

## A METHOD FOR UPDATING CHOLESKY FACTORIZATION OF A BAND MATRIX

Wei H. YANG

*Department of Applied Mechanics and Engineering Science, The University of Michigan,  
Ann Arbor, Michigan 48109, USA*

Received April 1976; Revised manuscript received June 1976

A method is presented for updating the Cholesky factorization of a band symmetric matrix modified by a rank-one matrix which has the same band width. Problems which could involve applications of such a method arise frequently in plasticity and structural optimization where repeated solutions of a band algebraic system with a changing matrix are needed. The Cholesky factorization of a stiffness matrix can be updated after modifying a local stiffness matrix which can be written as a sum of a few rank-one matrices. The number of operations required for the updating is of the order  $mn$  or less, where  $n$  is the dimension of the global matrix and  $m$  is its half band width (including the diagonal).

### 1. Introduction

The article [1] is an excellent summary of the methods developed in the past few years for updating the Cholesky factorization of a matrix after a rank-one modification. One method treats the special case of adding and deleting a row or a column of a matrix; it provides important improvements in the algorithms for linear [2] and nonlinear [3] programming problems.

The method presented here treats another special case where the matrices involved are banded. If a class of special problems arises frequently in applications, then an efficient method should be developed for this class. In the physical sciences described by partial differential equation band matrices frequently appear. The finite difference and the finite element approximations of the nonlinear differential equations involve variable matrices which need be updated and refractored as the computation proceeds. In particular, the need for solving the system of equations repeatedly after each change in a local stiffness matrix arises in plasticity, structural optimization and design modification problems [4]. Although many problems in these areas are regarded as solvable in principle, the task of numerical solution remains formidable or very costly at best. The method presented in this paper updates the Cholesky factors of the matrix directly and thus provides an efficient tool for further advances in these fields.

A local stiffness matrix is always a matrix of very small rank. It can be expressed as a sum of a few rank-one matrices, each having the same or smaller band width than the global matrix. When elements, however few, in a local stiffness matrix change, the solution is affected globally. The equations must be resolved for the new solution. The Cholesky factorization of the new matrix, a step toward the new solution, can be obtained directly by modifying the old Cholesky factorization with a sequence of a few rank-one matrices.

The problem of a rank-one matrix modification on a Cholesky factorization is stated as follows: Given a positive definite symmetric matrix  $A$  in the Cholesky factorized form [1]

$$A = LDL^t, \tag{1}$$

where  $L$  is a band lower triangular matrix with unit diagonal elements, and  $D$  is a positive diagonal matrix, find the matrices  $\bar{L}$  and  $\bar{D}$  of a new matrix  $\bar{A}$  given by

$$\bar{A} = LDL^t + \alpha zz^t, \tag{2}$$

such that  $\bar{A} = \bar{L}\bar{D}\bar{L}^t$ , where  $\alpha$  is a scalar and  $\mathbf{z}$  is a vector and  $\mathbf{z}\mathbf{z}^t$ , a rank-one matrix, has the same band width as that of  $A$ . All matrices have dimension  $n$ .

If  $\alpha > 0$ ,  $\bar{A}$  will be positive definite, and thus  $\bar{D}$  has positive diagonal elements. If  $\alpha < 0$ , there is possibility that  $\bar{A}$  may become indefinite beyond some value of  $\alpha$ . The algorithm developed in this paper detects such change of positive definiteness. In the updating procedure  $\bar{L}$  and  $\bar{D}$  are obtained directly from modifying  $L$  and  $D$  by the rank-one matrix  $\alpha\mathbf{z}\mathbf{z}^t$  with order  $n$  operations. Since the band width is preserved in the process, the method requires no additional storage.

### 2. Rank-one modification

We shall express  $\bar{A}$  in another form,

$$\bar{A} = [Lz]\hat{D}[Lz]^t, \quad \hat{D} = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \alpha \end{bmatrix}, \tag{3}$$

where  $\hat{D}$  is an  $(n+1) \times (n+1)$  diagonal matrix, and the  $(n+1) \times n$  matrix  $[Lz]^t$  looks like



$$\tag{4}$$

where  $2m+1$  is the full band width of  $A$ , and  $z_k \dots z_{k+m}$  are the only  $m+1$  nonzero elements of  $\mathbf{z}$  if  $\mathbf{z}\mathbf{z}^t$  has the same band width of  $A$ . There is no loss of generality with the assumption that the first nonzero element of  $\mathbf{z}$  is in the  $k$ th location since  $k$  may be assigned any integer value between 1 and  $n$ . The oblique band outlined in (4) represents the nonzero elements of  $L^t$ , and blank spaces in the matrix represent zero entries. In the computer program the oblique band is actually stored in a rectangular array, while the vector  $\mathbf{z}$  is stored separately. The presentation here in the regular matrix notation is for the convenience of analysis.

