A Stable Method for the LU-Decomposition of M-Matrices

by

ALAN ALBERT AHAC
B.Sc., University of Victoria, 1983

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

[redacted]

in the Department of Computer Science

We accept this thesis as conforming to the required standard

[redacted]

Dr. D. Dale Olesky

[redacted]

Dr. Frank D.K. Roberts

[redacted]

Dr. Denton E. Hewgill

[redacted]

Dr. Ahmed R. Sourour

UNIVERSITY OF VICTORIA
April 1985

Supervisor: Dr. D. Dale Olesky

Abstract

We present an algorithm for the LU-decomposition of M-matrices based upon Gaussian Elimination applied with a new pivoting strategy. At each step of the elimination, the pivoting strategy selects a column that is the most (column) diagonally dominant in the unreduced submatrix and exchanges it into the pivotal column position through a symmetric permutation on the matrix. We demonstrate that this approach is well-suited to M-matrices, and can be implemented efficiently. The stability of the method is shown by providing a bound on the growth factor associated with the backward error analysis of the Gaussian Elimination algorithm.

We provide background for our results by surveying the literature on M-matrices, describing characterizations of matrices of this type and noting previous work regarding the LU factorization of M-matrices. Some applications in which M-matrices occur are also given. Finally, we discuss the extension of our algorithm to a larger class of matrices known as H-matrices.

Examiners:

[redacted]

Dr. D. Dale Olesky

[redacted]

Dr. Frank D.K. Roberts

[redacted]

Dr. Denton E. Hewgill

[redacted]

Dr. Ahmed R. Sourour

Table of Contents

CHAPTER 1

Introduction

One of the most common problems encountered in numerical computing is to find a solution to a system of n linear equations with n unknown variables. A method of solution for this problem known as Gaussian Elimination has in fact been used for a number of centuries, and most people with any advanced mathematical training have encountered both the problem of solving the systems of linear equations along with the method of Gaussian Elimination. Many basic principles of matrix theory and linear algebra are related to the method of Gaussian Elimination.

A system of linear equations is commonly expressed as

$$Ax = b\,, \tag{1.1}$$

where A is an $n \times n$ *coefficient matrix,* b is the *right-hand side vector* of length n, and x is the *vector of unknowns* having length n. The matrix A and vector b consist of known values, and we are required to find the values for vector x. While there are a number of different methods for solving a system of linear equations, they can be categorized into two classes - iterative methods and direct methods. Gaussian Elimination is a type of direct method. Selection of an appropriate method for solving a given linear system usually depends on knowledge of the characteristics of the coefficient matrix A and on the accuracy and efficiency requirements of the problem. Computer programs implementing many different methods are available.

Our discussions will be directed at solving (1.1) when the coefficient matrix A is from the class of *M-matrices.* This class of matrices has a special structure and many characteristics that are useful in algebraic and matrix computation. M-matrices have a wide range of applications in such areas as numerical analysis, probability, economics and operations research. There has been considerable research in M-matrix theory, both in terms of characterization of the matrices in this class, as well as with the solution of problems involving M-matrices. A great deal of work has been done with iterative methods of solution for large sparse systems.

The primary purpose of this paper is to describe a direct method for the LU-decomposition of an M-matrix. The method we propose is a variant on the Gaussian Elimination algorithm that incorporates a pivoting strategy that is particularly well-suited to M-matrices. We will demonstrate that the method is stable as well as efficient compared with other known variations on Gaussian Elimination. The algorithm may be easily implemented in a computer program.

As well, this paper will provide a survey of M-matrix theory and a description of the applications where M-matrices are commonly found. Further, we will describe the extension of the class of M-matrices to a class known as *H-matrices,* and will discuss the possible extension of our algorithm to this larger class of matrices.

Chapter 2 will provide the rather extensive foundation of general matrix theory necessary for the development of our discussion of M-matrices. Definitions of required matrix terminology will be provided, and the concepts involved with methods of solution for systems of linear equations will be described in detail. Important issues in the use of Gaussian Elimination will be illustrated and conditions for the stability of the method will be provided.

Chapter 3 will survey the research that has been done in M-matrix theory and describe the applications in which M-matrices occur. The important characterizations of M-matrices will be described in some length. A brief historical perspective on the development of the theory of M-matrices is included. M-matrix applications are noted with brief description and references to other sources of these applications are given. Four of the more illustrative applications are described in more detail.

Chapter 4 will describe our main results regarding a variation on Gaussian Elimination for the LU-decomposition of M-matrices. We will describe the current state of research regarding LU-decompositions of M-matrices in the literature, and provide with proof, the theorems motivating our results. A rigorous proof of the stability of the method is also given, and illustrative examples are included.

Chapter 5 will briefly define and describe the class of matrices known as H-matrices. This class of matrices will be characterized, and we will discuss the extension of our method of solution from Chapter 4 to this larger class of matrices.

We will discuss and summarize the results of the research outlined in this paper in Chapter 6. Conclusions and suggestions for further work will be described.

CHAPTER 2

Terminology and Concepts

2.1. Definitions

Unless otherwise mentioned explicitly, we will let $A=(a_{i,j})$ and $B=(b_{i,j})$ be real matrices of order n. For the purposes of comparison of vector and matrix elements we will write

$$A \geq B \quad \text{if } a_{i,j} \geq b_{i,j} \quad \text{for all } i,j,$$

$$A > B \quad \text{if } A \geq B \text{ and } A \neq B,$$

and

$$A >> B \quad \text{if } a_{i,j} > b_{i,j} \quad \text{for all } i,j.$$

We will use the same notation for vectors as defined above for matrices. A matrix B such that $B \geq 0$ is called a *nonnegative matrix.*

We will denote the determinant of A by $\det(A)$, the transpose of A by A^T, and the spectral radius of A by $\rho(A)$.

If α,β are strictly increasing sequences in $<n>=\{1,2, \ldots, n\}$, then the *submatrix* $A[\alpha,\beta]$ consists of those elements in the rows of A specified by α and the columns of A specified by β.

A *principal submatrix* is a submatrix $A[\alpha,\beta]$ with $\alpha=\beta$. We denote the principal submatrix $A[\alpha,\alpha]$ as $A[\alpha]$.

A *leading principal submatrix* is a principal submatrix $A[\alpha]$ with $\alpha=\{1,2, \ldots, n-r\}$ for some $r \leq n$.

The determinant of a (leading) principal submatrix is called a *(leading) principal minor.*

A *permutation matrix* is a square matrix which has precisely one entry equalling unity in each row and each column, with all other entries zero. An *elementary permutation matrix* $P_{i,j}$ is the identity matrix with the rows i and j exchanged.

Given a matrix A and a permutation matrix P, the matrix PAP^T is called a *symmetric* or *simultaneous permutation* on A.

For $n \geq 2$, a matrix A is *reducible* if there exists a permutation matrix P such that

$$PAP^T = \begin{bmatrix} A_{1,1} & A_{1,2} \\ 0 & A_{2,2} \end{bmatrix},$$

where $A_{1,1}$ is an $r \times r$ submatrix and $A_{2,2}$ is an $(n-r)\times(n-r)$ submatrix, $1 \leq r < n$. If no such permutation exists, then A is *irreducible*. If A is of order 1, then A is irreducible if its entry is nonzero, and reducible otherwise.

Given a matrix A, we let $G(A)$ denote the *directed graph* on n distinct vertices associated with A. If $a_{i,j} \neq 0$ in A then there exists an arc from vertex i to vertex j in $G(A)$. More generally, there exists a *path* from vertex i to vertex j if there is a set

$$\{a_{k_r,k_{r+1}}\}_{r=1}^{l} \text{ with } l \geq 1, \ a_{k_r,k_{r+1}} \neq 0$$

and $k_1 = i$, $k_{l+1} = j$.

A directed graph $G(A)$ is *strongly connected* if, for all distinct vertices i and j, there exists a path from vertex i to vertex j. The significance of strong connectedness in matrix theory is that a matrix A is irreducible if and only if its directed graph $G(A)$ is strongly connected (see, for example, pp. 20 in Varga [1962]).

For a $n \times n$ matrix A, let $\tilde{A}$ denote the *reduced normal form* of A. That is,

$$\tilde{A} = PAP^T = \begin{bmatrix} \tilde{A}_{1,1} & \tilde{A}_{1,2} & \cdot & \cdot & \tilde{A}_{1,k} \\ & \tilde{A}_{2,2} & \cdot & \cdot & \tilde{A}_{2,k} \\ & & \cdot & & \cdot \\ & 0 & & \cdot & \cdot \\ & & & & \tilde{A}_{k,k} \end{bmatrix}, \tag{2.1}$$

where P is a permutation matrix and $\tilde{A}_{j,j}$ is either a square irreducible submatrix or a zero matrix of order one, $1 \le j \le k$.

An *elementary lower triangular* matrix of order n and index k is a matrix of the form

$$M_k = I_n - me_k^T, \tag{2.2}$$

where I_n denotes the $n \times n$ identity matrix, m is some vector and e_i is a vector with unity in position i and zero in all other positions and

$$e_i^T m = 0 \text{ for } 1 \le i \le k.$$

A matrix A is *row diagonally dominant* if

$$|a_{i,i}| \ge \sum_{\substack{j=1 \\ j \ne i}}^{n} |a_{i,j}| \qquad \text{for all } i. \tag{2.3}$$

A matrix is *strictly row diagonally dominant* if strict inequality holds in (2.3) for all i. Similarly, matrix A is *column diagonally dominant* if

$$|a_{j,j}| \ge \sum_{\substack{i=1 \\ i \ne j}}^{n} |a_{i,j}| \qquad \text{for all } j, \tag{2.4}$$

and *strictly column diagonally dominant* if strict inequality holds.

A symmetric matrix A is *positive definite* if

$$x \ne 0 \Rightarrow x^T Ax > 0,$$

and *positive semidefinite* if

$$x \ne 0 \Rightarrow x^T Ax \ge 0.$$

Let A be a matrix and $A[\alpha]$ a principal submatrix of A. Let $\hat{A}$ be a symmetric permutation of A so that

$$\hat{A} = PAP^T = \begin{bmatrix} A[\alpha] & \hat{A}_{1,2} \\ \hat{A}_{2,1} & A[<n>\setminus \alpha] \end{bmatrix}.$$

Let k denote the cardinality of α. If $A[\alpha]$ is nonsingular, then the *Schur Complement of* $A[\alpha]$ *in* A is the $(n-k)\times(n-k)$ matrix $(A/A[\alpha])$ given by

$$(A/A[\alpha]) = A[<n>\setminus\alpha] - \hat{A}_{2,1}(A[\alpha])^{-1}\hat{A}_{1,2}. \tag{2.5}$$

A matrix A is ***positive stable*** if the real part of each eigenvalue of A is positive.

A matrix A is *semipositive* if there exists a vector $x >> 0$ such that $Ax >> 0$.

A matrix A is ***inverse-positive*** if A^{-1} exists and $A^{-1} \geq 0$ ($a_{i,j} \geq 0$ for all i,j), or equivalently matrix A is *monotone* if

$$Ax \geq 0 \Rightarrow x \geq 0 \text{ for all } x \in R^n.$$

A is *almost monotone* if

$$Ax \geq 0 \Rightarrow Ax = 0 \text{ for all } x \in R^n.$$

So far we have described terminology relating to the characterization of matrices. We now describe some concepts and terms relating to algorithms.

In discussing the computer implementation of numerical algorithms, we will use the following functional notation. The numerical algorithm is defined by some mathematical formula or process that takes data from, say, the domain R^m and produces some result in a range R^n (for example). We can think of this algorithm, then, as a function $f: R^m \to R^n$. Unfortunately, the computer representation of the initial data $x \in R^m$ only approximates that data, and the computer implementation of the algorithm is not precisely the mathematical definition f, so at best we are only approximating the solution $f(x)$. We are interested in investigating conditions for the solution produced by a computer implementation of an algorithm to be, in some sense, "near" the actual solution $f(x)$.

