ELSEVIER Applied Mathematics and Computation 92 (1998) 49–58

APPLIED MATHEMATICS AND COMPUTATION

The generalized Cholesky factorization method for saddle point problems ☆

Jinxi Zhao

Department of Mathematics, Nanjing University, Nanjing 210 008, China

Abstract

Over the past 10 years, a variety of iterative methods for saddle point problems have been proposed. In this paper, we present a class of direct methods, the so-called generalized Cholesky factorization method, for solving linear systems arising from saddle point problems or discretization of the Stokes equations. Numerical results illustrate the efficiency of new methods given in this paper. © 1998 Elsevier Science Inc. All rights reserved.

Keywords: Saddle point problem; Generalized Cholesky factorization; Condition I; Generalized positive definite

1. Introduction

Consider the linear system of equations

$$\begin{pmatrix} A & B^{\mathrm{T}} \\ B & -C \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{1}$$

where A is symmetric positive definite, B of full row rank, and C symmetric positive semi-definite. Problems in this class arise frequently in the context of minimization of quadratic forms subject to linear constraints [8]. An important example arises from the numerical discretization of the Stokes equations. In particular, we are concerned with the discretization of

☆ Project supported by the Science Foundation of the State Education Commission of China and the "863" plan of China.

0096-3003/98/$19.00
PII: S0096-3003(97)10040-6

$$\begin{aligned} &-\Delta u-\nabla p=f,\\ &\nabla u=0, \quad \text{on } \Omega,\\ &u=0, \quad \text{on } \partial\Omega,\\ &\int_\Omega p=0, \end{aligned} \tag{2}$$

where Ω is a simply connected bounded domain in $\mathbb{R}^s, s=2$ or 3. This system of the Stokes equations is a fundamental problem arising in computational fluid dynamics [4]. Discretization of Eq. (2) by finite difference or finite element techniques leads to a linear system of equations of (1).

In recent years, a variety of iterative algorithms have been devised for solving saddle point problems [1,3,4,6,7]. In this paper, we have developed the generalized Cholesky factorization for four typical matrices arising in numerical optimization and computational fluid dynamics. Using the matrix factorization, we establish a class of direct methods for solving the corresponding linear system. New methods proposed in this paper remain main advantages of the classical Cholesky factorization for positive definite systems. Hence the new method is referred to as the generalized Cholesky factorization method.

In the following we always assume that matrices $A\in\mathbb{R}^{m\times m}$, $B\in\mathbb{R}^{n\times m}$, and $C\in\mathbb{R}^{n\times n}$ satisfy the following condition.

Condition I. A is symmetric positive definite, B is of full row rank and C is symmetric semi-positive definite.

2. Symmetric indefinite case

Let us assume that G_1 is an $(m+n)*(m+n)$ matrix and express it as

$$G_1=\begin{pmatrix} A & B^{\mathrm T}\\ B & -C\end{pmatrix}, \tag{3}$$

where A, B and C satisfy Condition I. It is easy to see that G_1 is symmetric indefinite. The purpose of this section is based on the matrix factorization of G_1 to give a new algorithm for solving the linear system (1).

Firstly, we can prove the following theorem.

Theorem 1. *Let G_1 be an $(m+n)*(m+n)$ matrix expressed by Eq. (3) and A, B, C satisfy Condition I. Then there always exists the factorization form*

$$G_1=\begin{pmatrix} A & B^{\mathrm T}\\ B & -C\end{pmatrix}=L_1L_1^d. \tag{4}$$

where

$$L_1 = \begin{pmatrix} L_A & 0 \\ L_B & L_w \end{pmatrix}, \qquad L_1^d = \begin{pmatrix} L_A^T & L_B^T \\ 0 & -L_w^T \end{pmatrix}$$

and $L_A \in \mathbb{R}^{m\times m}$ *is low triangular,* $L_w \in \mathbb{R}^{n\times n}$ *is low triangular,* $L_B \in \mathbb{R}^{n\times m}$.

