



Math-Net.Ru

Общероссийский математический портал

В. В. Воеводин, О методах сопряженных направлений,
Ж. вычисл. матем. и матем. физ., 1979, том 19, номер 5, 1313–1317

<https://www.mathnet.ru/zvmmf5368>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<https://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 18.97.14.87

17 мая 2025 г., 00:29:14



НАУЧНЫЕ СООБЩЕНИЯ

О МЕТОДАХ СОПРЯЖЕННЫХ НАПРАВЛЕНИЙ

В. В. ВОЕВОДИН

(Москва)

Исследуются обобщения методов сопряженных направлений на системы линейных уравнений с несимметричными матрицами. Указывается простое условие, которому удовлетворяют все наиболее известные варианты этих методов.

Методы сопряженных направлений [1] широко используются для решения систем линейных алгебраических уравнений

$$(1) \quad Ax=f$$

во многих задачах вычислительной математики.

Имеется большое число математически эквивалентных форм реализации методов сопряженных направлений. Их различие оказывается весьма значительным, особенно в отношении устойчивости к ошибкам округления. Понять причину возникновения таких различий из самих расчетных формул трудно.

Классический метод сопряженных градиентов применяется, как правило, к решению систем с вещественной симметричной положительно-определенной матрицей A . Принимая во внимание общий итог исследования устойчивости многих методов решения алгебраических задач [2], можно высказать предположение, что реально вычисленное решение будет точным решением некоторой возмущенной системы, скорее всего с несимметричным возмущением. Поэтому прямое распространение метода на случай систем с комплексными неэрмитовыми матрицами может показать, какие из вариантов методов сопряженных направлений являются основными, а какие — подчиненными. Конечно, подобное расширение значительно увеличит сферу применения методов этой группы.

Пусть R — комплексная матрица порядка n . Упорядоченную систему векторов s_1, \dots, s_p , $p \leq n$, назовем R -псевдоортогональной, если для $i < j \leq p$ выполняются условия

$$(Rs_j, s_i) = 0, \quad (Rs_j, s_j) \neq 0.$$

Любая R -псевдоортогональная система является линейно-независимой. Легко найти и разложение любого вектора x по базисной R -псевдоортогональной системе s_1, \dots, s_n . Если

$$x = \alpha_1 s_1 + \alpha_2 s_2 + \dots + \alpha_n s_n,$$

то для определения коэффициентов $\alpha_1, \dots, \alpha_n$ необходимо решить систему с левой треугольной матрицей:

$$\alpha_1 (Rs_1, s_1) = (Rx, s_1),$$

$$\alpha_1 (Rs_1, s_n) + \dots + \alpha_n (Rs_n, s_n) = (Rx, s_n).$$

Из любой линейно-независимой системы векторов r_0, \dots, r_{p-1} почти всегда можно построить R -псевдоортогональную систему s_1, \dots, s_p с помощью процесса, ана-

логичного процессу ортогонализации Грамма — Шмидта. Пусть $s_1=r_0$, и предположим, что из векторов r_0, \dots, r_{i-1} уже получены векторы s_1, \dots, s_i . Вектор s_{i+1} будем искать в виде

$$(2) \quad s_{i+1} = r_i + \sum_{k=1}^i \beta_{i+1,k} s_k.$$

Условия R -псевдоортогональности вектора s_{i+1} к векторам s_1, \dots, s_i снова дают для определения неизвестных коэффициентов $\beta_{i+1,k}$ левую треугольную систему. Если все величины $(R s_i, s_i)$ окажутся отличными от нуля, полученная система s_1, \dots, s_p будет R -псевдоортогональной.

Рассмотрим решение системы (1) с комплексной матрицей A порядка n . Следуя [1], возьмем $R=CAB$ для каких-либо матриц C, B , и пусть s_1, \dots, s_n — некоторая R -псевдоортогональная система. Обозначим через x_0 начальный вектор, и пусть

$$(3) \quad x = x_0 + B \sum_{j=1}^n a_j s_j, \quad x_i = x_0 + B \sum_{j=1}^i a_j s_j, \quad r_i = f - Ax_i.$$

Тогда из соотношений

$$(4) \quad x_i = x_{i-1} + a_i B s_i.$$

вытекает, что

$$(5) \quad r_i = r_{i-1} - a_i A B s_i.$$

Легко показать, что для выбранной CAB -псевдоортогональной системы s_1, \dots, s_n имеют место равенства

$$(6) \quad (C r_i, s_k) = 0, \quad 1 \leq k \leq i.$$

Действительно,

$$r_i = f - Ax_i = A(x - x_i) = \sum_{j=i+1}^n a_j A B s_j,$$

и далее

$$(C r_i, s_k) = \left(\sum_{j=i+1}^n a_j C A B s_j, s_k \right) = \sum_{j=i+1}^n a_j (C A B s_j, s_k) = 0$$

для всех $k \leq i$.

