О методах неполной факторизации с обобщенной компенсацией*

В.П. Ильин, К.Ю. Лаевский

УДК 519.63

Ильин В.П., Лаевский К.Ю. О методах неполной факторизации с обобщенной компенсацией // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 1998. — Т. 1, № 4. — С. 321–336.

Исследуются итерационные методы неполной факторизации, основанные на определении предобуславливающей матрицы B из обобщенного принципа компенсации $By_k = Ay_k$, $k = 1, \ldots, m$, где A – матрица исходной системы линейных алгебраических уравнений, а $\{y_k\}$ – совокупность алгоритмов при решении блочно-трехдиагональных систем стилтьесовского типа, а также условия положительной определенности предобуславливающих матриц для некоторых конкретных наборов пробных векторов. Выводятся оценки чисел обусловленности матричных произведений $B^{-1}A$, определяющие скорость сходимости итераций, в зависимости от свойств элементов исходных матриц.

Il'in V.P., Laevsky K.Yu. On incomplete factorization methods with generalized compensation // Siberian J. of Numer. Mathematics / Sib. Branch of Russ. Acad. of Sci. — Novosibirsk, 1998. — Vol. 1, № 4. — P. 321–336.

The iterative incomplete factorization methods are described on the base of definition of preconditioning B matrix from generalized compensation principle $By_k = Ay_k$, $k = 1, \ldots, m$, where A is the matrix of original system of linear algebraic equations and $\{y_k\}$ is the set of so called probe vectors. The correctness of such algorithms and conditions of positive definiteness of preconditioning matrices are investigated for solution to the Stieltjes type block-tridiagonal systems. The estimates of condition number of matrix product $B^{-1}A$, that define the iterative convergence rate, are derived in the terms of the properties of original matrices.

1. Введение

Итерационные алгоритмы неполной факторизации являются в настоящее время одними из самых эффективных и активно развиваемых подходов к решению систем линейных алгебраических уравнений высокого порядка

$$Au = f \tag{1}$$

с разреженными матрицами, возникающими при аппроксимации многомерных краевых задач сеточными методами (см. [1–3] и цитируемую там литературу). Главной проблемой здесь является построение факторизованных предобуславливающих матриц B, которые были бы легко обратимы и достаточно близкими, в некотором спектральном смысле, к исходной матрице системы A. Итерационный процесс в простейшем случае реализуется по формуле

$$B(u^n - u^{n-1}) = f - Au^{n-1}. \tag{2}$$

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант 96-01-01770), а также является частью исследований по проекту INTAS-93-377-ext.

Наиболее эффективные алгоритмы – чебышевского ускорения или сопряженных градиентов – также используют вычисления по формуле (2), но затем вектор очередного приближения u^n корректируется на основе спектральной или вариационной оптимизации итерационного процесса. При условии симметричности и положительной определенности матриц A, B число итераций, необходимое для уменьшения начальной ошибки в ε^{-1} раз $((Az^n, z^n)/(Az^0, z^0) \le \varepsilon$, $z^n = u - u^n)$, оценивается величиной $n(\varepsilon) \le 1/2|\ln\varepsilon|\kappa^{1/2} + 1$, где κ – число обусловленности матрицы $B^{-1/2}AB^{-1/2}$ в спектральной (евклидовой) норме, равное отношению максимального и минимального собственных чисел подобной ей матрицы $B^{-1}A$. Таким образом, проблема оптимизации итерационного процесса, в смысле минимизации оценки $n(\varepsilon)$, связана с минимизацией величины κ.

Для блочно-трехдиагональных матриц $A = D - L - U$, где $D = \mathrm{diag}\{D_k\}$ – блочно-диагональная, а $L = \{L_k\}$, $U = \{U_k\}$ – нижняя и верхняя строго треугольные матрицы (порядки N_k блоков D_k могут быть разные, а порядок матрицы A равен $N = N_1 + \cdots + N_M$), матрица B определяется в форме

$$B = (G - L)G^{-1}(G - U), \tag{3}$$

где $G = \mathrm{diag}\{G_k\}$ – блочно-диагональная матрица, блоки которой находятся следующим рекуррентным способом:

$$G_1 = D_1, \quad G_k = D_k - \overline{L_k G_{k-1}^{-1} U_{k-1}} - \theta C_k, \ \ k = 2, \dots, M. \tag{4}$$

Здесь $0 \le \theta \le 1$ – итерационный параметр, матрица $\overline{Q}$ означает приближение (в определенном смысле) к матрице Q, а C_k традиционно определяется как диагональная матрица, элементы которой находятся из равенства

$$C_1 = 0, \quad C_k e = \left(L_k G_{k-1}^{-1} U_{k-1} - \overline{L_k G_{k-1}^{-1} U_{k-1}}\right) e, \ \ k = 2, \dots, M, \tag{5}$$

где e – вектор с единичными компонентами.

Основной способ построения матричных “аппроксимаций” – ленточный:

$$\overline{Q} = Q^{(p)}, \tag{6}$$

где $Q^{(p)}$ означает “ленточную часть” шириной p матрицы Q (при $|i - j| \le (p-1)/2$, $p = 1, 3, 5, \dots$, элементы $Q^{(p)}$ те же, что и у Q, а остальные – нулевые).

В работах Н.И. Булеева, первооткрывателя методов неполной факторизации, $C = \mathrm{diag}\{C_k\}$, а матрицы G и B определялись эмпирически на основе аппроксимационных принципов. А именно, условие (5) соответствует тому, что при при $\theta = 1$ имеет место векторное равенство

$$Be = Ae, \tag{7}$$

названное условием полной компенсации. Поскольку вектор ошибки $z^1 = u - u^1$, определяемый из (2), удовлетворяет соотношению $Bz^1 = (B - A)z^0$, при выполнении условия (7) получаем $z^1 = 0$, т.е. точное решение достигается за одну итерацию, если вектор начальной ошибки z^0 имеет одинаковые компоненты. При этом делается естественное предположение: если вектор z^0 имеет “гладкие” компоненты, что зачастую имеет место при решении задач математической физики, то вектор u^1 будет близок к точному решению u.

Позднее Стоуном был предложен метод [4], в котором принцип компенсации развит так, что точное решение получалось за одну итерацию, если компоненты вектора начальной ошибки z^0 суть значения некоторой линейной функции. Однако конструируемая при этом предобуславливающая матрица B оказывается несимметричной, что затрудняет применение ускорения итерационного процесса. В работах других авторов [5] делались попытки построения аналогичных симметричных предобуславливателей, но при этом алгоритмы вычисления B оказывались неустойчивыми.

