ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ, 2001, том 37, № 7, с. 899–904

ЧИСЛЕННЫЕ МЕТОДЫ

УДК 519.632

ДВУХУРОВНЕВЫЕ ПРЕДОБУСЛОВЛИВАТЕЛИ СЕТОЧНЫХ УРАВНЕНИЙ В МЕТОДАХ ПСЕВДОДВОЙСТВЕННЫХ НАПРАВЛЕНИЙ

© 2001 г. С. С. Белявский, С. Г. Мулярчик

Введение. В настоящей работе рассматриваются двухуровневые методы решения сеточных уравнений, аппроксимирующих на прямоугольных сетках (или топологически инвариантных прямоугольным) двумерные краевые задачи для уравнения эллиптического типа

$$\nabla \cdot p\nabla u + q\nabla u = f \quad \text{в} \quad \Omega, \tag{1}$$

$$ru + su_n = g \quad \text{на} \quad \partial\Omega, \tag{2}$$

где Ω – двумерная область с кусочно-гладкой границей. После дискретизации на основной сетке с использованием метода конечных элементов или метода конечных разностей задача (1), (2) сводится к многократному решению систем линейных алгебраических уравнений большого порядка с разреженной матрицей коэффициентов

$$Ax = b, \tag{3}$$

где $A \in R^{n\times n}$, $x \in R^n$, $b \in R^n$, $\det A \neq 0$, и в общем случае $A \neq A^{\text{т}}$.

Многоуровневые (многосеточные) методы [1] предполагают использование вложенной последовательности грубых сеток и проведение некоторого числа итерационных шагов релаксационного метода для каждой из них с целью исключения высокочастотных компонент решения. Часто многосеточный метод рассматривают как предобусловливающую процедуру для системы (3) [2], применяя при этом эффективные алгоритмы ускорения сходимости из класса методов псевдодвойственных направлений [3].

В настоящей работе исследуются двухуровневые методы предобусловливания системы (3) в рамках подхода [4, 5]. При этом основное внимание уделяется раскрашиванию узлов сетки пространственной дискретизации. Развиваются методы двухуровневого решения с использованием r-линейных упорядочений применительно к методам псевдодвойственных направлений, в частности, к методу бисопряженных градиентов (БСГ). Эффективность предлагаемых алгоритмов иллюстрируется на примере численного решения алгебраизованных уравнений дрейфово-диффузионной модели переноса заряда в полупроводниковых средах.

Пусть имеется основная сетка ω. Рассмотрим грубую сетку $\tilde{\omega} \subset \omega$. В соответствии с подразделением узлов основной сетки на $\tilde{\omega}$ и $\omega \setminus \tilde{\omega}$ система (3) приобретает вид

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \tag{4}$$

где $A_{11} \in R^{k\times k}$, $x_1 \in R^k$, $b_1 \in R^k$ и размерность грубой сетки равна $n-k$.

Будем рассматривать предобусловливающие матрицы для дополнения Шура $S = A_{22} - A_{21}A_{11}^{-1}A_{12}$ матрицы A_{11}. Для этого аппроксимируем A_{11}^{-1} с помощью B таким образом, чтобы матрица $A_{22} - A_{21}BA_{12}$ имела простую структуру. Разрабатываемые далее двухсеточные методы основаны на возможности привлечения таких упорядочений, которые обеспечивают эффективное решение линейной системы с матрицей A_{11} прямыми методами. В этом случае результирующий алгоритм [6] решения системы (4) принимает следующий вид:

Алгоритм 1. 1) Решить систему $A_{11}v_1 = b_1$; 2) $v_2 \longleftarrow b_2 - A_{21}v_1$; 3) решить систему $Sx_2 = v_2$; 4) $y_1 \longleftarrow A_{12}x_2$; 5) решить систему $A_{11}z_1 = y_1$; 6) $x_1 \longleftarrow v_1 - z_1$.

