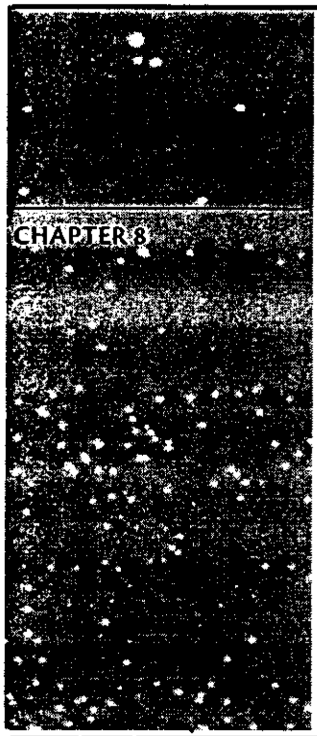

Elementary Numerical Methods in Reactor Statics

GETTING THE SOLUTION to the static diffusion equations rests necessarily on numerical methods. These methods are mostly iterative in nature. In one dimension, it is possible to solve in a direct fashion the matrix problems generated by finite differences. However, determining the K_{eff} is an iterative process. Only the simplest methods will be reviewed here. The resulting algorithms constitute the basis used in modern computer codes for neutronic analysis of the nuclear core.

Vector Norms

Vector norms were first introduced to provide a measure of the length of a vector, and also of the distance separating two vectors. Only the basic properties of vector norms are listed here:

$$\|u\| \geq 0$$

$$\|u\| = 0 \Leftrightarrow u = 0$$

$$\|\alpha u\| = |\alpha| \cdot \|u\|$$

$$\|u + v\| \leq \|u\| + \|v\| \text{ where } \alpha \text{ is a scalar}$$

The "p" norms are written $\|u\|_p = \sqrt[p]{\sum_{i=1}^{n} |u_i|^p}$

The most often used vector norms are :

- the 1 norm, $\|u\|_1 = \sum_{i=1}^{n} |u_i|^p$
- the 2 norm or euclidean norm, $\|u\|_2 = \sqrt{\sum_{i=1}^{n} |u_i|^2}$
- the ∞ norm, $\|u\|_\infty = \max|u_i|$,i=1,n

Power Method

The linear system that we must solve is given by equation (46which is repeated here

$$[A][\psi] = \frac{1}{K_{eff}}[B][\psi] \qquad \text{(EQ 47)}$$

whatever discretisation method is chosen. We have seen in the preceding chapter what are the matrix elements of $[A]$ and $[B]$ for either classical or mesh centered finite differences.

The determination of the eigenvalue K_{eff} is a relatively simple task. Equation (47) is a generalized eigenvalue-eigenvector problem. The right hand side includes the matrix $[B]$ which is usually absent in the standard eigenproblem. However, we can transform problem (47) into a standard eigenproblem, first by multiplying by K_{eff}, and the by multiplying by the inverse of $[A]$ to get

$$[A]^{-1}[B][\psi] = K_{eff}[\psi] \tag{EQ 48}$$

This is altogether identical to the initial problem (47). To simplify the discussion, we introduce a new matrix

$$[M] \equiv [A]^{-1}[B] \tag{EQ 49}$$

and then system (48)becomes

$$[M][\phi] = K_{eff}[\psi] \tag{EQ 50}$$

which is conventional eigenvalue-eigenvector problem, with K_{eff} as the eigenvalue, and the flux vector $[\psi]$ as eigenvector. But K_{eff} is the most positive eigenvalue, and to it corresponds to only eigenvector with positive elements everywhere. This is a very simple problem to solve with the power method.

Simple Method

Let us note by γ_e the set of eigenvalues of $[\mathbb{M}]$, and by $\vec{w}_e$ the corresponding eigenvectors. Furthermore, denote by γ_1 the most positive of the γ_e. In other words,

$$K_{eff} = \gamma_1 \qquad \text{(EQ 51)}$$

Now if we suppose that the eigenvectors $\vec{w}_e$ form a complete set, any vector can be written as a linear combination of these eigenvectors.