На алгебраическом языке условие (7) получило название принципа согласования строчных сумм. Его развитие – обобщенный принцип согласования строчных сумм – заключается в замене (7) на равенство $By = Ay$, $y > 0$, что приводит к единственной вариации в алгоритме – замене вектора e в определении C_k из (5) на произвольный положительный вектор y:

$$C_k y = \Big(L_k G_{k-1}^{-1} U_{k-1} - \overline{L_k G_{k-1}^{-1} U_{k-1}}\Big) y, \;\; y > 0. \tag{8}$$

Возможное развитие рассматриваемых алгоритмов – переход от диагональных матриц C_k к ленточным, для нахождения элементов которых привлекается не единственный "пробный" вектор y, а несколько:

$$C_k y_k^{(q)} = \Big(L_k G_{k-1}^{-1} U_{k-1} - \overline{L_k G_{k-1}^{-1} U_{k-1}}\Big) y_k^{(q)}, \quad q = 1, \ldots, m, \;\; k = 2, \ldots, M. \tag{9}$$

Отметим, что каждый вектор $y_k^{(q)}$ имеет порядок N_k, равный порядку соответствующих матриц D_k, G_k и C_k, а ширина ленты C_k, вычисляемой из условия (9), равна $2m - 1$. Равенства (9) соответствуют следующим соотношениям между исходной и предобуславливающей матрицами при $\theta = 1$:

$$Ay^{(q)} = By^{(q)}, \;\; q = 1, \ldots, m; \quad y^{(q)} = \{y_k^{(q)}, \;\; k = 1, \ldots, M\}, \tag{10}$$

которые будем называть обобщенным принципом компенсации. Некоторые алгоритмические аспекты построения матриц C_k и численные эксперименты с использованием двух или трех пробных векторов $y^{(q)}$ частного вида описаны в работах [2, 6, 7].

В п. 2 приводятся доказанные в [8] основные свойства методов, получаемых из обобщенного принципа компенсации для стилтьесовских систем уравнений. В частности показано, что если m векторов $y_k^{(q)}$, $q = 1, \ldots, m$, являются сильно линейно независимыми (определение дается ниже), то равенства (10) единственным образом определяют симметричные ленточные $(2m-1)$-диагональные матрицы C_k размерности N_k, которые дают матрицу $C = \mathrm{diag}\{C_k\}$, в общем случае не являющуюся положительно-определенной. Для некоторых частных случаев устанавливается положительная определенность предобуславливающей матрицы B. Далее, в п. 3 рассматриваются спектральные характеристики получаемых предобуславливателей и оценки числа обусловленности κ, определяющие скорость сходимости итераций.

2. Алгоритмы обобщенной компенсации

Итак, остановимся на алгоритме неполной факторизации, базирующемся, во-первых, на применении в определениях G_k и C_k из (4), (10) ленточных аппроксимаций матриц согласно (6) и, во-вторых, на использовании в (10) некоторого числа $m > 1$ пробных векторов $y^{(q)}$, $q = 1, \ldots, m$.

Обозначая через Y_k прямоугольную $N_k \times m$ матрицу, столбцы которой суть векторы $y_k^{(q)}$, перепишем соотношение (9) в матричной форме

$$C_k Y_k = R_k Y_k = V_k, \quad k = 2, \ldots, M, \tag{11}$$

где $R_k = L_k G_{k-1}^{-1} U_{k-1} - (L_k G_{k-1}^{-1} U_{k-1})^{(p)}$, а V_k – прямоугольная матрица той же структуры, что и Y_k (ее столбцы суть $v_k^{(q)} = R_k y_k^{(q)}$). Рекуррентные соотношения (4) тоже перепишем соответственно:

$$G_1 = D_1, \quad G_k = D_k - (L_k G_{k-1}^{-1} U_{k-1})^{(p)} - \theta C_k, \;\; k = 2, \ldots, M. \tag{12}$$

Реализация формул (11), (12) сталкивается с двумя нетрадиционными алгебраическими проблемами. Первая заключается в нахождении ленточной части матричного произведения, один из сомножителей которого есть матрица, обратная к ленточной. Алгоритм решения этой задачи, требующий выполнения $O(Nm^2)$ операций, описан в [1, 3].

Вторая задача, на которой останавливаемся ниже, заключается в нахождении ленточной матрицы C_k, удовлетворяющей условию (11). Здесь необходимо выделить три главных момента: построение рекуррентных соотношений для элементов матрицы $C = \{c_{i,j}\}$ (индексы k для краткости опускаем), определение условий на векторы $y^{(q)}$, достаточных для однозначной разрешимости данной задачи, и выяснение вопроса о симметричности получаемой матрицы C при наличии этого свойства у матрицы R.

Представим искомую матрицу C в виде

$$\begin{pmatrix} c_{1,1} & \cdots & c_{1,m} & & & & & \\ \vdots & \ddots & & c_{2,m+1} & & & \mathbf{0} & \\ c_{1,m} & & \ddots & & & & & \\ & c_{2,m+1} & & \ddots & & & & \\ & & & & c_{N-m+1,N-m+1} & \cdots & c_{N-m+1,N} \\ & & & & \vdots & & \vdots \\ & \mathbf{0} & & & c_{N,N-m+1} & \cdots & c_{N,N} \end{pmatrix}. \tag{13}$$

Предполагаем пока, что матрица C является “квазисимметричной”, т.е. ее элементы удовлетворяют условию $c_{i,j} = c_{j,i}$, если только $i, j \le N - m$. Другими словами, несимметричной предполагается только главная подматрица $C_m = \{c_{i,j};\;\; i,j \ge N - m + 1\}$ порядка m, находящаяся в правом нижнем углу матрицы C. Данное представление выбрано из соображения, что при этом количество неизвестных $c_{i,j}$ равно числу Nm уравнений системы

$$Cy^{(q)} = v^{(q)}, \quad v^{(q)} = Ry^{(q)}, \quad q = 1, \ldots, m, \tag{14}$$

где N – порядок векторов $y^{(q)}$, $v^{(q)}$ и искомой матрицы C.

Обозначим через c_l вектор-столбец порядка m, элементы которого для $l = 1, \ldots, N-m$ суть элементы $c_{l,l}, c_{l,l+1}, \ldots, c_{l,l+m-1}$ l-й строки матрицы C, находящиеся в ее верхней треугольной части, а для $l = N - m + 1, \ldots, N$ – представляют собой элементы $c_{l,N-m+1}, c_{l,N-m+2}, \ldots, c_{l,N}$ l-й строки матрицы C, расположенные в последних m ее столбцах. Через Y_l обозначим квадратную подматрицу порядка m, транспонированную к матрице, которая состоит из строк прямоугольной матрицы Y, с номерами от l до $l + m - 1$ включительно. Тогда уравнения (14) переписываются в следующей форме:

$$Y_1 c_1 = v_1;$$

$$Y_l c_l = v_l - c_{1,l} y_1 - \ldots - c_{l-1,l} y_{l-1}, \quad l = 2, \ldots, m - 1;$$

$$Y_l c_l = v_l - c_{l-m+1,l} y_{l-m+1} - \ldots - c_{l-1,l} y_{l-1}, \quad l = m, \ldots, N - m;$$

$$Y_{N-m+1}c_l = v_l - c_{l-m+1,l}y_{l-m+1} - \ldots - c_{N-m,l}y_{N-m}, \quad l = N-m+1,\ldots,N-1;$$
$$Y_{N-m+1}c_N = v_N, \tag{15}$$

где y_l и v_l обозначают вектор-столбцы m-го порядка с элементами из l-х строк матриц Y и V соответственно.