Как отмечено выше, система с A_{11} на шагах 1) и 5) будет решаться прямым методом, а система с S на шаге 3) – одним из БСГ-алгоритмов. При этом на каждом шаге БСГ-алгоритма необходимо будет решать прямым методом систему вида $A_{11}h = q$. Таким образом, практические реализации алгоритма 1 отличаются друг от друга способом представления системы (3) в виде (4), т.е. упорядочением узлов сетки пространственной дискретизации.

Известны случаи применения рассмотренного приема в сочетании с красно-черным и простейшим двухлинейным упорядочениями [7–11], которые приводят к матрице порядка примерно вдвое меньшего, чем порядок исходной матрицы. В работе [6] исследованы спектральные свойства матрицы S при трех- и четырехлинейном упорядочениях узлов сетки, но при практической реализации алгоритма 1 система

$$Sx_2 = v_2 \tag{5}$$

решается с использованием предобусловливающей процедуры по методу, предложенному в [12], т.е. предобусловливатель строился не к матрице S, а к матрице A. Эффективное построение предобусловливателя стало возможным в связи с принятой в настоящей работе нумерацией сеточных узлов, приводящей к матрице A_{11} с меньшей шириной полосы, чем в A.

r-Линейные упорядочения. Разобьем прямоугольную сетку вдоль оси OX на участки, содержащие r узлов, за исключением, быть может, последнего участка. Раскрасим сеточные узлы вдоль оси OY черным цветом для первых $r-1$ узлов каждого участка. Если число узлов последнего участка на оси OX меньше r, то все такие узлы раскрашиваем в черный цвет. Сеточные узлы линий вдоль оси OY, соответствующие узлам оси OX, кратным r, будем считать белыми. В отличие от [6] сначала в лексикографическом порядке (слева направо сверху вниз) нумеруются все черные узлы одного блока, затем таким же образом все последующие блоки, потом белые аналогичным образом. В результате матрица A_{11} имеет блочно-диагональную структуру $A_{11} = \operatorname{diag}\{F_1, F_2, \ldots, F_m\}$, где каждый блок F_i, $i = \overline{1, m-1}$, имеет порядок $m_y(r-1)$, m_y – число сеточных линий вдоль оси OY, а блок F_m может иметь порядок $m_y j$, $j \le r-1$, $j = m_x - (m-1)r$, m_x – число линий сетки вдоль оси OX, $m_x m_y = n$. Далее будем рассматривать блоки F_i, $i = \overline{1, m-1}$, и будем их обозначать через F. Все рассуждения будут справедливы и для блока F_m, если в них заменить r на j.

Для аппроксимации уравнения будем использовать стандартный пятиточечный шаблон. Тогда матрица системы (3) имеет пятидиагональную структуру, т.е. i-я строка представима в виде

$$(0, \ldots, 0,\ g_{i-m_x}\delta_{i-m_x, j_1},\ 0, \ldots, 0,\ f_{i-1}\delta_{i-1, j_2},\ a_i\delta_{i,j},\ d_i\delta_{i+1, j_3},\ 0, \ldots, 0,\ c_i\delta_{i+m_x, j_4},\ 0, \ldots, 0),$$

где $\delta_{i,j}$ – символ Кронекера, у которого второй индекс соответствует номеру столбца матрицы A. В этом случае A_{11} имеет пятидиагональную структуру, как и матрица A, с расстоянием $r-1$ до дальней диагонали, при этом индексы ее элементов соответствуют новой нумерации узлов, установленной при их упорядочении. Матрица A_{22} трехдиагональная, а матрицы A_{21} и A_{12} разреженные: A_{12} имеет не более двух ненулевых элементов в каждом столбце, A_{21} по структуре совпадает с A_{12}^{T}.

Приведем алгоритм БСГ-метода [3] решения системы (5), полагая $Sx = b$.