As noted, we usually have only an approximation x^* to the actual data x, so at best we can calculate only $f(x^*)$. If a problem for which x^* is "near" x results in $f(x)$ not "near" to

$f(x^*)$, we say the problem is *ill-conditioned.* If we attempt to solve an ill-conditioned problem starting with inexact data, the solution is likely to be inexact regardless of how it is computed.

The implementation of an algorithm to solve a mathematical problem described by f amounts to defining a new function f^*, that, given data x, produces an approximate solution $f^*(x)$ to $f(x)$. We would hope that if our problem is not ill-conditioned that $f^*(x)$ would be near $f(x)$. We say that an algorithm f^* is *stable* if for any $x \in R^m$ there is a nearby $x^* \in R^m$ such that $f(x^*)$ is near $f^*(x)$. In other words the algorithm is stable if it yields a solution that is near the exact solution of a slightly perturbed problem.

If we apply a stable algorithm to an ill-conditioned problem we have no guarantees that the computed solution will be near the actual solution. Thus, stability in an algorithm does not ensure accurate results in all circumstances. Nevertheless, when developing algorithms for use on a computer, we attempt to implement an algorithm that is stable, and we would hope to be able to detect ill-conditioning in the data. We want to be able to provide the user of the computer algorithm with the confidence that either his problem has been solved accurately, or that the nature of his problem suggests that an accurate solution may not be possible.

Admittedly, our definitions for the conditioning of data and the stability of algorithms are quite informal. We will be examining the stability question more rigorously in section 2.3 in the context of direct methods for the solution of linear systems of equations.

2.2. Solving Systems of Linear Equations

We are concerned with determining a solution vector x for the system of linear equations $Ax = b$, where matrix A of order n and vector b of length n contain known values. In general, two classes of methods for solving linear systems exist - iterative techniques and direct methods. While both classes of methods are in common use, iterative techniques are seldom used for solving systems of small dimension. For certain large sparse systems having special structure, the iterative methods can be efficient and accurate compared with corresponding direct methods. In particular, they avoid the fill-in of zero entries which usually occurs with direct methods of solution of sparse linear systems.

An iterative technique to solve the $n \times n$ linear system $Ax=b$ starts with an initial approximation $x^{(0)}$ to the solution x, and generates a sequence of vectors $\{x^{(k)}\}_{k=0}^{\infty}$. Most of these iterative techniques involve a process which converts the system $Ax=b$ into an equivalent system of the form $x=Tx+c$ for some $n \times n$ matrix T and vector c. Having selected the initial vector $x^{(0)}$, the sequence of approximate solution vectors is generated by computing

$$x^{(k)}=Tx^{(k-1)}+c$$

for each $k=1,2,3,\ldots$. If the method is convergent, then $\lim_{k\to\infty} x^{(k)}=x$.

While iterative techniques form an important class of solution methods for solving linear systems of equations, we will be focusing our attention on direct methods of solution. The method of *Gaussian Elimination* is the most common direct method, and will be the subject of most of our study.

2.2.1. Gaussian Elimination

Given the linear system $Ux=b$ where U is a nonsingular upper triangular matrix we can solve for x by the well-known method of backward substitution. Similarly, if U is singular we can easily determine if no solution exists, or if there exists many solutions. In general, matrix A of $Ax=b$ is not upper triangular, so we seek a method to factor matrix A into a product of a nonsingular lower triangular matrix L and an upper triangular matrix U such that $A=LU$. Then x can be determined by solving two triangular systems, $Ly=b$ and $Ux=y$. Such a factorization is called the *LU-decomposition* of A.

The method of Gaussian Elimination (we hereafter abbreviate this as GE) attempts to produce the LU-decomposition of a matrix A with L being unit lower triangular. The matrix is premultiplied by a sequence of elementary lower triangular matrices M_i, each chosen to introduce a column of zeros below the diagonal. Thus, letting $A = A_0$, each step of GE produces

$$A_k = M_k A_{k-1}$$

for each step k, $1 \le k \le n-1$, where matrix A_k has the first through kth columns all zero below the diagonal. Clearly A_{n-1} will be upper triangular and

$$U = A_{n-1} = M_{n-1} M_{n-2} \cdots M_1 A_0$$

so

$$A_0 = (M_1^{-1} M_2^{-1} \cdots M_{n-1}^{-1})U.$$

The product $M_1^{-1} M_2^{-1} \cdots M_{n-1}^{-1}$ is unit lower triangular so this is an LU-decomposition. The elements of the matrices M_i are easily calculable, and the inverse of an elementary lower triangular matrix is found by simply changing the signs of the off-diagonal elements (see Theorems 2.2, 2.3 in Stewart[1973], pp. 116).[1] The reduced matrix after k steps of GE will have the form of

$$A_k = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdot & \cdot & a_{1,k+1} & \cdot & \cdot & a_{1,n} \\ & a_{2,2}^{(1)} & \cdot & \cdot & a_{2,k+1}^{(1)} & \cdot & \cdot & a_{2,n}^{(1)} \\ & & \cdot & & \cdot & & & \cdot \\ & & & \cdot & \cdot & & & \cdot \\ & & & & a_{k+1,k+1}^{(k)} & \cdot & \cdot & a_{k+1,n}^{(k)} \\ & 0 & & & a_{k+2,k+1}^{(k)} & \cdot & & a_{k+2,n}^{(k)} \\ & & & & \cdot & & \cdot & \cdot \\ & & \cdot & & a_{n,k+1}^{(k)} & \cdot & \cdot & a_{n,n}^{(k)} \end{bmatrix} = \begin{bmatrix} \hat{U}_k & \hat{C}_k \\ 0 & \hat{A}_k \end{bmatrix} \tag{2.6}$$

where $\hat{U}_k$ is a $k \times k$ upper triangular matrix, $\hat{A}_k$ is a $(n-k)\times(n-k)$ matrix and $\hat{C}_k$ is a $k \times (n-k)$ matrix. We note that $\hat{A}_k$ is the Schur complement of A with respect to its $k \times k$ leading principal submatrix (if that submatrix is nonsingular).

The nonzero off-diagonal elements of M_{k+1} are given by $\frac{a_{i,k+1}^{(k)}}{a_{k+1,k+1}^{(k)}}$, for $k+2 \le i \le n$, so clearly the GE algorithm breaks down if the *pivot element* $a_{k+1,k+1}^{(k)} = 0$. If, for example, all lead-

[1]For a complete description of GE, see section 2 of Chapter 3 of Stewart[1973], pp. 113.

ing principal minors of A are nonzero then all pivotal elements are nonzero and the matrices M_k, $1 \leq k \leq n-1$, are uniquely determined (see, for example, Wilkinson [1965]). If we encounter $a_{k+1,k+1}^{(k)}=0$ and all $a_{i,k+1}^{(k)}=0$, $k+2 \leq i \leq n$, the reduction may continue but M_{k+1} is no longer unique and $U=A_{n-1}$ is singular. If some $a_{i,k+1}^{(k)} \neq 0$, $k+2 \leq i \leq n$, with $a_{k+1,k+1}^{(k)}=0$ then the reduction cannot be continued.

We have noted that GE may fail if a pivot element is zero. From the point of view of numerical computation using GE, we must also be concerned with small pivot elements. In fact, there are many documented examples of the instability of GE when small pivots are encountered. For this reason, strategies for avoiding small pivots are used.

2.2.1.1. Pivoting Strategies

A *pivoting strategy* is a means to avoid small pivot elements at the $(k+1)$th step of the GE process by interchanging rows and columns in a way that will leave a relatively large element in the pivotal position. This corresponds to the pre- and post-multiplication of A_k to produce

$$A'_k = P_{k+1,\rho_k} A_k P_{k+1,\gamma_k}, \tag{2.7}$$

where P_{k+1,ρ_k} and P_{k+1,γ_k} are elementary permutation matrices. Pivoting strategies differ based on the criterion used to select the new pivot element. The two most common pivoting strategies are *partial pivoting* and *complete pivoting*.

2.2.1.1.1. Partial Pivoting

At the kth step of GE, the partial pivoting strategy involves determining the largest element in modulus of $a_{i,k+1}^{(k)}$, $k+1 \leq i \leq n$. By exchanging the $(k+1)$th row with the row containing the largest element, we ensure that all the off-diagonal elements in column $k+1$ of M_{k+1} are ≤ 1 in modulus. Thus partial pivoting has the effect of producing A'_k in (2.7) with $P_{k+1,\gamma_k}=I$ and ρ_k corresponding to the row containing the element of maximum modulus below the diagonal in the $(k+1)$th column of A_k. If the maximum element in modulus of $a_{i,k+1}^{(k)}$ with $k+1 \leq i \leq n$ is zero, no reduction is necessary at this step, and the algorithm may proceed to the next step. Otherwise, GE can proceed as usual.

The advantage of partial pivoting is that it is relatively inexpensive to realize in a computer program. The disadvantage, as we will see, is that we still cannot guarantee the absolute stability of the GE process with this pivoting strategy. Nevertheless, GE with partial pivoting is one of the most commonly used direct methods for solving linear systems.

2.2.1.1.2. Complete Pivoting

The complete pivoting strategy determines ρ_k and γ_k in (2.7) so that

$$| a_{\rho_k,\gamma_k}^{(k)} | \geq | a_{i,j}^{(k)} |, \quad k+1 \leq i,j \leq n. \tag{2.8}$$

That is, the pivot element is chosen to be an element of maximum modulus in $\hat{A}_k$. If the maximum element is zero, we clearly need not proceed as A_k is then in upper triangular form.

In practice, the pivoting process at each step of the elimination may consume a good deal of time, since $(n-k)^2$ elements must be searched to determine the maximum element. On the other hand, GE with complete pivoting is a stable algorithm (as we will demonstrate later). Thus in comparison with partial pivoting, complete pivoting will guarantee stability but at the expense of increased cost.

2.2.2. Other Direct Methods for the Solution of Linear Systems

We have seen that GE produces an LU-decomposition of a matrix A with L unit lower triangular and U upper triangular. We have described conditions under which the LU-decomposition may not be unique or may not exist. Given an LU factorization of a nonsingular matrix A by GE (without pivoting) we observe that we can factor U such that $U=DU'$ where D is a diagonal matrix and U' is unit upper triangular. Then we see $A=LDU'$ is unique. In general, a matrix A has a unique LDU' decomposition if and only if its leading principal minors are nonzero (see Wilkinson [1965]).

So given that $A=LDU'$ exists, we can examine variants on the LU-decomposition by treating the diagonal matrix D in different ways. By associating D with U' we get the factorization

$$A=LU=L(DU')$$

produced by GE. But we could also associate D with L to get

$$A=L'U'=(LD)U'$$

where $L'=LD$ no longer has unit diagonal elements. This is known as the *Crout decomposition.*

When A is symmetric and has a unique LDU' decomposition the factorization must have the form

$$A=LDL^T.$$

If the elements $d_{i,i}$ of D are positive we can write $D^{1/2}=\text{diag}(d_{1,1}^{1/2}, \ldots, d_{n,n}^{1/2})$. Then A is positive definite and can be written as

$$A=(LD^{1/2})(D^{1/2}L^T).$$

This variant is known as the *Cholesky decomposition* of A.

While each of these decompositions could be calculated from the GE reduction on A, there are advantages to calculating the decompositions directly. The algorithm that directly factorizes a matrix into its Crout decomposition (called *Crout reduction*) requires the same amount of work as GE. The Crout algorithm usually requires partial pivoting and exhibits similar stability characteristics as GE with that same pivoting method.