A sequence of orthogonal and elementary matrices will be chosen to operate on  $[Lz]$  and reduce it to a lower triangular matrix. Let  $Q_1$  be an  $(n+1) \times (n+1)$  Givens matrix [5] of the form





where

$$[L^{(1)} z^{(1)}]^t = U_1^t [L z]^t, \quad (18)$$

which looks the same as (4) except that  $z_k^{(1)} = 0$ . Moreover,  $D^{(1)}$  differs from  $D$  only at the  $k$ th diagonal element such that

$$d_k^{(1)} = d_k + \alpha z_k^2 \quad (19)$$

and

$$\alpha^{(1)} = \alpha d_k (1 + z_k^2) / (d_k + \alpha z_k^2). \quad (20)$$

The updating process is then repeated with an appropriate matrix  $U_2$  such that the element  $z_{k+1}^{(1)}$  will be eliminated. The nonzero elements of the vector  $z^{(2)}$  may not be less than that of  $z^{(1)}$ . They merely shift one position down toward the end of the vector. Continuing the process with a proper sequence of matrices  $\{U_i | i = 3, 4, \dots, n-k\}$ , the nonzero elements of  $z$  will be first shifted to the last  $m+1$  locations and then reduced one by one until all elements of  $z$  become zero. This will take exactly  $n-k$  steps. At the end of the  $(n-k)$ th step, the matrix  $\bar{A}$  will have the desired form

$$\bar{A} = [\bar{L} \ 0] \begin{bmatrix} \bar{D} \\ \alpha^{(n-k)} \end{bmatrix} [\bar{L} \ 0]^t = \bar{L} \bar{D} \bar{L}, \quad (21)$$

where  $\bar{D}$  is diagonal and  $\bar{L}$  is lower triangular with unit diagonal.

Note that the elements in  $\bar{D}$  are updated sequentially starting from  $d_k$ . If a diagonal element of  $D$  becomes negative or zero at any stage of the updating procedure,  $\bar{A}$  will not be positive definite. This usually corresponds to a physical instability phenomenon. The updating procedure should be terminated and the solution examined. If all elements in  $\bar{D}$  remain positive, the updating is regarded successful and the new Cholesky factorization is given by eq. (21).

### 3. Local stiffness matrix

A local stiffness matrix is first used by engineers to relate the forces and deformation of a single structural member in its most general configuration. The assemblage of local stiffness matrices of all members in a structure forms a global matrix equation governing the mechanical behavior of the structure. This idea is generalized for a continuous structure such as a plate or a shell [6] with an artificial set of grid system to discretize the continuous media. Further development leads to the finite element method for partial differential equations [7, 8]. The concept of a local stiffness matrix is still of great importance, although it is not usually considered in modern finite element analysis.

An example is given here to show that a local stiffness matrix can be expressed as a sum of a few rank-one matrices similar to the one in eq. (2). A similar situation arises in a partial differential equation when local values of the coefficient function change.

The first example is the constant-strain-triangle element in the plane stress problem of elasticity.

The  $6 \times 6$  local stiffness matrix [6] may be given by the product

$$\mathbf{K}^{(e)} = \mathbf{G} \mathbf{E} \mathbf{G}^t, \quad (22)$$

where  $\mathbf{G}$  is a  $6 \times 3$  matrix containing geometric information of the triangle in a cartesian coordinate system, and  $\mathbf{E}$  is a  $3 \times 3$  matrix containing physical information including the area, thickness and elastic moduli of the triangular element. When  $\mathbf{K}^{(e)}$  is expressed in  $n$ -dimensional space, we may embed the nonzero elements of  $\mathbf{G}$  in an  $n \times 3$  zero matrix. Hence,  $\mathbf{K}^{(e)} = \mathbf{G} \mathbf{E} \mathbf{G}^t$  may be regarded as an  $n \times n$  matrix and  $\mathbf{G}$  and  $n \times 3$  matrix with most of the elements equal to zero. The nonzero elements of  $\mathbf{G}$  fall in a certain band width. In the  $n$ -dimensional expression,  $\mathbf{E}$  remains unchanged. For obvious physical reasons,  $\mathbf{E}$  is positive definite. It can be diagonalized by an orthogonal transformation such that

$$\mathbf{E} = \mathbf{Q} \begin{bmatrix} \alpha_1 & & \\ & \alpha_2 & \\ & & \alpha_3 \end{bmatrix} \mathbf{Q}^t, \quad (23)$$

where  $\mathbf{Q}$  is a  $3 \times 3$  orthogonal matrix, and  $\alpha_1, \alpha_2, \alpha_3$  are positive numbers. Let  $\mathbf{Z}$  be an  $n \times 3$  matrix such that

$$\mathbf{Z} = \mathbf{G} \mathbf{Q} = [\mathbf{z}_1 \ \mathbf{z}_2 \ \mathbf{z}_3], \quad (24)$$

where  $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3$  are the three columns vectors of  $\mathbf{Z}$ .

The local stiffness matrix can be rewritten in  $n$ -dimensional space as

$$\mathbf{K}^{(e)} = \mathbf{Z} \begin{bmatrix} \alpha_1 & & \\ & \alpha_2 & \\ & & \alpha_3 \end{bmatrix} \mathbf{Z}^t = \sum_{i=1}^3 \alpha_i \mathbf{z}_i \mathbf{z}_i^t, \quad (25)$$

which is a sum of three rank-one matrices.

