Numerical methods for the linear algebraic systems with M-matrices

B.Sc. Thesis

by

Imre Fekete

Mathematics B.Sc., Mathematical Analyst

Supervisor:

István Faragó

Associate Professor and

Head of the

Department of Applied Analysis and Computational Mathematics

Eötvös Loránd University

Budapest

2010

Acknowledgement

I am heartily thankful to my supervisor, István Faragó, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject.

Contents

1 Introduction

The idea of solving large systems of linear equations by directly or iterative methods is certainly not new, dating back at least to Gauss[1]. Very often problems in the biological, physical, economic and social sciences can be reduced to problems involving matrices which have some special structure. One of the most common situations is where the matrix A in question has nonpositive off-diagonal and positive diagonal entries, that is, A is a finite matrix of the type

$$A = \begin{pmatrix} a_{11} & -a_{12} & -a_{13} & \cdots \\ -a_{21} & a_{22} & -a_{23} & \cdots \\ -a_{31} & -a_{32} & a_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{1.1}$$

where a_{ij} are nonnegative.

Such matrix (1.1) often occur in the real life modelling. Thus, our main aim is to get the hang of different methods for system of linear algebraic equations. This thesis consists of four main sections. The first section includes the mathematical background on the system of linear algebraic equations. In this section we show that the reliable numerical solution of some system of linear algebraic equations depends on the condition number of the system. We also show that the condition number is important for the realization of the method on computer.

We get acquainted with nonnegative matrices. We consider all the reducible and irreducible cases and we add bounds for the spectral radius of nonegative matrcies based on Perron-Frobenius theorem. We show why the spectral radius is important to solve these equations. We realize that these type of matrices play very important role to set up the theory of M-matrices.

Afterwards we formulate a lot of properties of M-matrices and we give 50 equivalent conditions for A being a nonsingular M-matrix. These statements help us in the understanding the qualitative properties of the iterative methods. Finally we formulate different iterative methods for M-matrices.

For the better understanding in each single section we give examples.

[1] Johann Carl Friedrich Gauss (1777-1855) was a German mathematician and scientist who contributed significantly to many fields. Gauss is ranked as one of history's most influential mathematicians. He referred to mathematics as the 'Queen of Sciences'.

2 Mathematical Background

2.1 About the systems of linear algebraic equations

A general system of k linear algebraic equations (SLAE) with n unknows can be written as

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n &= b_k, \end{aligned} \tag{2.1}$$

where $x_1, x_2, \ldots, x_n$ are the unknows, the numbers $a_{11}, a_{12}, \ldots, a_{kn}$ (the coefficients of the system) and $b_1, b_2, \ldots, b_k$ (the right hand side) are given. We call the SLAE homogeneous, when the right hand side equals the null vector, otherwise SLAE is inhomogeneous. We deal with the case $k = n$, because then the numbers of unknows equal to the numbers of equations.
Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}$, and we seek the solution of the system of linear equations

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \qquad 1 \le i \le n, \tag{2.2}$$

which we can written in matrix notation as

$$A\mathbf{x} = \mathbf{b}, \tag{2.2'}$$

where $\mathbf{b}$ is a given column vector.

Theorem 2.1. *An $A \in \mathbb{R}^{n\times n}$ is nonsingular, (i.e., invertible) if and only if $detA \neq 0$.*

Theorem 2.2. *If A is nonsingular, there exists nonsingular $P \in \mathbb{R}^{n\times n}$ (so called permutation matrix) such that $PA\boldsymbol{x} = P\boldsymbol{b}$ and $diag(PA)$ doesn't contain zero elements.*

The solution vector $\mathbf{x}$ exits and it is unique if and only if A is nonsingular, and this solution vector is given explicitly by

$$\mathbf{x} = A^{-1}\mathbf{b}. \tag{2.3}$$

Hence, without loss of generality, we may assume that A is nonsingular and its diagonal entries are nonzero.

2.2 Numerical methods for the solution of the systems of linear algebraic equations

There are two general schemes for solving linear systems: Direct and Iterative Methods. The Direct Methods (DM) attempt to solve the problem by a finite sequence of operations, and in the absence of rounding errors, would deliver an exact solution. The most popular methods are the following: Gaussian elimination, LU factorization, Cholesky factorization. These will be discussed in the section 5.

The other methods are the Iterative Methods (IM) which attempt to solve the problem by finding successive approximations to the solution starting from an initial guess.

Our aim is to show the convergence of the vector sequence $\mathbf{x}^{(m)}$ to solution of the systems of linear algebraic equation. To this we define the convergence of the sequence of vectors. The concept of vector norms, matrix norms, and the spectral radii of matrices play an important role in iterative numerical analysis. Just as it is convenient to compare two vectors in terms of their lenght, it will be similarly convenient to compare two matrices by norm.

To begin with, let $V_n(\mathbb{R})$ be the n-dimensional vector space over the field of real numbers $\mathbb{R}$ of column vectors $\mathbf{x}$.

Definition 2.1. Let be $V_n(\mathbb{R})$ given vector space with the norms: $\|\cdot\|$ and $|||\cdot|||$. The two norms called equavilent, in notation: $\|\cdot\| \cong |||\cdot|||$, if there exist $M \geq m \geq 0$ such that $m\|\mathbf{x}\| \leq |||\mathbf{x}||| \leq M\|\mathbf{x}\|$ for all $\mathbf{x} \in V_n(\mathbb{R})$.

Theorem 2.3. *If $\boldsymbol{x}$ and $\boldsymbol{y}$ are vectors of $V_n(\mathbb{R})$, then*

$$\begin{gathered} \|\boldsymbol{x}\| > \boldsymbol{0}, \quad \textit{unless} \quad \boldsymbol{x} = \boldsymbol{0}; \\ \textit{if } \alpha \textit{ is a scalar, then } \|\alpha\boldsymbol{x}\| = |\alpha| \cdot \|\boldsymbol{x}\|; \\ \|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|. \end{gathered} \tag{2.4}$$

Definition 2.2. Let $\mathbf{x}$ be a (column) vector of $V_n(\mathbb{R})$. Then the $\|\mathbf{x}\|_p$ is defined as

$$\begin{aligned} \|\mathbf{x}\|_p &= (\sum_{i=1}^{n} |x_i|^p)^{1/p}, \quad 1 \leq p < \infty \\ \|\mathbf{x}\|_\infty &= \max_{1 \leq i \leq n} |x_i|, \ \textit{when } p = \infty. \end{aligned} \tag{2.5}$$

The case $p = 2$ is called the *Euclidean*[2] *norm*. In throughout the $\|\cdot\|_p$ (where p is arbitrary) denoted by $\|\cdot\|$. With this definition, the following results are well known.

Theorem 2.4. *In finite-dimensional normed spaces the norms are equivalent.*

Definition 2.3. The $\mathbf{x}^{(m)}$ converges to $\mathbf{x}$ if and only if:

$$\lim_{m\to\infty} \|\mathbf{x}^{(m)} - \mathbf{x}\| = \mathbf{0}.$$

Corollary 2.1. If we have an infinite sequence $\mathbf{x}^0, \mathbf{x}^1, \mathbf{x}^2, \ldots$ of vectors of $V_n(\mathbb{R})$, this sequence converges to a vector $\mathbf{x}$ of $V_n(\mathbb{R})$ if

$$\lim_{m\to\infty} \mathbf{x}_j^{(m)} = x_j, \text{ for all } 1 \le j \le n,$$

where $x_j^{(m)}$ and x_j are the jth components of the vectors $\mathbf{x}^{(m)}$ and $\mathbf{x}$ respectively. Similarly, under the convergence of an infinite series $\sum_{m=0}^{\infty} \mathbf{y}^{(m)}$ of vectors of $V_n(\mathbb{R})$ to a vector $\mathbf{y}$ of $V_n(\mathbb{C})$, we mean that

$$\lim_{N\to\infty} \sum_{m=0}^{N} y_i^{(m)} = y_i, \quad \text{for all } 1 \le i \le n.$$

Definition 2.4. Let $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ with eigenvalues λ_i, $1 \le i \le n$. Then,

$$\rho(A) \equiv \max_{1\le i\le n} |\lambda_i| \tag{2.6}$$

is called the *spectral radius* of the matrix A.

Definition 2.5. Let $A : \mathbb{R}^n \to \mathbb{R}^n$ a linear mapping, i.e,

$$A(x_1 + x_2) = A(x_1) + A(x_2)$$

$$A(\lambda x_1) = \lambda \cdot A(x)$$

for all $x_1, x_2 \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. If A is linear and continuous(i.e., $A \in Lin(\mathbb{R}^n, \mathbb{R}^n)$), then

$$\sup_{\substack{\mathbf{x}\in\mathbb{R}^n \\ \mathbf{x}\neq\mathbf{0}}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$$

is finite. Introducing the notation

$$\|A\| := \sup_{\substack{\mathbf{x}\in\mathbb{R}^n \\ \mathbf{x}\neq\mathbf{0}}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \in \mathbb{R}, \tag{2.7}$$

[2]Euclid (360 a.C.-295 a.C.) was a Greek mathematician and is often referred to as the 'Father of Geometry'.

we can prove the following:

Theorem 2.5. *The function* $\|\cdot\| : Lin(\mathbb{R}^n, \mathbb{R}^n) \to \mathbb{R}$ *defines a norm.*

Remark 2.1. Every real $k \times n$ matrix yields a linear map from $\mathbb{R}^n$ to $\mathbb{R}^k$. Each such choice of norms gives rise to an operator norm and therefore yields a norm on the space of all $k \times n$ matrices. If $k = n$ and one uses the same norm on the domain and the range, then the induced operator norm is a sub-multiplicative matrix norm.

Example 2.1. The operator norm corresponding to the $p - norm$[3] for vectors is:

$$\|A\|_p = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}.$$

In the case of $p = 1$ and $p = \infty$, the norms can be defined as:

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}|$$

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|.$$

For instance, for the matrix

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 6 & 8 & 9 \end{pmatrix},$$

we get $\|A\|_1 = 18$ and $\|A\|_\infty = 25$.

Remark 2.2. In the special case of $p = 2$ (the Euclidean norm) and $m = n$ (square matrices), the induced matrix norm is the *spectral norm*.

$$\|A\|_2 = \sqrt{\lambda_{max}(A^* A)},$$

where A^* denotes the conjugate transpose of A. Hence, for the symmetric matrices: $\|A\|_2 = \rho(A)$. In Example 2.1: $\|A\|_2 = 16.4701$.

[3] In Matlab we can compute $\|A\|_p$ as $norm(A, p)$, where A denotes the given matrix, $p = \{1, 2 \text{ and } inf\}$ the given norm.