Обозначая через w_l правые части в уравнениях (15), сведем нахождение неизвестных векторов c_l к последовательному решению систем $Y_l c_l = w_l$, $l = 1,\ldots,N$, где компоненты w_l выражаются рекуррентно через уже вычисленные значения компонент векторов c_k, $k < l$. Очевидно, что для однозначной разрешимости задачи (15) при выбранной "квазисимметричной" структуре матрицы C требуется невырожденность всех матриц Y_l.

Определение 1. Прямоугольную $N \times m$-матрицу Y будем называть матрицей сильного ранга m, если все ее подматрицы Y_l m-го порядка (состоящие из последовательно расположенных строк $y_l, y_{l+1},\ldots,y_{l+m-1}$ матрицы Y) являются невырожденными.

Введенное понятие можно перефразировать в терминах свойств векторов $y^{(q)}$, $q = 1,\ldots,m$, – столбцов матрицы Y.

Определение 2. Векторы $y^{(q)}$ N-го порядка будем называть сильно (или равномерно) линейно независимыми, если они образуют прямоугольную $N \times m$-матрицу ($m < N$) сильного ранга m.

Напомним, что для "обычной" линейной независимости набора векторов $y^{(q)}$, $q = 1,\ldots,m$, необходимым и достаточным условием является наличие у матрицы Y хотя бы одного отличного от нуля минора m-го порядка (ранг матрицы Y равен m). При этом Y называется матрицей полного ранга.

Определяемая соотношениями (14), (15) матрица C в блочном представлении второго порядка

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_m \end{pmatrix} \tag{16}$$

обладает свойствами "квазисимметричности" в смысле выполнения условий

$$C_{11} = C'_{11}, \quad C_{21} = C'_{12}. \tag{17}$$

Полученный результат можно сформулировать следующим образом.

Теорема 1. *Если векторы $y^{(q)}$, $q = 1,\ldots,m$, являются сильно линейно независимыми, то ленточная матрица C вида* (13) *со свойствами* (16), (17) *определяется однозначно из соотношений* (14), (15) *для любого набора векторов $v^{(q)}$.*

Очевидно, что сильно линейно независимые векторы $y^{(q)}$ можно строить многими способами. Один из них состоит, например, в выборе некоторой невырожденной матрицы Y_1 со строками $y_1, y_2,\ldots,y_m$ и в определении остальных строк периодически: $y_l = y_{l-m}$ при $l > m$. В работе [8] доказана

Теорема 2. *Пусть выполняются условия теоремы 1, а векторы $v^{(q)}$ определяются соотношениями* (14) *с симметричной матрицей R. Тогда матрица C является симметричной.*

Теоремы 1, 2 устанавливают фактически минимальные условия существования и единственности алгоритма обобщенной компенсации, т.е. вычислимости симметричной матрицы C по условиям (9).

Пусть теперь H_k – квадратная невырожденная матрица порядка m, а Y_k – прямоугольная $N_k \times m$-матрица сильного ранга m. Тогда матрица

$$\overline{Y}_k = Y_k H_k \tag{18}$$

также имеет сильный ранг m. Отсюда следует, что матричное уравнение

$$C_k \overline{Y}_k = R_k \overline{Y}_k \tag{19}$$

однозначно разрешимо, и его решение совпадает с матрицей C_k из уравнения (11). Полученное утверждение можно сформулировать (опуская для краткости индекс k) следующим образом:

Теорема 3. *При условиях теоремы 1 матрица C инвариантна относительно любого невырожденного преобразования векторов $y^{(q)}$ вида* (18).

Другими словами, если $\overline{y}^{(q)}$, $q = 1, \dots, m$, – столбцы матрицы $\overline{Y}$, а h_{pq} – элементы матрицы H, то матричное соотношение $\overline{Y} = YH$ эквивалентно линейному преобразованию пробных векторов $\overline{y}^{(q)} = h_{1q} y^{(1)} + \ldots + h_{mq} y^{(m)}$. Таким образом, если предобуславливающая матрица B построена по формулам (3), (4), (9) при некотором наборе векторов $y^{(q)}$, $q = 1, \dots, m$, то ее вид не изменится при их любом невырожденном линейном преобразовании.

Для обоснования обобщенного принципа компенсации главный вопрос заключается в доказательстве положительной определенности матриц G и B. В основе этого лежит

Лемма 1 [8]. *Пусть $A = D - L - U$ – блочно-трехдиагональная матрица Стилтьеса ($L = U'$), удовлетворяющая условиям*

$$D_1 e_1 \geq U_1 e_2, \qquad D_k e_k \geq L_k e_{k-1} + U_k e_{k+1}, \quad k = 2, \dots, M-1, \tag{20}$$

$$D_k e_k > L_k e_{k-1}, \quad k = 2, \dots, M, \tag{21}$$

где e_k – единичный вектор размерности N_k. Тогда, если вычисленные по формулам (11), (12) *матрицы G_k имеют неположительные внедиагональные элементы, то они являются матрицами Стилтьеса.*

В случае стилтьесовости матрицы G_{k-1} для неположительности внедиагональных элементов матрицы G_k достаточно неотрицательности внедиагональных элементов матрицы $L_k G_{k-1}^{-1} U_{k-1} + C_k$ при $\theta \in [0, 1]$, и в частности, неотрицательности внедиагональных элементов матрицы C_k.

Для случая $m = 2$ рассмотрим векторы

$$y_k^{(1)} = e_k, \quad y_k^{(2)} = \{y_{k,1}^{(2)}, \dots, y_{k,N_k}^{(2)}\}, \quad k = 2, \dots, M, \tag{22}$$

где $y_k^{(2)}$ – строго монотонный вектор, т.е. при $i = 2, \dots, N_k$ либо $y_{k,i}^{(2)} > y_{k,i-1}^{(2)}$, либо $y_{k,i}^{(2)} < y_{k,i-1}^{(2)}$. Как нетрудно видеть, такие векторы сильно линейно независимы. Опуская ради краткости индекс k, можно установить следующие свойства внедиагональных элементов матрицы C, полученной из условий (14), (22).