Из (2) заключаем, что если последовательность векторов s_i получена путем CAB -псевдоортогонализации векторов r_i , то r_k есть линейная комбинация векторов s_1, \dots, s_{k+1} . Следовательно, скалярное произведение $(C r_i, r_k)$ — линейная комбинация чисел $(C r_i, s_1), \dots, (C r_i, s_{k+1})$, и оно равно нулю в соответствии с (6) для $k < i$, т. е.

$$(7) \quad (C r_i, r_k) = 0, \quad k < i.$$

Это означает, что если $(C r_i, r_i) \neq 0$ для всех i , то r_0, \dots, r_{n-1} есть C -псевдоортогональная система.

В общем случае векторы s_1, \dots, s_n находятся по рекуррентному соотношению (2) и для определения коэффициентов $\beta_{i+1,k}$ приходится решать систему с треугольной матрицей. Коэффициенты a_i из (4) всегда определяются очень просто. Согласно (5) — (7), имеем

$$(8) \quad a_i = \frac{(C r_{i-1}, r_{i-1})}{(C A B s_i, r_{i-1})} = \frac{(C r_{i-1}, s_i)}{(C A B s_i, s_i)}.$$

В классической схеме метода сопряженных градиентов для вещественной симметричной матрицы A среди коэффициентов $\beta_{i+1,k}$ из (2) только $\beta_{i+1,i}$ отличен от нуля. Аналогичным свойством обладает семейство матриц, удовлетворяющих условию

$$(9) \quad (CABC^{-1})^* = \alpha E + \beta AB$$

для некоторых чисел α, β . Пусть

$$(10) \quad s_{i+1} = r_i + f_i s_i,$$

причем

$$(11) \quad f_i = - \frac{(CABr_i, s_i)}{(CABs_{i+1}, s_i)} = - \bar{\beta} \frac{(Cr_i, ABs_i)}{(CABs_i, s_i)}.$$

(Здесь и ниже чертой сверху обозначаются комплексно-сопряженные величины.) Условие $(CABs_{i+1}, s_i) = 0$ выполняется из-за выбора коэффициента f_i . Принимая во внимание (9), получаем для $k < i$ из (5) - (7), что

$$\begin{aligned} (CABs_{i+1}, s_k) &= ((CABC^{-1})Cr_i, s_k) = (Cr_i, (CABC^{-1})^* s_k) = \\ &= (Cr_i, (\alpha E + \beta AB)s_k) = \bar{\alpha} (Cr_i, s_k) + \frac{\bar{\beta}}{a_k} \{ (Cr_i, r_{k-1}) - (Cr_i, r_k) \} = 0. \end{aligned}$$

Если матрица AB из (9) имеет хотя бы два различных собственных значения λ_i, λ_j , то $|\beta| = 1$. Действительно, приведем матрицы левой и правой частей соотношения (9) одновременно к каноническому виду Жордана и приравняем между собой диагональные элементы. Тогда либо при некоторой нумерации собственных значений будут выполняться равенства

$$(12) \quad \bar{\lambda}_1 = \alpha + \beta \lambda_1, \quad \bar{\lambda}_2 = \alpha + \beta \lambda_2,$$

где $\lambda_1 \neq \lambda_2$, либо при какой-то другой нумерации собственных значений будем иметь

$$(13) \quad \bar{\lambda}_1 = \alpha + \beta \lambda_2, \quad \bar{\lambda}_2 = \alpha + \beta \lambda_3, \dots, \bar{\lambda}_k = \alpha + \beta \lambda_1,$$

где снова $\lambda_1 \neq \lambda_2$. Исключая α в каждой из групп соотношений, заключаем, что либо $|\lambda_1 - \lambda_2| = |\beta| |\lambda_1 - \lambda_2|$, либо $|\lambda_1 - \lambda_2| = |\beta|^k |\lambda_1 - \lambda_2|$. Так как $\lambda_1 \neq \lambda_2$, то $|\beta| = 1$ в обоих случаях.