To apply the Cholesky decomposition algorithm to a matrix A, we require that A be positive definite. This algorithm takes approximately half as a many operations as GE or Crout reduction, and demonstrates a particularly stable behaviour.

2.2.3. Stability Considerations for Direct Methods

We are interested in determining conditions for the stability of direct methods for the solution of systems of linear equations. The algorithms in the previous section are used to compute an LU-decomposition for some matrix $A = A_0$. Using floating point arithmetic, computed triangular factors are found such that

$$LU = A_0 + E.$$

We wish to find a bound on the size of elements $e_{i,j}$ of E. If the elements of E can be shown to be small, then the algorithm can be considered stable.

Stability conditions on the GE algorithm derived by Wilkinson [1961] and later modified by Reid [1971] suggest that control of the size of the growth of the elements of the reduced matrices A_k is of prime importance. As noted above, if GE with floating point arithmetic is applied to a general $n \times n$ matrix A_0, then triangular factors L and U are determined such that

$$LU = A_0 + E,$$

where L is unit lower triangular and U is upper triangular. The elements $e_{i,j}$ of E are bounded by

$$|e_{i,j}| \leq \begin{cases} (3.01)\epsilon a\,(i-1), & j \geq i \\ (3.01)\epsilon a j, & j < i \end{cases}$$

where ϵ specifies the relative accuracy of the floating point computation and a is the largest number in any of the matrices A_k, $0 \leq k \leq n-1$, encountered during the reduction. That is,

$$a = \max_{i,j,k} |a_{i,j}^{(k)}|.$$

It is common, and often more meaningful, to consider the *growth factor* γ defined as

$$\gamma = \frac{\max\limits_{i,j,k} |a_{i,j}^{(k)}|}{\max\limits_{i,j} |a_{i,j}|}$$

as a measure of the perturbation with respect to the magnitude of the elements $a_{i,j}$ of the original matrix A_0.

For GE with partial pivoting, the bound on the growth factor γ is

$$\gamma \le 2^{n-1}.$$

This function is fast growing in relation to n, so we cannot assert that GE with partial pivoting is necessarily stable. In fact, examples are known for which the bound can be attained.

GE with complete pivoting has a bound of

$$\gamma < (n \cdot 2^1 3^{1/2} 4^{1/3} \cdots n^{1/(n-1)})^{1/2},$$

and the proof that establishes it shows that it cannot be attained. This bound increases rather slowly with n and thus GE with complete pivoting is a stable algorithm.

The error analysis for Crout reduction is similar to that for GE. The growth factor γ in the bounds for Crout reduction is the same as the growth factor for GE. Since partial pivoting is the only convenient pivoting method for Crout, we observe the same bound as for GE with partial pivoting.

The error analysis for the Cholesky decomposition indicates that there is no growth, so $\gamma=1$. It should be mentioned, however, that there is no growth when GE or Crout reduction (with no pivoting) is performed on positive definite matrices. Regardless, we see that the Cholesky algorithm for positive definite matrices is unconditionally stable.

CHAPTER 3

A Survey on M-matrices

In this chapter we will be defining M-matrices and describing many of the characteristics associated with these matrices. The special structure of M-matrices occurs in many diverse applications, including problems in partial differential equations, linear and non-linear systems, dynamic systems, economic modeling, operations research and Markov processes in probability and statistics. We will provide descriptions of some of these application areas in this chapter as well.

3.1. Definition and Characterizations of M-matrices

A commonly occurring matrix structure found in many applications has nonpositive off-diagonal and nonnegative diagonal entries. An $n \times n$ matrix A of this type can be written as

$$A = \begin{bmatrix} a_{1,1} & -a_{1,2} & -a_{1,3} & \cdot & \cdot & \cdot \\ -a_{2,1} & a_{2,2} & -a_{2,3} & \cdot & \cdot & \cdot \\ -a_{3,1} & -a_{3,2} & a_{3,3} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \cdot & \\ \cdot & \cdot & \cdot & & & \cdot \end{bmatrix}, \tag{3.1}$$

where $a_{i,j} \geq 0$ for all i,j. Obviously, we can express A in the form

$$A = sI - B \quad , \quad s > 0, \quad B \geq 0. \tag{3.2}$$

We adopt the usual notation of the literature and define

$$Z^{n \times n} = \{A = (a_{i,j}) \in R^{n \times n} : \ a_{i,j} \leq 0, \ i \neq j\}.$$

We are interested in a subclass of $Z^{n \times n}$ called *M-matrices* defined as a restriction on (3.2).

Definition: A matrix A is called an *M-matrix* if it can be written as $A = sI - B$ where B is a nonnegative matrix and s is a scalar such that $s \geq \rho(B)$, the spectral radius of B.

Clearly A is a nonsingular M-matrix if and only if $s > \rho(B)$.

There appears to be no clear picture of the history of the development of M-matrix theory. The first reference to the term M-matrix seems to have been by Ostrowski [1937] in describing the work of Minkowski [1900,1907]. In some areas the terms Minkowski matrix and (usually, nonsingular) M-matrix are used interchangably. Following the work by Ostrowski, M-matrix research was done primarily by two groups - mathematicians and economists. Fan and Householder [1958, 1959, 1960] studied M-matrices primarily for the purpose of giving topological proofs for certain theorems on matrices with nonnegative elements. The economists studied M-matrices in connection with Leontief's input-output analysis (Leontief [1936, 1941]) and stability of a general equilibrium (Hawkins and Simon [1949]).

The first systematic effort to present a unified account of the properties and applications of M-matrices was by Fiedler and Ptak [1962]. They gave 13 equivalent conditions for a matrix A in $Z^{n \times n}$ to be a nonsingular M-matrix. Some of these conditions are

(1) A is semipositive;

(2) A is positive stable;

(3) all principal minors of A are positive;

(4) there exists a permutation matrix P such that $PAP^T = LU$, where L is a lower triangular M-matrix with positive diagonals, and U is an upper triangular M-matrix with positive diagonals;

(5) A is inverse-positive.

As well, Fiedler and Ptak gave equivalent conditions for a matrix $A \in Z^{n \times n}$ to be an M-matrix. These conditions include

(1) all principal minors of A are nonnegative;

(2) the real part of all eigenvalues of A are nonnegative;

(3) $A+\epsilon I$ is a nonsingular M-matrix whenever $\epsilon>0$.

Further, Fiedler and Ptak noted some properties of singular, irreducible M-matrices which will be important in some of our later results. In particular, if A is a singular, irreducible M-matrix of order n, then

(1) A has rank $n-1$;

(2) there exists a vector $x>>0$ such that $Ax=0$;

(3) each principal submatrix of A other than A itself is a nonsingular M-matrix;

(4) A is almost monotone.

Poole and Boullion [1974] presented a further survey on M-matrices. Plemmons [1977] combined and extended this survey to include the work of Varga [1976], and stated 40 equivalent conditions for a matrix A in $Z^{n\times n}$ to be a nonsingular M-matrix. Some important conditions not contained in the work of Fiedler and Ptak are

(1) A is monotone;

(2) A has a ***convergent regular splitting,*** that is, A has a representation

$$A=M-N, \quad M^{-1}\geq 0, \quad N\geq 0$$

with $M^{-1}N$ convergent (Varga [1962]);

(3) there exists a positive diagonal matrix D such that

$$AD+DA^T$$

is positive definite;

(4) A has all diagonal elements positive, and there exists a positive diagonal matrix D such that AD is strictly row diagonally dominant.

The Plemmons paper also gives 4 conditions for a matrix A to be a nonsingular M-matrix given that A is an arbitrary matrix in $R^{n \times n}$. These conditions are

(1) $A + D$ is inverse-positive for each nonnegative matrix D;

(2) $A + \alpha I$ is inverse-positive for each scalar $\alpha \geq 0$;

(3) each principal submatrix of A is inverse-positive;

(4) each principal submatrix of A of orders 1, 2 and n is inverse-positive.

Conditions (1) and (2) are due to Willson [1971], and (3) and (4) to Cottle and Vienott [1972].

Related work regarding nonsingular M-matrices was done by Schröder [1978], who listed some characterizations using operator theory and partially ordered linear spaces, and Kaneko [1978], who noted characterizations and applications of nonsingular M-matrices in terms of linear complementarity problems in operations research.

The work of Plemmons was expanded by Berman and Plemmons [1979], particularly in Chapter 6 of that book. They list 50 equivalent conditions for a matrix in $Z^{n \times n}$ to be a nonsingular M-matrix, and give some classifications of general M-matrices. The work of Berman and Plemmons remains as probably the most comprehensive collection on M-matrix theory. In addition, Chapters 7-10 in this book describe applications in which M-matrices occur.

Finally, Neumann and Plemmons [1980] extended the work of Berman and Plemmons to include some new characterizations relating to stability and monotonicity.

3.2. Applications of M-matrices

As mentioned in the introduction to this chapter, there are many diverse applications for M-matrices. This has lead to some duplication and parallel development in many aspects of M-matrix theory. It is often difficult to determine an original author of a particular M-matrix result for this reason; similarly, different groups of researchers in different application contexts have, on occasion, obtained common results independently.

We will describe four of the more common applications of M-matrices - namely, finding stationary distribution vectors of ergodic Markov chains, determining equilibrium population levels in the Lotka-Volterra model, finding solutions to linear complementarity problems, and determining gross output in the Leontief input-output model. For additional information on M-matrices and finite difference methods for partial differential equations, see Varga [1962] or Wendroff [1966]; in convergence of iterative methods for solving linear systems of equations see Chapter 7 of Berman and Plemmons [1979] or Varga [1962]; for the design of multivariable control systems, see Araki and Nwokah [1974]; for stability of control systems using Lyapunov methods (see Araki [1975]) and non-Lyapunov methods, see Cook [1974]; for stabilization by feedback, see Šiljak and Vukcevic [1974]; and for arms race stability (and dynamic systems in general), see Šiljak [1978].

3.2.1. Markov Chains

The computation of the stationary distributions of a Markov chain is of widespread interest. For example, computing the stationary distribution vector p of an ergodic Markov chain with probability transition matrix Q amounts to solving

$$Q^T p = p \, , \tag{3.3}$$

where we note that Q is an $n \times n$ irreducible (nonnegative) row stochastic matrix, and p is an n-vector of positive probabilities. Solving for p in (3.3) is important in many areas, including queueing theory (see Kaufman [1983]) and in compartmental analysis tracer models (Funderlic and Mankin [1981]). Harrod and Plemmons [1984] describe more applications of this problem, and provide a comparison of solution methods.

Clearly $A = I - Q^T$ is a singular M-matrix, and solving (3.3) is equivalent to solving $Ap = 0$. A is irreducible so rank(A) = $n-1$. But A has an LU-decomposition where L is a unit lower triangular M-matrix and U is an upper triangular M-matrix of rank $n-1$ with $u_{n,n} = 0$ (see Kuo

[1977]). Solution of $Ap = 0$ is then possible, and algorithms for solving this problem are described in Funderlic and Mankin.

3.2.2. Lotka-Volterra Model

The Lotka-Volterra equations attempt to model the interaction between the populations of n species in a common environment. (See Šiljak [1978], Chapter 5, for a detailed description of the Lotka-Volterra model and related topics.) There exists a predator-prey relationship amongst the species, as for example, species i might prey on species j, while species k might be a predator on species i. The absence of species j from the system might mean the extinction of species i and k due to starvation, while the removal of species k from the environment might mean explosive growth of the species i population. We seek to determine the equilibrium populations of the system.