We first choose an initial arbitrary flux guess (but different from zero)

$$[\psi]^0 = \sum_e a_e \vec{w}_e \qquad \text{(EQ 52)}$$

and we apply to it the matrix $[\mathbb{M}]$, to obtain a new vector $[\psi]^1$,

$$[\psi]^1 = [\mathbb{M}][\psi]^0 = [\mathbb{M}]\sum_e a_e \vec{w}_e$$

$$[\psi]^1 = \sum_e a_e [\mathbb{M}]\vec{w}_e$$

$$[\psi]^1 = \sum_e a_e \gamma_e \vec{w}_e$$

We repeat this process a second time, to obtain a new flux $[\psi]^2$,

$$[\psi]^2 = [M][\psi]^1 = [M]\sum_e a_e\gamma_e\vec{w}_e$$

$$[\psi]^2 = \sum_e a_e[M]\gamma_e\vec{w}_e$$

$$[\psi]^2 = \sum_e a_e\gamma_e^2\vec{w}_e$$

Repeating this k + 1 times,

$$[\psi]^{k+1} = \sum_e a_e\gamma_e^{k+1}\vec{w}_e$$

Now let us isolate the first eigenvalue from the sum,

$$[\psi]^{k+1} = \gamma_1^{k+1}\sum_e a_e\left(\frac{\gamma_e}{\gamma_1}\right)^{k+1}\vec{w}_e$$

But γ_1 is the largest of the eigenvalues. It then follows that the ratio γ_e/γ_1 is smaller than 1 for $e \neq 1$, and equal to 1 for $e = 1$. If k is sufficiently large, we will have

$$\left(\frac{\gamma_e}{\gamma_1}\right)^{k+1} \to 0$$

and it follows that

$$[\psi]^{k+1} = \gamma_1^{k+1}a_1\vec{w}_1 + \sum_{e>1} a_e\left(\frac{\gamma_e}{\gamma_1}\right)^{k+1}\vec{w}_e$$

$$[\psi]^{k+i} = \gamma_1^{k+1}a_1\vec{w}_1$$

Let us now take a norm of $[\psi]^{k+1}$,

$$\left\|[\psi]^{k+1}\right\| = \left\|\gamma_1^{k+1}a_1\vec{w}_1\right\|$$
$$\left\|[\psi]^{k+1}\right\| = \left|\gamma_1^{k+1}\right|\left|a_1\right|\left\|\vec{w}_1\right\|$$
$$\left\|[\psi]^{k+1}\right\| = \left|\gamma_1\right|^{k+1}\left|a_1\right|\left\|\vec{w}_1\right\|$$

If we take the ratio of the norms of this vector $[\psi]^{k+1}$ and of the preceding vector $[\psi]^k$

$$\frac{\left\|[\psi]^{k+1}\right\|}{\left\|[\psi]^{k}\right\|} = \frac{\left|\gamma_1\right|^{k+1}\left|a_1\right|\left\|\vec{w}_1\right\|}{\left|\gamma_1\right|^{k}\left|a_1\right|\left\|\vec{w}_1\right\|}$$

and it follows that

$$\gamma_1 = \frac{\left\|[\psi]^{k+1}\right\|}{\left\|[\psi]^{k}\right\|} \qquad \text{(EQ 53)}$$

Thus, K_{eff} is obtained as the ratio of two successive norms of the series of vectors generated the application of the matrix $[M]$.

It must be remembered that the application of the matrix $[M]$ to a vector is identical to the following series of operations,

$$\begin{aligned}[u]^k &= [B][\psi]^k \\ [A][\psi]^{k+1} &= [u]^k\end{aligned} \qquad \text{(EQ 54)}$$

It is thus not necessary to construct the $[M]$ matrix in the process leading to the largest eigenvalue $K_{eff} = \gamma_1$ It suffices to solve a linear system instead of evaluating the matrix inverse in the definition of $[\mathbb{M}]$.

We also note that the dominance ratio is defined as

$$r = \frac{\gamma_1}{|\gamma_2|}$$

If this ratio is very close to 1, a large number of iterations, or a very large k, will be necessary before the contribution of the second eigenvector is effectively removed from the vector $[\psi]^{k+1}$. To counteract this problem, different acceleration techniques may be applied. The most popular of these are preconditioning techniques, and convergence acceleration based on Tchebychev polynomials.

Method with Iterative Eigenvalue Calculation

The procedure described in the last section works perfectly well in theory. However, with each iteration, the solution vector may increase or decrease in amplitude, according to whether the eigenvalue is greater or smaller than 1. If many iterations are required, the components of the solution vector may become so large or so small that their representation in computer memory may become difficult; we may have to go to double precision, for example. It is preferable to prevent such an amplification process.