Proof. Since A is symmetric positive definite, there always exists the Cholesky factorization

$$A = L_A L_A^T,$$

where $L_A \in \mathbb{R}^{m\times m}$ is nonsingular low triangular. Take

$$L_B = B(L_A^T)^{-1}, \tag{5}$$

so that

$$B = L_B L_A^T.$$

It follows that L_B is full row rank because B is full row rank. Because C is symmetric semi-positive definite, the matrix $C + L_B L_B^T$ must be symmetric positive definite. Hence we have the Cholesky factorization

$$L_w L_w^T = C + L_B L_B^T, \tag{6}$$

also let

$$L_1 = \begin{pmatrix} L_A & 0 \\ L_B & L_w \end{pmatrix}, \qquad L_1^d = \begin{pmatrix} L_A^T & L_B^T \\ 0 & -L_w^T \end{pmatrix}.$$

Thus we have

$$L_1 L_1^d = \begin{pmatrix} L_A L_A^T & L_A L_B^T \\ L_B L_A^T & L_B L_B^T - L_w L_w^T \end{pmatrix} = \begin{pmatrix} A & B^T \\ B & -C \end{pmatrix}.$$

Remark 1. From Theorem 1, we set that matrices L_1 and L_1^d can be obtained conveniently as long as submatrices L_A, L_B and L_w have been computed. This is why our method is as fast as the classical Cholesky factorization for symmetric positive definite.

We now discuss the realization of the generalized Cholesky factorization (4) of G_1. Let

$$A = [a_{ij}] \in \mathbb{R}^{m\times m}, \quad L_A = [l_{ij}].$$

Using the Cholesky factorization of A, the elements of L_A can be computed from

$$l_{ij} = \begin{cases} 0 & i < j, \\ \left(a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2\right)^{1/2} & i = j, \\ \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk}\right) / l_{jj}, & i > j, \end{cases} \tag{7}$$

$$i, j = 1:n.$$

Set

$$L_B = [g_{ij}] = B(L_A^{\mathrm{T}})^{-1} \in \mathbb{R}^{n \times m}.$$

Then all elements g_{ij} can be obtained from the following:

$$g_{ij} = \left(b_{ij} - \sum_{k=1}^{j-1} g_{ik} l_{jk}\right) / l_{jj}, \quad i = 1:n, \quad j = 1:m. \tag{8}$$

Since

$$L_w L_w^{\mathrm{T}} = C + L_B L_B^{\mathrm{T}},$$

let

$$L_w = [v_{ij}] \in \mathbb{R}^{n \times n}$$

be low triangular.

Hence

$$v_{ij} = \begin{cases} 0, & i < j, \\ \left(c_{ii} + \sum_{k=1}^{m} g_{ik}^2 - \sum_{p=1}^{i-1} v_{ip}^2\right)^{1/2}, & i = j, \\ \left(c_{ij} + \sum_{k=1}^{m} g_{ik} g_{jk} - \sum_{p=1}^{j-1} v_{ip} v_{jp}\right) / v_{jj}, & i > j, \end{cases}$$

$$i = 1:n, \quad j = 1:n. \tag{9}$$

From the above discussion, the triangular factor of G_1 can be obtained provided that submatrices L_A, L_B and L_w have been formed, i.e.,

$$L_1 = \begin{pmatrix} L_A & 0 \\ L_B & L_w \end{pmatrix}$$

and

$$L_1^d = \begin{pmatrix} L_A^{\mathrm{T}} & L_B^{\mathrm{T}} \\ 0 & -L_w^{\mathrm{T}} \end{pmatrix}.$$

Using the factorization expression

$$G_1 = L_1 L_1^d.$$

it is easy to give the solution procedure of the linear system (1). Here we describe the following.

Algorithm 1.
1. Given $A = [a_{ij}] \in \mathbb{R}^{m\times m}$, $B = [b_{ij}] \in \mathbb{R}^{n\times m}$ and $C = [C_{ij}] \in \mathbb{R}^{n\times n}$ satisfying Condition I and given $f \in \mathbb{R}^m, g \in \mathbb{R}^n$.
2. $L_A = [l_{ij}] = \mathrm{chol}(A)$ or using expression (7).
3. Computing $L_B = [g_{ij}]$ from $L_B L_A^{\mathrm{T}} = B$ or using expression (8).
4. $L_w = [v_{ij}] = \mathrm{chol}(C + L_B L_B^{\mathrm{T}})$ or using expression (9).
5. Computing

$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

from

$$\begin{pmatrix} L_A & 0 \\ L_B & L_w \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}.$$

6. Obtained the final solution of the linear system (1) from

$$\begin{pmatrix} L_A^{\mathrm{T}} & L_B^{\mathrm{T}} \\ 0 & -L_w^{\mathrm{T}} \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

As a special case, if we have $C = 0$ in the linear system (1), i.e.,

$$G_2 = \begin{pmatrix} A & B^{\mathrm{T}} \\ B & 0 \end{pmatrix},$$

then we can obtain the analogous factorization form.