The environmental system is described by the equations

$$\dot{x}_i = x_i(c_i + \sum_{j=1}^{n} b_{i,j} x_j), \quad 1 \le i \le n, \tag{3.4}$$

where x_i represents the population of species i, and $\dot{x}_i$ is the first derivative of x_i with respect to time (that is, the rate of change of the population of species i with respect to time). The coefficients $b_{i,j}$ represent the interaction between species i and j, and the c_i represent growth rates for population i. Equilibrium populations are determined by solving the equations when $\dot{x}_i = 0$, so

$$x_i(c_i + \sum_{j=1}^{n} b_{i,j} x_j) = 0.$$

We are interested in a solution with $x \neq 0$, that is, we must solve

$$c + Bx = 0 \tag{3.5}$$

and obtain $x = (-B)^{-1} c >> 0$. Since $c_i > 0$, $x >> 0$ if $-B$ is inverse-positive. A special case of the Lotka-Volterra model has the matrix $-B \in Z^{n \times n}$ and thus a solution to (3.5) exists if $-B$ is a nonsingular M-matrix.

3.2.3. Linear Complementarity Problems

M-matrices play a role in the *linear complementarity problem* (LCP) encountered in mathematical programming. The problem is stated as follows: for a given $r \in R^n$ and $M \in R^{n \times n}$ find (or conclude there is no) $z \in R^n$ such that

$$r + Mz \geq 0,$$

$$z \geq 0, \quad z^T(r + Mz) = 0. \tag{3.6}$$

We denote this problem by the symbol (r, M).

Problems posed in the form of (3.6) are found in (linear and) complex quadratic programming, in finding the Nash equilibrium point of a bimatrix game (see Cottle and Dantzig [1968], Lemke [1965]), and also in a number of free boundary problems of fluid mechanics (Cryer [1971]).

The set of feasible vectors z associated with the LCP (r, M) is defined as

$$X(r, M) = \{z \in R^n : \; r + Mz \geq 0, \; z \geq 0\}.$$

The off-diagonal elements of M are nonpositive ($M \in Z^{n \times n}$) if and only if $X(r, M)$ has a least element which is a solution of (r, M) for each r such that $X(r, M) \neq \emptyset$. Matrix M is a nonsingular M-matrix if and only if $X(r, M)$ has a least element which is the unique solution of (r, M). These results relating to the least element of $X(r, M)$ are important features in determining if the LCP (r, M) can be solved by a single linear program. (See Cottle [1975] and Kaneko [1978] for surveys on the usage of nonsingular M-matrices with LCP's.)

3.2.4. Input-Output Analysis

M-matrix results have played a large role in the construction and analysis of input-output models in economics. First introduced by Leontief [1936], input-output analysis has been used as a tool of analysis for a wide variety of economic problems and as a guide for the implementation of various kinds of economic policies. Input-output models have been used in other areas as well, as with certain corporate planning problems (see Stone [1970] or Sandberg [1974]), and in studies of environmental pollution (Gutmanis [1972]).

Leontief's input-output analysis deals with determining what level of output each of n interdependent industries should achieve to satisfy the demand for the product of each industry. The analysis relies on the assumption that each industry requires varying numbers of output commodities from the other industries as input, and produces a single output commodity. So, for the jth industry to output one unit of its commodity, it requires $t_{i,j}$ units produced by the ith industry, $1 \leq i \leq n$. These $t_{i,j}$ are called *input coefficients,* and are assumed to be constant.

If we let x_i denote the output of the ith industry per fixed unit of time, a portion of the *gross output* from industry i is needed as input into the n industries. This portion of the gross output can be expressed as

$$\sum_{j=1}^{n} t_{i,j} x_j .$$

The remainder of the output from industry i can be given as

$$d_i = x_i - \sum_{j=1}^{n} t_{i,j} x_j . \tag{3.6}$$

The vector $d = (d_i)$ can be thought of as the contribution of the *open sector* of the economy where consumer purchases, etc. are taken into account.

Letting x and d be n-vectors with elements x_i and d_i respectively, we can express (3.6) as

$$(I-T)x = d ,$$

where $A = I-T$ is obviously in $Z^{n \times n}$, since the $t_{i,j}$ are positive (and usually scaled to be ≤ 1). If the economic system is to be *feasible,* A^{-1} must exist and $A^{-1} \geq 0$ (A is inverse-positive). In other words, A must be a nonsingular M-matrix. Thus, given any demand vector d, we can solve for the gross output vector x.

We have described the *open Leontief model,* where the open sector lies outside the model. The *closed Leontief model* treats the open sector as just another industry in the system, thus final demand does not appear. This way all commodities are produced simply to satisfy the input requirements for the industries. Mathematically, we are solving a homogeneous system of equations with the coefficient matrix in $Z^{n \times n}$ again.

Recently, there has been interest in sensitivity analysis of solutions to linear equations involving M-matrices as applied to the Leontief model (see Sierksma [1979], Fujimoto, Herrero and Villar [1985]).

CHAPTER 4

The LU-Decomposition of M-Matrices Using Gaussian Elimination

In discussing the LU-decomposition of M-matrices by Gaussian Elimination we shall adopt the following notation. An $n \times n$ M-matrix A will be written as

$$A = A_0 = \begin{bmatrix} a_{1,1} & -a_{1,2} & -a_{1,3} & \cdot & \cdot & \cdot \\ -a_{2,1} & a_{2,2} & -a_{2,3} & \cdot & \cdot & \cdot \\ -a_{3,1} & -a_{3,2} & a_{3,3} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \cdot & \\ \cdot & \cdot & \cdot & & & \cdot \end{bmatrix}, \tag{4.1}$$

where $a_{i,j} \geq 0$ for all i,j. The reduced matrix after k steps of Gaussian Elimination (which we continue to abbreviate as GE) we denote by

$$A_k = \begin{bmatrix} a_{1,1} & -a_{1,2} & \cdot & \cdot & -a_{1,k+1} & \cdot & \cdot & -a_{1,n} \\ & a_{2,2}^{(1)} & \cdot & \cdot & -a_{2,k+1}^{(1)} & \cdot & \cdot & -a_{2,n}^{(1)} \\ & & \cdot & & \cdot & & & \cdot \\ & & & \cdot & \cdot & & & \cdot \\ & & & & a_{k+1,k+1}^{(k)} & \cdot & \cdot & -a_{k+1,n}^{(k)} \\ & 0 & & & -a_{k+2,k+1}^{(k)} & \cdot & & -a_{k+2,n}^{(k)} \\ & & & & \cdot & & \cdot & \cdot \\ & & & & -a_{n,k+1}^{(k)} & \cdot & \cdot & a_{n,n}^{(k)} \end{bmatrix} = \begin{bmatrix} \hat{U}_k & \hat{C}_k \\ 0 & \hat{A}_k \end{bmatrix} \tag{4.2}$$

where $\hat{U}_k$ is a $k \times k$ upper triangular M-matrix, $\hat{A}_k$ is a $(n-k)\times(n-k)$ M-matrix and $\hat{C}_k$ is a $k\times(n-k)$ matrix of nonpositive elements. We note that $\hat{A}_k$ is the Schur complement of A with respect to its $k \times k$ leading principal submatrix (if that submatrix is nonsingular).

A sufficient condition for the existence of an LU-decomposition of an M-matrix A is given in Theorem 1 in Funderlic and Plemmons [1981], and Varga and Cai [1981] showed that this condition is both necessary and sufficient for the LU-decomposition of PAP^T for every permutation matrix P. That is, an M-matrix A satisfies the generalized diagonal dominance condition

$$y^T A \geq 0 \text{ for some vector } y >> 0, \tag{4.3}$$

if and only if for every permutation matrix P, there exists a lower triangular nonsingular M-matrix L with unit diagonal and an upper triangular M-matrix U such that

$$PAP^T = LU.$$

All nonsingular M-matrices satisfy (4.3), as do all singular irreducible M-matrices (see Berman and Plemmons [1979], pp. 136, 156). In addition, nonsingular M-matrices have all principal minors positive so the application of GE will produce nonzero pivot elements. This (theoretically speaking) ensures that the application of GE to a nonsingular M-matrix will produce an LU-decomposition without the need for pivoting. From a computational standpoint, though, this may not be a satisfactory approach.

4.1. Stability Considerations

We are concerned with establishing conditions for the stability of an algorithm producing an LU-decomposition of a nonsingular M-matrix. We should note that if an M-matrix A is column diagonally dominant (this implies that y^T may be chosen as $(1, \ldots, 1)$ in (4.3)), we can apply GE without pivoting and have a growth factor $\gamma \leq 1$.[1] Similarly, if a vector y satisfying (4.3) is known *a priori*, with $D = \text{diag}(y_1, \ldots, y_n)$ then DA is column diagonally dominant. Under these conditions GE is a stable algorithm. In practice, though, such a vector is usually unknown before the decomposition.

GE is also guaranteed to be stable when applied to a symmetric M-matrix. A symmetric nonsingular M-matrix is called a *Stieltjes matrix*. It is well-known that given a symmetric matrix $A \in Z^{n \times n}$, A is a Stieltjes matrix if and only if A is positive definite (see Berman and Plemmons [1979], pp. 141). Obviously, Cholesky's method could be applied in these circumstances.

---

[1] In general, $\gamma \leq 2$ for a diagonally dominant matrix (see Wendroff [1966]). Since growth in an M-matrix can only occur in an off-diagonal element, $\gamma \leq 1$.

The use of partial or complete pivoting with GE could certainly be used when decomposing a general M-matrix to ensure the usual bounds on the growth factor. But such an approach has the disadvantage of potentially destroying the M-matrix structure. The additional structure and information that we gain by knowing our matrix is an M-matrix is lost. Fan [1960] showed that if GE is applied to an $n \times n$ nonsingular M-matrix A, the remaining unreduced submatrices $\hat{A}_k$ are nonsingular M-matrices. Similarly, Lemma 2 in Varga and Cai [1981] implies that if an M-matrix A that admits an LU-decomposition with L nonsingular is irreducible, each submatrix $\hat{A}_k$ is also an irreducible M-matrix. These results suggest that it would be valuable to maintain the M-matrix structure through the elimination.

Discussion regarding direct methods for the LU-decomposition of M-matrices in the literature usually suggest the algorithms described in Funderlic and Plemmons [1981] or Funderlic and Mankin [1981]. Essentially, these algorithms use GE without pivoting, and assume knowledge of the vector y in the generalized diagonal dominance condition (4.3). As noted above though, the vector y is usually unknown *a priori*. In addition, GE without pivoting has no guarantees of stability. We will see an example shortly of large growth in the LU-decomposition of an M-matrix using GE with no pivoting.

Kuo [1977] shows that there exists at least one permutation matrix P for every M-matrix A such that PAP^T has an LU-decomposition with nonsingular L. As mentioned earlier, the result of Varga and Cai [1981] shows that an LU-decomposition exists for *all* P given that A satisfies the generalized diagonal dominance condition (4.3). All of these results suggest that the logical method for pivoting during GE would be the simultaneous row and column interchanges produced by PAP^T. But this leads to the obvious question: what should be the criteria for determining which columns and rows should be interchanged?

Let us firstly make some observations based on the generalized diagonal dominance condition. If $y^T=(1, \ldots, 1)$ in (4.3), then A is diagonally dominant and GE is stable. But there exist M-matrices that are not diagonally dominant and which can subsequently have large growth of individual elements when GE is applied with no pivoting. For example, consider the nonsingular irreducible M-matrix

$$A = \begin{bmatrix} 2 & 0 & -100 \\ -100 & 100 & -1 \\ 0 & -1 & 100 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ -50 & 1 & 0 \\ 0 & -0.01 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & -100 \\ 0 & 100 & -5001 \\ 0 & 0 & 49.99 \end{bmatrix}. \tag{4.4}$$

The growth factor γ is 5001/100=50.01. Now with a symmetric permutation of the first and second rows and columns we have

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} A \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 100 & -100 & -1 \\ 0 & 2 & -100 \\ -1 & 0 & 100 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -0.01 & -.5 & 1 \end{bmatrix} \begin{bmatrix} 100 & -100 & -1 \\ 0 & 2 & -100 \\ 0 & 0 & 49.99 \end{bmatrix}.$$

The growth factor γ is 1 for GE applied to the permuted matrix. Obviously, the symmetric permutation allowed us to avoid a large multiplier in the reduction step. By observation, we were able to pick a column that would create small multipliers for the elimination. To be more precise - and to extract an analogy with column diagonal dominance - we chose the pivotal columns to be those columns of A that were, in a sense, the "most" diagonally dominant. For matrices with the M-matrix sign pattern this is equivalent to choosing a column with the maximal column sum.