Алгоритм 2. 1) Построение предобусловливателя H; 2) инициализация $x_0 = x^*$, $u_0 = b - Sx_0$, $\widehat{u}_0 = u_0$, $p_{-1} = 0$, $\widehat{p}_{-1} = 0$, $\beta_0 = 0$; 3) итерационное уточнение $(i = 0, 1, 2, \ldots)$: $3_1)$ $\beta_i = \widehat{u}_i^{\mathrm{T}} H^{-1} u_i / (\widehat{u}_{i-1}^{\mathrm{T}} H^{-1} u_{i-1})$ $(i > 0)$; $3_2)$ $p_i = H^{-1}u_i + \beta_i p_{i-1}$; $3_3)$ $\widehat{p}_i = (H^{-1})^{\mathrm{T}}\widehat{u}_i + \beta_i \widehat{p}_{i-1}$; $3_4)$ $\alpha_i = \widehat{u}_i^{\mathrm{T}} H^{-1} u_i / (\widehat{p}_i^{\mathrm{T}} S p_i)$; $3_5)$ $x_{i+1} = x_i + \alpha_i p_i$; $3_6)$ $u_{i+1} = u_i - \alpha_i S p_i$; $3_7)$ $\widehat{u}_{i+1} = \widehat{u}_i - \alpha_i S^{\mathrm{T}} \widehat{p}_i$.

Для организации умножения вектора p на матрицу S и S^{T} нет необходимости вычислять S явно. Из ее структуры следует, что самой сложной вычислительной операцией является умножение матрицы A_{11}^{-1} на вектор $q = A_{12}p$. Результат умножения z можно рассматривать как решение системы $A_{11}z = q$. Решать эту систему предполагается прямым методом, например, можно провести полную факторизацию матрицы $A_{11} = LQU$. Так как матрица A_{11} ленточная с шириной полосы $2r-1$, то при малых r такая процедура не будет обременительной по числу операций [13, с. 151] и не требуется дополнительная память для хранения элементов матриц L, Q и U. Так, в случае трехлинейного упорядочения требуется хранить

$2n$ новых элементов, но поскольку массивы p, $\widehat{p}$, u и $\widehat{u}$ заполнены только на $1/3$ по сравнению с их размерами при решении системы (3) непосредственно БСГ-методом, то их можно разместить в каких-либо трех из них.

Наиболее важным для ускорения сходимости БСГ-метода является построение предобусловливателя H для матрицы S. Рассмотрим этот вопрос более подробно.

Будем предполагать, что A является M-матрицей с диагональным преобладанием. Следовательно, это свойство присуще и матрицам A_{11} и A_{22}.

Так как матрица A_{11} блочно-диагональная, то будем находить приближение ее обращения для одного блока F, который имеет блочную структуру и состоит из m_y^2 блоков размеров $l \times l$, где $l = r - 1$. Для построения приближения матрицы F^{-1} воспользуемся неполной факторизацией матрицы F в виде $\widehat{L}\widehat{V}\widehat{U}$ [14], где матрицы $\widehat{L}$ и $\widehat{U}$ соответственно нижняя и верхняя треугольные, а матрица $\widehat{V}$ диагональная. Известно [15], что выражение $\widehat{L}\widehat{V}\widehat{U}$ является хорошим приближением матрицы F, если $\widehat{V} = \operatorname{diag}\{\widehat{a}_1, \widehat{a}_2, \ldots, \widehat{a}_s\}$, $\widehat{L} = G+F+E$, $\widehat{U} = E+D+C$. Здесь $s = lm_y$, E – единичная матрица, $G = \|g_{i-l}\delta_{i-l,j}/\widehat{a}_{i-l}\|$, $F = \|\widehat{f}_{i-1}\delta_{i-1,j}/\widehat{a}_{i-1}\|$, $D = \|\widehat{d}_i\delta_{i+1,j}/\widehat{a}_i\|$, $C = \|c_i\delta_{i+l,j}/\widehat{a}_i\|$, $\widehat{a}_i = a_i - \widehat{f}_{i-1}\widehat{d}_{i-1}/\widehat{a}_{i-1} - c_{i-l}g_{i-l}/\widehat{a}_{i-l}$, $\widehat{d}_i = d_i - \widehat{f}_{i-1}c_{i-1}/\widehat{a}_{i-1}\omega(l)$, $\widehat{f}_i = d_i - \widehat{d}_{i-1}g_{i-1}/\widehat{a}_{i-1}\omega(l)$. В последних формулах $\omega(l) = 1$ при $l = 2$ и $\omega(l) = 0$ при $l > 2$.