Modification of a global matrix by a local stiffness matrix is usually associated with changes in modulus or thickness of the element. It may be written in the form

$$\bar{\mathbf{K}} = \mathbf{K} \pm c \mathbf{K}^{(e)}, \quad (26)$$

where  $c$  is a constant. For the  $\mathbf{K}^{(e)}$  given in eq. (25) this is a rank-three matrix modification – or, equivalently, three rank-one modifications.

The second example is concerned with the finite difference approximation of the partial differential equation

$$-\nabla \cdot a(x, y) \nabla u(x, y) = f(x, y), \quad (27)$$

where  $\nabla$  is the 2-dimensional gradient operator,  $a(x, y)$  and  $f(x, y)$  are known functions, and  $u(x, y)$  is the unknown. The scheme of a staggered mesh with a constant mesh size is used, where the functions  $u(x, y)$  and  $f(x, y)$  are discretized at the mesh points (nodes) and the coefficient

$a(x, y)$  is discretized at the mid-point between two adjacent nodes. We have the finite difference equations in matrix form

$$\mathbf{A} \mathbf{u} = \mathbf{f}, \quad (28)$$

where the elements of the matrix  $\mathbf{A}$  are constants in terms of values of  $a(x, y)$  at the mid-points between every two adjacent nodes, and  $\mathbf{u}$  and  $\mathbf{f}$  are vectors. In eq. (28), proper boundary conditions on the differential equation have been assumed. Let the value of  $a(x, y)$  at the mid-point between the pair of nodes  $i$  and  $j$  be  $a_{ij}$ . A change of value from  $a_{ij}$  to  $a'_{ij}$  will cause four elements in  $\mathbf{A}$  to change values. The new matrix will have the form

$$\bar{\mathbf{A}} = \mathbf{A} + (a'_{ij} - a_{ij}) \mathbf{p} \mathbf{p}^t, \quad (29)$$

where  $\mathbf{p}^t = [0 \dots 0 \ 1 \ 0 \dots 0 \ -1 \ 0 \dots 0]$  is a row vector with only two nonzero elements: 1 in the  $i$ th and  $-1$  in the  $j$ th locations, respectively. Thus, a change of coefficient function at each point corresponds to a rank-one modification on the matrix  $\mathbf{A}$ .

Changes in the local stiffness matrix or in the local value of the coefficient of a differential equation are induced by changes of physical or geometrical properties of material. An example of physical property change is demonstrated in plasticity problems where local stress exceeds yield limit. The geometric property change is often encountered in design problems such as thickness changes of a plate. This type of small rank perturbations of a matrix arise in many other nonlinear and design problems.

#### 4. Final remarks

In the updating procedure the orthogonal matrices are preferred for reasons of numerical stability. They may be replaced by the elementary matrices to give a slight advantage in efficiency. The operation count when orthogonal matrices are used for a rank-one update is less than

$$(n-k)[(m-1)(2M+A) + 10M+2A+R], \quad (30)$$

where  $M$ ,  $A$ ,  $R$  denote multiplication, addition and square root respectively.

The efficiency of the method is clearly demonstrated by (30), considering the operations required for refactoring the band matrix, which is of the order  $nm^2$ . In some problems where the location of the changing local matrices can be anticipated, such as in the stress concentration problem in plasticity, the integer  $k$  may be maximized to dramatize the efficiency of the new method. Even with the most unfavorable choice of  $k$  the cost ratio favors the updating method by a wide margin when compared to other methods used in plasticity literature.

Another advantage of the method is its everreadiness of monitoring the positive definiteness of the changing matrix by the signs of the elements in  $\mathbf{D}$ . This is accomplished during the updating procedure with very little effort.

## Acknowledgement

The author acknowledges the support from the NSF Grant Eng. 75--17487.

## References

- [1] P.E. Gill, G.H. Golub, W. Murray and M. Saunders, Methods for modifying matrix factorizations. *Math. Comp.* 28 (1974) 505–535.
- [2] M.A. Saunders, Fast stable implementation of the simplex method using Bartel-Golub updating, *Proceedings of the Symposium on Sparse Matrix Computation* (Sept. 9–11, 1975).
- [3] W. Murray, An algorithm for indefinite quadratic programming (National Physical Laboratory, DNAC Report No. 1, Teddington, England, 1971).
- [4] J.H. Argyris and J.R. Roy, General treatment of structural modifications, *J. Struct. Div. Proc. ASCE* 98 (1972) 465–492.
- [5] J.H. Wilkinson, *The algebraic eigenvalue problem* (Oxford Univ. Press, 1965).
- [6] H.C. Martin and G.F. Carey, *Introduction to finite element analysis, theory and application* (McGraw-Hill, 1973).
- [7] J.T. Oden, *Finite elements of nonlinear continua* (McGraw-Hill, 1972).
- [8] G. Strang and G.F. Fix, *An analysis of the finite element method* (Prentice-Hall, 1973).