4.2. Column Diagonal Dominance in M-matrices

The following theorem states that we can always find at least one column of an M-matrix that satisfies the column diagonal dominance condition.

Theorem 4.1: Given an $n \times n$ M-matrix A, there exists at least one subscript j such that

$$a_{j,j} \geq \sum_{\substack{i=1 \\ i \neq j}}^{n} a_{i,j} \quad . \tag{4.5}$$

Proof: Assume without loss of generality that A is in reduced normal form. Then if $\tilde{A}_{1,1}$ is a zero matrix of order one, trivially

$$0 = a_{1,1} \geq \sum_{i=2}^{n} a_{i,1} = 0$$

and the condition is satisfied. If $\tilde{A}_{1,1} \neq 0$, it is sufficient to show that the condition is true for an irreducible M-matrix A.

If $A = sI - B$ is an irreducible M-matrix with $B \geq 0$ and $s \geq \rho(B)$, B is irreducible. Let j be a column of B such that

$$\sum_{i=1}^{n} b_{i,j} = \min_{1 \leq l \leq n} \left(\sum_{i=1}^{n} b_{i,l} \right).$$

Then by a well-known result of Perron-Frobenius theory (see Lemma 2.5 of Varga [1962], pp. 31)

$$\rho(B) \geq \sum_{i=1}^{n} b_{i,j}$$

and $s \geq \rho(B)$ implies

$$s - b_{j,j} \geq \sum_{\substack{i=1 \\ i \neq j}}^{n} b_{i,j}$$

which is (4.5). ■

Theorem 4.1 provides us with assurance that, given an M-matrix A, we can always find a column with a nonnegative column sum. By exchanging that column into the pivotal position through symmetric permutations on A, we are assured that the sum of the multipliers at that step is ≤ 1. This would at least ensure that growth as occurred in the example (4.4) could not happen and would suggest improved stability for the GE process.

We now state the special case of Theorem 4.1 for nonsingular M-matrices.

Theorem 4.2: Given a $n \times n$ nonsingular M-matrix A, there exists at least one subscript j such that

$$a_{j,j} > \sum_{\substack{i=1 \\ i \neq j}}^{n} a_{i,j} \quad . \tag{4.6}$$

Proof: Since A is nonsingular, the reduced normal form of A cannot have $\tilde{A}_{1,1}$ as a zero matrix of order one, so we need only consider the irreducible (nonsingular) case.

The result now follows from the proof of Theorem 4.1 when we note that $s > \rho(B)$ in the nonsingular case. ■

The inequality condition of Theorem 4.1 can also be extended to strict inequality for singular, irreducible M-matrices if we know that there exists a column where the sum of the off-diagonal elements is strictly *greater* (in modulus) than the diagonal element. Theorem 4.3 formalizes this observation.

Theorem 4.3: Let A be an $n \times n$ singular, irreducible M-matrix. There exists a column j such that

$$a_{j,j} > \sum_{\substack{i=1 \\ i \neq j}}^{n} a_{i,j} \tag{4.7}$$

if and only if there exists a column r such that

$$a_{r,r} < \sum_{\substack{i=1 \\ i \neq r}}^{n} a_{i,r} \quad . \tag{4.8}$$

Proof: Assume firstly that (4.7) is satisfied. If there is no column r as in (4.8) then

$$a_{r,r} \geq \sum_{\substack{i=1 \\ i \neq r}}^{n} a_{i,r}$$

for all $r \neq j$. But then $(1, \ldots, 1)A \geq 0$, and since A is *almost monotone* (see Berman and Plemmons [1979], pp.156), this implies $(1, \ldots, 1)A = 0$, which contradicts (4.7).

To prove the converse, assume that (4.8) is satisfied and that there is no j for which (4.7) holds. A contradiction is obtained as $(-1, \ldots, -1)A = 0$. ■

4.3. GE With Cdd-Pivoting

We now define the pivoting process we have described. Assume $A = A_0$ is an $n \times n$ M-matrix and let A_k be the reduced M-matrix (4.2) after k steps of GE. Assume that column j_k $(k < j_k \leq n)$ has the maximal column sum in $\hat{A}_k$. By Theorem 4.1 we know this sum is nonnegative. Then *column diagonal dominant pivoting (cdd-pivoting)* is the process of exchanging the $(k+1)$th and j_kth columns and rows of A_k prior to the $(k+1)$th step of GE $(0 \leq k < n-2)$. The interchanging is equivalent to forming

$$P_{k+1,j_k} A_k P_{k+1,j_k}^T$$

where P_{k+1,j_k} is an elementary permutation matrix.

Certainly if all the column sums are nonnegative or if at some stage in the elimination process all column sums of the remaining unreduced submatrix $\hat{A}_k$ become nonnegative, we no longer need to pivot (since $\hat{A}_k$ will be diagonally dominant). Thus we may not need to use cdd-pivoting through all the n-1 steps of GE.

The calculation of the column sums of an $n \times n$ matrix requires on the order of n^2 additions. To do this at each step of the elimination would entail a substantial computational expense. In order to reduce this cost, we consider two possibilities:

(1) Can we determine the necessary ordering for the rows and columns of our matrix prior to any steps of GE?

(2) Can we find an inexpensive computation to calculate the column sums of the unreduced submatrix $\hat{A}_k$ at each step based on information available prior to that step?

Consider the 4×4 nonsingular irreducible M-matrix

$$A_0 = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & -1 & 1 & -1 \\ -1/4 & 0 & -1/4 & 1 \end{bmatrix}$$

which has column sums decreasing from columns 1 through 4. One step of GE gives the reduced matrix

$$A_1 = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & -1/2 & 1 \end{bmatrix},$$

and the second column of the unreduced submatrix $\hat{A}_1$ of order 3 has the only positive column sum. Thus we see the original ordering based on column sums is not invariant under GE, and it does not seem to be possible to order the columns of A_0 *a priori* so that no further interchanges would be necessary to implement the cdd-pivoting strategy.

Theorem 4.4 relates the column sums of the unreduced submatrix $\hat{A}_1$ to the column sums of $A = A_0$. This result provides an easy method for determining the column sums of the unreduced submatrix $\hat{A}_k$ at the kth step in the elimination based on information obtained at the previous step.

Theorem 4.4: Let A be an $n \times n$ M-matrix, and define

$$s_j = a_{j,j} - \sum_{\substack{i=1 \\ i \neq j}}^{n} a_{i,j} \quad , \qquad 1 \leq j \leq n. \tag{4.9}$$

Let A_1 be the result of one step of GE applied to A (cf. (4.2)), and

$$s_j^{(1)} = a_{j,j}^{(1)} - \sum_{\substack{i=2 \\ i \neq j}}^{n} a_{i,j}^{(1)} \quad , \qquad 2 \leq j \leq n. \tag{4.10}$$

Then,

$$s_j^{(1)} = s_j + s_1 \left(\frac{a_{1,j}}{a_{1,1}} \right) \quad , \qquad 2 \leq j \leq n. \tag{4.11}$$

Proof: Firstly we note from (4.9)

$$\sum_{\substack{i=1 \\ i \neq j}}^{n} \frac{a_{i,j}}{a_{j,j}} = 1 - \frac{s_j}{a_{j,j}} \quad , \qquad 1 \leq j \leq n. \tag{4.12}$$

From (4.10),

$$\begin{aligned} s_j^{(1)} &= a_{j,j} - \frac{a_{j,1}}{a_{1,1}} a_{1,j} - \sum_{\substack{i=2 \\ i \neq j}}^{n} \left(a_{i,j} + \frac{a_{i,1}}{a_{1,1}} a_{1,j} \right) \\ &= a_{j,j} - \sum_{\substack{i=2 \\ i \neq j}}^{n} a_{i,j} - a_{1,j} \sum_{i=2}^{n} \frac{a_{i,1}}{a_{1,1}} \\ &= a_{j,j} - \sum_{\substack{i=2 \\ i \neq j}}^{n} a_{i,j} - a_{1,j} \left(1 - \frac{s_1}{a_{1,1}} \right) \\ &= a_{j,j} - \sum_{\substack{i=1 \\ i \neq j}}^{n} a_{i,j} + s_1 \frac{a_{1,j}}{a_{1,1}} \\ &= s_j + s_1 \frac{a_{1,j}}{a_{1,1}} \quad , \qquad 2 \leq j \leq n. \quad \blacksquare \end{aligned}$$

As noted earlier, Kuo [1977] showed that for any M-matrix A, there always exists at least one permutation matrix P such that PAP^T has an LU-decomposition with nonsingular L. By applying GE with cdd-pivoting we will always find such a permutation matrix P, thus our algorithm will never fail to determine L and U.

In general, GE (without pivoting) will fail if the elimination process results in $a_{k+1,k+1}^{(k)}=0$ and $a_{l,k+1}^{(k)}\neq 0$ for some $l\in\{k+2, \ldots, n\}$. Using cdd-pivoting on M-matrices ensures that the pivotal column sum is nonnegative so if a zero pivot element is encountered, then $a_{l,k+1}^{(k)}=0$ for all $l\in\{k+2, \ldots, n\}$. No computation is needed at that step, and we can continue on to the $(k+1)$th step. Thus, GE with cdd-pivoting will not fail for any M-matrix, and the resultant LU-decomposition will be an LU-decomposition of PAP^T for some permutation matrix P.

We see then that by using cdd-pivoting with GE we maintain the M-matrix structure, we ensure relatively small multipliers at each step, and we can do this with a minimum of extra cost. But what assurances do we have regarding the stability of our method?

4.3.1. Stability of GE With Cdd-Pivoting

To examine the stability of our algorithm we need to find a bound on the growth factor γ for the method. Thus we must examine the growth of the elements of the intermediate matrices A_k, $1\leq k\leq n-1$ (cf. (4.2)). The interchanges required for the pivoting strategy complicate the expressions for the A_k, making analysis of the algorithm difficult. Fortunately, the following lemma shows that using GE with cdd-pivoting is equivalent to doing the interchanges first and then applying GE without any pivoting.

Lemma 4.1: Let $A'=A_0'$ be an $n\times n$ M-matrix and let $1\leq k\leq n-1$. Let M_i', $1\leq i\leq k$, denote elementary lower triangular matrices and P_i, $1\leq i\leq k$, denote elementary permutation matrices such that

$$A_i'=M_i' \, P_i A_{i-1}' \, P_i^T, \qquad 1\leq i\leq k,$$

denote the reduced matrices obtained by applying k steps of GE with cdd-pivoting to A'. Let $A_0=P_k P_{k-1}\cdots P_1 A_0' \, P_1^T \cdots P_{k-1}^T P_k^T$. Then GE without pivoting may be applied to A_0 (and if some pivot $a_{i,i}^{(i-1)}=0$, then $a_{j,i}^{(i-1)}=0$ for $i+1\leq j\leq n$). If M_i, $1\leq i\leq k$, denote the elementary lower triangular matrices such that

$$A_i = M_i A_{i-1}, \quad 1 \le i \le k,$$

are the reduced matrices obtained by applying k steps of GE without pivoting to A_0, then

$$A_k = A_k' .$$

Proof: See Theorem 2.9 in Stewart[1973], pp. 125. ■

Using Lemma 4.1 we obtain the following result which gives an upper bound, dependent upon values contained in $A = A_0, \ldots, A_{k-2}$ for the elements of the intermediate matrix A_k, $1 \le k \le n-1$.