Using matrices L_A and L_B obtained in Theorem 1, and L_B full row rank, we have the Cholesky factorization of $L_B L_B^{\mathrm{T}}$, i.e.,

$$L_x L_x^{\mathrm{T}} = L_B L_B^{\mathrm{T}},$$

where $L_x \in \mathbb{R}^{n\times n}$ is low triangular. Hence the following theorem is proved.

Theorem 2. *Let*

$$G_2 = \begin{pmatrix} A & B^{\mathrm{T}} \\ B & 0 \end{pmatrix}$$

and matrices A and B satisfy Condition I, then there always exists a factorization

$$G_2 = L_2 L_2^d, \tag{10}$$

where

$$L_2 = \begin{pmatrix} L_A & 0 \\ L_B & L_x \end{pmatrix}, \quad L_2^d = \begin{pmatrix} L_A^{\mathrm{T}} & L_B^{\mathrm{T}} \\ 0 & -L_x^{\mathrm{T}} \end{pmatrix}.$$

$L_A \in \mathbb{R}^{m\times m}$ is low triangular and the Cholesky factor of A, $L_B = B(L_A^{\mathrm{T}})^{-1} \in \mathbb{R}^{n\times m}$, $L_s \in \mathbb{R}^{n\times n}$ is low triangular and the Cholesky factor of $L_B L_B^{\mathrm{T}}$.

3. Generalized semi-positive definite case

Consider the linear system

$$\begin{pmatrix} A & -B^{\mathrm{T}} \\ B & C \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}, \tag{11}$$

where matrices A, B and C satisfy Condition I.

Let

$$G_3 = \begin{pmatrix} A & -B^{\mathrm{T}} \\ B & C \end{pmatrix} \tag{12}$$

and obviously, the matrix G_3 is nonsymmetric. But the symmetric part of G_3, i.e., $(G_3^{\mathrm{T}} + G_3)/2$, is symmetric semi-positive definite [2,5]. So we call G_3 the generalized semi-positive definite.

As in the above discussion, we can always form the low triangular matrix L_A from symmetric positive definite matrix A and

$$L_B = B(L_A^{\mathrm{T}})^{-1}.$$

Since C is symmetric semi-positive definite and L_B full row rank, there always exists the Cholesky factorization, i.e.,

$$L_s L_s^{\mathrm{T}} = C + L_B L_B^{\mathrm{T}}.$$

where $L_s \in \mathbb{R}^{n\times n}$ is low triangular. Thus we have proved the following result.

Theorem 3. *Assume that G_3 is of*

$$G_3 = \begin{pmatrix} A & -B^{\mathrm{T}} \\ B & C \end{pmatrix},$$

where A, B and C satisfy Condition I. Then there always exists the generalized Cholesky factorization,

$$G_3 = L_3 L_3^{d}, \tag{13}$$

where

$$L_3 = \begin{pmatrix} L_A & 0 \\ L_B & L_s \end{pmatrix}, \qquad L_3^{d} = \begin{pmatrix} L_A^{\mathrm{T}} & -L_B^{\mathrm{T}} \\ 0 & L_s^{\mathrm{T}} \end{pmatrix}.$$

$L_A \in \mathbb{R}^{n\times n}$ is low triangular and the Cholesky factor of A, $L_s \in \mathbb{R}^{n\times n}$ is low triangular and the Cholesky factor of $C + L_B L_B^{\mathrm{T}}$, $L_B = B(L_A^{\mathrm{T}})^{-1} \in \mathbb{R}^{n\times m}$.

Since the realization procedures of submatrices L_A, L_B and L_y are the same as in Section 2. Hence we can propose an algorithm for solving the linear system

$$\begin{pmatrix} A & -B^{\mathrm{T}} \\ B & C \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}.$$

Algorithm 2.