Лемма 2. *Если симметричная матрица R имеет неотрицательные элементы, то у определяемой по формулам* (14), (22) *матрицы C внедиагональные элементы также неотрицательны.*

Таким образом, из лемм 1 и 2 следует стилтьесовость матрицы G в случае $m = 2$ для векторов (22). Тем самым установлена положительная определенность матрицы G, а в силу представления $B = (G - U')G^{-1}(G - U)$, и матрицы B, т.е. справедлива

Теорема 4. *При выполнении условий леммы 1 относительно матрицы A и использовании двух пробных векторов, один из которых имеет постоянные компоненты, а второй – строго монотонные, предобуславливающая матрица B положительно определена.*

Рассмотрим теперь другую пару пробных векторов:

$$y_k^{(1)} = e_k, \quad y_k^{(2)} = (1, 0, 1, 0, \ldots)', \quad k = 2, \ldots, M, \tag{23}$$

являющихся сильно линейно независимыми.

Для анализа предобуславливающей матрицы (3) условие обобщенной компенсации (10) сформулируем с помощью следующего представления для матриц G_k:

$$G_1 = D_1, \quad G_k = D_k - (1-\theta)L_k\left(G_{k-1}^{-1}\right)^{(p)}U_{k-1} - \theta X_k, \; k = 2, \ldots, M, \tag{24}$$

где X_k – трехдиагональные матрицы, найденные из условий

$$X_k y_k^{(q)} = Q_k y_k^{(q)}, \quad q = 1, 2, \tag{25}$$

а матрица Q_k определяется равенством $Q_k = L_k G_{k-1}^{-1} U_{k-1} = \{q_{ij}^{(k)}\}$. Очевидно, что матрицы X_k связаны с C_k из (9) соотношениями $X_k = C_k + L_k(G_{k-1}^{-1})^{(p)}U_{k-1}$, обеспечивающими совпадение определяемых из (12) и (24) матриц G_k.

Рассмотрим внедиагональные элементы матрицы X_k, опуская далее для краткости индекс k. При этом сделаем дополнительное предположение, что элементы матрицы Q_k убывают при их удалении от главной диагонали:

$$q_{ij} \geq q_{i,j+1} \geq q_{i-1,j+1} \geq 0, \quad j \geq i. \tag{26}$$

Замечание 1. Предположение о справедливости неравенств (26) является естественным, поскольку, например, они выполняются, если матрица G_{k-1} стилтьесова, а элементы матриц L_k и U_{k-1} являются постоянными.

Лемма 3. *Если элементы симметричной матрицы Q удовлетворяют условиям* (26), *то у определяемой по формулам* (23)–(25) *трехдиагональной матрицы X внедиагональные элементы неотрицательны.*

Из неотрицательности внедиагональных элементов матриц X_k следует неположительность внедиагональных элементов матриц G_k, определяемых формулами (12), (24) при $0 \leq \theta \leq 1$, и в силу леммы 1 – их стилтьесовость. Отсюда уже непосредственно следует результат, аналогичный теореме 4.

Теорема 5. *При выполнении условий леммы 3 предобуславливающая матрица B, определяемая формулами* (3)–(5), (23), *положительно определена при* $0 \leq \theta \leq 1$.

Рассмотрим случай $m = 3$ для следующего набора пробных векторов:

$$y_k^{(1)} = e_k, \quad y_k^{(2)} = y_k, \quad y_k^{(3)} = z_k = (1,0,1,0,\ldots)', \quad k = 2,\ldots,M, \tag{27}$$

где y_k – строго монотонный вектор, т.е. при $i = 2,\ldots,N_k$ либо $y_{k,i} > y_{k,i-1}$, либо $y_{k,i} < y_{k,i-1}$. Определим пятидиагональную матрицу X_k по формулам, аналогичным (25), используя при этом три вектора из (27):

$$X_k y_k^{(q)} = Q_k y_k^{(q)}, \quad q = 1,2,3. \tag{28}$$

Как и раньше, индекс k в дальнейшем будем опускать.

Лемма 4. *Если элементы симметричной матрицы Q удовлетворяют условиям* (26), *то у определяемой по формулам* (27), (28) *пятидиагональной матрицы X внедиагональные элементы неотрицательны.*

Аналогично теореме 5, из лемм 1, 4 имеем следующее утверждение:

Теорема 6. *При выполнении условий леммы 4 предобуславливающая матрица B, определяемая формулами* (3), (27), (28), *положительно определена при $0 \le \theta \le 1$.*

Замечание 2. В соответствии с теоремой 3, теоремы 4–6 справедливы и при использовании любых пробных векторов, получаемых из исходных невырожденным линейным преобразованием.

Замечание 3. Отметим следующий важный факт. Полученные в частных случаях теоремы о положительной определенности предобуславливающих матриц B не могут быть обобщены для произвольных сильно линейно независимых векторов даже в случае $m = 2$.

3. Анализ скорости сходимости итераций

Для верхней оценки κ важна следующая

Теорема 7 [3]. *Пусть A, G, D – стилтьесовские матрицы, предобуславливающая матрица B определяется соотношениями* (3), (4), (9) *и пусть выполняются условия*

$$L = U' \ge 0, \quad Ay \ge 0, \quad By \ge (1-\tau)Ay,$$
$$\tau = \max_i\{(G^{-1}Uy)_i/y_i\} < 1, \quad \mathrm{offdiag}[(1+\tau)G - D] \le 0,$$

где y – некоторый вектор с положительными компонентами, $\mathrm{offdiag}(\cdot)$ *– внедиагональная часть матрицы. Тогда справедливо неравенство $\lambda(B^{-1}A) \le (1-\tau)^{-1}$.*

В некоторых характерных для практических задач случаях величину τ можно оценить с помощью следующего утверждения, несложно доказываемого по индукции.

Лемма 5. *Пусть A – блочно-диагональная стилтьесовская матрица, блоки которой удовлетворяют соотношениям*

$$L_k e_{k-1} \ge g_k U_k e_{k+1}, \quad g_k > 0, \quad k = 2,\ldots,m-1,$$
$$D_k e_k \ge t_k(L_k e_{k-1} + U_k e_{k+1}), \quad t_k \ge 1, \quad k = 1,\ldots,m-1,$$

где e_k *– векторы порядка* N_k *с единичными компонентами. Тогда для* $\tau_k = \|G_k^{-1}U_k\|_\infty$, *где матрицы* G_k *являются стилтьесовскими и определяются соотношениями* (4), (9) *при* $y_k^{(q)} = e_k$, *справедливы неравенства* $\tau_k \le [t_k + g_k(t_k - \tau_{k-1}]^{-1}$, $k = 2, \ldots, m$.