Lemma 4.2: Let A' be an $n \times n$ nonsingular M-matrix and let $1 \le k \le n-2$. Let $A = A_0$ and A_i, $1 \le i \le k$, be defined as in Lemma 4.1. Then for $1 \le t \le k-1$,

$$a_{i,j}^{(k)} < a_{i,j} + \sum_{s=k-t+1}^{k} a_{s,j} + a_{1,j}\left(\frac{a_{i,1}}{a_{1,1}} + \sum_{s=k-t+1}^{k} \frac{a_{s,1}}{a_{1,1}}\right) + \sum_{l=1}^{k-t-1} a_{l+1,j}^{(l)}\left(\frac{a_{i,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} + \sum_{r=k-t+1}^{k} \frac{a_{r,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}}\right) \tag{4.13}$$

where $k+1 \le i \le n$, $k+1 \le j \le n$ and $i \ne j$.

Proof: We first note that all elements which appear in denominators in (4.13) are nonzero since the reduced matrices A_k are also nonsingular M-matrices.

The off-diagonal elements of A are defined by

$$-a_{i,j}^{(k)} = -a_{i,j}^{(k-1)} - \frac{a_{i,k}^{(k-1)}}{a_{k,k}^{(k-1)}} a_{k,j}^{(k-1)} ,$$

from which it follows that

$$a_{i,j}^{(k)} = a_{i,j} + \frac{a_{i,1}}{a_{1,1}} a_{1,j} + \sum_{l=1}^{k-1} a_{l+1,j}^{(l)} \frac{a_{i,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} . \tag{4.14}$$

The lemma is now proved by induction on t.

From (4.14) we obtain

$$a_{i,j}^{(k)} = a_{i,j} + \frac{a_{i,1}}{a_{1,1}} a_{1,j} + \frac{a_{i,k}^{(k-1)}}{a_{k,k}^{(k-1)}} a_{k,j}^{(k-1)} + \sum_{l=1}^{k-2} a_{l+1,j}^{(l)} \frac{a_{i,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}}$$

$$< a_{i,j} + \frac{a_{i,1}}{a_{1,1}} a_{1,j} + a_{k,j}^{(k-1)} + \sum_{l=1}^{k-2} a_{l+1,j}^{(l)} \frac{a_{i,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}}$$

since cdd-pivoting guarantees that $\frac{a_{i,k}^{(k-1)}}{a_{k,k}^{(k-1)}} < 1$. On expanding $a_{k,j}^{(k-1)}$ using (4.14),

$$a_{i,j}^{(k)} < a_{i,j} + \frac{a_{i,1}}{a_{1,1}} a_{1,j} + a_{k,j} + \frac{a_{k,1}}{a_{1,1}} a_{1,j} + \sum_{l=1}^{k-2} a_{l+1,j}^{(l)} \frac{a_{k,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} + \sum_{l=1}^{k-2} a_{l+1,j}^{(l)} \frac{a_{i,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}}$$

$$= a_{i,j} + a_{k,j} + a_{1,j} \left(\frac{a_{i,1}}{a_{1,1}} + \frac{a_{k,1}}{a_{1,1}} \right) + \sum_{l=1}^{k-2} a_{l+1,j}^{(l)} \left(\frac{a_{i,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} + \frac{a_{k,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} \right) .$$

Thus, the condition is true for $t = 1$.

Now, on assuming that (4.13) holds for $t = m-1 < k-1$, we have

$$a_{i,j}^{(k)} < a_{i,j} + \sum_{s=k-m+2}^{k} a_{s,j} + a_{1,j} \left(\frac{a_{i,1}}{a_{1,1}} + \sum_{s=k-m+2}^{k} \frac{a_{s,1}}{a_{1,1}} \right)$$

$$+ \sum_{l=1}^{k-m} a_{l+1,j}^{(l)} \left(\frac{a_{i,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} + \sum_{r=k-m+2}^{k} \frac{a_{r,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} \right)$$

$$= a_{i,j} + \sum_{s=k-m+2}^{k} a_{s,j} + a_{1,j} \left(\frac{a_{i,1}}{a_{1,1}} + \sum_{s=k-m+2}^{k} \frac{a_{s,1}}{a_{1,1}} \right)$$

$$+ a_{k-m+1,j}^{(k-m)} \left(\frac{a_{i,k-m+1}^{(k-m)}}{a_{k-m+1,k-m+1}^{(k-m)}} + \sum_{r=k-m+2}^{k} \frac{a_{r,k-m+1}^{(k-m)}}{a_{k-m+1,k-m+1}^{(k-m)}} \right)$$

$$+ \sum_{l=1}^{k-m-1} a_{l+1,j}^{(l)} \left(\frac{a_{i,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} + \sum_{r=k-m+2}^{k} \frac{a_{r,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} \right).$$

Cdd-pivoting implies that

$$\left(\frac{a_{i,k-m+1}^{(k-m)}}{a_{k-m+1,k-m+1}^{(k-m)}} + \sum_{r=k-m+2}^{k} \frac{a_{r,k-m+1}^{(k-m)}}{a_{k-m+1,k-m+1}^{(k-m)}} \right) < 1,$$

and expansion of $a_{k-m+1,j}^{(k-m)}$ using (4.14) yields

$$a_{i,j}^{(k)} < a_{i,j} + \sum_{s=k-m+2}^{k} a_{s,j} + a_{1,j}\left(\frac{a_{i,1}}{a_{1,1}} + \sum_{s=k-m+2}^{k}\frac{a_{s,1}}{a_{1,1}}\right) + a_{k-m+1,j} + \frac{a_{k-m+1,1}}{a_{1,1}}a_{1,j}$$

$$+ \sum_{l=1}^{k-m-1} a_{l+1,j}^{(l)}\frac{a_{k-m+1,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} + \sum_{l=1}^{k-m-1} a_{l+1,j}^{(l)}\left(\frac{a_{i,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} + \sum_{r=k-m+2}^{k}\frac{a_{r,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}}\right)$$

$$= a_{i,j} + \sum_{s=k-m+1}^{k} a_{s,j} + a_{1,j}\left(\frac{a_{i,1}}{a_{1,1}} + \sum_{s=k-m+1}^{k}\frac{a_{s,1}}{a_{1,1}}\right)$$

$$+ \sum_{l=1}^{k-m-1} a_{l+1,j}^{(l)}\left(\frac{a_{i,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}} + \sum_{r=k-m+1}^{k}\frac{a_{r,l+1}^{(l)}}{a_{l+1,l+1}^{(l)}}\right)$$

which is (4.13) with $t = m$, completing the proof. ■

We can now state our result regarding the stability of GE with cdd-pivoting on M-matrices.

Theorem 4.5: Let A be an $n \times n$ nonsingular M-matrix. Then the growth factor γ resulting from the application of GE with cdd-pivoting to A is bounded by

$$\gamma = \frac{\max\limits_{i,j,k} a_{i,j}^{(k)}}{\max\limits_{i,j} a_{i,j}} < n-1. \tag{4.15}$$

Proof: Since growth in the elements of the reduced matrices produced by GE only occurs in off-diagonal positions, we need only consider bounding the growth of off-diagonal elements of A_k, $1 \le k \le n-2$.

Letting $t = k-1$ in (4.13) we have

$$a_{i,j}^{(k)} < a_{i,j} + \sum_{s=2}^{k} a_{s,j} + a_{1,j}\left(\frac{a_{i,1}}{a_{1,1}} + \sum_{s=2}^{k}\frac{a_{s,1}}{a_{1,1}}\right) ,$$

where $k+1 \le i \le n$, $k+1 \le j \le n$ and $i \ne j$. Since $\left(\frac{a_{i,1}}{a_{1,1}} + \sum_{s=2}^{k}\frac{a_{s,1}}{a_{1,1}}\right) < 1$ when using cdd-pivoting,

$$a_{i,j}^{(k)} < a_{i,j} + \sum_{s=1}^{k} a_{s,j}$$

$$< (k+1) * \max_{s \in \{1,\ldots,k,i\}} a_{s,j} \ .$$

The result now follows on letting $k=n-2$, since no growth in off-diagonal elements can occur when $k=n-1$. ■

The matrix

$$A_0 = \begin{bmatrix} 1 & 0 & \cdot & \cdot & 0 & 0 & -1 \\ -1+\epsilon & 1 & \cdot & & 0 & 0 & -1 \\ & -1+\epsilon & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & \cdot & \cdot & 0 & 0 & -1 \\ & & & \cdot & 1 & -1 & -1 \\ & 0 & & & -1+\epsilon & 1 & -1 \\ & & & & 0 & -\hat{\epsilon} & 1 \end{bmatrix} \tag{4.16}$$

is an $n \times n$ nonsingular irreducible M-matrix for $0<\epsilon\leq 1$ and for sufficiently small $\hat{\epsilon}>0$. Applying GE with cdd-pivoting results in

$$A_0 = \begin{bmatrix} 1 & & & & & & \\ -1+\epsilon & 1 & & & 0 & & \\ 0 & -1+\epsilon & \cdot & & & & \\ 0 & 0 & & \cdot & & & \\ \cdot & \cdot & & & 1 & & \\ \cdot & \cdot & & & -1+\epsilon & 1 & \\ 0 & 0 & \cdot & \cdot & 0 & -\hat{\epsilon}/\epsilon & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 & -1 \\ & 1 & & & & 0 & -2+\epsilon \\ & & \cdot & & & \cdot & -3+\delta_3(\epsilon) \\ & & & \cdot & & 0 & -4+\delta_4(\epsilon) \\ & & & & \cdot & -1 & \cdot \\ & 0 & & & & \epsilon & -(n-1)+\delta_{n-1}(\epsilon) \\ & & & & & & 1-\frac{\hat{\epsilon}}{\epsilon}[n-1-\delta_{n-1}(\epsilon)] \end{bmatrix}.$$

$$\text{where } \delta_i(\epsilon) = \frac{i(i-1)}{2}\epsilon + O(\epsilon^2).$$

Thus for sufficiently small $\epsilon>0$, we can obtain growth in the $(n-1,n)$-position of U approaching $n-1$ times the maximum element in magnitude in A_0.

The strict inequality in (4.15) for nonsingular M-matrices is also applicable to singular, irreducible M-matrices. By Theorem 4.3, if A is a singular irreducible M-matrix, it has all column sums equal to zero, or it has at least one negative column sum and at least one positive column sum. For maximal growth to occur, the latter must be true. But this implies that there exists a pivotal column with positive column sum, indicating that the sum of the multipliers at that step will be strictly less than one. This ensures that the growth factor γ cannot equal $n-1$. Note that by choosing $\hat{\epsilon}$ in (4.16) so that the (n,n)-element of U is zero, A_0 is a singular, irreducible M-matrix and the upper bound of $n-1$ on γ is seen to be tight for this case.

We observe that the 5×5 singular reducible M-matrix

$$A_0 = \begin{bmatrix} 1 & 0 & 0 & -1 & -1 \\ -1 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 & -1 \\ 0 & 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

can be decomposed using GE with cdd-pivoting to return the lower and upper triangular factors

$$\begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ 0 & -1 & 1 & & \\ 0 & 0 & -1 & 1 & \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 & -1 \\ & 1 & 0 & -1 & -2 \\ & & 1 & -1 & -3 \\ & & & 0 & -4 \\ & & & & 1 \end{bmatrix}.$$

The growth factor γ is 4 in this example, so our bound of Theorem 4.5 can be attained for singular reducible M-matrices. This is consistent with our expectations considering that Theorem 4.1 shows that the sum of the multipliers at every step of GE can be equal to one. Note that the above example generalizes to the $n \times n$ case.

