Appl. Math. Lett. Vol. 6, No. 5, pp. 79–84, 1993
Printed in Great Britain.
0893-9659/93 \$6.00 + 0.00

AN OPTIMUM PARTITION FOR INVERTING A NONSINGULAR MATRIX

JENN-CHING LUO
Department of Civil Engineering
National Chiao Tung University, Hsinchu, Taiwan 30050

(Received and accepted March 1993)

Abstract—A new strategy for inverting a nonsingular matrix is evaluated in this paper. The essential concept of this strategy is to partition a nonsingular matrix into blocks, then to apply the author's decomposition to the block-based procedure. Since different partitions require different costs, finding an economical partition is necessary. This paper studies the optimum partition so that the complexity for the inverse of a nonsingular matrix may be significantly reduced.

1. INTRODUCTION

The author's decomposition [1,2] provides a special feature in inverting a nonsingular matrix $[A]$, which decomposes $[A]$ directly into $[A]^{-1}$. For a general nonsingular matrix, the inverse of $[A]$ is written in the form $[A]^{-1} = [L][D][U]$, in which $[L]$ is a lower triangular matrix with unit coefficients, $[D]$ is a diagonal matrix, and $[U]$ is an upper triangular matrix with unit coefficients. The procedure for decomposing $[A]$ of order $(n \times n)$ into $[A]^{-1} = [L][D][U]$ is as follows [2]:

For $j = n \to 1$ with step (-1), do

(a) for $i = j+1 \to n$ with step 1, do

$$A_{ji} \leftarrow A_{ji} + \sum_{k=i+1}^{n} A_{jk} * A_{ki}; \tag{1}$$

(b) for $i = n \to j+1$ with step (-1), do

$$A_{ji} \leftarrow -A_{ji} * A_{ii} - \sum_{k=j+1}^{i-1} A_{jk} * A_{kk} * A_{ki}; \tag{2}$$

(c) $$A_{jj} \leftarrow \frac{1}{A_{jj} + \sum_{k=j+1}^{n} A_{jk} * A_{kj}}; \tag{3}$$

(d) for $i = j+1 \to n$ with step 1, do

$$A_{ij} \leftarrow A_{ij} + \sum_{k=i+1}^{n} A_{ik} * A_{kj}; \tag{4}$$

(e) for $i = n \to j+1$ with step (-1), do

$$A_{ij} \leftarrow -A_{ii} * A_{ij} - \sum_{k=j+1}^{i-1} A_{ik} * A_{kk} * A_{kj}. \tag{5}$$

After the computation of equations (1)–(5), the lower triangular part of $[A]$ is $[L]$, the diagonal part of $[A]$ is $[D]$, and the upper triangular part of $[A]$ is $[U]$. It has been proved [3] that the complexity in equations (1)–(5) can be reduced by a block-based procedure. The purpose of this work is to study the optimum number of blocks so that the complexity in inverting a nonsingular matrix may be significantly reduced.

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

2. A BLOCK-BASED PROCEDURE

If the matrix $[A]$ is partitioned into a $(N \times N)$-block matrix, the block-based version of the author's decomposition is written as:

For $J = N \to 1$ with step (-1), do

(a) for $I = J+1 \to N$ with step 1, do

$$[A_{JI}] \leftarrow [A_{JI}] + \sum_{K=I+1}^{N} [A_{JK}][A_{KI}]; \tag{6}$$

(b) for $I = N \to J+1$ with step (-1), do

$$[A_{JI}] \leftarrow -[A_{JI}][A_{II}] - \sum_{K=J+1}^{I-1} [A_{JK}][A_{KK}][A_{KI}]; \tag{7}$$

(c) $$[A_{JJ}] \leftarrow \left[[A_{JJ}] + \sum_{K=J+1}^{N} [A_{JK}][A_{KJ}] \right]^{-1}; \tag{8}$$