Theorem 2.6. *Let $A, B \in \mathbb{R}^{n\times n}$ arbitrary matrices. Then the following relations are true:*

$$\|A \cdot B\| \leq \|A\| \cdot \|B\|, \tag{2.8}$$

$$\|A \cdot \boldsymbol{x}\| \leq \|A\| \cdot \|\boldsymbol{x}\|, \tag{2.9}$$

$$\rho(A) \leq \|A\|, \tag{2.10}$$

where $\|\cdot\|$ denotes any norm is $\mathbb{R}^n$.

Proof. The (2.8) and (2.9) statements are evident due to the definition. We prove (2.10): if λ is any eigenvalue of A, and is $\mathbf{x}$ is an eigenvector associated with the eigenvalue λ, then $A\mathbf{x} = \lambda\mathbf{x}$. Thus, from Theorem 2.3 and (2.9),

$$|\lambda| \cdot \|\mathbf{x}\| = \|\lambda\mathbf{x}\| = \|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|, \tag{2.11}$$

from which we conclude that $\|A\| \geq |\lambda|$ for all eigenvalues of A, which proves (2.10).

Remark 2.3. In addition to (2.9), we note that for any $A \in \mathbb{R}^{n\times n}$ there exists a vector $\mathbf{x} \in \mathbb{R}^n$ such that $\|A\mathbf{x}\| = \|A\| \cdot \|\mathbf{x}\|$.

Definition 2.6. An $A \in \mathbb{R}^{n\times n}$ is convergent matrix, if

$$\lim_{m\to\infty} A^m = O.$$

Theorem 2.7. *The matrix $A \in \mathbb{R}^{n\times n}$ is convergent if and only if $\rho(A) < 1$.*

Proof. It can be seen in [6] as Theorem 1.4.

Corollary 2.2. If $A \in \mathbb{R}^{n\times n}$ and $\|A\| < 1$, then A is convergent.

Proof. Using (2.10) of Theorem 2.6 and Theorem 2.7 it is evident.

2.3 Connection between the condition numbers and the systems of linear algebraic equations

Mainly we meet large SLAE. To solve (2.2') we need computers. But, in practice, as the vector $\mathbf{b}$ also A matrix are given inaccurately, charged with errors: for example input errors. For this reason we will investigate that how serious error makes these ones in the solution vector. The result of our investigation is that we will recognise: the condition number determines that can we solve on a given computer a given system of linear algebraic equations with a given error is acceptable or not.
Consider now (2.2') in that case, when A is $n \times n$ nonsingular matrix and $\mathbf{b} \neq \mathbf{0}$. Applying the norm to both sides, we get

$$0 < \|\mathbf{b}\| = \|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|. \tag{2.12}$$

2.3.1 When b is charged with error

Let us estimate the error, when we consider $\mathbf{b} + \delta\mathbf{b}$ instead of $\mathbf{b}$ on the right side. Denoting by $\mathbf{x} + \delta\mathbf{x}$ the solution of the new perturbed system, we have

$$\mathbf{b} + \delta\mathbf{b} = A(\mathbf{x} + \delta\mathbf{x}) = A\mathbf{x} + A\delta\mathbf{x}. \tag{2.13}$$

Since $A\mathbf{x} = \mathbf{b}$, therefore (2.13) implies the relation $\delta\mathbf{x} = A^{-1}\delta\mathbf{b}$. Applying again the norm, we obtain the estimation

$$\|\delta\mathbf{x}\| \leq \|A^{-1}\| \cdot \|\delta\mathbf{b}\|. \tag{2.13'}$$

Hence, the relations (2.13') and (2.12) together imply:

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|A^{-1}\| \cdot \|\delta\mathbf{b}\|}{\|\mathbf{x}\|} \leq \frac{\|A^{-1}\| \cdot \|\delta\mathbf{b}\|}{\|\mathbf{b}\|/\|A\|} = \|A^{-1}\| \cdot \|A\| \cdot \frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}. \tag{2.14}$$

The (2.14) estimation is sharp, because in (2.14) the equality can be obtained. (For example, the choice $A = I$ is good.)

Definition 2.7. The number $cond(A) := \|A^{-1}\| \cdot \|A\|$ is called condition number of matrix. Since the condition number depends on the choosen norm, sometimes we use the notation:

$$cond_p(A) := \|A^{-1}\|_p \cdot \|A\|_p.$$

Hence, for the relative error in (2.14) we have

$$\frac{\|\delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq cond(A) \cdot \frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|}. \tag{2.15}$$

On the right side of (2.15) we can see the the relative error of $\mathbf{b}$ (for example this can be input error) and on the left side the relative error of solution. Thus $cond(A)$ plays an important role, because if it is big, then the relative error in the input data may cause a significant increase in the error of the solution.
Knowing condition number is important for solving these equations on computer. In general, it is too difficult to compute the condition number of a matrix A exactly because it is far too expensive for large matrices. Avoiding this we can give approximation algorithm for the condition number.

Remark 2.4. The condition number cannot be less then one in the case of induced norm, because $1 = \|I\| = \|AA^{-1}\| \leq \|A\| \cdot \|A^{-1}\| = cond(A)$.

Remark 2.5. Let A is nonsingular, i.e., $\lambda = 0$ is not an eigenvalue of A. Then, due to the relation

$$A\mathbf{v} = \lambda \mathbf{v}$$

(where $\mathbf{v} \in \mathbb{R}^n$ is the eigenvector), we got $\frac{1}{\lambda}\mathbf{v} = A^{-1}\mathbf{v}$, which means that $\frac{1}{\lambda}$ is the eigenvalue of A^{-1} and therefore $\|A^{-1}\| \geq |\lambda_{min}(A^{-1})| = \frac{1}{|\lambda_{min}(A)|}$.
Hence

$$cond(A) \geq \left| \frac{\lambda_{max}(A)}{\lambda_{min}(A)} \right|.$$

Example 2.2.

$$A = \begin{pmatrix} 1 & 2 & 5 & 7 \\ 3 & 6 & 4 & 11 \\ 1 & 4 & 5 & 9 \\ 6 & 5 & 16 & 2 \end{pmatrix},$$

For different p : $cond_1(A) \approx 57.7130$, $cond_2(A) \approx 36.3474$ and $cond_\infty(A) \approx 54.4888$. Applying the estimation of $cond(A)$[4] of Remark 2.2: $cond(A) \geq 15.8691$.
If A is symmetric matrix, then the estimation is sharp for $cond_2(A)$.

Remark 2.6. Let be $\frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|} = \epsilon_1$, where ϵ_1 is the machine epsilon, ϵ_1 significance consists in it, that it is constituted absolutely, relative error bar at the input

[4] In Matlab we can compute $cond(A)$ as $cond(A, p)$ where A denotes the given matrix, $p = \{\ 1, 2 \text{and } inf\}$ the given norm.

and at the four fundamental operations.

$$cond(A) \geq \frac{1}{\epsilon_1}.$$

Then (2.14) shows us that the relative error of solution can be quite huge:

$$cond(A)\frac{\|\delta \mathbf{b}\|}{\|\mathbf{b}\|} \geq 1.$$

Such a system is called ill-conditioned.

Example 2.3. Let $H_n \in \mathbb{R}^{n\times n}$ the so called Hilbert[5] matrix, defined as

$$H_n = \left(\frac{1}{i+j-1}\right)_{i,j=1}^{n}$$

For instence, for $n = 4$ it is defined as:

$$H_4 = \begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 & 1/5 \\ 1/3 & 1/4 & 1/5 & 1/6 \\ 1/4 & 1/5 & 1/6 & 1/7 \end{pmatrix}.$$

The following table shows the behavior of the $cond_2(H_n)$ for some n:

n	4	7	10	18	25
$cond_2(H_n)$	$1.55 \cdot 10^4$	$4.75 \cdot 10^8$	$1.60 \cdot 10^{13}$	$2.39 \cdot 10^{18}$	$6.14 \cdot 10^{18}$

One can see that $cond_2(H_n)$ increases very rapidly. Therefore, the usual solver of the SLAE cannot work efficiently on the test-problems with H_n-matrices.

The reliability of the solution under given computer architecture of some given SLAE is independent on the determinant and it depends mainly on the condition number.

2.3.2 When A is charged with error

Let us consider the problem when A is charged with error:

$$(A + \delta A) \cdot (x + \delta x) = b \tag{2.16}$$

First, we need the condition under which $A + \delta A$ is nonsingular. (We have assumed only the regularity of A.)

[5]David Hilbert (1862-1943) was a German mathematician. He formulated the theory of Hilbert spaces, one of the foundations of functional analysis.

Lemma 2.1. (Perturbation lemma): Let $S := I + R$ and $\|R\| =: q < 1$. Then S is nonsingular and $\|S^{-1}\| \leq \frac{1}{1-q}$.

Proof. It can be seen in [5] in section 2.3.3.
As A nonsingular, we can rewrite (2.16) as

$$(I + A^{-1}\delta A) \cdot (x + \delta x) = A^{-1}b. \tag{2.16'}$$

Using Lemma 2.1 to (2.16') we assume: $\|A^{-1}\delta A\| < 1$. Then $\|\delta A\| < \frac{1}{\|A^{-1}\|}$ is a sufficient condition for that $I + A^{-1}\delta A$ be nonsingular, because using (2.8) of Theorem 2.6 :

$$\|A^{-1}\delta A\| \leq \|A^{-1}\| \cdot \|\delta A\| < 1. \tag{2.17}$$

Obviously, we can write $(A + \delta A)$ as $A(I + A^{-1}\delta A)$. In case of (2.17) and with the help of Lemma 2.1, we can estimate $\|(A + \delta A)^{-1}\|$ as follows

$$\|(A + \delta A)^{-1}\| = \|(I + A^{-1}\delta A)^{-1} \cdot A^{-1}\| \leq \|(I + A^{-1}\delta A)^{-1}\| \cdot \|A^{-1}\| \leq$$
$$\leq \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \cdot \|\delta A\|}. \tag{2.18}$$

Hereupon we can estimate the distinction of (2.2') and (2.16). As $A\mathbf{x} = \mathbf{b} =$ $= (A + \delta A) \cdot (\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + (A + \delta A) \cdot \delta\mathbf{x}$, so $\delta\mathbf{x} = -(A + \delta A)^{-1}\delta A\mathbf{x}$ and using (2.18) we get:

$$\|\delta\mathbf{x}\| \leq \|(A + \delta A)^{-1}\| \cdot \|\delta A\| \cdot \|\mathbf{x}\| \leq \frac{\|A^{-1}\| \cdot \|\delta A\|}{1 - \|A^{-1}\| \cdot \|\delta A\|} \cdot \|\mathbf{x}\|,$$
$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{cond(A)\frac{\|\delta A\|}{\|A\|}}{1 - cond(A)\frac{\|\delta A\|}{\|A\|}}. \tag{2.19}$$

In this case the relative error of result expressible with the help of condition number, and with the relative error of the data of A.