The following algorithm will be used instead of the previous one.We first compute

$$[A][\psi]^{k+1} = \frac{1}{\lambda^k}[B][\psi]^k \qquad \text{(EQ 55)}$$

followed by the evaluation of

$$\lambda^{k+1} = \lambda^k \frac{\|\psi^{k+1}\|}{\|\psi^k\|} \qquad \text{(EQ 56)}$$

where the index k or k + 1 on the λ indicate the iteration number, not an exponent. We choose an initial guess vector $[\psi]^0$ and an initial estimate of K_{eff}, labelled λ^0.

Analysis. We must prove that the series of λ^k converge to K_{eff}, and that the $[\psi]^k$ converge to the fundamental mode. We reformulate the problem as in the preceding section,

$$[\psi]^{k+1} = \frac{1}{\lambda^k}[A]^{-1}[B][\psi]^k$$

which becomes

$$[\psi]^{k+1} = \frac{1}{\lambda^k}[M][\psi]^k$$

$$\lambda^{k+1} = \lambda^k \frac{\|\psi^{k+1}\|}{\|\psi^k\|}$$

Therefore, in the sequence of successive iterations for the vectors $[\psi]^k$, we find that,

$$
\begin{aligned}
&[\psi]^0 \\
[\psi]^1 &= \frac{1}{\lambda^0}[\mathbb{M}][\psi]^0 \\
[\psi]^2 &= \frac{1}{\lambda^1}[\mathbb{M}][\psi]^1 \\
&= \frac{1}{\lambda^1}[\mathbb{M}]\frac{1}{\gamma^0}[\mathbb{M}][\psi]^0 \\
&= \frac{1}{\lambda^1\lambda^0}[\mathbb{M}]^2[\psi]^0 \\
&\quad\cdot \\
[\psi]^{k+1} &= \frac{1}{\lambda^k\ldots\lambda^1\lambda^0}[\mathbb{M}]^{k+1}[\psi]^0 \\
[\psi]^{k+1} &= \left(\prod_{\mathcal{K}=0}^{k}(1/\lambda^{\mathcal{K}})\right)[\mathbb{M}]^{k+1}[\psi]^0
\end{aligned}
$$

If we perform the expansion of the $[\psi]^0$ vector in terms of the eigenvectors of the matrix $[\mathbb{M}]$,

$$
\begin{aligned}
[\psi]^{k+1} &= \left(\prod_{\mathcal{K}=0}^{k}(1/\lambda^{\mathcal{K}})\right)[\mathbb{M}]^{k+1}\sum_{e=1}^{N}a_e\vec{w}_e \\
[\psi]^{k+1} &= \left(\prod_{\mathcal{K}=0}^{k}(1/\lambda^{\mathcal{K}})\right)\sum_{e=1}^{N}a_e[\mathbb{M}]^{k+1}\vec{w}_c \\
[\psi]^{k+1} &= \left(\prod_{\mathcal{K}=0}^{k}(1/\lambda^{\mathcal{K}})\right)\sum_{e=1}^{N}a_e\gamma_e^{k+1}\vec{w}_e
\end{aligned}
$$

When k becomes very large, this becomes

$$[\psi]^{k+1} = \left(\prod_{\mathcal{K}=0}^{k}(1/\lambda^{\mathcal{K}})\right)\gamma_1^{k+1}\sum_{e=1}^{N} a_e\left(\frac{\gamma_e^{k+1}}{\gamma_1^{k+1}}\right)\vec{w}_e$$

$$[\psi]^{k+1} = \left(\prod_{\mathcal{K}=0}^{k}(1/\lambda^{\mathcal{K}})\right)\gamma_1^{k+1}a_1\vec{w}_1$$

which shows that the series of solution vectors does converge on the fundamental eigenvector.