Отсюда, используя обозначения $\tau_1 = \|D_1^{-1}U_1\|_\infty < 1$, $g_1 = \tau_1^{-1} - 1$ и полагая $\tau = \max\limits_k \{\tau_k\}$, $t_k \equiv 1$ (кроме $t_1 > 1$), получаем следующий результат:

Теорема 8. *В условиях теоремы 7 и леммы 5 для собственных чисел матрицы* $B^{-1}A$ *справедливо неравенство*

$$\lambda(B^{-1}A) \le \max_k \left\{ \frac{1 + g_1 + g_1 g_2 + \ldots + g_1 \times \ldots \times g_k}{g_1 \times \ldots \times g_k} \right\}.$$

В частности, для случая $g_k \ge 1$ имеет место неравенство

$$\lambda(B^{-1}A) \le (1 - \tau)^{-1} \le M + 1. \tag{29}$$

Отметим, что рекуррентные соотношения для τ_k из леммы 5 с помощью обозначений $l_k = u_k g_k$, $u_k = [t_k(1 + g_k)]^{-1}$ можно переписать в виде неравенств

$$\tau_1 \le \frac{1}{t_1}, \quad \tau_k \le \frac{u_k}{1 - l_k \tau_{k-1}}, \quad k = 2, \ldots, m - 1,$$

откуда при дополнительных предположениях

$$l_k u_{k-1} \le l_{k+1} u_k \le \frac{1}{4}, \quad k = 2, \ldots, m - 1,$$

по аналогии с [3] получаем оценки

$$\tau_k \le \frac{2u_k}{1 + \sqrt{1 - 4l_{k+1}u_k}}, \quad \tau_k \le 2u_k \frac{k}{k+1}, \quad k = 2, \ldots, m - 1,$$

которые могут быть эффективно использованы при некоторых характерных свойствах элементов исходной матрицы A. Примеры получения соответствующих оценок скорости сходимости итераций при решении диффузионно-конвективных или параболических уравнений описаны в [3].

Нижняя оценка $\lambda(B^{-1}A)$ может быть установлена следующим образом:

Лемма 6. *Пусть предобуславливающая матрица* B *определяется соотношениями* (3), (4), (9), *причем* A, D, G *– стилтьесовские матрицы. Пусть, кроме того, выполняются условия*

$$Ae \ge (1 - \tau)Be, \quad \mathrm{offdiag}[\tau D + \theta(1 - \tau)C] \le 0, \tag{30}$$

где $0 < \tau < 1$. *Тогда справедливо неравенство*

$$\lambda(B^{-1}A) \ge 1 - \tau. \tag{31}$$

Доказательство. Из (4) следует представление $B = A + R - \theta C$, в соответствии с которым имеют место равенства

$$Q = A - (1 - \tau)B = \tau A - (1 - \tau)R + \theta(1 - \tau)C = \tau D + \theta(1 - \tau)C - \tau L - \tau U - (1 - \tau)R.$$

В силу стилтьесовости матрицы G элементы матрицы R неотрицательны. Из второго условия в (30) следует неположительность внедиагональных элементов матрицы Q, а из первого – наличие диагонального преобладания. Таким образом, матрица Q положительно полуопределена, и следовательно, имеет место оценка (31). □

Для оценки числа обусловленности κ с помощью неравенств (29) или (31) зачастую с помощью соотношения

$$B = A + (1-\theta)C - \Big[C - (LG^{-1}U - (LG^{-1}U)^{(p)}\Big] \tag{32}$$

удается дополнительно установить, что единица является нижней или верхней границей спектра $\lambda(B^{-1}A)$. Например, если y – единственный и положительный пробный вектор, а матрица C – диагональная, то при $\theta = 1$ из (32) следует $\lambda(B^{-1}A) \geq 1$ и $\kappa \leq (1-\tau)^{-1}$. Вопросы оптимизации параметра θ (с точки зрения минимизации κ) рассмотрены для такого случая в [3].

При использовании двух пробных векторов оценку κ можно вывести с помощью утверждения, являющегося обобщением теоремы 8.23 из [2].

Лемма 7. *Пусть $R = \{r_{ij}\}$ – симметричная матрица с неотрицательными внедиагональными элементами r_{ij}, а $C = \{c_{ij}\}$ – трехдиагональная матрица, определяемая из условий* (9) *с двумя пробными векторами* (22). *Тогда справедливо неравенство*

$$(Cu, u) \leq (Ru, u) \tag{33}$$

для любого вектора u.

Доказательство. Обозначим $y = Te = \{y_i = y^{(2)}_{k,i}\}$, $T = \operatorname{diag}\{y_i\}$. Из соотношений (14), (22) следует, что $(CTe)_i - y_i(Ce)_i = (Se)_i$, $i = 1, \ldots, N_k$, где $S = RT - TR$ – кососимметричная матрица, элементы которой выражаются через элементы матрицы $R = \{r_{ij}\}$ по формуле: $s_{ij} = (y_j - y_i)r_{ij}$. Отсюда следуют уравнения для внедиагональных элементов матрицы C:

$$(y_2 - y_1)c_{1,2} = (Se)_1, \quad (y_{i+1} - y_i)c_{i,i+1} - (y_i - y_{i-1})c_{i-1,i} = (Se)_i, \quad i = 2, \ldots, N_k - 1.$$

Частичное суммирование этих уравнений приводит к равенствам

$$c_{i,i+1} = \frac{1}{y_{i+1} - y_i} \sum_{j=1}^{i} (Se)_j, \quad i = 1, \ldots, N_k - 1.$$

Поскольку матрица S кососимметрична, то в последней сумме присутствуют элементы только ее верхней треугольной части. Отсюда получаем

$$c_{i,i+1} = \frac{1}{y_{i+1} - y_i} \sum_{j=1}^{i} \sum_{l=i+1}^{N_k} (y_l - y_j) r_{jl}, \quad i = 1, \ldots, N_k - 1. \tag{34}$$

Далее, для квадратичной формы симметричной матрицы R нетрудно получить представление

$$(Ru, u) = \sum_{i=1}^{N} u_i^2 \sum_{j=1}^{N} r_{ij} - \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} r_{ij}(u_i - u_j)^2.$$

Поскольку по условию леммы $\sum_{j=1}^{N} r_{ij} = \sum_{j=1}^{N} c_{ij}$, справедливо равенство

$$((R-C)u,u) = \sum_{i=1}^{N-1} c_{i,i+1}(u_{i+1}-u_i)^2 - \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} r_{ij}(u_j-u_i)^2. \tag{35}$$

Положим для определенности $y^{(2)}_{i+1} > y^{(2)}_i$. Из неравенства Коши–Буняковского следует, что при $j > i$

$$(u_j-u_i)^2 = \left(\sum_{k=i}^{j-1}\sqrt{y_{k+1}-y_k}\,\frac{u_{k+1}-u_k}{\sqrt{y_{k+1}-y_k}}\right)^2 \le (y_j-y_i)\sum_{k=i}^{j-1}\frac{(u_{k+1}-u_k)^2}{y_{k+1}-y_k}.$$