3 Nonnegative Matrices

3.1 Bounds for the spectral radius of a matrix

It is generally difficult to determine precisely the spectral radius of a given matrix. Nevertheless, upper bounds can easily be found from the following theorem.

Theorem 3.1. *(Gerschgorin[6] theorem) Let $A = (a_{ij})$ be an arbitrary $n \times n$ matrix, and let*

$$\Lambda_i \equiv \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \quad 1 \leq i \leq n. \tag{3.1}$$

Then, all the eigenvalues λ of A lie in the union of the disks

$$|z - a_{ii}| \leq \Lambda_i \quad 1 \leq i \leq n.$$

Proof. Let λ be any eigenvalue of the matrix A, and let $\mathbf{x}$ be the corresponding normalized eigenvector. (We normalize the vector $\mathbf{x}$ so that its largest component in modulus is unity.) By definition,

$$(\lambda - a_{ii})x_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij}x_j, \quad 1 \leq i \leq n. \tag{3.2}$$

In particular, if $|x_r| = 1$, then

$$|\lambda - a_{rr}| \leq \sum_{\substack{j=1 \\ j \neq r}}^{n} |a_{rj}| \cdot |x_j| \leq \sum_{\substack{j=1 \\ j \neq r}}^{n} |a_{rj}| = \Lambda_r.$$

Thus, the eigenvalues λ lies in the disk $|z - a_{rr}| \leq \Lambda_r$. But since λ was an arbitrary eigenvalue of A, it follows that all the eigenvalues of the matrix A lie in the union of the disks $|z - a_{ii}| \leq \Lambda_i, 1 \leq i \leq n$, completing the proof.

Since the disk $|z - a_{ii}| \leq \Lambda_i$ is a subset of the disk $|z| \leq |a_{ii}| + \Lambda_i$, we have the immediate result of

[6] Semyon Aranovich Gerschgorin (1901-1933) was a Russian mathematician. More information about his theorem in the book: Richard S. Varga: Gerschgorin and His Circles.

Corollary 3.1. If $A = (a_{ij})$ is an arbitrary $n \times n$ matrix, and

$$v \equiv \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|, \tag{3.3}$$

then $\rho(A) \le v = \|A\|_1$.

But as A and A^T have the same eigenvalues, the application of Corollary 3.1 to A^T gives us

Corollary 3.2. If $A = (a_{ij})$ is an arbitrary $n \times n$ matrix, and

$$v' \equiv \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|, \tag{3.4}$$

then $\rho(A) \le v' = \|A\|_\infty$.

To improve further on these results, we use now the fact that similarity transformations on a matrix leave its eigenvalues invariant.

Corollary 3.3. If $A = (a_{ij})$ is an arbitrary $n \times n$ matrix, and $x_1, x_2, \cdots, x_n$ are any n positive real numbers, let

$$v \equiv \max_{1 \le i \le n} \left\{ \frac{\sum_{j=1}^{n} |a_{ij}| x_j}{x_i} \right\}; \quad v' \equiv \max_{1 \le j \le n} \left\{ x_j \sum_{i=1}^{n} \frac{|a_{ij}|}{x_i} \right\} \tag{3.5}$$

Then, $\rho(A) \le min(v, v')$.

Proof. Let $D = diag(x_1, x_2, \cdots, x_n)$ and apply Corollary 3.1 and 3.2 to the matrix $D^{-1}AD$, whose spectral radius necessarily coincides with that of A.

Example 3.1. Let

$$A = \begin{pmatrix} 0 & 0 & 1/4 & 1/4 \\ 0 & 0 & 1/4 & 1/4 \\ 1/4 & 1/4 & 0 & 0 \\ 1/4 & 1/4 & 0 & 0 \end{pmatrix},$$

if $x_1 = x_2 = x_3 = x_4 = 1$, then $v = v' = \frac{1}{2}$, which is the exact value of $\rho(A)$. This shows that equality is possible in Corollary 3.1, 3.2 and 3.3.

Remark 3.1. In an attempt to improve the Corollary 3.1, suppose that the row sums of the moduli of the entries of the matrix A were not equal to v of (3.3). Could we hope to conclude that $\rho(A) < v$? The counterexample given by the matrix

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix},$$

shows this to be false, since v of (3.3) is 3, but $\rho(B) = 3$ also. This leads us to the following important definition.

Definition 3.1. For $n \geq 2$, an $n \times n$ matrix A is reducible if there exits an $n \times n$ permutation matrix P such that

$$PAP^T = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

where A_{11} is an $r \times r$ submatrix and A_{22} is an $(n-r) \times (n-r)$ submatrix, where $1 \leq r \leq n$. If no such permutation matrix exits, then A is irreducible. If A is a 1×1 matrix, then A is irreducible[7] if its single entry is nonzero, and reducible otherwise.

With the concept of irreducibility, we can sharpen the result of Theorem 3.1 as follows:

Theorem 3.2. *Let $A = (a_{ij})$ be an irreducible $n \times n$ matrix, and assume that λ, an eigenvalue of A, is a boundary point of union of the disks $|z - a_{ii}| \leq \Lambda_i$. Then, all the n circles $|z - a_{ii}| = \Lambda_i$ pass through the point λ.*

Proof. It can be seen in [6] as Theorem 1.7.
This sharpening of Theorem 3.1 immediately gives rise to the sharpened form of Corollary 3.3 of Theorem 3.1.

Corollary 3.4. Let $A = (a_{ij})$ be an irreducible $n \times n$ matrix, and $x_1, x_2, \cdots, x_n$ be any n positive real numbers. If

$$\frac{\sum_{j=1}^{n} |a_{ij}| x_j}{x_i} \leq v \tag{3.6}$$

[7] The term *irreducible(unzerlegbar)* was introduced by the German mathematician, Ferdinand Georg Frobenius (1849-1917) in 1912; it is also called *unreduced* and *indecomposable* in the literature.

for all $1 \leq i \leq n$, with strict inequality for at least one i, then $\rho(A) < v$. Similarly, if

$$x_j \frac{\sum_{i=1}^{n} |a_{ij}|}{x_i} \leq v' \tag{3.7}$$

for all $1 \leq j \leq n$, with strict inequality for at least one j, then $\rho(A) < v'$.

Example 3.2. The following matrix

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 4 & 1 \\ 1 & 0 & 3 \end{pmatrix} \text{ and } \mathbf{x} = (1, 1, 1)$$

is good example for Corollary 3.4, in this case $v = 5$, $v' = 6$ and $\rho(A) = 4.4142$. We get strict inequality for $i = 1$ in (3.6) and for $j = 3$ in (3.7).

3.2 Spectral radius of nonnegative matrices

In investigations of the rapidity of convergence of various iterative methods, the spectral radius $\rho(A)$ of the corresponding iteration matrix A plays a major role. For the practical estimation of this constant $\rho(A)$, we have thus far only upper bounds for $\rho(A)$ by the extensions of Gerschgorin's Theorem 3.1. In this section, we shall look closely into the Perron-Frobenius theory of square matrices having nonnegative real numbers as entries. Theory of Perron-Frobenius plays an important role in the study of M-matrices. First, we define an (partial) ordering for the matrices.

Definition 3.2. Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $n \times r$ matrices. Then, $A \geq B (> B)$ if $a_{ij} \geq b_{ij} (> b_{ij})$ for all $1 \leq i \leq n, 1 \leq j \leq r$. If O is the nullmatrix and $A \geq O (> O)$, we say that A is a *non-negative (positive) matrix*. Finally, if $B = (b_{ij})$ is an arbitrary $n \times r$ matrix, then $|B|$ denotes the matrix with $|b_{ij}|$.

In developing the Perron-Frobenius theory, we shall first establish a series of lemmas on non-negative irredducible square matrices.

Lemma 3.1. If $A \geq O$ is an irreducible $n \times n$ matrix, then $(I + A)^{n-1} > O$.

Proof. It can be seen in [6] as Lemma 2.1.

If $A = (a_{ij}) \geq O$ is an irreducible $n \times n$ matrix and $\mathbf{x} \geq \mathbf{0}$ is any nonzero vector, and let

$$r_{\mathbf{x}} = \min \left\{ \frac{\sum_{j=1}^{n} a_{ij} x_i}{x_i} \right\}, \tag{3.8}$$

where the minimum is taken over all i for which $x_i > 0$. Clearly, $r_{\mathbf{x}}$ is a non-negative real number and is the suprerum of all numbers $\psi \geq 0$ for which

$$A\mathbf{x} \geq \psi \mathbf{x}. \tag{3.9}$$

We now consider the non-negative quantity r defined by

$$r = \sup_{\substack{\mathbf{x} \geq \mathbf{0} \\ \mathbf{x} \neq \mathbf{0}}} \{r_{\mathbf{x}}\}. \tag{3.10}$$

As $r_{\mathbf{x}}$ and $r_{\alpha \mathbf{x}}$ have the same value for any scalar $\alpha > 0$ (due to (3.8)), we need consider only the set P of vectors $\mathbf{x} \geq \mathbf{0}$ with $\|\mathbf{x} = 1\|$, and we correspondingly let Q be the set of all vectors $y = (I + A)^{n-1}\mathbf{x}$, where $\mathbf{x} \in P$. From Lemma 3.1, Q consists only of positive vectors. Multiplying both sides of the inequality $A\mathbf{x} \geq r_{\mathbf{x}}\mathbf{x}$ by $(I + A)^{n-1}$, we obtain

$$A\mathbf{y} \geq r_{\mathbf{x}}\mathbf{y},$$

and we conclude from (3.9) that $r_{\mathbf{y}} \geq r_{\mathbf{x}}$. Therefore, the quantity r of (3.10) can be defined equivalenty as

$$r = \sup_{\mathbf{y} \in Q} \{r_{\mathbf{y}}\}. \tag{3.10'}$$

As P is a compact set of vectors, so is Q, and as $r_{\mathbf{x}}$ is a continuous function on Q, there neccessarily exists a positive vector $\mathbf{z}$ for which

$$A\mathbf{z} \geq r\mathbf{z}, \tag{3.11}$$

and no vector $\mathbf{w} \geq \mathbf{0}$ exists for which $A\mathbf{w} > r\mathbf{w}$. We shall call all non-negative nonzero vectors $\mathbf{z}$ satisfying (3.11) *extremal vectors* of the matrix A.

Lemma 3.2. If $A \geq 0$ is an irreducible $n \times n$ matrix, the quantity r of (3.10) is positive. Moreover, each extremal vector $\mathbf{z}$ is a positive eigenvector of the matrix A with corresponding eigenvalue r, i.e., $A\mathbf{z} = r\mathbf{z}$, and $\mathbf{z} > \mathbf{0}$.