Изучим некоторые свойства матрицы L^{-1}, в частности, найдем оценку ее элементов.

Обозначим через q модуль максимального внедиагонального элемента матрицы L. Построим матрицу Q, которая получается из матрицы L заменой ненулевых внедиагональных элементов на $-q$.

Лемма. *Имеет место неравенство*

$$0 \le \widehat{L}^{-1} \le Q^{-1}. \tag{6}$$

Действительно, из представления матриц $Q = E - Q_1$ и $L = E - L_1$ имеем $Q^{-1} = E + \sum_{k=1}^{l-1} Q_1^k$ и $L^{-1} = E + \sum_{k=1}^{l-1} L_1^k$. Отсюда и из того, что $0 \le L_1 \le Q_1$, следует неравенство (6).

Матрица Q имеет блочный нижне-треугольный вид и состоит из m_y^2 блоков размеров $l \times l$. Не уменьшая общности, можно считать, что блоки Q_{ik} имеют вид

$$Q_{ik} = \begin{cases} E - qI, & \text{если } i = k, \\ -qE, & \text{если } i = k+1, \\ O, & \text{если } i > k+1 \text{ или } i < k, \end{cases}$$

где матрица O нулевая. Тогда матрица Q^{-1} имеет также нижнюю блочно-треугольную форму. Чтобы описать составляющие ее блоки, которые будем обозначать через $(Q^{-1})_{ik}$, определим функцию Z_s^p, принимающую целочисленные значения для $p \in N$ и $s \in N_0$, с помощью рекуррентных соотношений. Положим $Z_0^p = 1$ и для $s > 0$ последовательно найдем $Z_s^p = 1 + \sum_{k=1}^{p} Z_{s-1}^k$.

Справедлива следующая

Теорема. *Ненулевые блоки матрицы Q^{-1} имеют вид*

$$(Q^{-1})_{ik} = q^{i-k}\left(E + \sum_{p=1}^{l-1} Z_{i-k}^p q^p I^p\right), \quad i = \overline{1, m_y}, \quad k = \overline{1, i}. \tag{7}$$

Доказательство проведем по индукции.

1) При $m_y = 1$ теорема следует из представления для обратной матрицы для $Q = E - qI$. Действительно, так как I^k является нулевой для $k \ge l$, то $Q^{-1} = (E - qI)^{-1} = E + \sum_{p=1}^{l-1} q^p I^p$.

2) Предположим, что утверждение верно для $m_y = m$, и докажем его для $m_y = m+1$. Представим матрицу Q в виде $Q = \begin{bmatrix} M_{ml\times ml} & O_{ml\times l} \\ N_{l\times ml} & R_{l\times l} \end{bmatrix}$. Тогда, согласно формуле Фробениуса [16, с. 60], $Q^{-1} = \begin{bmatrix} M^{-1} & O \\ -R^{-1}NM^{-1} & R^{-1} \end{bmatrix}$. Так как матрица M имеет структуру матрицы Q и состоит из m^2 блоков, то, согласно индуктивному предположению, блоки матрицы Q^{-1} представимы в виде (7). Обращение матрицы $R = Q_{ii}$ получено в 1).

Из правила умножения блочных матриц следует, что

$$-R^{-1}NM^{-1} = \left(E + \sum_{p=1}^{l-1} q^p I^p\right) qE \times$$

$$\times \left[q^{m-1}\left(E + \sum_{p=1}^{l-1} Z^p_{m-1} q^p I^p\right), q^{m-2}\left(E + \sum_{p=1}^{l-1} Z^p_{m-2} q^p I^p\right), \ldots, E + \sum_{p=1}^{l-1} Z^p_1 q^p I^p\right]. \tag{8}$$

Перемножая матрицы из (8), убеждаемся, что представления блоков $Q_{m+1,1}$, $Q_{m+1,2}$, … …, $Q_{m+1,m}$ имеют вид (7).