1. Given $A = [a_{ij}] \in \mathbb{R}^{m\times m}$, $B = [b_{ij}] \in \mathbb{R}^{n\times m}$ and $C = [c_{ij}] \in \mathbb{R}^{n\times n}$ satisfying Condition I and given $f \in \mathbb{R}^m$, $g \in \mathbb{R}^n$.
2. $L_A = [l_{ij}] = \mathrm{chol}(A)$ or using expression (7).
3. Computing $L_B = [g_{ij}]$ from $L_B L_A^{\mathrm{T}} = B$ or using expression (8).
4. $L_y = [v_{ij}] = \mathrm{chol}(C + L_B L_B^{\mathrm{T}})$ or using expression (9).
5. Computing
$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$
from
$$\begin{pmatrix} L_A & 0 \\ L_B & L_y \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}.$$
6. Obtained the final solution of the linear system (1) from
$$\begin{pmatrix} L_A^{\mathrm{T}} & -L_B^{\mathrm{T}} \\ 0 & L_y^{\mathrm{T}} \end{pmatrix}\begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Consider the special case where $C = 0$ in Eq. (12), i.e.,

$$G_4 = \begin{pmatrix} A & -B^{\mathrm{T}} \\ B & 0 \end{pmatrix}.$$

Obviously, we can prove the following.

Theorem 4. *If*

$$G_4 = \begin{pmatrix} A & -B^{\mathrm{T}} \\ B & 0 \end{pmatrix}$$

and A, B satisfy Condition I, then we have

$$G_4 = L_4 L_4^d, \tag{14}$$

where

$$L_4 = \begin{pmatrix} L_A & 0 \\ L_B & L_z \end{pmatrix}, \qquad L_4^d = \begin{pmatrix} L_A^{\mathrm{T}} & -L_B^{\mathrm{T}} \\ 0 & L_z^{\mathrm{T}} \end{pmatrix}.$$

Here matrices L_A and L_B are the same as in Theorem 3, and L_z is the Cholesky factor of $L_B L_B^{\mathrm{T}}$. Since L_B is of full row rank, there always exists the Cholesky factor L_z of $L_B L_B^{\mathrm{T}}$.

4. Discussion

The method described in this paper retains the main advantages of the classical Cholesky factorization. During the course of the computation, $N = m + n$ square roots must be taken. Condition I assures us that the arguments of these square roots will be positive. About $N^3/6$ flops are needed beyond n square roots. Finally, because A is symmetric positive definite, the elements of L_A will be controllable. In fact, we have the following relation,

$$l_{ik} \leqslant \sqrt[2]{a_{ii}}, \quad i = 1:m, \quad k = 1:i. \tag{15}$$

Since

$$L_B = B(L_A^{\mathrm{T}})^{-1},$$

it follows that

$$c_{ii} + \sum_{k=1}^{m} g_{ik}^2 = \sum_{p=1}^{i} v_{ip}^2.$$

If we take

$$M = \max_{1 \leqslant j \leqslant n} (g_{j1}^2 + \cdots - g_{jm}^2),$$

then

$$v_{ip} \leqslant \sqrt[2]{c_{ii} + M}, \quad i = 1:n, \quad j = 1:i. \tag{16}$$

That is, the elements of L_w (or L_x, L_y, L_z) cannot become too large.

5. Numerical results

We now present the results of numerical experiments for solving Eqs. (1) and (11). All experiments were performed in MATLAB on a PC-386 computer.

Take

$$A_m = [a_{ij}] = H_m + I_m \in \mathbb{R}^{m \times m},$$

where

$$H_m = \left[\frac{1}{i + j - 1}\right]$$

is an $m \times m$ Hilbert matrix and I_m is an $m \times m$ unit matrix.

Also take

$$B = [b_{ij}] = [\max(i, j)] \in \mathbb{R}^{n \times m}$$

and

$$C_n = [c_{ij}] = U_n \sigma_n U_n^{\mathrm{T}} \in \mathbb{R}^{n \times n}.$$

where

$$U_n = I_n - \frac{2}{w^T w} w w^T, \quad w = (1:n)^T$$

and

$$\sigma_n = \mathrm{diag}(1, 2, \ldots, n-1, 0).$$

Thus, matrices A_m, B and C_n satisfy Condition I.