Proof. If ζ is the positive vector whose components are all unity, then since the matrix A is irreducible, no row of A can vanish, and consequently no component of $A\zeta$ can vanish. Thus, $r_\zeta > \mathbf{0}$, proving that $\mathbf{r} > \mathbf{0}$. For the second part of this lemma, let $\mathbf{z}$ be an extremal vector with $A\mathbf{z} - r\mathbf{z} = \eta$, where $\eta \geq \mathbf{0}$. If $\eta \neq \mathbf{0}$, then some component of η is positive; multiplying through by the matrix $(I + A)^{n-1}$, we have

$$A\mathbf{w} - r\mathbf{w} > \mathbf{0}$$

where $\mathbf{w} = (I+A)^{n-1}\mathbf{z} > \mathbf{0}$. It would then follow that $r_{\mathbf{w}} > r$, contradicting the definition of r in (3.10'). Thus, $A\mathbf{z} = r\mathbf{z}$, and since $\mathbf{w} > \mathbf{0}$ and $\mathbf{w} = (1+r)^{n-1}\mathbf{z}$, then $\mathbf{z} > \mathit{0}$, completing the proof.

Lemma 3.3. Let $A = (a_{ij}) \geq 0$ be an irreducible $n \times n$ matrix, and let $B = (b_{ij})$ be an $n \times n$ matrix with $|B| \leq A$. If β is any eigenvalue of B, then

$$|\beta| \leq r,$$

where r is the positive constant of (3.10). Moreover, equality is valid, i.e., $\beta = re^{i\Phi}$, if and only if $|B| = A$, and where B has the form

$$B = e^{i\Phi}DAD^{-1},$$

and D is a diagonal matrix whose diagonal entries have modulus unity.

Proof. It can be seen in [6] as Lemma 2.3.
Setting $B = A$ in Lemma 3.3 immediately gives us

Corollary 3.5. If $A \geq 0$ is an irreducible $n \times n$ matrix, then the positive eigenvalue r of Lemma 3.2 equals $\rho(A)$ of A.

Lemma 3.4. If $A \geq 0$ is an irreducible $n \times n$ matrix, and B is any principal square submatrix of A, then $\rho(B) < \rho(A)$.

Proof. If B is any principal submatrix of A, then there is an $n \times n$ permutation matrix P such that $B = A_{11}$, where

$$C \equiv \begin{pmatrix} A_{11} & 0 \\ 0 & 0 \end{pmatrix}; \quad PAP^T \equiv \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Here, A_{11} and A_{22} are, respectively, $m \times m$ and $(n-m) \times (n-m)$ principal square submatrices of $PAP^T, 1 \leq m < n$. Clearly, $O \leq C \leq PAP^T$, and

$\rho(C) = \rho(B) = \rho(A_{11})$, but as $C = |C| \neq PAP^T$, the conclusion follows immediately from Lemma 3.3 and Corollary 3.5.
We now collect above results into the following main theorem.

Theorem 3.3. *(Perron[8]-Frobenius) Let $A \geq O$ be an irreducible $n \times n$ matrix. Then,*
1. A has a positive real eigenvalue equal to its spectral radius.
2. $\rho(A)$ there correnspponds an eigenvector $\boldsymbol{x} > \boldsymbol{0}$.
3. $\rho(A)$ increases when any entry of A increases.
4. $\rho(A)$ is a simple eigenvalue of A.

Proof. Parts 1 and Parts 2 follow immediately from Lemma 3.2 and the Corollary 3.5 to Lemma 3.3. To prove part 3, suppose we increase some entry of the matrix A, giving us a new irreducible matrix $\widetilde{A}$, where $\widetilde{A} \geq A$ and $\widetilde{A} \neq A$. Applying Lemma 3.3, we conclude that $\rho(\widetilde{A}) > \rho(A)$. To prove that $\rho(A)$ is a simple eigenvalue of A, i.e., $\rho(A)$ is a zero of multiplicity one of the characteristic polynomial $\Phi(t) = det(tI - A)$, we make use of the fact that $\Phi^{'}(t)$ is the sum of the determinant of the principal $(n-1) \times (n-1)$ submatrices of $tI - A$. If A_i is any principal submatrix of A, then from Lemma 3.4, $det(tI - A_i)$ cannot vanish for any $t \geq \rho(A)$. From this it follows that

$$det(\rho(A)I - A_i) > 0,$$

and thus

$$\Phi^{'}(\rho(A)) > 0.$$

Consequently, $\rho(A)$ cannot be a zero of $\Phi(t)$ of multiplicity greater then one, and thus $\rho(A)$ is a simple eigenvalue of A.

Remark 3.2. This result of Theorem 3.3 incidentally shows us that if $A\mathbf{x} = \rho(A)\mathbf{x}$, where $\mathbf{x} > \mathbf{0}$ and $\|\mathbf{x}\| = 1$, we cannor find another eigenvector $\mathbf{y} \neq \gamma\mathbf{x}$ (γ a scalar) of A with $A\mathbf{y} = \rho(A)\mathbf{y}$, so that the eigenvector $\mathbf{x}$ so normalized is uniquely determined.

We now return to the problem of finding bounds for the spectral radius of a matrix. For non-negative matrices, the Perron-Frobenius theory gives us

[8]Historically, the German mathematician Perron (1880-1975) proved in 1907 this theorem assuming that $A > O$. Later (1912), Frobenius extended most of Perron's results to the class of non-negative irreducible matrices.

a nontrivial lower-bound estimates of A. Coupled with Gerschgorin's Theorem 3.1, we thus obtain lemma for the spectral radius of a non-negative irreducible square matrix.

Lemma 3.5. If $A = (a_{ij}) \geq O$ is an irreducible $n \times n$ matrix, then either

$$\sum_{j=1}^{n} a_{ij} = \rho(A) \ for \ all \ 1 \leq i \leq n, \tag{3.12}$$

or

$$\min_{1 \leq i \leq n} \left(\sum_{j=1}^{n} a_{ij} \right) < \rho(A) < \max_{1 \leq i \leq n} \left(\sum_{j=1}^{n} a_{ij} \right). \tag{3.13}$$

Proof. First, suppose that all the row sums of A are equal to σ. If ζ is the vector with all components unity, then obviously $A\zeta = \sigma\zeta$, and $\sigma \leq \rho(A)$ by Definition 2.2. But, Corollary 3.1 shows us that $\rho(A) \leq \sigma$, and we conclude that $\rho(A) = \sigma$, which is the result of (3.12). If all the row sums of A are not equal, we can construct a non-negative irreducible $n \times n$ matrix $B = (b_{ij})$ by decreasing certain positive entries of A so that for all $1 \leq i \leq n$,

$$\sum_{j=1}^{n} b_{ij} = \alpha = \min_{1 \leq i \leq n} \left(\sum_{j=1}^{n} a_{ij} \right),$$

where $O \leq B \leq A$ and $B \neq A$. As all the row sums of B are equal to α, we can apply the result of (3.12) to the matrix B, and thus $\rho(B) = \alpha$. Now, by the statment 3 of Theorem 3.3, we must have that $\rho(B) < \rho(A)$, so that

$$\min_{1 \leq i \leq n} \left(\sum_{j=1}^{n} a_{ij} \right) < \rho(A).$$

On the other hand, we can similarly construct a non-negative irreducible $n \times n$ matrix $C = (c_{ij})$ by incerasing certain of the positive entries of A so that all the row sums of C are equal, where $C \geq A$ and $C \neq A$. It follows that

$$\rho(A) < \rho(C) = \max_{1 \leq i \leq n} \left(\sum_{j=1}^{n} a_{ij} \right),$$

and the combination of these inequalities gives the desired result of (3.13).

Example 3.3. Provable, that

$$A(\alpha) = \begin{pmatrix} 1 & 4 & 1+0.9\alpha \\ 3 & 0.8\alpha & 3 \\ \alpha & 5 & 1 \end{pmatrix}, \ \alpha \geq 0$$

is irreducible. For different α, we determine $\rho(A[\alpha)]$, lower and upper bounds for $\rho[A(\alpha)]$. As $A(\alpha)$ is irreducible we know by the statment 3 of Theorem 3.3 that as α increases $\rho[A(\alpha)]$ is strictly increasing.

α	0	1	2	3	5	10.283
$\rho[A(\alpha)]$	6	6.8878	7.7757	8.6643	10.4438	15.1617
Lower bound	6	6.8	7.6	8.4	10	14.2264
Upper bound	6	7	8	9	11	16.2830

For $\alpha = 0$ the row sums of $A(0)$ are equal, thus by the first part of Lemma 3.5 we know $\rho[A(0)]$, lower and upper bound are equal.

3.3 Reducible Matrices

In the previous sections of this chapter we have investigated the Perron-Frobenius theory of non-negative irreducible square matrices. We now consider extensions of these basic results which make no assumption of irreducibility.
Let A be a reducible $n \times n$ matrix. By Definition 3.1, there exits an $n \times n$ permutation matrix P_1 such that

$$P_1 A P_1^T = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

where A_{11} is an $r \times r$ submatrix and A_{22} is an $(n-r) \times (n-r)$ submatrix, where $1 \leq r \leq n$. We again ask if A_{11} and A_{22} are irreducible, and if not, we reduce them in the manner we initially reduced the matrix. Thus, there exists an $n \times n$ permutation matrix P such that

$$PAP^T = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ O & R_{22} & \cdots & R_{2m} \\ \vdots & & \ddots & \vdots \\ O & O & \cdots & R_{mm} \end{pmatrix},$$

where each square submatrix $R_{jj}, 1 \leq j \leq m$ (3.14), is either irreducible or a 1×1 null matrix. We shall say that (3.14) is the *normal form*[9] of a reducible matrix A. Clearly, the eigenvalues of A are the eigenvalues of the square submatrices $R_{jj}, 1 \leq j \leq m$. With (3.14), we prove the following generalization of Theorem 3.3.

Theorem 3.4. *Let $A \geq O$ be an $n \times n$ matrix. Then,*
1. A has a non-negative real eigenvalue equal to its spectral radius. Moreover, this eigenvalue is positive unless A is reducible and the normal form of A is strictly upper triangular.
2. To $\rho(A)$, there correnspponds an eigenvector $\boldsymbol{x} \geq \boldsymbol{0}$.
3. $\rho(A)$ does not decrease when any entry of A is increased.

Proof. If A is irreducible, the conclusions follow immediately from Theorem 3.3. If A is reducible, assume that A is in the normal form of (3.14). If any submatrix R_{jj} of (3.14) is irreducible, then R_{jj} has a positive eigenvalue equal to its spectral radius. Similarly, if R_{jj} is a 1×1 null matrix, its single eigenvalue is zero. Clearly, A then has a non-negative real eigenvalue equal to its spectral radius. If $\rho(A) = 0$, then each R_{jj} of (3.14) is a 1×1 null matrix, which proves that the matrix of (3.14) is strictly upper triangular. The remaining statements of this theorem follow by applying a continuity argument to the result of Theorem 3.3.
Using the notation of Definition 3.2, we include the following generalization of Lemma 3.3.