Следствие. *Если* $r = 3$, *то ненулевые блоки матрицы* Q^{-1} *имеют вид*

$$(Q^{-1})_{ik} = q^{i-k}\left(E + (i-k+1)qI\right), \qquad i = \overline{1, m_y}, \quad k = \overline{1, i}. \tag{9}$$

Следуя [15], при построении предобусловливателя H оставим в матрицах $\widehat{L}^{-1}$ и $\widehat{U}^{-1}$ наиболее существенные диагонали. Их расположение будет зависеть от q, r и m_y. В частности, для $\widehat{L}^{-1}$ при $r = 3$ из леммы и (9) имеем, что наиболее существенными будут две поддиагонали, ближайшие к главной диагонали, если $iq^{i-1} < 1$, $\quad i = \overline{2, m_y}$. Тогда на основании [15] в матрицах $\widehat{L}^{-1}$ и $\widehat{U}^{-1}$ можно оставить только три диагонали. В приведенном ниже численном эксперименте использовано такое представление для аппроксимации A^{-1}.

Таким образом, доказанная теорема дает возможность поиска наиболее существенных диагоналей при построении приближения для A_{11}^{-1}.

Численное исследование. В настоящем пункте алгоритм 1 исследуется при решении линейной системы, порожденной системой уравнений дрейфово-диффузионной модели переноса заряда в полупроводниковых средах [14, с. 24]:

$$\Delta\psi = n - p - C_N, \tag{10}$$

$$\nabla \cdot [\mu_n(\nabla n - n \nabla \psi)] = R, \tag{11}$$

$$\nabla \cdot [\mu_p(\nabla p + p \nabla \psi)] = R, \tag{12}$$

где подвижности μ_n, μ_p являются нелинейными функциями пространственных переменных, градиента электростатического потенциала ψ и концентрации носителей заряда (n либо p соответственно); скорость генерации-рекомбинации R также нелинейная функция ψ, n, p и пространственных переменных.

Нетрудно видеть, что каждое из уравнений непрерывности (11), (12) имеет вид (1). В то же время уравнение Пуассона (10) представляет собой частный случай уравнения (1), соответствующий отсутствию конвективного члена. Данная система замыкается граничными условиями Дирихле на омических, поликремниевых и выпрямляющих контактах. На остальной части границы предполагается справедливость условий Неймана. При использовании концентраций носителей заряда n, p в качестве переменных разностная пятиточечная аппроксимация каждого из уравнений непрерывности (11), (12) с применением интегро-интерполяционного метода [17, с. 144] и линеаризации приводит к линейной системе с несимметрической пятидиагональной матрицей. Системы такого типа имеют большой порядок (десятки тысяч и даже миллионы) и плохо обусловлены, поэтому при численном их решении возникают серьезные трудности.

Кроме того, для квадратичной сходимости внешнего цикла решения нелинейной задачи (10)–(12) [14] в режиме слабой инжекции порождаемые линейные системы необходимо решать с высокой точностью.

Все численные эксперименты выполнены на IBM Pentium 166 с использованием двойной значности и пакета двумерного моделирования KFSM [14]. В качестве тестовой структуры выбран планарный транзистор [12]. Размеры неравномерной разностной сетки для него составляют $39 \times 63 = 2457$ узла. Внешние смещения соответствуют режиму слабой инжекции. Для решения редуцированной системы (5) алгебраизованных уравнений непрерывности используется алгоритм 2, предобусловленный с помощью неполного разложения [15] и описанных в настоящей работе двухуровневых процедур.

В табл. 1 приведено число итераций N и нормализованное время, соответствующие решению систем, порожденных уравнениями (11) и (12) для планарного транзистора на первой внешней итерации, а также общее время их решения на всех внешних итерациях последовательного алгоритма решения нелинейной задачи (10)–(12). Для построения двухуровневого предобусловливателя использовано трехлинейное упорядочение и для предобусловливания матрицы S сохранено семь диагоналей для ее приближения. Решение системы считалось полученным при выполнении условия $\|r_N\|_\infty/\|r_0\|_\infty < \varepsilon$, в экспериментах $\varepsilon = 10^{-12}$.