The norm of the solution vector at iteration $k+1$ will be

$$\|\psi^{k+1}\| = \left(\prod_{\mathcal{K}=0}^{k}(1/\lambda^{\mathcal{K}})\right)\gamma_1^{k+1}|a_1|\|\vec{w}_1\|$$

Let us examine the series of λ^k. We have after iteration $k+1$,

$$\lambda^{k+1} = \lambda^k\frac{\|\psi^{k+1}\|}{\|\psi^k\|}$$

$$\therefore \lambda^{k+1} = \lambda^k\frac{\left(\prod_{\mathcal{K}=0}^{k}(1/\lambda^{\mathcal{K}})\right)\gamma_1^{k+1}|a_1|\|\vec{w}_1\|}{\left(\prod_{\mathcal{K}=0}^{k-1}(1/\lambda^{\mathcal{K}})\right)\gamma_1^{k}|a_1|\|\vec{w}_1\|}$$

$$\lambda^{k+i} = \frac{\lambda^k(1/\lambda^k)\left(\prod_{\mathcal{K}=0}^{k-1}(1/\lambda^{\mathcal{K}})\right)}{\left(\prod_{\mathcal{K}=0}^{k-1}(1/\lambda^{\mathcal{K}})\right)}\gamma_1$$

$$\lambda^{k+1} = \gamma_1$$

Hence, we have the result that the sequence of the λ^k does indeed converge to γ_1, which is the K_{eff} of the discretised problem.

The process of evaluating the eigenvalue can be stopped by using a very simple convergence criterion, such as

$$\frac{|\lambda^{k+1} - \lambda^k|}{\lambda^k} \leq \epsilon_K \qquad \text{(EQ 57)}$$

where ϵ_K is chosen arbitrarily according to the requirements of the analysis, and is normally given values from 1×10^{-4} to 1×10^{-6} in the majority of cases.

Solution Strategy

There is still left to solve the complete flux problem, in the spatial variable

$$[A][\psi]^{k+1} = \frac{1}{\lambda^k}[B][\psi]^k \qquad \text{(EQ 58)}$$

which is followed by

$$\lambda^{k+1} = \lambda^{k}\frac{\|\psi^{k+1}\|}{\|\psi^{k}\|} \qquad \text{(EQ 59)}$$

The power method being used to determine the λ, there is still a linear algebraic problem to solve for the $[\psi]$ vectors.

If the flux $[\psi]^{k}$ is known, the right hand side of (58) will be known, since it is easy to evaluate the product of the matrix by the vector, and to divide by the current value of λ^{k}. We have seen in chapter 7, *Statics*, page 63, that the matrix $[A]$ is made of tri-diagonal blocks, with other contributions which are block diagonal matrices containing scattering terms from one energy group to another.

Since all neutrons appear from fission in the fast groups, and slow down towards the thermal groups, we propose the following method to solve the linear system:

1. Start from the fastest group, and go down in the groups sequentially
2. For a given group, put with the neutron source the scattering terms comprising down scattering and up scattering towards the group of interest
3. Solve the reduced linear system for the fluxes
4. Evaluate the new estimate for the eigenvalue
5. Go back to step until convergence

Just as for the eigenvalue, a stopping criterion is used for the fluxes, such as

$$\max_{1 \le i \le N}\left(\frac{|\psi_i^{k+1} - \psi_i^{k}|}{\psi_i^{k+1}}\right) \le \epsilon_{\phi} \qquad \text{(EQ 60)}$$

where ϵ_{ϕ} is also chosen arbitrarily according to particular needs, and is usually taken from 10^{-4} to 10^{-6} in the vast majority of calculations. It should be noted that the fluxes are usually much slower than the eigenvalue to reach convergence, and we often see a factor of 10 between the instantaneous errors on the fluxes and on the eigenvalue.I

The iterative process in determining the flux is known as "internal iterations", and that of the determination of the eigenvalue, "external iterations".

Matrix Norms

On the most abstract level, a vector is simply a member of a vector space. Vector norms are then only a mapping of a vector unto a scalar, with a set of rules.

Matrices can also be considered as an abstract vector space. This space may be normed, and matrix norms then appear as a natural extension of vector norms. Following are important properties of matrix norms:

$$\|A\| \geq 0$$

$$\|A\| = 0 \Leftrightarrow A = 0$$

$$\|\alpha A\| = |\alpha| \cdot \|A\| \text{ where} \alpha \text{ is a scalar}$$

$$\|A + B\| \leq \|A\| + \|B\|$$

$$\|AB\| \leq \|A\| \cdot \|B\|$$

Furthermore if $\|Au\| \le \|A\| \cdot \|u\|$, we say that the matrix norm is consistent with the vector norm.