Example 1. Solve the linear system

$$\begin{pmatrix} A_m & B^T \\ B & -C_n \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix},$$

where

$$f_i = \sum_{j=1}^{m} a_{ij} \times j + \sum_{k=1}^{n} b_{ki} \times (m+k), \quad i = 1:m,$$

$$g_i = \sum_{j=1}^{m} b_{ij} \times j - \sum_{k=1}^{n} c_{ki} \times (m+k), \quad i = 1:n.$$

The vectors

$$x = \begin{pmatrix} u \\ p \end{pmatrix}, \quad x^* = \begin{pmatrix} u^* \\ p^* \end{pmatrix}$$

denote the computed solution and the exact solution, respectively.

Example 2. Consider the linear system

$$\begin{pmatrix} A_m & -B^T \\ B & C_n \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}.$$

Table 1
The result of Example 1 [a]

Order		Algorithm 1		Gauss elimination	
m	n	Flops	Norm $(x-x^*)$	Flops	Norm $(x-x^*)$
10	10	3998	9.4259×10^{-12}	8207	2.8856×10^{-12}
20	10	11 533	3.4882×10^{-11}	24 125	1.0046×10^{-11}
30	20	50 321	4.7859×10^{-10}	100 187	2.2251×10^{-10}
50	30	188 938	6.1818×10^{-9}	384 447	1.7993×10^{-9}
50	40	267 548	1.7401×10^{-8}	540 643	4.4792×10^{-9}
50	50	344 524	2.0480×10^{-8}	733 965	9.4886×10^{-9}

[a] The Gauss elimination is provided by MATLAB.

Table 2
The result of Example 2 [a]

Order		Algorithm 2		Gauss elimination	
m	n	Flops	Norm $(x-x^*)$	Flops	Norm $(x-x^*)$
10	10	3998	6.7242×10^{-12}	8207	1.8781×10^{-12}
20	10	11 533	2.5209×10^{-11}	24 125	1.0121×10^{-11}
30	20	50 321	5.2676×10^{-10}	100 187	1.9155×10^{-10}
50	30	188 938	6.3810×10^{-9}	384 447	1.3570×10^{-9}
50	40	267 548	8.7125×10^{-9}	540 643	8.1077×10^{-9}
50	50	344 524	1.0074×10^{-8}	733 965	1.5418×10^{-8}

[a] The Gauss elimination is provided by MATLAB.

where:

$$f_i = \sum_{j=1}^{m} a_{ij} \times j - \sum_{k=1}^{n} b_{ki} \times (m+k), \quad i = 1:m,$$

$$g_i = \sum_{j=1}^{m} b_{ij} \times j - \sum_{k=1}^{n} c_{ki} \times (m+k), \quad i = 1:n.$$

From the above result we can see that the generalized Cholesky method presented in this paper will be efficient enough also for practical application. In fact, it would be still efficient when $C_n = 0$ in linear systems (1) or (11).

References

[1] R.E. Bank, B.D. Welfert, H. Yserentant, A class of iterative methods for solving saddle point problems, Numer. Math. 56 (1990) 645–666.

[2] A. Björck, Least squares methods, in: P.G. Ciarlet, J.L. Lions (Eds.), Handbook of Numerical Analysis vol. 1: Solution of Equations in R^n, Elsevier, Amsterdam, 1990.

[3] J.H. Bramble, J.E. Pasciak, A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems, Math. Comp. 50 (1988) 1–17.

[4] H.C. Elmain, G.H. Golub, Inexact and preconditioned Uzawa algorithms for saddle point problems, Technical Report NA-93-02, Computer Science Department, Stanford University, 1993.

[5] G.H. Golub, C.F. Van Loan, Matrix Computations, 2nd ed., The Johns Hopkins University Press, Baltimore, 1989.

[6] G.H. Golub, M.L. Overton, The convergence of inexact Chebyshev and Richardson iterative methods for solving linear systems, Numer. Math. 53 (1988) 571–593.

[7] I. Gustafsson, A class of first order factorizations, BIT 18 (1978) 142–156.

[8] Y. Yuan, Numerical methods for nonlinear programming, Modern Mathematics Series, Shanghai Scientific and Technical Publishers, 1993.