4.4. The GE With Cdd-Pivoting Algorithm

Following is an outline of the algorithm for GE with cdd-pivoting:

```
for j=1 to n do          //* calc. column sums
   c_j = a_{1,j}
   for i=2 to n do
      c_j = c_j + a_{i,j}
   end for
   if c_j < min_{1<=t<=j-1} c_t then
      s = j
   end if
end for

for k=1 to n-1 do        //* pivot if necessary
   if c_s <= 0 then
      find r such that c_r = max_{k<=i<=n} c_i
        & s such that c_s = min_{k<=i<=n} c_i
   end if
   if r != k and c_s <= 0 then
      exchange r th and k th columns and rows of A
      exchange b_r and b_k
   end if

   if a_{i,i} > tol do
      for i=k+1 to n do         //* do GE step
         m = a_{i,k} = a_{i,k} / a_{i,i}
         for j=k+1 to n do
            a_{i,j} = a_{i,j} - m*a_{k,j}
         end for
         b_i = b_i - m*b_k
      end for
   end if

   if k < n-1 and c_s <= 0 then
      for l=k+1 to n do    //* update column sums for next step
         c_l = c_l + c_k *(a_{k,l} / a_{k,k})
      end for
   end if

end for
```

The operation count for GE (with partial or no pivoting) is

$$\frac{1}{3}n^3 + \frac{1}{2}n^2 - \frac{5}{6}n \quad \text{multiplications/divisions}$$

and

$$\frac{1}{3}n^3 - \frac{1}{3}n \quad \text{additions/subtractions.}$$

Using GE with cdd-pivoting, the additional column sum calculations increase the operation count to

$$\frac{1}{3}n^3 + \frac{3}{2}n^2 - \frac{11}{6}n - 6 \quad \text{multiplications/divisions}$$

and

$$\frac{1}{3}n^3 + \frac{3}{2}n^2 - \frac{11}{6}n - 3 \quad \text{additions/subtractions.}$$

CHAPTER 5

Extensions of Gaussian Elimination With Cdd-Pivoting

In the last chapter we have shown the usefulness of using cdd-pivoting with Gaussian Elimination when determining the LU-decomposition of an M-matrix. In this chapter we will investigate the application of our algorithm to a larger, more general class of matrices known as H-matrices.

5.1. H-matrices

If $B=(b_{i,j})$ is a ***complex*** $n \times n$ matrix, we define the ***comparison matrix*** $M(B)=(\delta_{i,j})$ by

$$\delta_{i,j} = \begin{cases} \mid b_{j,j} \mid & \text{if } i=j, \\ -\mid b_{i,j} \mid & \text{if } i \neq j. \end{cases}$$

We will define matrix B to be an *H-matrix* if its comparison matrix is an M-matrix.[1] Given any complex matrix B, if there exists scalars $d_i > 0$, $1 \leq i \leq n$, such that

$$d_j \mid b_{j,j} \mid \geq \sum_{\substack{i=1 \\ i \neq j}}^{n} d_i \mid b_{i,j} \mid \quad , \qquad 1 \leq j \leq n, \tag{5.1}$$

matrix B is said to be ***generalized diagonally dominant.*** Equation (5.1) is another form of the generalized diagonal dominance condition (4.3) we encountered for M-matrices. Clearly, an H-matrix is generalized diagonally dominant if its comparison matrix is nonsingular, or singular and irreducible. By our definition of H-matrices, there obviously exist some H-matrices that are not generalized diagonally dominant.

The generalized diagonal dominance condition is important in determining the existence of LU-decompositions of M-matrices as well as in our development of the cdd-pivoting strategy.

[1]There appears to be no standard definition for H-matrices in the literature. Frequently, H-matrices are defined as having a *nonsingular* comparsion matrix (for example, Varga [1976] or Neumann and Plemmons [1984]). We adopt the definition of Berman and Plemmons [1979], pp. 184.

Thus, we seek to extend some of that theory to the more general H-matrices.

Clearly, the set of M-matrices is a subset of the set of H-matrices. We would expect then that some of the theory developed for M-matrices can be generalized to H-matrices. Berman and Plemmons [1979] note some characterizations of H-matrices in terms of the convergence of iterative methods. Varga [1976] gives 9 equivalent conditions for a complex matrix with nonzero diagonal elements to be an H-matrix (with a *nonsingular* comparison matrix).

Funderlic, Neumann and Plemmons [1982] investigated the existence and stability of LU-decompositions of generalized diagonally dominant matrices by GE without pivoting. They present an existence theorem for LU-decompositions of these matrices that extends Theorem 1 in Varga and Cai [1981] and Theorem 1 in Funderlic and Plemmons [1981]. But firstly, for an M-matrix A (in the form of (3.1)) we define the set Ω_A of complex matrices by

$$\Omega_A = \{B:\ A \leq M(B)\}.$$

Thus, $B = (b_{i,j}) \in \Omega_A$ if and only if

$$|b_{j,j}| \geq a_{j,j}, \quad 1 \leq j \leq n,$$

and

$$|b_{i,j}| \leq a_{i,j}, \quad i \neq j, \ 1 \leq i,j \leq n.$$

So, when $B \in \Omega_A$, B is at least as diagonally dominant as A. In discussing LU-decompositions we are interested in M-matrices A which are generalized diagonally dominant. Theorem 1 of Funderlic *et al.* [1982] states that the following are equivalent given that A is an M-matrix:

(1) $y^T A \geq 0$ for some $y >> 0$;

(2) each $B \in \Omega_A$ is generalized diagonally dominant;

(3) for each $B \in \Omega_A$ and each permutation matrix P, $PM(B)P^T$ has an LU-decomposition where L and U (dependent on P) are M-matrices;

(4) for each $B \in \Omega_A$ and each permutation matrix P, PBP^T has an LU-decomposition dependent on P.

As was the case for M-matrices, investigations into the LU-decomposition of H-matrices have focused on applying GE with no pivoting to those matrices. As was also the case with M-matrices, it is easy to construct examples of H-matrices where large growth could occur during the decomposition. Again, we are motivated to find a stable method for producing such an LU factorization.

5.2. GE With Cdd-Pivoting for H-matrices

The results mentioned above suggest that we may be able to extend our theorems of Chapter 4 to H-matrices. In fact, Theorem 4.1 can be extended directly.

Theorem 5.1: Given an $n \times n$ H-matrix B, there exists at least one subscript j such that

$$| b_{j,j} | \geq \sum_{\substack{i=1 \\ i \neq j}}^{n} | b_{i,j} | . \tag{5.2}$$

Proof: The proof of Theorem 4.1 establishes Theorem 5.1 when all references to elements $a_{i,j}$ of A in Theorem 4.1 are replaced by $| b_{i,j} |$. ■

Theorem 4.2 does *not* extend to H-matrices, as demonstrated by the nonsingular H-matrix

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

We note (see Funderlic *et al.* [1982]) that the application of GE to a generalized diagonally dominant matrix B results in the unreduced matrix $\hat{B}_k$ after k steps of GE (cf. 2.6) being generalized diagonally dominant. If we let $A = M(B)$, then there exists $y >> 0$ such that $y^T A \geq 0$. Letting A_k be the result of k steps of GE applied to A, and $y^{(k)} = (0, \ldots, 0, y_{k+1}, \ldots, y_n)^T$ then

$$y^T A = (y^{(0)})^T A_0 \geq (y^{(1)})^T A_1 \geq \cdots \geq (y^{(n-1)})^T A_{n-1} = (y^{(n-1)})^T U \geq 0.$$

Further, for any $B \in \Omega_A$ and each k, $1 \leq k \leq n-1$,

$$(y^{(k)})^T M(B_k) \geq (y^{(k)})^T A_k .$$

Thus, while the degree to which $\hat{B}_k$ is generalized diagonally dominant may deteriorate during the decomposition, we are assured that the unreduced matrix remains generalized diagonally dominant. Given an H-matrix, we can see that the above argument holds if we consider the matrix to be in reduced normal form.

The above argument combined with the result of Theorem 5.1 assures us that GE with cdd-pivoting can be applied to any H-matrix $B = B_0$ to produce an LU factorization, if we use the column sums of $M(\hat{B}_k)$ to determine the pivot at the $(k+1)$th step of the elimination. As was the case with M-matrices, the algorithm will never fail, since if a zero pivot element is encountered, we are guaranteed that all entries in the column below that pivot element are zero. If that is the case, we may skip to the next elimination step.

Unfortunately, Theorem 4.4 cannot be extended to H-matrices since it depends on the M-matrix sign pattern. To determine the "most" diagonally dominant column would then entail calculating

$$| b_{j,j}^{(k)} | - \sum_{\substack{i=k+1 \\ i \neq j}}^{n} | b_{i,j}^{(k)} |$$

after the k th step of GE for the unreduced H-matrix $\hat{B}_k$ (cf.2.6). This would add $O(n^3)$ addition operations to the operation count for the GE algorithm. While this is relatively expensive, the user can expect improved stability with this method as opposed to GE with no pivoting.

We do not provide a bound on the growth factor for GE with cdd-pivoting on H-matrices. However, as shown in the proof of Theorem 2 of Funderlic *et al.* [1982], if an M-matrix A is the comparison matrix for H-matrix B such that A is generalized diagonally dominant, then the application of k steps of GE (without pivoting) results in $\hat{A}_k \leq M(\hat{B}_k)$. In fact, this will always be the case for any M-matrix A as long as GE without pivoting can be applied to A up to the kth step. This shows that the magnitude of the off-diagonal elements of the reduced matrix B_k are bounded above by their corresponding elements in A_k. This observation leads to Theorem 5.2, which extends Theorem 2 of Funderlic *et al.* [1982].

Theorem 5.2: Let A be an M-matrix of order n. Let $B \in \Omega_A$ and let $PAP^T = L' U'$ be the LU-decomposition of A by GE with cdd-pivoting, where P is the permutation matrix representing the product of the elementary permutation matrices produced at each step by the pivoting method. Then, PBP^T can be solved by GE with no pivoting and $PBP^T = LU$ with

$$|l_{i,j}| \leq |l_{i,j}'| \leq 1 \quad \text{for all } i,j,$$

$$|u_{i,j}| \leq |u_{i,j}'| \leq (n-1) \max_{s,t} a_{s,t} \quad \text{for all } i \neq j,$$

and

$$|u_{j,j}| \geq |u_{j,j}'|.$$

Proof: We define

$$t_j^{(k)} = a_{j,j}^{(k)} - \sum_{\substack{i=k+1 \\ i \neq j}}^{n} a_{i,j}^{(k)}$$

and

$$s_j^{(k)} = |b_{j,j}^{(k)}| - \sum_{\substack{i=k+1 \\ i \neq j}}^{n} |b_{i,j}^{(k)}|$$

for $k+1 \leq j \leq n$, using the notation of (4.2) and (2.6) to represent A_k and B_k after k steps of Gaussian Elimination have been applied. We note that $t_j^{(k)} \leq s_j^{(k)}$ for all j at any step k (this is essentially described in Funderlic *et al.* [1982]). Since the maximal column sum $t_j^{(k)}$ in $\hat{A}_k$ is nonnegative, the corresponding $s_j^{(k)}$ in $\hat{B}_k$ is also nonnegative. This ensures that the multipliers at that next step will be ≤ 1 when using cdd-pivoting.