При $y^{(2)}_{i+1} < y^{(2)}_i$ достаточно заменить $\sqrt{y_{k+1}-y_k}$ на $\sqrt{y_k-y_{k+1}}$. Используя это неравенство и производя несложные преобразования сумм, получим

$$\begin{aligned}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} r_{ij}(u_j-u_i)^2 &\le \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(y_j-y_i)r_{ij}\sum_{k=i}^{j-1}\frac{(u_{k+1}-u_k)^2}{y_{k+1}-y_k}\\ &= \sum_{i=1}^{N-1}(u_{i+1}-u_i)^2\frac{1}{y_{i+1}-y_i}\sum_{j=1}^{i}\sum_{k=i+1}^{N}(y_k-y_j)r_{jk}.\end{aligned}$$

Подстановка данного неравенства в (35) и использование формулы (34) для $c_{i,i+1}$ приводит к положительной полуопределенности матрицы $R-C$. □

Теперь можно сформулировать следующий результат:

Теорема 9. *Пусть выполняются условия лемм 6 и 7. Тогда при определении предобуславливающей матрицы B по постоянному и строго монотонному векторам при $\theta = 1$ справедлива оценка $\kappa \le (1-\tau)^{-1}$ для числа обусловленности матрицы B.*

Для доказательства, в силу леммы 6, достаточно показать справедливость соотношения $\lambda(B^{-1}A) \le 1$. Последнее же следует из равенства $B = A + R - C$ при $\theta = 1$ и леммы 7.

Для получения более конкретных оценок с учетом свойств исходной задачи необходимы детальные "штучные" исследования.

Приведем пример такого результата для метода с двумя пробными векторами – "постоянным" и "линейным":

$$y^{(1)} = e, \quad y^{(2)} = \{y^{(2)}_{k,i} = i\} \tag{36}$$

в применении к решению системы пятиточечных сеточных уравнений, аппроксимирующих задачу Дирихле для уравнения Пуассона на квадратной сетке с шагом $h = 1/(N+1)$ в прямоугольной области $[0,1]\times[0,(M+1)h]$. В этом случае матрицы N-го порядка L_k, D_k, U_k имеют следующий вид (E – единичная матрица):

$$L_2 = \ldots = L_M = U_1 = \ldots = U_{M-1} = E, \quad D_k = \mathrm{tridiag}\{-1,4,-1\}, \quad k = 1,\ldots,M, \tag{37}$$

а матрицы G_k определяются рекуррентными соотношениями

$$G_1 = D, \quad G_k = D - G^{-1}_{k-1} + R_k - C_k, \;\; k = 2,\ldots,M,$$

где $R_k = G^{-1}_{k-1} - (G^{-1}_{k-1})^{(3)}$.

В этом случае при $\theta = 1$, в силу леммы 7 и неотрицательности внедиагональных элементов матриц R, C, для любого вектора $u = \{u_{k,i} : k = 1,\ldots,M; i = 1,\ldots,N\}$ имеем

$$0 \le ((R-C)u,u) \le c_0 \sum_{k=1}^{M} \sum_{i=1}^{N-1} (u_{k,i+1}-u_{k,i})^2, \tag{38}$$

где c_0 – максимальный из внедиагональных элементов $c_{i,i+1}^{(k)}$ матрицы $C = \mathrm{diag}\{C_k\}$. Поскольку для рассматриваемой матрицы A легко показывается неравенство

$$(Au,u) \ge \sum_{k=1}^{M} \sum_{i=1}^{N-1} (u_{k,i+1}-u_{k,i})^2,$$

то в данном случае, с учетом леммы 7, имеем

$$(Au,u) \le (Bu,u) \le (1+c_0)(Au,u), \tag{39}$$

и для оценки числа обусловленности κ нужно оценить сверху значение c_0.

Обозначая через $g_{j,e}^{(k)}$ элементы матрицы G_k^{-1}, в силу определения $R_k = \{r_{j,l}^{(k)}\}$ имеем равенства $r_{j,l}^{(k)} = g_{j,l}^{(k-1)}$ для $|j-l| \ge 2$ и $r_{j,l}^{(k)} = 0$ при $|j-l| \le 1$.

Отсюда с помощью формулы (34) получаем

$$c_{i,i+1}^{(k)} = \sum_{j=1}^{i} \sum_{l=i+1}^{N} (l-j) r_{j,l}^{k} \le \hat{c}_{i,i+1}^{(k)} = \sum_{j=1}^{i} \sum_{l=i+1}^{N} (l-j) g_{j,l}^{(k-1)}, \quad i = 1,\ldots,N-1,$$

где величина $\hat{c}_{i,i+1}^{(k)}$ может быть определена как

$$\hat{c}_{i,i+1}^{(k)} = \sum_{j=1}^{i} z_{j,j}^{k-1}, \quad z_{j,l}^{(k)} = \sum_{p=1}^{N} (p-j) g_{l,p}^{(k)}. \tag{40}$$

Лемма 8. *Пусть предобуславливающая матрица B определяется соотношениями (3), (4), (9) при $\theta = 1$, $m = 2$ и условиях (36), (37). Тогда справедливы следующие неравенства:*

$$\begin{aligned}
\sum_{l=1}^{N} l g_{j,l}^{(k)} &\le \frac{k}{k+1} j, \quad k = 1,\ldots,N-1, \;\; j = 1,\ldots,N, \\
z_{j,l}^{(k)} &\le \frac{k}{k+1}(l-j) + \frac{j}{k+1}\Big[k g_{l,1}^{(k)} + (k-1) \sum_{l-1=1}^{N} g_{l,l_1}^{(k)} g_{l_1,l}^{(k-1)} + \ldots + \\
&\qquad 2 \sum_{l_1=1}^{N} \sum_{l_{k-2}=1}^{N} g_{l,l_1}^{(k)} g_{l_1,l_2}^{(k-1)} \times \ldots \times g_{l_{k-1},l}^{(1)} \Big], \\
\hat{c}_{i,i+1}^{(k)} &\le \frac{1}{3}(k-1), \quad k = 2,\ldots,M, \;\; i = 1,\ldots,N-1.
\end{aligned} \tag{41}$$

Доказательство. Из очевидного равенства $y_k^{(2)} = G_k^{-1} G_k u_k^{(2)}$, рекуррентного выражения для матрицы G_k и условия компенсации $R_k y_k^{(2)} = C_k y_k^{(2)}$ получаем соотношение

$$y_1^{(2)} = G_1^{-1} D_1 y_1^{(2)}, \quad y_k^{(2)} = G_k^{-1}(D_k - G_{k-1}^{-1}) y_k^{(2)}, \;\; k = 2,\ldots,M. \tag{42}$$

Поскольку $(D_k y_k^{(2)})_j = 2j$ для $j = 1,\ldots,N-1$ и $(D_k y_k^{(2)})_N = 3N+1$, при $k = 1$ справедливо равенство

$$j = 2\sum_{l=1}^{N} l g_{j,l}^{(1)} + (N+1) g_{j,N}^{(1)}. \tag{43}$$

В силу $g_{j,N}^{(1)} > 0$ справедливо неравенство $\sum_{l=1}^{N} l g_{j,l}^{(1)} \le j/2$. Предположим теперь, что для некоторого k выполняется $\sum_{l=1}^{N} l g_{j,l}^{(k-1)} \le (k-1)j/k$.