Не ограничивая существенно общности, можно считать, что $\beta = 1$. Выполнения этого условия легко достичь умножением матрицы A на множитель $\beta^{1/2}$. Поэтому в дальнейшем можно было бы рассматривать класс матриц, удовлетворяющих такому условию:

$$(14) \quad (CABC^{-1})^* = \alpha E + AB.$$

Из соотношений (12), (13) следует, что число α в (14) чисто мнимое. Кроме этого, нетрудно установить некоторые свойства спектра матрицы AB из (14). Пусть λ является собственным значением AB из (14). Если $\text{Im } \lambda = \alpha/2$, то при существовании собственного значения $\bar{\lambda}$ его кратность обязательно меньше кратности λ . Если $\text{Im } \lambda \neq \alpha/2$, то наряду с λ существует собственное значение вида $\bar{\lambda} - \alpha/2$ такой же кратности. Других ограничений на распределение собственных значений матрицы AB и на число α соотношение (14) не накладывает.

Отметим некоторые из важнейших классов матриц, для которых условие (9) выполняется: A, B, C - эрмитовы, $C=B$; CAB, C - эрмитовы; матрица C перестановочна с AB , AB - нормальная и ее спектр лежит на прямой линии; $A=M+N$, где $M=M^*, N=-N^*$, при этом $C=B=M^{-1}$ или $C=B=N^{-1}$.

При выполнении соотношения (9) общий процесс метода сопряженных направлений может осуществляться по следующему предписанию:

$$s_1 = r_0, \quad r_i = r_{i-1} - a_i ABs_i, \quad s_{i+1} = r_i + f_i s_i, \quad x_i = x_{i-1} + a_i Bs_i.$$

Здесь x_0 — произвольный начальный вектор, коэффициенты a_i, f_i вычисляются согласно (8), (11). Если обозначить $u_i = Bs_i$, то процесс будет таким:

$$u_i = Br_0, \quad r_i = r_{i-1} - a_i Au_i, \quad u_{i+1} = Br_i + f_i u_i, \quad x_i = x_{i-1} + a_i u_i,$$

при этом

$$a_i = \frac{(Cr_{i-1}, r_{i-1})}{(CAu_i, r_{i-1})} = \frac{(B^{-1}Cr_{i-1}, u_i)}{(B^{-1}CAu_i, u_i)},$$

$$f_i = -\frac{(B^{-1}CABr_i, u_i)}{(B^{-1}CAu_i, u_i)} = -\bar{\beta} \frac{(Cr_i, Au_i)}{(B^{-1}CAu_i, u_i)}.$$

В частных случаях данные формулы могут оказаться более простыми. Например, при $C=E, B=A^*$ имеем

$$a_i = \frac{(r_{i-1}, r_{i-1})}{(Au_i, r_{i-1})} = \frac{(r_{i-1}, r_{i-1})}{(u_i, u_i)} > 0,$$

$$f_i = -\frac{(r_i, Au_i)}{(u_i, u_i)} = \frac{(r_i, r_i)}{(r_{i-1}, r_{i-1})} > 0.$$

Если же $C=AA^*, B=A^*$, то

$$a_i = \frac{(A^*r_{i-1}, A^*r_{i-1})}{(Au_i, Au_i)} > 0, \quad f_i = \frac{(A^*r_i, A^*r_i)}{(A^*r_{i-1}, A^*r_{i-1})} > 0.$$

Как уже отмечалось, последовательность векторов r_i является C -псевдоортогональной. Следовательно, невязка r_n заведомо будет нулевой, а вектор x_n — точным решением системы (1). Единственное, что может помешать реализации метода, — это обращение в нуль одного из скалярных произведений $(CABs_i, s_i)$. Однако данной ситуации всегда можно избежать путем выбора нового начального приближения x_0 .

Векторы r_i и s_{i+1} — линейные комбинации векторов $r_0, AB r_0, \dots, (AB)^i r_0$. Поэтому если в разложении вектора по каноническому базису Жордана матрицы AB присутствуют не все составляющие, то обращение невязки в нуль произойдет раньше. Процесс оканчивается особенно быстро, если матрица AB к тому же имеет простую структуру и большое число совпадающих собственных значений. Именно, если в разложении вектора r_0 по собственным векторам матрицы AB ненулевые составляющие соответствуют m попарно различным собственным значениям, то $r_m = 0$.