By noting that Lemma 4.1 extends to H-matrices, and by Theorem 2 of Funderlic *et al.* [1982] we can establish the bounds on $| l_{i,j} |$ and $| u_{i,j} |$. ■

This theorem does not directly suggest a bound on the growth factor γ since the $| u_{j,j} |$ are not bounded. The knowledge that the absolute value of the sum of the multipliers at any step is bounded by one suggests that large growth could not occur in the diagonal position.

CHAPTER 6

Conclusions

In this paper we have presented a new pivoting strategy for computing the LU-decomposition of M-matrices. Column diagonal dominant pivoting is based on the observation that at least one column sum of an M-matrix is nonnegative. It uses symmetric permutations, which insures that all reduced matrices of the Gaussian Elimination process remain M-matrices.

We note that the application of GE with cdd-pivoting to any M-matrix will *always* produce an LU-decomposition with L nonsingular, even if that M-matrix is singular. It is known that for an M-matrix A, there always exists at least one permutation matrix P such that $PAP^T = LU$ with L nonsingular. GE with cdd-pivoting will generate one of those matrices P.

We have shown that the growth factor γ from the backward error analysis of GE is bounded by $n-1$ for GE with cdd-pivoting applied to M-matrices. Compared with the exponential growth factor bound for the widely used partial pivoting strategy, the cdd-pivoting bound is very favourable. The user of GE with cdd-pivoting for the LU-decomposition of an M-matrix has the assurance of numerical stability.

The cdd-pivoting variation on the GE algorithm increases the operation count for GE (with no pivoting) by about $O(n^2)$ additions and multiplications. As is often the case, the price for improved stability of an algorithm is increased cost. Considering that GE without pivoting itself uses $O(n^3)$ additions and multiplications, the increased cost due to cdd-pivoting is quite acceptable.

We have shown that GE with cdd-pivoting can be extended to a larger class of matrices known as H-matrices, of which M-matrices are a subset. While the pivoting method and the pivoting criteria extend easily, we noted that we could not use the results of Theorem 4.4 to calculate the column sums of the comparison matrix of the unreduced submatrix at any step in the elimination. This suggests we must perform that calculation explicitly, which means our H-matrix algorithm is more costly than the M-matrix version.

We also demonstrated that the growth of the off-diagonal entries of an H-matrix B are bounded by the off-diagonals of an M-matrix A for which $B \in \Omega_A$. We did not provide a bound, however, for the diagonal element growth of B. Hence, we cannot claim that GE with cdd-pivoting is stable for H-matrices, but we hypothesize that such a bound exists. The derivation of a growth factor bound for GE with cdd-pivoting on H-matrices is a potential area for further work. As well, we believe much of the work in the paper of Funderlic, Neumann and Plemmons [1982] could be extended based on the use of GE with cdd-pivoting. We illustrated such an extension with Theorem 5.2.

Comparing GE with cdd-pivoting to the current methods for the LU-decomposition of M-matrices, we see a number of advantages with our method. Current algorithms sometimes rely on knowledge of the scaling vector in the generalized diagonal dominance condition *a priori* to produce the decomposition. In general, this vector is unknown before the computation. The current algorithms usually use GE without pivoting to produce the triangular factors, but we have demonstrated that this approach can be unstable for M-matrices. As well, our algorithm will work for the LU-decomposition of M-matrices that are not necessarily generalized diagonally dominant. Current algorithms depend on the generalized diagonal dominance condition being satisfied.

Given a linear system $Ax = b$ with A a nonsingular M-matrix, if A is diagonally dominant, then GE without pivoting may be used to determine x, and the algorithm is stable. Similarly, if A is a symmetric M-matrix, then the Cholesky algorithm (with its guaranteed stability) may be used to compute x. For a general M-matrix A, we believe that GE with cdd-pivoting is a reliable and efficient method for the solution of $Ax = b$. Compared with the current approach of using GE with no pivoting to factor A, we believe the cdd-pivoting approach is more reliable and robust. Similarly, we believe that the cdd-pivoting algorithm could produce accurate results when applied to H-matrices.

Bibliography

M. ARAKI [1975], *Analysis of M-matrices to the stability problems of composite dynamical systems,* J. Math. Anal. Appl., **52,** pp. 309-321.

M. ARAKI and O. NWOKAH [1974], *Computing and Control,* Report 74/76, Imperial College, London.

A. BERMAN and R. PLEMMONS [1979], *Nonnegative Matrices in the Mathematical Sciences,* Series on Computer Science and Applied Mathematics, Academic Press, New York.

P. COOK [1974], *On the stability of interconnected systems,* Int. J. Control, **20,** pp. 407-415.

R. COTTLE [1975], *On Minkowski matrices and the linear complementarity problem,* in *Optimization and Optimal Control,* Lecture Notes in Mathematics, No. 477, Springer, Berlin.

R. COTTLE and G. DANTZIG [1968], *Complementary pivot theory of mathematical programming,* Linear Algebra Appl., **1,** pp. 103-125.

R. COTTLE and A. VIENOTT [1972], *Polyhedral sets having a least element,* Math. Progr. **3,** pp. 238-249.

C. CRYER [1973], *The LU-factorization of totally positive matrices,* Linear Algebra Appl., **7,** pp. 83-92.

K. FAN [1958], *Topological proofs for certain theorems on matrices with non-negative elements,* Monatsh. Math., **62,** pp. 219-237.

K. FAN [1960], *Note on M-matrices,* Quart. J. Math. Oxford Ser. (2), **11,** pp. 43-49.

K. FAN and A. HOUSEHOLDER [1959], *A note concerning positive matrices and M-matrices,* Monatsh. Math., **63,** pp. 265-270.

M. FIEDLER and V. PTAK [1962], *On matrices with nonpositive off-diagonal elements and positive principal minors,* Czech. Math. J., **12,** pp. 382-400.

T. FUJIMOTO, C. HERRERO and A. VILLAR [1985], *A sensitivity analysis for linear systems involving M-matrices and its application to the Leontief model,* Linear Algebra Appl., **64,** pp.85-91.

R. FUNDERLIC and J. MANKIN [1981], *Solution of homogeneous systems of linear equations arising from compartmental models,* SIAM J. Sci. Stat. Comput., **2,** pp.375-383.

R. FUNDERLIC, M. NEUMANN and R. PLEMMONS [1982], *LU decompositions of generalized* *diagonally dominant matrices,* Numer. Math., **40,** pp. 57-69.

R. FUNDERLIC and R. PLEMMONS [1981], *LU decomposition of M-matrices by elimination without pivoting,* Linear Algebra Appl., **41,** pp.99-110.

G. GUTMANIS [1971], *Environmental implications of economic growth in the United States, 1970-2000; An input-output analysis,* Proc. IEEE Conf. Decision and Control, New Orleans.

W. HARROD and R. PLEMMONS [1984], *Comparison of some direct methods for computing stationary distributions of Markov chains,* SIAM J. Sci. Stat. Comput., **5,** pp. 453-469.

D. HAWKINS and H. SIMONS [1949], *Note: Some conditions of macroeconomic stability,* Econometrica, **17,** pp. 245-248.

K. KANEKO [1978], *Linear complementarity problems and characteristics of Minkowski matrices,* Linear Algebra Appl., **20,** pp. 111-130.

L. KAUFMAN [1983], *Matrix methods for queueing problems,* SIAM J. Sci. Stat. Comput., **4,** pp. 525-552.

I. KUO [1977], *A note on factorizations of singular M-matrices,* Linear Algebra Appl., **16,** pp. 217-220.

C. LEMKE [1965], *Bimatrix equilibrium points and mathematical programming,* Management Sci., **11,** pp. 681-689.

W. LEONTIEF [1936], *Quantitative input and output relations in the economic system of the United States,* Rev. Econ. Statist., **18,** pp. 100-125.

W. LEONTIEF [1941], *The Structure of the American Economy,* Harvard Univ. Press, Cambridge, Mass.

H. MINKOWSKI [1900], *Zur Theorie der Einkerten in der algebraischen Zahlkorper* Nachr. K. Ges. Wiss. Gött, Math.-Phys. Klasse, pp. 90-93.

H. MINKOWSKI [1907], *Diophantische Approximationen,* Teubner, Leipzig.

M. NEUMANN and R. PLEMMONS [1980], *M-matrix characteristics II: General M-matrices,* Lin. Multi. Alg., **9,** pp. 211-225.

M. NEUMANN and R. PLEMMONS [1984], *Backward error analysis for linear systems associated with inverses of H-matrices,* BIT, **24,** pp. 102-112.

A. OSTROWSKI [1937], *Uber die Determinanten mit uberwiegender Hauptdiagonale,* Comment. Math. Helv., **10,** pp.69-96.

R. PLEMMONS [1977], *M-matrix characterizations I: Nonsingular M-matrices,* Linear Algebra Appl., **18,** pp. 175-188.

G. POOLE and T. BOULLION [1974], *A survey on M-matrices,* SIAM Rev., **16,** pp. 419-427.

J. REID [1971], *A note on the stability of Gaussian Elimination,* J. Inst. Maths Applics., **8**, pp. 374-375.

I. SANDBERG [1974], *A global non-linear extension of LeChatelier-Samuelson principle for linear Leontief models,* J. of Economic Th., **7**, pp. 40-52.

J. SCHRODER [1978], *M-matrices and generalizations,* SIAM Rev., **20**, pp. 213-244.

G. SIERKSMA [1979], *Non-negative matrices: the open Leontief model,* Linear Algebra Appl., **26**, pp. 175-201.

D. ŠILJAK [1978], *Large-Scale Dynamic Systems: Stability and Structure,* North-Holland, New York.

D. ŠILJAK and M. VUKCEVIC [1974], *Proc. 8th Asilomar Conf. Circuits, Systems, Computers,* Pacific Grove, California.

G. STEWART [1973], *Introduction to Matrix Computations,* Series on Computer Science and Applied Mathematics, Academic Press, New York.

D. STONE [1970], *An Economic Approach to Planning the Conglomerate of the 70's,* Auerbach Publ., New York.

R. VARGA [1962], *Matrix Iterative Analysis,* Prentice-Hall, Englewood Cliffs, N.J.

R. VARGA [1976], *On recurring theorems on diagonal dominance,* Linear Algebra Appl., **13**, pp. 1-9.

R. VARGA and D.-Y. CAI [1981], *On the LU factorization of M-matrices,* Numer. Math., **38**, pp. 179-192.

B. WENDROFF [1966], *Theoretical Numerical Analysis,* Academic Press, New York.

J. WILKINSON [1961], *Error analysis of direct methods of matrix inversion,* J. ACM, **8**, pp. 281-330.

J. WILKINSON [1963], *Rounding Errors in Algebraic Processes,* Prentice-Hall, Englewood Cliffs, New Jersey.

J. WILKINSON [1965], *The Algebraic Eigenvalue Problem,* Oxford Univ. Press (Clarendon), London and New York.

A. WILLSON [1971], *A useful generalization of the P_0 matrix concept,* Numer. Math., **17**, pp. 62-70.

VITA

Surname:	AHAC	Given Names:	ALAN ALBERT
Place of Birth:	Kelowna, B.C.	Date of Birth:	March 28, 1960

Educational Institutions Attended, with Dates of Entering and Leaving:

Okanagan College	1978	to	1980
University of Victoria	1980	to	1983
University of Victoria	1983	to	1985

Degrees, Diplomas, Etc., Awarded, with Dates and Names of Institutions:

B.Sc. (Honors)	1983	University of Victoria

Honors and Awards:

University of Victoria Graduate Fellowship, 1983-1985

Publications:

A. A. AHAC and M. R. LEVY, *The Use of Modern Software Design Principles in Developing Mathematical Library Software,* Congressus Numerantium Series, **46,** *to appear.*

Partial Copyright License

A STABLE METHOD FOR THE LU-DECOMPOSITION OF M-MATRICES

Author: [redacted]

APRIL 17, 1985.

Date