Аналогично (42) из (41) для произвольного k имеем

$$j = \sum_{l=1}^{N} g_{j,l}^{(k)} \left(2l - \sum_{p=1}^{N} p g_{l,p}^{(k-1)}\right) + (N+1) g_{j,N}^{(k)}.$$

Отсюда по сделанному предположению получаем соотношение

$$j \ge \sum_{l=1}^{N} g_{j,l}^{(k)} \left(2l - \frac{k-1}{k}\right) = \frac{k+1}{k} \sum_{l=1}^{N} l g_{j,l}^{(k)},$$

которое и доказывает первое из неравенств (41).

Далее, умножая каждую компоненту векторного равенства $y_1^{(1)} = G_1^{-1} D_1 y_1^{(1)}$, представляющего собой условие компенсации, на величину j, получаем выражение

$$j = 2\sum_{p=1}^{N} j g_{l,p}^{(1)} + j(g_{l,1}^{(1)} + g_{l,N}^{(1)}), \tag{44}$$

при выводе которого учтено, что $(D_1 y_{(1)}^{(1)})_1 = (D_1 y_1^{(1)})_N = 3$ и $(D_1 y_1^{(1)}))_l = 2$ для $l = 2, \ldots, N-1$. Заменим теперь в (43) индексы j, l соответственно на l, p и почленно вычтем из него (44):

$$l - j = 2\sum_{p=1}^{N} (p-j) g_{l,p}^{(1)} + (N+1-j) g_{1,N}^{(1)} - j g_{l,1}^{(1)}. \tag{45}$$

Отсюда с использованием обозначений (40) приходим к неравенству

$$z_{j,l}^{(1)} \le \frac{1}{2}(l-j) + \frac{j}{2} g_{l,1}^{(1)},$$

которое означает второе из неравенств (40) при $k = 1$. Проводя аналогичные сделанным при выводе (45) преобразования (с заменой $g_{l,p}^{(1)}$ на $g_{l,p}^{(k)}$, $k > 1$), получаем соотношения

$$l - j = \sum_{p=1}^{N} g_{l,p}^{(k)} \left[2(p-j) - \sum_{q=1}^{N} (g-j) g_{p,q}^{(k-1)}\right] + (N+1-j) g_{l,N}^{(k)} - j g_{l,1}^{(k)}. \tag{46}$$

Предположим теперь, что второе неравенство в (41) выполняется при замене индекса k на $k-1$, что в результате подстановки его в (46) дает соотношение

$$l - j \ge \sum_{p=1}^{N} g_{l,p}^{(k)} \left[2(p-j) - \frac{k-1}{k}(p-j) - \frac{k-1}{k} g_{p,1}^{(k-1)} - \ldots - \right.$$
$$\left. \frac{1}{k} j \sum_{p_1=1}^{N} \cdots \sum_{p_{k-2}=1}^{N} g_{p,p_1}^{(k-1)} \times \ldots \times g_{p_{k-2},1}^{(1)}\right] - j g_{l,1}^{(k)},$$

из которого следует доказательство неравенства для $z_{j,l}^{(k)}$ в (41).

Установим теперь последнее утверждение леммы. В силу легко проверяемого соотношения $\sum_{j=1}^{i}\sum_{l=1}^{i}(l-j)g_{j,l}^{(k-1)}=0$ и обозначений (40), можно записать

$$\hat{c}_{i,i+1}^{(k)}=\sum_{j=1}^{i}z_{j,j}^{(k-1)}\le\sum_{j=1}^{N}z_{j,j}^{(k-1)} \tag{47}$$

Применяя в (47) второе неравенство (41) при $l=j$ и меняя порядок суммирования, получаем

$$\hat{c}_{i,i+1}^{(k)}\le\frac{1}{k}\Big[(k-1)\sum_{j=1}^{N}jg_{j,1}^{(k-1)}+(k-2)\sum_{l_1=1}^{N}g_{l_1,1}^{(k-2)}\sum_{j=1}^{N}jg_{j,l_1}^{(k-1)}+\ldots+$$
$$2\sum_{l_{k-3}=1}^{N}g_{l_{k-3},1}^{(2)}\times\ldots\times\sum_{j=1}^{N}jg_{j,l_1}^{(k-1)}+\sum_{l_{k-2}=1}^{N}g_{l_{k-2},1}^{(1)}\times\ldots\times\sum_{j=1}^{N}jg_{j,l_1}^{(k-1)}\Big]. \tag{48}$$

Используя далее первое из неравенств (41), приходим последовательно к оценкам

$$\begin{aligned}
\sum_{j=1}^{N}jg_{j,1}^{(k-1)}&\le\frac{k-1}{k},\\
\sum_{l_1=1}^{N}g_{l_1,1}^{(k-2)}\sum_{j=1}^{N}jg_{j,l_1}^{(k-1)}&\le\frac{k-1}{k}\sum_{l_1=1}^{N}l_1g_{l_1,1}^{(k-2)}\le\frac{k-2}{k},\\
\sum_{l_{k-2}=1}^{N}g_{l_{k-2},1}^{(1)}\times\ldots\times\sum_{j=1}^{N}jg_{j,l_1}^{(k-1)}&\le\frac{k-1}{k}\sum_{l_{k-2}=1}^{N}g_{l_{k-2},1}^{(1)}\times\ldots\times\sum_{l_1=1}^{N}l_1g_{l_1,l_2}^{(k-2)}\le\ldots\\
&\le\frac{2}{k}\sum_{l_{k-2}=1}^{N}l_{k-2}g_{l_{k-2},1}^{(1)}\le\frac{1}{k},
\end{aligned}$$

после подстановки которых в (48) имеем итоговый результат:

$$\hat{c}_{i,i+1}^{(k)}\le\frac{1}{k^2}\sum_{n=1}^{k-1}n^2=\frac{1}{3}(k-1)(1-(2k)^{-1}),\quad k=2,\ldots,M.$$

Отсюда получаем последнее неравенство в (41). □

Непосредственно из (39), (41) устанавливается следующий результат:

Теорема 10. *В условиях леммы 8 для числа обусловленности матрицы $B^{-1}A$ выполняется оценка*

$$\kappa\le(M+2)/3. \tag{49}$$

Отметим, во-первых, что эта оценка не зависит от N, и во-вторых, что она лучше известных аналогичных результатов для других методов неполной факторизации [2, 3].