(d) for $I = J+1 \to N$ with step 1, do

$$[A_{IJ}] \leftarrow [A_{IJ}] + \sum_{K=I+1}^{N} [A_{IK}][A_{KJ}]; \tag{9}$$

(e) for $I = N \to J+1$ with step (-1), do

$$[A_{IJ}] \leftarrow -[A_{II}][A_{IJ}] - \sum_{K=J+1}^{I-1} [A_{IK}][A_{KK}][A_{KJ}]; \tag{10}$$

where $[A_{IJ}]$ is a submatrix of $[A]$. The results obtained from equations (6)–(10) represent that

$$[L] = \begin{bmatrix} [I] & & & \\ [A_{21}] & [I] & & \\ \vdots & \vdots & \ddots & \\ [A_{N1}] & [A_{N2}] & \cdots & [I] \end{bmatrix}, \tag{11}$$

$$[D] = \begin{bmatrix} [A_{11}] & & & \\ & [A_{22}] & & \\ & & \ddots & \\ & & & [A_{NN}] \end{bmatrix}, \tag{12}$$

$$[U] = \begin{bmatrix} [I] & [A_{12}] & \cdots & [A_{1N}] \\ & [I] & \cdots & [A_{2N}] \\ & & \ddots & \vdots \\ & & & [I] \end{bmatrix}, \tag{13}$$

and the product of $[L][D][U]$ is $[A]^{-1}$. The inverse operation of the right side in equation (8) first implements the element-based procedure as shown in equations (1)–(5) so that $[A]^{-1}$ is in the form of $[L][D][U]$, and then performs the product of $[L][D][U]$ to obtain a single matrix representation for $[A]^{-1}$.

This work partitions matrix $[A]$ of order $(M \times M)$ into a $(N \times N)$-block matrix. Certainly, there exists a variety of partition patterns. This work will try a balance partition, i.e., each block has as equal size as possible. Generally, M is not a multiple of N. Because of balance partition, the difference between any blocks is at most 1; for example, some blocks may be of size n, while others may be of size $n+1$. In this work, we just write the block size n as

$$n = \frac{M}{N}. \tag{14}$$

It can be counted that the procedure in equations (6)–(10) requires N inverse operations, $\frac{2}{3}N^3 - \frac{1}{2}N^2 - \frac{1}{6}N$ matrix additions/subtractions, and $N^3 - \frac{3}{2}N^2 + \frac{1}{2}N$ matrix multiplications.

3. THE OPTIMUM PARTITION

RECALL 1. *The number of arithmetic operations in inverting an asymmetric matrix* $[A]$ *of order* $(n \times n)$ *into* $[A]^{-1}$ *is* $\frac{7}{3}n^3 - 2n^2 + \frac{2}{3}n$ *(see [3]).*

FACT 1. The number of arithmetic operations for the product of 2 matrices of order $(n \times n)$ is $2n^3 - n^2$, and the number of arithmetic operations for the addition of 2 matrices of order $(n \times n)$ is n^2.

Since a $(N \times N)$-block procedure has N inverse operations, $\frac{2}{3}N^3 - \frac{1}{2}N^2 - \frac{1}{6}N$ addition/subtraction operations, and $N^3 - \frac{3}{2}N^2 + \frac{1}{2}N$ multiplication operations, the cost in implementing a $(N \times N)$-block procedure is

$$\begin{aligned} C &= N\left(\frac{7}{3}n^3 - 2n^2 + \frac{2}{3}n\right) + \left(\frac{2}{3}N^3 - \frac{1}{2}N^2 - \frac{1}{6}N\right)n^2 + \left(N^3 - \frac{3}{2}N^2 + \frac{1}{2}N\right)(2n^3 - n^2) \\ &= \left(2N^3 - 3N^2 + \frac{10}{3}N\right)n^3 + \left(-\frac{1}{3}N^3 + N^2 - \frac{8}{3}N\right)n^2 + \left(\frac{2}{3}N\right)n. \end{aligned} \tag{15}$$

Substituting equation (14) into (15) yields

$$C = -\frac{M^2}{3}N - \left(3M^3 + \frac{8}{3}M^2\right)\frac{1}{N} + \left(\frac{10}{3}M^3\right)\frac{1}{N^2} + \left(2M^3 + M^2 + \frac{2}{3}M\right). \tag{16}$$

The optimum partition may be analyzed by the first and second order derivatives of C with respect to N, which can be written as

$$\frac{dC}{dN} = -\frac{M^2}{3} + \left(3M^3 + \frac{8}{3}M^2\right)\frac{1}{N^2} - \frac{20M^3}{3}\frac{1}{N^3}, \tag{17}$$

$$\frac{d^2C}{dN^2} = \frac{2}{3}\frac{M^2}{N^4}(30M - 9MN - 8N). \tag{18}$$

Then, the optimum N is the solution of $\frac{dC}{dN} = 0$ subject to the condition that $\frac{d^2C}{dN^2} > 0$. The condition that $\frac{dC}{dN} = 0$ yields

$$N^3 - (9M + 8)N + 20M = 0. \tag{19}$$

Generally, the solution N to equation (19) may not be an integer. The optimum partition N is the integer closest to the solution of equation (19). The condition that $\frac{d^2C}{dN^2} > 0$ leads to

$$N < \frac{30M}{9M + 8}. \tag{20}$$

Then, the optimum partition N may be written in Lemmas 1–3.

