4.1 приведены результаты расчетов в случае, когда $a = 10$, $b = 19$. Здесь рассмотрены два типа дважды циркулянтных предобусловливателей: $\tilde{C}$ и C. Матрица C есть решение задачи типа (3.4.1), изученное в лемме 3.4.3. Матрица $\tilde{C}$ строится следующим образом: компоненты $\tilde{c}_{i_1 i_2}$ ее первого столбца выражаются через компоненты $a_{i_1 i_2}$ первого столбца матрицы A по формулам

$$\tilde{c}_{i_1 i_2} = \begin{cases} a_{i_1 i_2}, & 0 \le i_1 \le \frac{n_1}{2}, \quad 0 \le i_2 \le \frac{n_2}{2}, \\ a_{i_1, n - i_2}, & 0 \le i_1 \le \frac{n_1}{2}, \quad \frac{n_2}{2} < i_2 \le n_2 - 1, \\ a_{n_1 - i_1, i_2}, & \frac{n_1}{2} < i_1 \le n_1 - 1, \quad 0 \le i_2 \le \frac{n_2}{2}, \\ a_{n_1 - i_1, n_2 - i_2}, & \frac{n_1}{2} < i_1 \le n_1 - 1, \quad \frac{n_2}{2} < i_2 \le n_2 - 1. \end{cases} \qquad (3.4.32)$$

Число итераций отвечает моменту падения евклидовой нормы невязки в 10^5 раз.

Таблица 3.4.1

Дважды циркулянтное предобусловливание

n	n_1	n_2	Число итераций	
			$\tilde{C}$	C
256	16	16	25	18
1024	32	32	37	32
4096	64	64	36	30

8957

I2-2