Таблица 1

Предобусловливание	(11)		(12)		(10)–(12)	
	N	t,c	N	t,c	N	t,c
Для матрицы A	66	0.55	42	0.33	615	4.98
Для матрицы S	55	0.42	38	0.27	529	4.30

Как и ожидалось, алгоритмы, использующие предобусловливание для S, демонстрируют в экспериментах более высокую эффективность. Аналогичные результаты получены при моделировании широкого ряда кремниевых структур современной микроэлектроники. Это можно объяснить двумя причинами: во-первых, как хорошо известно [4, 6, 11], спектральные свойства дополнения Шура предпочтительнее свойств матрицы исходной системы (по крайней мере для рассматриваемого класса задач), во-вторых, одним из проявлений свойства оптимальности многосеточных методов является возможность численной реализации итерации БСГ-метода посредством меньшего числа арифметических операций [12]. Из табл. 1 можно заключить, что одним из путей дальнейшего повышения эффективности БСГ-методов решения алгебраизованных уравнений математической физики может являться использование линейных упорядочений в двухуровневых процедурах рассмотренного в настоящей работе класса.

Замечание. Предложенная схема с двухуровневым предобусловливанием может быть использована в других итерационных методах псевдодвойственных направлений [4], в частности, CGS [18], CGSTAB [19].

СПИСОК ЛИТЕРАТУРЫ

1. *Федоренко Р.П.* // Журн. вычислит. математики и мат. физики. 1961. Т. 1. № 5. С. 922–927.
2. *Hackbuch W.* Multi-Grid Methods and Applications. Berlin, 1985.
3. *Fletcher R.* // Lecture Notes Math. 1976. № 506. P. 73–89.
4. *Axelsson O.* Iterative Solution Methods. Cambridge, 1996.
5. *Axelsson O., Gustafsson I.* // Math. Comp. 1983. V. 40. № 161. P. 219–242.
6. *Белявский С.С., Мулярчик С.Г., Попов А.В.* // Дифференц. уравнения. 1998. Т. 34. № 7. С. 929–934.
7. *Brand C.W.* // Numer. Math. 1992. V. 61. № 2. P. 433–454.
8. *Foresti S., Brusino G., Hassanzaden S., Sonnand V.* // Comput. Phys. Communinic. 1989. V. 53. P. 349–355.
9. *Axelsson O., Gustafsson I.* // J. Comput. Phys. 1980. V. 35. № 2. P. 284–289.
10. *Axelsson O., Eijkhout V.* // SIAM J. Sci. Statist. Comput. 1991. V. 12. P. 1373–1400.
11. *Elman H.C., Golub G.H.* // SIAM J. Sci. Statist. Comput. 1992. V. 13. № 1. P. 339–363.
12. *Mulyarchik S.G., Bielawski S.S., Popov A.V.* // J. Comput. Phys. 1994. V. 110. № 2. P. 201–211.

13. *Golub G.H., van Loan C.F.* Matrix Computations. Baltimore, 1996.
14. *Мулярчик С.Г.* Численное моделирование микроэлектронных структур. Мн., 1989.
15. *Meijerink J.A., van der Vorst H.A.* // Math. Comp. 1977. V. 31. № 137. P. 148–162.
16. *Гантмахер Ф.Р.* Теория матриц. М., 1988.
17. *Самарский А.А.* Теория разностных схем. М., 1983.
18. *Sonneveld P.* // SIAM J. Sci. Stat. Comput. 1989. V. 10. № 1. P. 36–52.
19. *Van der Vorst H.A.* // SIAM J. Sci. Stat. Comput. 1992. V. 13. № 2. P. 631–644.

Белорусский государственный университет,
г. Минск

Поступила в редакцию
01.03.2001 г.