Theorem 3.5. *Let A and B be two $n \times n$ matrices with $O \leq |B| \leq A$. Then,*

$$\rho(B) \leq \rho(A).$$

Proof. If A is irreducible, then we know from Lemma 3.3 and its Corollary 3.5 that $\rho(B) \leq \rho(A)$. On the other hand, if A is reducible, we note that the property $O \leq |B| \leq A$ is invariant under similarity transformations by permutation matrices. Putting A into its normal form (3.14), we now simply apply the argument above to the submatrices $|R_{jj}(B)|$ and $R_{jj}(A)$ of the matrices $|B|$ and A, respectively.

[9] This expression is introduced in 1959 by the Russian mathematician Ganthmakher.

4 M-matrices

Consider the matrix A in (1.1). Since A can be expressed in the form

$$A = sI - B,\ s > 0,\ B \geq O, \tag{4.1}$$

it should come no suprise that the theory of non-negative matrices plays a dominant role in the study of certain of these matrices. Matrices of the form (4.1) often occur in relation to systems of linear algebraic equations or eigenvalue problems in a wide variety of areas including finite differnce methods for partial differencial equations, input-output production and growth models in economics and Markov process in probality and statistics.
We adopt here the traditional notation by letting

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

Our aim is to give a systematic treatment of a certain subclass of matrices in $Z^{n\times n}$ called M-matrices.

Definition 4.1. Any matrix A of the form (4.1) for which $s \geq \rho(B)$ is called an M-matrix.

In the following section we consider nonsingular M-matrices, that is, those of the form (4.1) for which $s > \rho(B)$. Characterization theorems are given for $A \in Z^{n\times n}$ and $A \in \mathbb{R}^{n\times n}$ to be a nonsingular M-matrix and the symmetric and irreducible cases are considered.

4.1 Nonsingular M-matrices

Before proceeding to the main characterization theorem for nonsingular M-matrices, it will be convenient to have available the following lemma, which is the matrix version of the Neumann[10] lemma for convergent series.

Lemma 4.1. The non-negative matrix $T \in Z^{n\times n}$ is convergent; that is, $\rho(T) < 1$, if and only if $(I - T)^{-1}$ exists and

$$(I - T)^{-1} = \sum_{k=0}^{\infty} T^k \geq 0. \tag{4.2}$$

[10] János Neumann (1903-1957) was a Hungarian mathematician, who made major contributions to a vast range of fields. He is generally regarded as one of the greatest mathematicians in modern history. Anecdotes from his life in „ P.R.Halmos: The legend of John von Neumann" article.

Proof. If T is convergent then (4.2) follows from the identity

$$(I-T)(I+T+T^2\cdots T^k) = I - T^{k+1},\ k \geq 0,$$

by letting k approach infinity.

For the converse let $T\mathbf{x} = \rho(T)\mathbf{x}$ for some $\mathbf{x} > \mathbf{0}$. (Such an $\mathbf{x}$ exists by the Perron-Frobenius Theorem 3.3.) Then $\rho(T) \neq 1$ since $(I-T)^{-1}$ exists and thus

$$(I-T)\mathbf{x} = [1-\rho(T)]\mathbf{x}$$

implies that

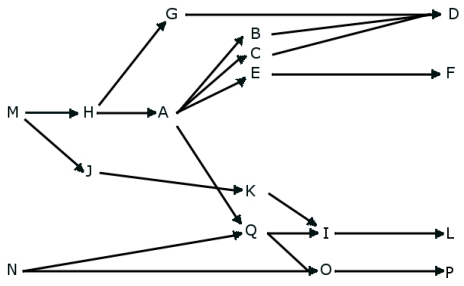
$$(I-T)^{-1}\mathbf{x} = \frac{\mathbf{x}}{1-\rho(T)}.$$

Since $\mathbf{x} > \mathbf{0}$ and $(I-T)^{-1} > 0 \Rightarrow \rho(T) < 1$.

For practical purposes in characterizing nonsingular M-matrices, it is evident that we can often begin by assuming that $A \in Z^{n\times n}$. However, many of the statements of these characterizations are equivalent without this assumption. We have attempted here to group together all such statements into certain categories. Moreover, certain other implications follow without assuming that $A \in Z^{n\times n}$ and we point out such implications in the following inclusive theorem. All matrices and vectors considered in this theorem are real.

4.2 Theorem 4.1.

Theorem 4.1. [2] *Let $A \in \mathbb{R}^{n\times n}$. Then for each fixed letter ϑ representing one of the following conditions, conditions ϑ_i are equivalent for each i. Moreover, letting ϑ then represent any of the equivalent conditions ϑ_i, the following implication tree holds:*



Finally, if $A \in Z^{n\times n}$ then each of the following 50 conditions is equivalent to the statement: " A is a nonsingular M-matrix".

Positivity of principal minors

(A_1) *All of the principal minors of A are positive.*

(A_2) *Every real eigenvalue of each principal submatrix of A is positive.*

(A_3) *For each $\boldsymbol{x} \neq \boldsymbol{0}$ there exists a positive diagonal matrix D such that*

$$\boldsymbol{x}^T AD\boldsymbol{x} > 0.$$

(A_4) *For each $\boldsymbol{x} \neq \boldsymbol{0}$ there exists a nonnegative diagonal matrix D such that*

$$\boldsymbol{x}^T AD\boldsymbol{x} > 0.$$

(A_5) *A does not reverse the sign of any vector; that is, if $\boldsymbol{x} \neq \boldsymbol{0}$ and $\boldsymbol{y} = A\boldsymbol{x}$, then for some subscript i,*

$$x_i y_i > 0.$$

(A_6) *For each* signature matrix *S (here S is diagonal with diagonal entries ± 1), there exists an $\boldsymbol{x} >> \boldsymbol{0}$ such that*

$$SAS\boldsymbol{x} > \boldsymbol{0}.$$

(B_7) *The sum of all the $k \times k$ principal minors of A is positive for $k = 1, \cdots, n$.*

(C_8) *A is nonsingular and all the principal minors of A are non-negative.*

(C_9) *A is nonsingular and every real eigenvalue of each principal submatrix of A is non-negative.*

(C_{10}) *A is nonsingular and $A + D$ is nonsingular for each positive diagonal matrix D.*

(C_{11}) *$A + D$ is nonsingular for each non-negative diagonal matrix D.*

(C_{12}) *A is nonsingular and for each $\boldsymbol{x} \neq \boldsymbol{0}$ there exists a non-negative diagonal matrix D such that*

$$\boldsymbol{x}^T D\boldsymbol{x} \neq 0 \text{ and } \boldsymbol{x}^T AD\boldsymbol{x} \geq 0.$$

(C_{13}) *A is nonsingular and if $\boldsymbol{x} \neq \boldsymbol{0}$ and $\boldsymbol{y} = A\boldsymbol{x}$, then for some subscript i, $x_i \neq 0$, and*

$$x_i y_i \geq 0.$$

(C_{14}) *A is nonsingular and for each signature matrix S there exists a vector $\boldsymbol{x} > \boldsymbol{0}$ such that*

$$SAS\boldsymbol{x} \geq \boldsymbol{0}.$$

(D_{15}) *$A + \alpha I$ is nonsingular for each $\alpha \geq 0$.*

(D_{16}) *Every real eigenvalue of A is positive.*

(E_{17}) *All the leading principal minors of A are positive.*

(E_{18}) *There exist lower and upper triangular matrices L and U, respectively, with positive diagonals such that*

$$A = LU.$$

(F_{19}) *There exists a permutation matrix P such that PAP^T satisfies condition (E_{17}) or (E_{18}).*

Positive Stability

(G_{20}) *A is positive stable; that is, the real part of each eigenvalue of A is positive.*

(G_{21}) *There exists a symmetric positive definite matrix W such that $AW + WA^T$ is positive definite.*

(G_{22}) *$A + I$ is nonsingular and*

$$G = (A + I)^{-1}(A + I)$$

is convergent,

(G_{23}) *$A + I$ is nonsingular and for*

$$G = (A + I)^{-1}(A + I)$$

there exists a positive definite matrix W such that

$$W - G^T W G$$

is positive definite.

(H_{24}) *There exists a positive diagonal matrix D such that*

$$AD + DA^T$$

is positive definite.

(H_{25}) *There exists a positive diagonal matrix E such that for $B = E^{-1}AE$, the matrix*

$$\frac{B + B^T}{2}$$

is positive definite.

(H_{26}) *For each positive semidefinite matrix Q, the matrix QA has a positive*

diagonal element.

Semipositivity and Diagonal Dominance

(I_{27}) *A is semipositive; that is, there exists* $\boldsymbol{x} >> \boldsymbol{0}$ *with* $A\boldsymbol{x} >> \boldsymbol{0}$.
(I_{28}) *There exists* $\boldsymbol{x} > \boldsymbol{0}$ *with* $A\boldsymbol{x} >> \boldsymbol{0}$.
(I_{29}) *There exists a positive diagonal matrix D such that AD has all positive row sums.*
(J_{30}) *There exists* $\boldsymbol{x} >> \boldsymbol{0}$ *with* $A\boldsymbol{x} > \boldsymbol{0}$ *and*

$$\sum_{j=1}^{i} a_{ij}\boldsymbol{x}_j > 0, \quad i = 1, \ldots, n.$$

(K_{31}) *There exists a permutation matrix P such that* PAP^T *satisfies* (J_{30}).
(L_{32}) *There exists* $\boldsymbol{x} >> \boldsymbol{0}$ *with* $\boldsymbol{y} = A\boldsymbol{x} > \boldsymbol{0}$ *such that if* $y_{i_0} = 0$, *then there exists a sequence of indices* $i_1, \ldots, n$ *with* $a_{i_{j-1}i_j} \neq 0, j = 1, \ldots, r$, *and with* $y_{i_r} \neq 0$.
(L_{33}) *There exists* $\boldsymbol{x} >> \boldsymbol{0}$ *with* $\boldsymbol{y} = A\boldsymbol{x} > \boldsymbol{0}$ *such that the matrix* $\widetilde{A} = \widetilde{(a_{ij})}$ *defined by*

$$\widetilde{(a_{ij})} = \begin{cases} 1 & \text{if } a_{ij} \neq 0 \text{ or } y_i \neq 0, \\ 0 & \text{otherwise} \end{cases}$$

is irreducible. (M_{34}) *There exists* $\boldsymbol{x} >> \boldsymbol{0}$ *such that for each signature matrix S,*

$$SAS\boldsymbol{x} >> \boldsymbol{0}.$$

(M_{35}) *A has all positive diagonal elements and there exists a positive diagonal matrix D such that AD is strictly diagonally dominant; that is,*

$$a_{ii}d_i > \sum_{j \neq i} |a_{ij}|d_j, \ i = 1, \ldots, n.$$

(M_{36}) *A has all positive diagonal elements and there exists a positive diagonal matrix E such that* $E^{-1}AE$ *is strictly diagonally dominant.*
(M_{37}) *A has all positive diagonal elements and there exists a positive diagonal matrix D such that AD is lower semistrictly diagonally dominant; that is,*

$$a_{ii}d_i \geq \sum_{j \neq i} |a_{ij}|d_j, \ i = 1, \ldots, n,$$

and

$$a_{ii}d_i > \sum_{j=1}^{i-1} |a_{ij}|d_j,\ i = 2, \ldots, n.$$

Inverse-Positivity and Splittings

(N_{38}) *A is inverse-positive; that is, A^{-1} exists and*

$$A^{-1} \geq 0.$$

(N_{39}) *A is monotone; that is, $A\boldsymbol{x} \geq \boldsymbol{0} \rightarrow \boldsymbol{x} \geq \boldsymbol{0}$ for all $\boldsymbol{x} \in \mathbb{R}^n$.*