Сделаем теперь следующее любопытное замечание. Если для данной модельной задачи в качестве второго пробного вектора брать не “линейный”, а строго монотонный вектор, компоненты которого удовлетворяют неравенствам

$$\frac{y_{k,l} - y_{k,j}}{l - j} \le \alpha(y_{k,i+1} - y_{k,i}), \quad l \ge i + 1, \ \ j \ge i, \ \ k = 1, \dots, M,$$

(легко показать, что при этом обязательно $\alpha \ge 1$), то в силу формулы (34) вместо (41) получаем $\hat{c}^k_{i,i+1} \le \alpha(k - 1)/3$, откуда для числа обусловленности следует оценка

$$\kappa \le \frac{\alpha(M - 1) + 3}{3}. \tag{50}$$

Таким образом, в смысле оценок (49), (50) “линейный” вектор оказывается наилучшим из всех строго монотонных пробных векторов $y^{(2)}$.

Анализ предложенных методов проведен в условиях точных вычислений. Учет приближенного представления матриц и погрешностей округлений требует, строго говоря, перехода к исследованию “возмущенных” методов неполной факторизации и сопряженных градиентов, что представляет собой самостоятельную малоизученную проблему. Для оценки же практической эффективности описанных алгоритмов необходимо рассмотреть их устойчивость к ошибкам округлений. В принципе понятно, что выбор пробных векторов должен обеспечивать не очень сильный рост чисел обусловленности матриц Y_l и норм векторов y_l, c_l, однако эти вопросы требуют специальных исследований.

Для иллюстрации возникающих вопросов и подходов проведем прямой анализ накопления погрешностей округлений при реализации формул (15) без учета возможных возмущений векторов v_l и матриц Y_l.

Обозначая через $\tilde{c}_{i,j} = c_{i,j} + \delta_{i,j}$, $\tilde{c}_l = c_l + \delta_l$ вычисляемые приближенные величины, для векторов ошибок $\delta_l = \{\delta_{l,j},\, j = l, \dots, l + m - 1\}$ можно записать

$$\begin{aligned}
\delta_1 &= \varepsilon_1;\\
\delta_l &= -Y_l^{-1}(\delta_{1,l}y_1 + \ldots + \delta_{l-1,l}y_{l-1}) + \varepsilon_l, \quad l = 2, \dots, m - 1;\\
\delta_l &= -Y_l^{-1}(\delta_{l-m+1,l}y_{l-m+1} \ldots \delta_{l-1,l}y_{l-1}) + \varepsilon_l, \quad l = m, \dots, N - m;\\
\delta_l &= -Y_{N-m+1}^{-1}(\delta_{l-m+1,l}y_{l-m+1} + \ldots + \delta_{N-m,l}y_{N-m}) + \varepsilon_l, \quad l = N - m + 1, \dots, N - 1;\\
\delta_N &= \varepsilon_N.
\end{aligned} \tag{51}$$

Здесь ε_l – векторы m-го порядка, компоненты которых суть суммарная погрешность арифметических операций, выполняемых при вычислении одного вектора c_l по формулам (15). Из (51) получаем следующие неравенства для равномерных (кубических) норм:

$$\bar{\delta}_l \le \|Y_l^{-1}\| \cdot \|Y\| B \bar{\delta}_{l-1} + \varepsilon e, \quad l = 2, \dots, N - 1, \tag{52}$$

где $\varepsilon = \max\limits_l\{\|\varepsilon_l\|\}$, $\bar{\delta}_l = (\|\delta_l\|, \dots, \|\delta_{l+m-1}\|)'$, $\|Y\| = \max\limits_l\{\|y_l\|\}$ – максимальная сумма модулей элементов строки прямоугольной матрицы Y, B – матрица Фробениуса вида

$$\begin{bmatrix} 1 & 1 & 1 & \ldots & 1 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}.$$

Вводя далее обозначение $\mu(Y) = \max\limits_l \|Y_l^{-1}\| \cdot \|Y\|$ и учитывая легко проверяемую оценку для нормы степеней матрицы (46): $\|B^k\| \le m\, 2^{k-1}$, приходим к следующему неравенству:

$$\|\bar{\delta}_l\| \le \varepsilon(1 + 2m\mu + 4m\mu^2 + \ldots + 2^{l-1}m\mu^{l-1}). \tag{53}$$

Определение 3. Величину $\mu(Y)$ будем называть числом обусловленности системы векторов $y^{(q)}$, $q = 1, \ldots, m$, или прямоугольной матрицы Y.

Как видно из (53), ошибка округления может быстро накапливаться с ростом l, и выбираемые системы векторов (или матрицы Y) должны быть хорошо обусловлены. Под этим подразумевается, что значение $\mu(Y)$ относительно невелико и, в частности, не зависит от порядка векторов $y^{(q)}$. Как видно из предыдущего анализа, рассмотренные примеры систем пробных векторов как раз удовлетворяют этому свойству.

Замечание 4. Введенное понятие числа обусловленности системы векторов $\mu(Y) \geq 1$ базируется на обычном определении числа обусловленности квадратной матрицы. Отметим, что в работе [9] дано определение числа обусловленности базиса Крылова другим образом, допускающим, в частности, его обращение в нуль.

Литература

[1] **Il'in V.P.** Iterative Incomplete Factorization Methods. — Singapore: World Scientific Publishing Co., 1992.

[2] **Axelsson O.** Iterative Solution Methods. — Cambridge University Press, 1994.

[3] **Ильин В.П.** Методы неполной факторизации для решения алгебраических систем. — М.: Наука, 1995.

[4] **Stone H.L.** Iterative solution of implicit approximations of multidimensional partial differential equations // SIAM J. Numer. Anal. — 1968. — Vol. 5. — P. 530–558.

[5] **Lin Avi.** Towards generalization and optimization of implicit methods // Intern. J. for Numer. Methods in Fluids. — 1985. — Vol. 5. — P. 357–380.

[6] **Axelsson O., Polman B.** Block preconditioning and domain decomposition methods. I. — Nijmegen: Catholic University, 1987. — (Report; 8735).

[7] **Axelsson O., Polman B.** Block preconditioning and domain decomposition methods. II. — Nijmegen: Catholic University, 1988. — (Report; 8807).

[8] **Il'in V.P., Laevskii K.Yu.** Generalized compensation principle in incomplete factorization methods // Rus. J. Num. Anal. Math. Mod. — 1997. — Vol. 12, № 5. — P. 399–420.

[9] **Carpraux J.F., Godunov S.K., Kuznetsov S.V.** Condition number of the Krylov bases and subspaces // Linear algebra and its applications. — 1996. — Vol. 248. — P. 137–160.

В.П. Ильин,
Институт вычислительной математики
и математической геофизики СО РАН,
просп. Акад. М.А. Лаврентьева, 6,
Новосибирск, 630090
E-mail: ilin@sscc.ru

К.Ю. Лаевский,
Eindhoven University of Technology,
5600 MB Eindhoven,
The Netherlands
E-mail: laevsky@tue.win.nl

Статья поступила
26 января 1998 г.
Переработанный вариант
22 апреля 1998 г.