LEMMA 1. *The optimum partition* N *is defined as follows:*

(a) *N is the integer closest to the solution of* $N^3 - (9M + 8)N + 20M = 0$,

(b) $1 \le N < \dfrac{30M}{9M + 8}$,

in which $M \ge 2$ *is the order of matrix.*

LEMMA 2. *For a given* $M \ge 2$, *the optimum partition* N *exists.*

PROOF. Let $F(N) = N^3 - (9M + 8)N + 20M$. Then, $F(1) = 11M - 7 > 0$, since $M \ge 2$. Furthermore,

$$F\left(\frac{30M}{9M + 8}\right) = \frac{-1}{(9M + 8)^3}\left(M^3(7290M - 7560) + 17280M^2 + 5120M\right),$$

in which the value in each () is positive. This means that $F(30M/(9M+8)) < 0$. Therefore, $F(1) * F(30M/(9M+8)) < 0$, which implies that there exist one or three roots of $F(N) = 0$ between 1 and $30M/(9M+8)$. However, by *Descartes's rule of signs*, there exists one negative root so that it is impossible to have three roots between 1 and $30M/(9M+8)$. This implies that there exists exactly one root between 1 and $30M/(9M+8)$. ∎

LEMMA 3. *For a given $M \geq 2$, the optimum partition N is 2.*

PROOF. By Lemma 2, there exists exactly one root between 1 and $30M/(9M+8)$ such that $F(N) = 0$. Since

$$\frac{30M}{9M+8} - \frac{20}{9} = \frac{90M-160}{9(9M+8)} > 0, \tag{21}$$

where $M \geq 2$, so that $1 < \frac{3}{2} < \frac{20}{9} < 30M/(9M+8)$. Furthermore, $F(\frac{20}{9}) = -(4960/9^3) < 0$, and $F(\frac{3}{2}) = (52M-69)/8 > 0$ since $M \geq 2$. This shows that $F(\frac{3}{2}) * F(\frac{20}{9}) < 0$, and implies that the optimum partition N is between $\frac{3}{2}$ and $\frac{20}{9}$. Therefore, the closest integer N to the interval between $\frac{3}{2}$ and $\frac{20}{9}$ is 2. This completes the proof. ∎

4. DISCUSSION

This paper proves that for a given nonsingular matrix, the optimum partition is $N = 2$, such that equation (16) becomes $C = \frac{4}{3}M^3 - M^2 + \frac{2}{3}M$ and the block procedure as shown in equations (6)–(10) is simplified as:

$$\text{(a) } [A_{22}] \leftarrow [A_{22}]^{-1}, \tag{22}$$
$$\text{(b) } [A_{12}] \leftarrow -[A_{12}][A_{22}], \tag{23}$$
$$\text{(c) } [A_{11}] \leftarrow \left[[A_{11}] + [A_{12}][A_{21}]\right]^{-1}, \tag{24}$$
$$\text{(d) } [A_{21}] \leftarrow -[A_{22}][A_{21}]. \tag{25}$$

The optimum partition also requires two inverse operations as shown in equations (22) and (24). Certainly, we may apply other optimum partitions to invert submatrices, and we may repeatedly apply an optimum partition to the inverse of submatrices in each level. This forms a recurrent optimum strategy for the inverse of a nonsingular matrix. A recurrent optimum partition may make a significant reduction in arithmetic operations. For the example of $[A_{22}]$ of order $(\frac{M}{2} \times \frac{M}{2})$ in equation (22), applying an optimum partition for inverting $[A_{22}]$ into $[A_{22}]^{-1}$ in a single matrix representation requires