(N_{40}) *There exist inverse-positive matrices B_1 and B_2 such that*

$$B_1 \leq A \leq B_2.$$

(N_{41}) *There exists an inverse-positive matrix $B \geq A$ such that $I - B^{-1}A$ is convergent.*

(N_{42}) *There exists an inverse-positive matrix $B \geq A$ and A satisfies $(I_{27}), (I_{28})$ or (I_{29}).* (N_{43}) *There exists an inverse-positive matrix $B \geq A$ and a nonsingular M-matrix C, such that*

$$A = BC.$$

(N_{44}) *There exists an inverse-positive matrix B and a nonsingular M-matrix C such that*

$$A = BC.$$

(N_{45}) *A has a convergent regular splitting; that is, A has a representation*

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

where $M^{-1}N$ is convergent.

(N_{46}) *A has a convergent weak regular splitting; that is, A has a representation*

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

where $M^{-1}N$ is convergent.

(O_{47}) *Every weak regular splitting of A is convergent.*

(P_{48}) *Every regular splitting of A is convergent.*

Linear Inequalities

(Q_{49}) *For each* $\boldsymbol{y} \geq \boldsymbol{0}$ *the set*

$$S_{\boldsymbol{y}} = \{\boldsymbol{x} \geq \boldsymbol{0} : A^T \boldsymbol{x} \leq \boldsymbol{y}\}$$

is bounded, and A *is nonsingular.*
(Q_{50}) $S_{\boldsymbol{0}}$ *that is; the inequalities* $A^T\boldsymbol{x}$ *and* $\boldsymbol{x} \geq \boldsymbol{0}$ *have only the trivial solution* $\boldsymbol{x} = \boldsymbol{0}$*, and* A *is nonsingular.*

Proof. We show now that each of the 50 conditions, $(A_1) - (Q_{50})$, can be used to characterize a nonsingular M-matrix A, beginning with the assumption that $A \in Z^{n\times n}$. Suppose first that A is a nonsingular M-matrix. Then in view of the implication tree in the statement of the theorem, we need only show that conditions (M) and (N) hold, for these conditions together imply each of the remaining conditions in the theorem for arbitrary $A \in \mathbb{R}^{n\times n}$. Now by Definition 4.1, A has the representation $A = sI - B,\ B \geq O$ with $s \geq \rho(B)$. Moreover, $s > \rho(B)$, since A is nonsingular. Letting $T = \frac{B}{s}$, it follows that $\rho(T) < 1$ so by Lemma 4.1,

$$A^{-1} = (I - T)^{-1}s \geq 0.$$

Thus condition (N_{38}) holds. In addition, A has all positive diagonals since the inner product of the ith row of A with the ith column of A^{-1} is one for $i = 1, \ldots, n$. Let $x = A^{-1}e$, where $e = (1, \ldots, 1)^T$. Then for $D = diag(x_1, x_2, \ldots, x_n)$, D is a positive diagonal matrix and

$$AD\mathbf{e} = A\mathbf{x} = \mathbf{e} >> \mathbf{0},$$

and thus AD has all positive row sums. But since $A \in Z^{n\times n}$, this means that AD is strictly diagonally dominant and thus M_{35} holds. We show next that if $A \in Z^{n\times n}$ satisfies any of the conditions $(A) - (Q)$, then DA is a nonsingular M-matrix. Once again, in view of the implication tree, it suffices to consider only conditions $(D), (F), (L)$ and (P).
Suppose condition (D_{16}) holds for A and let $A = sI - B,\ B \geq O,\ s > 0$. Suppose that $s \leq \rho(B)$. Then if $B\mathbf{x} = \rho(B)\mathbf{x},\ \mathbf{x} \neq \mathbf{0}$,

$$Ax = [s - \rho(B)]\mathbf{x},$$

so that $s - \rho(B)$ would be a nonpositive real eigenvalue of A, contradicting (D_{16}). Now suppose condition (F_{19}) holds for A. Thus suppose $PAP^T = LU$ where L is lower triangular with positive diagonals and U is upper triangular

with positive diagonals. We first show that the off-diagonal elements of both L and U are nonpositive. Let $L = (r_{ij}),\ U = (s_{ij})$ so that

$$r_{ij} = 0 \text{ for } i < j \text{ and } s_{ij} = 0 \text{ for } i > j \text{ and } r_{ii} > 0,\ s_{ii} > 0 \text{ for } 1 \le i, j \le n.$$

We shall prove the inequalities $r_{ij} \le 0,\ s_{ij} \le 0$ for $i \ne j$ by induction on $i + j$. Let

$$A' = PAP^T = (a'_{ij}).$$

If $i + j = 3$ the inequalities $r_{21} \le 0$ and $s_{21} \le 0$ follow from $a'_{12} = r_{11}s_{12}$ and $a'_{21} = r_{21}s_{11}$. Let $i + j > 3,\ i \ne j$, and suppose the inequalities $r_{kl} \ge 0$ and $s_{kl} \ge 0,\ k \ne l$, are valid if $k + l < i + j$. Then if $i < j$ in the realtion

$$a_{ij} = r_{ii}s_{ij} + \sum_{k<i} r_{ik}s_{kj},$$

we have

$$a_{ij} \le 0,\ \sum_{k<i} r_{ik}s_{kj} \ge 0 \text{ since } r_{ik} \le 0,\ s_{kj} \le 0$$

according to $i + k < i + j,\ k + j < i + j$. Thus $s_{ij} \le 0$. Analogously, for $i > j$ the inequality $r_{ij} \le 0$ can be proved. It is easy to see then that L^{-1} and U^{-1} exist and are nonnegative. Thus

$$A^{-1} = (P^T LUP)^{-1} = P^T U^{-1} L^{-1} P \ge O.$$

Now letting $A = sI - B,\ s > 0,\ B \ge O$ it follows that $(I - T)^{-1} \ge O$, where $T = \frac{B}{s}$. Then $\rho(B) < 1$ by Lemma 4.1, and thus $s > \rho(B)$ and A is a nonsingular M-matrix. Next, assume that condition (L_{33}) holds for A. We write $A = sI - B,\ s > 0,\ B \ge O$ and let $T = \frac{B}{s}$. Then since $\mathbf{y} = A\mathbf{x} > \mathbf{0}$ for some $\mathbf{x} >> \mathbf{0}$, it follows that $T\mathbf{x} < \mathbf{x}$. Now define $\hat{T} = (\hat{t}_{ij})$ by

$$\hat{t}_{ij} = \begin{cases} t_{ij} & \text{if } t_{ij} \ne 0, \\ \epsilon & \text{if } t_{ij} = 0 \text{ and } y_i \ne 0, \\ 0 & \text{otherwise.} \end{cases}$$

It follows then that T is irreducible since $\widetilde{A}$ defined in (L_{33}) is irreducible. Moreover, for sufficiently small $\epsilon > 0$,

$$\hat{T}\mathbf{x} < \mathbf{x},$$

so that $\rho(\hat{T}) < 1$ by the Perron-Frobenius Theorem 3.3. Finally, since $T \le \hat{T}$ it follows that

$$\rho(T) \le \rho(\hat{T}) < 1,$$

by Theorem 3.5. Thus A is a nonsingular M-matrix. Next, assume that condition (P_{48}) holds for A. Then since A has a regular splitting of the form $A = sI - B,\ s > 0,\ B \geq O$, it follows from (P_{48}) that $T = \frac{B}{s}$ is convergent, so that $s > \rho(B)$ and A is a nonsinguIar M-matrix. This completes the proof of Theorem 4.1 for $A \in Z^{n\times n}$.

5 Iterative methods for the linear algebraic systems with M-matrices

As we have seen in chapter two, there are two general schemes for solving SLAE of the form (2.2'): direct and iterative methods. We mention two DM. The following two general schemes applicable for M-matrices.
We prove the Gaussian elimination and the LU factorization exist for M-matrices. The LU factorization is a matrix decomposition which writes a matrix as the product of a lower triangular L matrix and an upper triangular U matrix.

Theorem 5.1. *If $A \in \mathbb{R}^{n\times n}$ is an M-matrix, then the LU factorization exists and the Gaussian elimination is executable.*

Proof. To prove this theorem we use the fact there exist a unique LU factorization if and only if the pricipal minors of A are nonzero (It can be seen in [3]). This theorem is valid for M-matrices by the statement of (A_1) of Theorem 4.1.
The LU factors of an M-matrix are guaranteed to exist and can be stably computed without need for numerical pivoting, also have positive diagonal entries and non-positive off-diagonal entries.
In this chapter we apply nonnegativity to the study of iterative methods for solving SLAE of the form (2.2'). We shall study various iterative methods for approximating $\mathbf{x}$. Such methods are usually ideally suited for problems involving large sparse matrices, much more so in most cases than direct methods such as Gaussian elimination. A typical iterative method involves the selection of an initial approximation $\mathbf{x}^{(0)}$ to the solution $\mathbf{x}$ to (2.2') and the determination of a sequence $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots$ according to some specified algorithm which, if the method is properly chosen, will converge to the exact solution $\mathbf{x}$ of (2.2'). Since

$\mathbf{x}$ is unknown, a typical method for terminating the iteration might be whenever

$$\frac{|\mathbf{x}_i^{(k+1)} - \mathbf{x}_i^{(k)}|}{|\mathbf{x}_i^{(k+1)}|} < \epsilon \quad i = 1, \ldots, n, \tag{5.1}$$

where ϵ is some pre-chosen small number, depending upon the precision of the computer being used. Essentially, the stopping criteria (5.1) for the iteration is that we will choose $\mathbf{x}^{(k+1)}$ as the approximation to the solution $\mathbf{x}$ whenever the relative difference between successive approximations $\mathbf{x}_i^{(k)}$ and $\mathbf{x}_i^{(k+1)}$ becomes sufficiently small for $i = 1, \ldots, n$.