Классический метод сопряженных градиентов является не только конечным, но и итерационным, так как на каждом шаге минимизирует функционал ошибок. Описанная группа методов в общем случае таким свойством не обладает. Но если матрица CAB эрмитова и положительно-определенная, то на каждом шаге будет минимизироваться обобщенный функционал ошибок с матрицей $B^{-1}CA$. Доказательство этого факта ничем не отличается от соответствующего доказательства в [1].

Если матрица A вырожденная, то метод всегда позволяет получить одно из решений, а в случае положительной определенности матрицы CAB — решение с минимальным обобщенным функционалом ошибок.

Как и в методе сопряженных градиентов, в данной группе методов имеют место трехчленные соотношения для векторов s_i, r_i . Из (5), (10) вытекает, что

$$s_{i+1} = (1 + f_i) s_i - a_i AB s_i - f_{i-1} s_{i-1},$$

$$r_{i+1} = \left(1 + \frac{f_i a_{i+1}}{a_i}\right) r_i - a_{i+1} AB r_i - \frac{f_i a_{i+1}}{a_i} r_{i-1}.$$

Отсюда можно получить и другие соотношения, например такое:

$$(15) \quad x_{i+1} = x_{i-1} + \omega_{i+1} (\alpha_i B r_i + x_i - x_{i-1}),$$

где ω_{i+1} , α_i — подходящим образом выбранные числа.

Все наиболее известные варианты методов сопряженных направлений с точностью до обозначений и порядка выполнения операций укладываются в описанную выше схему. Рассмотрим некоторые из примеров.

1. A — положительно-определенная эрмитова матрица, $C=B=E$. Классический метод сопряженных градиентов, его вещественный вариант описан в [1]. На каждом шаге минимизируется функционал ошибок с матрицей A .

2. A — произвольная матрица, $C=E$, $B=A^*$. Математически метод совпадает с методом AA^* -минимальных итераций при специальном выборе начального вектора. Вещественный вариант описан в [1]. На каждом шаге минимизируется функционал ошибок с матрицей E , т. е. евклидова норма самой ошибки.

3. A — произвольная матрица, $C=AA^*$, $B=A^*$. Математически метод совпадает с методом A^*A -минимальных итераций при специальном выборе начального вектора. Вещественный вариант описан в [1]. На каждом шаге минимизируется функционал ошибок с матрицей A^*A , т. е. евклидова норма вектора невязки.

4. A , B — эрмитовы матрицы, $C=B$. Вещественный вариант метода в форме (15) для случая положительно-определенных матриц A , B описан в [3]. На каждом шаге минимизируется функционал ошибок с матрицей A , если $A \geq 0$.

5. A — произвольная матрица. Если $A=M+N$, где $M=M^*$, $N=-N^*$, то $B=C=M^{-1}$ или $B=C=N^{-1}$. Вещественный вариант метода в форме (15) для случая положительно-определенной матрицы M описан в [4].

6. A — нормальная матрица, спектр которой лежит на прямой, $C=A^*$, $B=E$. Вещественный вариант метода для симметричной матрицы A в несколько иной форме описан в [5]. На каждом шаге минимизируется функционал ошибок с матрицей A^*A , т. е. евклидова норма вектора невязки.

Если матрица A прямоугольная или квадратная неэрмитова, то можно переходить к системам с неотрицательно-определенными эрмитовыми матрицами с помощью первой или второй трансформации Гаусса. Однако заметим, что в случае отказа от явного вычисления матриц A^*A и AA^* при реализации метода сопряженных градиентов расчетные формулы оказываются совпадающими с формулами для методов A^*A - и AA^* -минимальных итераций.

Поступила в редакцию 20.04.1979

Цитированная литература

1. Д. К. Фаддеев, Н. В. Фаддеева. Вычислительные методы линейной алгебры. М., «Наука», 1963.
2. В. В. Воеводин. Вычислительные основы линейной алгебры. М., «Наука», 1977.
3. P. Concus, G. Golub, D. O'Leary. A generalized conjugate method for the numerical solution of elliptic partial differential equations. In «Sparse Matrix Computations». New York, Acad. Press, 1976, 309–332.
4. P. Concus, G. Golub. A generalized conjugate method for nonsymmetric systems of linear equations. Berkeley, Univ. of California, Techn. Rept, 1975.
5. O. Axelsson. Solution of linear systems of equations: iterative methods. In «Sparse Matrix Techniques». New York, Springer, 1977, 1–51.