$$\left[\frac{4}{3}\left(\frac{M}{2}\right)^3 - \left(\frac{M}{2}\right)^2 + \frac{2}{3}\left(\frac{M}{2}\right)\right] + \left[\frac{2}{3}\left(\frac{M}{2}\right)^3 - \frac{2}{3}\left(\frac{M}{2}\right)\right] + 2\left[\frac{2}{3}\left(\frac{M}{4}\right)^3 - \frac{2}{3}\left(\frac{M}{4}\right)\right] = \frac{13}{6}\left(\frac{M}{2}\right)^3 - \left(\frac{M}{2}\right)^2 - \frac{2}{3}\left(\frac{M}{2}\right) \tag{26}$$

arithmetic operations, in which the first [•] is the cost for an optimum partition, the second [•] is the cost for converting $[A_{22}]^{-1}$ from the form $[L][D][U]$ into a single matrix representation (see [3]), and the third [•] is the cost for converting the inverse of two submatrices into single matrix representations. Apparently, applying an optimum partition to equation (22) may further reduce the arithmetic operations in the amount

$$\left[\frac{7}{3}\left(\frac{M}{2}\right)^3 - 2\left(\frac{M}{2}\right)^2 + \frac{2}{3}\left(\frac{M}{2}\right)\right] - \left[\frac{13}{6}\left(\frac{M}{2}\right)^3 - \left(\frac{M}{2}\right)^2 - \frac{2}{3}\left(\frac{M}{2}\right)\right] = \frac{1}{6}\left(\frac{M}{2}\right)^3 - \left(\frac{M}{2}\right)^2 + \frac{4}{3}\left(\frac{M}{2}\right). \tag{27}$$

Therefore, repeatedly applying an optimum partition to the inverse of each submatrix in each level may reduce the arithmetic operations in the amount

$$\begin{aligned}
&2\left[\frac{1}{6}\left(\frac{M}{2}\right)^3-\left(\frac{M}{2}\right)^2+\frac{4}{3}\left(\frac{M}{2}\right)\right]+2^2\left[\frac{1}{6}\left(\frac{M}{2^2}\right)^3-\left(\frac{M}{2^2}\right)^2+\frac{4}{3}\left(\frac{M}{2^2}\right)\right]\\
&\quad+\cdots+2^{\log_2(M/2)}\left[\frac{1}{6}\left(\frac{M}{2^{\log_2(M/2)}}\right)^3-\left(\frac{M}{2^{\log_2(M/2)}}\right)^2+\frac{4}{3}\left(\frac{M}{2^{\log_2(M/2)}}\right)\right]\\
&=\frac{1}{18}M^3-M^2+\frac{17}{18}M+\frac{4}{3}M\log_2 M,
\end{aligned} \tag{28}$$

and the cost for inverting $[A]$ becomes

$$\begin{aligned}
C&=\left[\frac{4}{3}M^3-M^2+\frac{2}{3}M\right]-\left[\frac{1}{18}M^3-M^2+\frac{17}{18}M+\frac{4}{3}M\log_2 M\right]\\
&=\frac{23}{18}M^3-\frac{5}{18}M-\frac{4}{3}M\log_2 M.
\end{aligned} \tag{29}$$

Since Gaussian elimination takes $\frac{8}{3}M^3+O(M^2)$ operations for the inverse of a nonsingular matrix, the presented method is superior to the Gaussian elimination for the inverse. Comparing with the Strassen algorithm [4] for the inverse of a matrix requiring approximately $5.64M^{2.8}$ arithmetic operations, the presented method is also superior to Strassen's algorithm for a certain range of M, for example, $M \le 1675$. A combination of Strassen algorithm (for $M > 1675$) and the presented method (for $M \le 1675$) may provide a mixed procedure for the inverse of a nonsingular matrix. However, in practical applications we always try to avoid inverting a dense large-scaled matrix, but we used to invert small matrices for large-scaled applications; for an example to general iterative method with a tridiagonal block matrix $[C]$ as

$$[C]=\begin{bmatrix}[C_{11}] & [C_{12}] & & & \\ [C_{21}] & [C_{22}] & [C_{23}] & & \\ & [C_{32}] & [C_{33}] & [C_{34}] & \\ & & \ddots & \ddots & \ddots\end{bmatrix},$$

where each diagonal submatrix $[C_{ii}]$ is full and of a small order, a preconditioner $[P]$ may be defined as

$$[P]=\begin{bmatrix}[C_{11}]^{-1} & & \\ & [C_{22}]^{-1} & \\ & & \ddots\end{bmatrix},$$

in which each inverse can be implemented by the presented method.