5.1 Basic iterative methods

5.1.1 Jacobi and Gauss-Seidel method

We begin this section by describing two iterative formulas for solving the linear system (2.2'). Here we assume that the coefficient matrix $A = (a_{ij})$ for (2.2') is nonsingular and has all nonzero diagonal entries.
Assume that the kth approximating vector $\mathbf{x}^{(k)}$ to $\mathbf{x} = A^{-1}\mathbf{b}$ has been computed. Then the *Jacobi*[11] *method* for computing $\mathbf{x}^{(k+1)}$is given by

$$\mathbf{x}_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j \neq i} a_{ij}\mathbf{x}_j^{(k)}\right), \quad i = 1, \ldots, n. \tag{5.2}$$

Now let $D = diag(A)$ and $-L$ and $-U$ be the strictly lower and strictly upper triangular parts of A, respectively; that is,

$$L = -\begin{pmatrix} 0 & & & 0 \\ a_{21} & & & \\ \vdots & & \ddots & \\ a_{n1} & \ldots & a_{nn-1} & 0 \end{pmatrix}, \quad U = -\begin{pmatrix} 0 & a_{12} & \ldots & a_{1n} \\ & & & \vdots \\ & & \ddots & a_{n-1n} \\ 0 & & & 0 \end{pmatrix}.$$

Then, clearly, (5.2) may be written in matrix form as

$$\mathbf{x}^{(k+1)} = D^{-1}(L + U)\mathbf{x}^{(k)} + D^{-1}\mathbf{b}, \quad k = 0, 1, \ldots \tag{5.3}$$

A closely related iteration may be derived from the following observation. If we assume that the computations of (5.2) are done sequentially for $i = 1, \ldots, n$, then

[11] It was originally considered by the Prussian mathematician Carl Gustav Jacob Jacobi (1804-1851).

at the time we are ready to compute $\mathbf{x}_i^{(k+1)}$ the new components $\mathbf{x}_1^{(k+1)}, \ldots, \mathbf{x}_{i-1}^{(k+1)}$ are already available, and it would seem reasonable to use them instead of the old components; that is, we compute

$$\mathbf{x}_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} \mathbf{x}_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} \mathbf{x}_j^{(k)} \right), \quad i = 1, \ldots, n. \tag{5.4}$$

This is the *Gauss-Seidel*[12] *method*. It is easy to see that (5.4) may be written in the form

$$D\mathbf{x}^{(k+1)} = \mathbf{b} + L\mathbf{x}^{(k+1)} + U\mathbf{x}^{(k)},$$

so that the matrix form of the Gauss-Seidel method (5.4) is given by

$$\mathbf{x}^{(k+1)} = (D - L)^{-1} U \mathbf{x}^{(k)} + (D - L)^{-1} \mathbf{b}, \quad k = 0, 1, \ldots \tag{5.5}$$

Of course (5.2) and (5.4) should be used rather than their matrix formulations (5.3) and (5.5) in programming these methods. We shall return to these two fundamental procedures later, but first we consider the more general iterative formula

$$\mathbf{x}^{(k+1)} = H\mathbf{x}^{(k)} + \mathbf{c}, \quad k = 0, 1, \ldots \tag{5.6}$$

The matrix H is called the iteration matrix for (5.6) and it is easy to see that if we split A into

$$A = M - N, \qquad M \text{ nonsingular},$$

then for $H = M^{-1}N$, $\mathbf{c} = M^{-1}\mathbf{b}$ and $\mathbf{x} = H\mathbf{x} + \mathbf{c}$ if and only if $A\mathbf{x} = \mathbf{b}$. Clearly, the Jacobi method is based upon the choice

$$M = D, \qquad N = L + U,$$

while the Gauss-Seidel method is based upon the choice

$$M = D - L, \quad N = U.$$

Then for the Jacobi method $H = M^{-1}N = D^{-1}(L + U)$ for the Gauss-Seidel method $H = M^{-1}N = (D - L)^{-1}U$. We next prove the basic convergence lemma for (5.6).

[12] It is named after the German mathematicians Carl Friedrich Gauss and Philipp Ludwig von Seidel (1821-1896). In numerical linear algebra also known as the *Liebmann method* or the *method of successive displacement*.

Lemma 5.1. Let $A = M - N$ is an arbitrary $n \times n$ matrix with A and M nonsingular. Then for $H = M^{-1}N$ and $\mathbf{c} = M^{-1}\mathbf{b}$, the iterative method (5.6) converges to the solution $\mathbf{x} = A^{-1}\mathbf{b}$ to (2.2') for each $\mathbf{x}^{(0)}$ if and only if $\rho(H) < 1$.

Proof. If we subtract $\mathbf{x} = H\mathbf{x} + \mathbf{c}$ from (5.6), we obtain the error equation

$$\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} = H(\mathbf{x}^{(k)} - \mathbf{x}) = \ldots = H^{k+1}(\mathbf{x}^{(0)} - \mathbf{x}).$$

Hence the sequence $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots$, converges to $\mathbf{x}$ for each $\mathbf{x}^{(0)}$ if and only if

$$\lim_{k \to \infty} H^k = O,$$

that is, if and only if $\rho(H) < 1$, by considering the Jordan form for H, which proves the lemma.

In short we shall say that a given iterative method converges if the iteration (5.6) associated with that method converges to the solution to the given linear system for every $\mathbf{x}^{(0)}$. Now we discuss the general topic of rates of convergence.

Definition 5.1. For $H \in \mathbb{C}^{n \times n}$ assume that $\rho(H) < 1$ and let $\mathbf{x} = H\mathbf{x} + \mathbf{c}$. Then for

$$\alpha = \sup \left\{ \lim_{k \to \infty} \|\mathbf{x}^{(k)} - \mathbf{x}\|^{\frac{1}{k}} : \mathbf{x}^{(0)} \in \mathbb{C}^n \right\} \tag{5.7}$$

the number

$$R_\infty(H) = -\ln(\alpha)$$

is called the *asymptotic rate of convergence* of the iteration (5.6).

The supremum is taken in (5.7) so as to reflect the worst possible rate of convergence for any $\mathbf{x}^{(0)}$. Clearly, the larger the $R_\infty(H)$, the smaller the α and thus the faster the convergence of the process. Also note that α is independent of the particular vector norm used in (5.7). In the next section, the Gauss-Seidel method is modified using a relaxation parameter ω and it is shown how, in certain instances, to choose ω in order to maximize the asymptotic rate of convergence for the resulting process.

Remark 5.1. Let $H \in \mathbb{R}^{n \times n}$ assume that $\rho(H) < 1$. Then, the asymptotic rate of convergence of (5.6) is

$$R_\infty(H) = -\ln \rho(H).$$

5.1.2 SOR method

Here we investigate in some detail a procedure that can sometimes be used to accelerate the convergence of the Gauss-Seidel method. We first note that the Gauss-Seidel method can be expressed in the following way. Let $\overline{\mathbf{x}}_i^{(k+1)}$ be the ith component of $\mathbf{x}^{(k+1)}$ computed by the formula (5.4) and set

$$\Delta \mathbf{x}_i = \overline{\mathbf{x}}_i^{(k+1)} - \mathbf{x}_i^{(k)}.$$

Then for $\omega = 1$, the Gauss-Seidel method can be restated as

$$\mathbf{x}_i^{(k+1)} = \mathbf{x}_i^{(k)} + \omega \Delta \mathbf{x}_i, \quad i = 1, \ldots, n, \quad \omega > 0. \tag{5.8}$$

It was discovered during the years of hand computation (probably by accident) that the convergence is often faster if we go beyond the Gauss-Seidel correction $\Delta \mathbf{x}_i$. If $\omega > 1$ we are *overcorrecting* while if $\omega < 1$ we are *undercorrecting*. As just indicated, if to $\omega = 1$ we recover the Gauss-Seidel method (5.4). In general the method (5.8) is called the *successive overrelaxation* (SOR[13]) method. Of course the problem here is to choose the relaxation parameter ω to so as to maximize the asymptotic rate of convergence of (5.8).
In order to write this procedure in matrix form, we replace $\overline{\mathbf{x}}_i^{(k+1)}$ in (5.8) by the expression in (5.4) and rewrite (5.8) as

$$\mathbf{x}_i^{(k+1)} = (1-\omega)\mathbf{x}_i^{(k)} + \frac{\omega}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} \mathbf{x}_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij} \mathbf{x}_j^{(k)} \right). \tag{5.9}$$

Then (5.9) can be rearranged into the form

$$a_{ii}\mathbf{x}_i^{(k+1)} + \omega \sum_{j=1}^{i-1} a_{ij} \mathbf{x}_j^{(k+1)} = (1-\omega) a_{ii} \mathbf{x}_i^{(k)} - \omega \sum_{j=i+1}^{n} a_{ij} \mathbf{x}_j^{(k)} + \omega \mathbf{b}_i.$$

This relation of the new iterates $\mathbf{x}_i^{(k+1)}$ to the old $\mathbf{x}_i^{(k)}$ holds for $i = 1, \ldots, n$, and by means of the decomposition $A = D - L - U$, we can written as

$$D\mathbf{x}^{(k+1)} - \omega L \mathbf{x}^{(k+1)} = (1-\omega) D \mathbf{x}^{(k)} + \omega U \mathbf{x}^{(k)} + \omega \mathbf{b}$$

or, under the assumption that $a_{ii} \neq 0,\ i = 1, \ldots, n$,

$$\mathbf{x}^{(k+1)} = H_\omega \mathbf{x}^{(k)} + \omega (D - \omega L)^{-1} \mathbf{b}, \quad k = 0, 1, \ldots, \tag{5.10}$$

[13] This iterative method is also called the *accelerated Liebmann* method, the *extrapolated Gauss-Seidel* method and the method of *systematic overrelation*.

where

$$H_\omega = (D - \omega L)^{-1}[(1-\omega)D + \omega U]. \tag{5.11}$$

We first prove a result that gives the maximum range of values of $\omega > 0$ for which the SOR iteration can possibly converge.

Theorem 5.2. *Let A an arbitrary $n \times n$ matrix have all nonzero diagonal elements. Then the SOR method (5.8) converges only if*

$$0 < \omega < 2. \tag{5.12}$$

Proof. Let H_ω be given by (5.11). In order to establish (5.12) under the assumption that $\rho(H_\omega) < 1$, it suffices to prove that

$$|\omega - 1| \leq \rho(H_\omega). \tag{5.13}$$

Because L is strictly lower triangular, $detD^{-1} = det(D - \omega L)^{-1}$. Thus

$$detH_\omega = det(D - \omega L)^{-1} det[(1-\omega)D + \omega U] =$$
$$= det[(1-\omega)I + \omega D^{-1}U] = det[(1-\omega)I] = (1-\omega)^n,$$

because $D^{-1}U$ is strictly upper triangular. Then since $detH_\omega$, is the product of its eigenvalues, (5.12) must hold and the theorem is proved.