This paper proposes a smart procedure as shown in equations (22)–(25) for the inverse of a nonsingular matrix, by which the complexity can be significantly reduced. The procedures in equations (22)–(25) are clear and easy to program. The inverse of submatrices in equation (22) and (24) can be efficiently and accurately implemented by the original version of the author's decomposition as shown in equations (1)–(5); for example, let us consider an ill-conditioned matrix $[A]$ of order (500×500) in which $A_{ii} = 1.0$, $A_{ij} = 1.0 + \epsilon$ $(i > j)$, and $A_{ij} = 1.0 - \epsilon$ $(i < j)$ with a tiny ϵ. We understand that if $\epsilon = 0$, then matrix $[A]$ is singular, and for a tiny ϵ the matrix $[A]$ in this example is ill-conditioned. In this example, let $\epsilon = 10^{-7}$. The 4-byte computation with 7 digits (single precision) is not suited for this example because round-off error will truncate $(1.0 \pm \epsilon)$ into 1.000000 such that $[A]$ becomes singular. We have to use 8-byte computation (double precision) for this example. Define $b_i = \sum_{j=1}^{500} A_{ij}$ the i^{th} coefficient of $\{B\}$. Then, the exact solution $\{X\}$ to $[A]\{X\} = \{B\}$ is with unit coefficients. The author's decomposition accurately solves this example, and all the obtained 500 coefficients of $\{X\}$ are unit. More experimental tests with pseudo random procedures for generating matrix $[A]$ show that double precision may provide a high accurate result of the new class of decomposition. Another way to improve accuracy can be achieved by the equations of round-off error [5]. The inverse of a matrix is very useful, but expensive. As emphasized in [6], it may be worthwhile to calculate $[A]^{-1}$ explicitly for the solution of $[A]\{X\} = \{B\}$ when exploiting parallelism. Many practical applications will lead to a sparse matrix $[A]$. Certainly, $[A]^{-1}$ may be a full matrix, even though $[A]$ is in a sparse configuration. The conventional method, i.e., Gaussian elimination and Strassen algorithm, can not take advantage of sparsity for dealing with the inverse. Because of writing $[A]^{-1} = [L][D][U]$, the author's decomposition also provides a possibility to keep $[A]$, $[L]$, and $[U]$ in the same sparse configuration. It has been found [7,8] that a stairs-shape sparsity is well-suited for the author's decomposition, which can be an efficient sparse solver. The comparisons [9] between the author's decomposition and the Gaussian elimination in solving sparse systems of linear equations showed that the author's decomposition also provides a faster solution. The author's decomposition not only can provide an optimum partition for dense matrices, but also can take advantage of sparsity for the inverse. It may be a fundamental tool in scientific and engineering computing.

REFERENCES

1. J.-C. Luo, A new class of decomposition for symmetric systems, *Mechanics Research Communications* **19**, 159–166 (1992).
2. J.-C. Luo, A new class of decomposition for inverting asymmetric and indefinite matrices, *Computers Math. Applic.* **25** (4), 95–104 (1993).
3. J.-C. Luo, A proof in favor of block-based procedure, *Appl. Math. Lett.* **5** (5), 75–78 (1992).
4. V. Strassen, Gaussian elimination is not optimal, *Numer. Math.* **13**, 354–356 (1969).
5. J.-C. Luo, Analysis of round-off errors, *Appl. Math. Lett.* **5** (4), 77–80 (1992).
6. I.S. Duff, A.M. Erisman and J.K. Reid, *Direct Methods for Sparse Matrices*, Oxford University Press, New York, (1986).
7. J.-C. Luo, A sparsity for decomposing a symmetric matrix, *Computers Math. Applic.* **25** (5), 83–90 (1993).
8. J.-C. Luo, Extended concept of stair-shape sparsity for the inverse of an asymmetric matrix, *Computers Math. Applic.* **25** (5), 113–122 (1993).
9. J. Luo, Parallel algorithms for the solution of a sparse system, Presented at the *Second Symposium on Parallel Computational Methods For Large-Scale Structural Analysis and Design*, Marriot Hotel, Norfolk, VA, February 24–25, 1993.
10. A. Householder, *The Numerical Treatment of a Single Nonlinear Equation*, McGraw-Hill, Highstown, NJ, (1970).