It will be shown in following section that for certain important classes of matrices, (5.12) is also a sufficient condition for convergence of the SOR method. For other classes this method converges if and only if $0 < \omega < c$ for some fixed $c > 0$ which depends upon A.

5.2 Convergence

In this section we investigate several important convergence criteria for the iterative methods for solving SLAE in the form (1.1). We now investigate certain types of general splittings, discussed first in Theorem 4.1, in terms of characterizations of nonsingular M-matrices.

Definition 5.2. The splitting $A = M - N$ with A and M nonsingular is called a *regular splitting* if $M^{-1} \geq O$ and $N > O$. It is called a *weak regular splitting* if $M^{-1} \geq O$ and $M^{-1}N \geq O$.

Clearly, a regular splitting is a weak regular splitting. The next result relates the convergence of (5.4) to the inverse-positivity of A. We shall call it the (weak) regular splitting theorem.

Remark 5.2. If $A \in \mathbb{R}^{n\times n}$ is an M-matrix, then it has been [weak] regular splitting by the statments (N_{45}) and (N_{46}) of Theorem 4.1.

$$M = D \qquad N = L + U \qquad \text{(Jacobi method)}$$
$$M = D - L \qquad N = U \qquad \text{(Gauss-Seidel method)}$$
$$M = D - \omega L \quad N = (1-\omega)L + U \qquad \text{(SOR method)}$$

We know by Definition 5.2 that $A = M - N$ and we receive the desired result if we do the extractions.

Theorem 5.3. *Let $A = M - N$ be a weak regular splitting of A. Then the following statements are equivalent:*
(1) $A^{-1} \geq O$ that is, A is inverse-positive
(2) $A^{-1}N \geq O$
(3) $\rho(H) = \frac{\rho(A^{-1}N)}{1+\rho(A^{-1}N)}$ so that $\rho(H) < 1$.

Proof. It can be seen in [2] as Theorem 5.6.

Corollary 5.1. Let A be inverse-positive and let $A = M_1 - N_1$ and $A = M_2 - N_2$ be two regular splittings of A where $N_2 \leq N_1$. Then for $H^1 = M_1^{-1}N_1$ and $H^2 = M_2^{-1}N_2$,

$$\rho(H^2) \leq \rho(H^1) < 1$$

so that

$$R_\infty(K) \geq R_\infty(H).$$

Proof. The proof follows from (3) of Theorem 5.3, together with the fact that $\alpha(1+\alpha)^{-1}$ is an increasing function of α for $\alpha \geq 0$, which proves the theorem.

As indicated earlier, every [weak] regular splitting of a nonsingular M-matrix is convergent by the statments (N_{45}) and (N_{46}) of Theorem 4.1. Clearly for such matrices, the Jacobi and Gauss-Seidel methods defined by (5.3) and (5.5) are based upon regular splittings. Moreover, if $0 < \omega \leq 1$, then the SOR method defined by (5.10) is based upon a regular splitting and in this case the SOR

method is convergent by the Kahan[14]-theorem. These concepts will now be extended to an important class of complex matrices.
Let $A \in \mathbb{R}^{n\times n}$, have all nonzero diagonal elements and let $A = D - L - U$, where as usual $D = diag(A)$ and where $-L$ and $-U$ represent the lower and the upper parts of A, respectively.
We return to the case where $A \in \mathbb{R}^{n\times n}$ and A is a nonsingular M-matrix. In the following analysis we may assume, without loss of generality, that $D = diag(A) = I$. In this case the Jacobi iteration matrix for A is given by

$$J = L + U,$$

while the SOR iteration matrix for A is given by

$$H_\omega = (I - \omega L)^{-1}[(1-\omega)I + \omega U].$$

Theorem 5.4. *Let $A = I - L - U \in \mathbb{R}^{n\times n}$ where $L \geq O$ and $U \geq O$ are strictly lower and upper triangular, respectively. Then for $0 < \omega < 1$,*
(1) $\rho(J) < 1$ if and only if $\rho(H_\omega) < 1$.
(2) $\rho(J) < 1$ if and only if A is a nonsingular M-matrix, in which case

$$\rho(H_\omega) \leq 1 - \omega + \omega\rho(J).$$

(3) if $\rho(J) \geq 1$ then $\rho(H_\omega) \geq 1 - \omega + \omega\rho(J) \geq 1$.

Proof. It can be seen in [2] as Theorem 5.21.

Corollary 5.2. Let A be as in Theorem 5.4. Then
(1) $\rho(J) < 1$ if and only if $\rho(H_1) < 1$.
(2) $\rho(J) < 1$ and $\rho(H_1) < 1$ if and only if A is a nonsingular M-matrix, moreover, $\rho(J) < 1$ then

$$\rho(H_1) \leq \rho(J).$$

(3) if $\rho(J) \geq 1$ then $\rho(H_1) \geq \rho(J) \geq 1$.

[14]William Morton Kahan (1933-) is a Canadian mathematician and computer scientist whose main area of contribution has been numerical analysis.

Example 5.1. We use the Jacobi and the Gauss-Seidel method to solve the linear system

$$A \cdot \mathbf{x} = \begin{pmatrix} 18 & -4 & 0 & -8 \\ -3 & 18 & -3 & -6 \\ -5 & -5 & 18 & -9 \\ -2 & 0 & -7 & 18 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \\ 5 \\ 4 \end{pmatrix},$$

and let the starting vector be $\mathbf{p} = (0, -1, 1, 1)^T$. Since the $\rho(A)$ is $0.700678 < 1$, methods in this example will converge. A tolerance can be supplied to either the Jacobi or Gauss-Seidel method which will permit it to exit the loop if convergence has been achieved. We use different tolerances and a maximum of 1000 iterations.

Tolerance	10^{-2}	10^{-3}	10^{-4}	10^{-6}	10^{-10}	10^{-18}
Jacobi	7	13	19	32	58	99
Gauss-Seidel	6	9	12	18	30	48

We next give a comparison theorem of the convergence rates of the SOR method for nonsingular M-matrices.

Theorem 5.5. *Let A be a nonsingular M-matrix and let $0 < \omega_1 \leq \omega_2 \leq 1$. Then*

$$\rho(H_{\omega_2}) \leq \rho(H_{\omega_1}) < 1,$$

so that

$$R_\infty(H_{\omega_2}) \geq R_\infty(H_{\omega_1}).$$

Proof. Let $A = D - L - U$. Then

$$H_\omega = M_\omega^{-1} N_\omega,$$

where

$$M_\omega = \omega^{-1} D - L, \qquad N_\omega = (\omega^{-1} - 1)D + U.$$

But $M_\omega^{-1} \geq O$ and $N_\omega \geq O$ for $0 < \omega \leq 1$ as before, and $A = M_\omega - N_\omega$ is a regular splitting of A. Now since to ω^{-1} is a decreasing function of ω for $0 < \omega \leq 1$, it follows that if $0 < \omega_1 \leq \omega_2 \leq 1$, then

$$N_{\omega_2} \leq N_{\omega_1}.$$

The result then follows from Corollary 5.1, which proves Theorem 5.5.

Now if A is an M-matrix, then $\rho(J) < 1$ by Theorem 5.4, and by Theorem 5.5, $\rho(H_\omega)$ is a nonincreasing function of ω in the range $0 < \omega \leq 1$. Moreover, $\rho(H_\omega) < 1$. By the continuity of $\rho(H_\omega)$ as a function of ω, it must follow that $\rho(H_\omega) < 1$ for $0 < \omega \leq \alpha$ with some $\alpha > 1$.

Theorem 5.6. *If A is an M-matrix then*

$$\rho(H_\omega) < 1$$

for all ω satisfying

$$0 < \omega < \frac{2}{1 + \rho(J)}. \tag{5.14}$$

Proof. The convergence follows from Theorem 5.4, whenever $0 < \omega \leq 1$, so assume that $\omega > 1$. Assume that $D = diag(A) = I$, as usual, and define the matrix

$$T_\omega = (I - \omega L)^{-1}[\omega U + (\omega - 1)I].$$

Then clearly

$$T_\omega \geq O \quad \text{and} \quad |H_\omega| \leq T_\omega.$$

Let $\lambda = \rho(T_\omega)$. Then for some $\mathbf{x} > \mathbf{0}$, we have $T\mathbf{x} = \lambda\mathbf{x}$ by Theorem 3.3, and so

$$(\omega U + \omega\lambda L)\mathbf{x} = (\lambda + 1 - \omega)\mathbf{x}.$$

Hence

$$\lambda + 1 - \omega \leq \rho(\omega U + \omega\lambda L)$$

and if $\lambda \geq 1$, then

$$\lambda + 1 - \omega \leq \omega\lambda\rho(J)$$

since $\lambda L \geq L$. In this case then

$$\omega \geq \frac{1 + \lambda}{1 + \lambda\rho(J)} \geq \frac{2}{1 + \rho(J)}.$$

Hence if (5.14) hold then we must have $\lambda < 1$. Then from $|H_\omega| \leq T_\omega$ it follows that $\rho(H_\omega) \leq \lambda < 1$, and the theorem is proved.

6 Summary

In this thesis we have studied different numerical methods for solving SLAE with M-matrices. Each single method is important from a viewpoint of the solvability: knowing of the condition number for solving these equations on computer, the knowledge of bounds for spectral radius can guarantee that the method will be convergent, the theory of nonnegative matrices set up the theory of M-matrices and finally the different iterative methods for SLAE with M-matrices help to accelerate the convergence.

We got different special properties for M-matrices. As we have seen in part of DM, the proof of the existence and uniqueness of LU factorization is simple and we can stably computed without need for numerical pivoting in this case. We can easily got the regular splittings of various iterative methods. In the end we considered an extension of Kahan's theorem.

References

[1] Richard Bellman : *Introduction to matrix analysis*, Mcgraw-Hill, 1960.

[2] Abraham Berman, Robert J. Plemmons : *Nonnegative matrices in the mathematical sciences*, Academic Press, 1979.

[3] István Faragó: *Alkalmazott Analízis I-II.*, ELTE kézirat.

[4] Pál Rózsa : *Lineáris algebra és alkalmazásai*, Tankönyvkiadó, 1974.

[5] Gisbert Stoyan : Numerikus matematika mérnököknek és programozóknak, Typotex, 2007.

[6] Richard S. Varga : *Matrix Iterative Analysis*, Prentice-Hall, 1962.