The most often used matrix norms are

- the 1 norm, which is the maximum of the sum of the absolute values of the matrix along the columns, $\|A\|_1 = \max_j\left(\sum_{i=1}^{n}|a_{ij}|\right)$
- the ∞ norm, the maximum of the sum of the absolute values of the matrix elements along the lines, $\|A\|_\infty = \max_i\left(\sum_{j=1}^{n}|a_{ij}|\right)$
- for the 2 norm or euclidean norm, we must first define the spectral radius of A,

  $$\rho(A) = \max_i|\lambda_i(A)|$$

  The 2 norm is then $\|A\|_2 = \sqrt{\rho(A^TA)}$. It can be shown that $\rho(A) \le \|A\|$ for any matrix A and any norm $\|\ \|$.

LU Decomposition

In the algorithm just proposed, there will necessarily be linear systems to solve. In the one dimensional case, these linear systems are of tri-diagonal form, which makes the solution relatively easy to obtain. The LU decomposition permits to solve such problems quite efficiently. We illustrate the process with a 4×4 case.

The idea behind LU decomposition is to write the original matrix as a product of two triangular matrices, a "lower triangular" matrix, $[L]$, and an "upper triangular" matrix, $[U]$, in the following

$$[A] = [L][U]$$

and the system to be solved becomes

$$\begin{aligned}[A][\psi] &= [b]\\ [L][U][\psi] &= [b]\end{aligned}$$

which can be solved in two steps,

$$\begin{aligned}[L][z] &= [b]\\ [U][\psi] &= [z]\end{aligned} \qquad \text{(EQ 61)}$$

In the case of triangular matrices, the factorization and solution processes are very simple and efficient. First, the elements of the two matrices $[L]$ and $[U]$ are calculated. Then the solution process (61) is performed, needing only simple steps.

Decomposition Example

Consider a 4×4 case, which is

$$\begin{bmatrix} \ell_{11} & 0 & 0 & 0 \\ \ell_{21} & \ell_{22} & 0 & 0 \\ 0 & \ell_{32} & \ell_{33} & 0 \\ 0 & 0 & \ell_{43} & \ell_{44} \end{bmatrix} \begin{bmatrix} 1 & u_1 & 0 & 0 \\ 0 & 1 & u_2 & 0 \\ 0 & 0 & 1 & u_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & a_{41} & a_{44} \end{bmatrix}$$

we find

$$\begin{aligned}
\ell_{11} &= a_{11} \\
\ell_{11}u_1 &= a_{12} \Rightarrow u_1 = a_{12}/\ell_{11} \\[1ex]
\ell_{21} &= a_{21} \\
\ell_{21}u_1 + \ell_{22} &= a_{22} \Rightarrow \ell_{22} = a_{22} - \ell_{21}u_1 \\
\ell_{22}u_2 &= a_{23} \Rightarrow u_2 = a_{23}/\ell_{22} \\[1ex]
\ell_{32} &= a_{32} \\
\ell_{32}u_2 + \ell_{33} &= a_{33} \Rightarrow \ell_{33} = a_{33} - \ell_{32}u_2 \\
\ell_{33}u_3 &= a_{34} \Rightarrow u_3 = a_{34}/\ell_{33} \\[1ex]
\ell_{43} &= a_{43} \\
\ell_{43}u_3 + \ell_{44} &= a_{44} \Rightarrow \ell_{44} = a_{44} - \ell_{43}u_3
\end{aligned}$$

Once all elements calculated, it is very simple to solve first for the system $[L][z] = [b]$, by eliminating the unknowns, proceeding from top to bottom, a process known as "forward elimination"; we find all elements of $[z]$, from the first to last one,

$$\begin{aligned}
\ell_{11}z_1 &= b_1 \Rightarrow z_1 = b_1/\ell_{11} \\[1ex]
\ell_{21}z_1 + \ell_{22}z_2 &= b_2 \Rightarrow z_2 = (b_2 - \ell_{21}z_1)/\ell_{22} \\[1ex]
\ell_{32}z_2 + \ell_{33}z_3 &= b_3 \Rightarrow z_3 = (b_3 - \ell_{32}z_2)/\ell_{33} \\[1ex]
\ell_{43}z_3 + \ell_{44}z_4 &= b_4 \Rightarrow z_4 = (b_4 - \ell_{43}z_3)/\ell_{44}
\end{aligned}$$

The the system $[U][\psi] = [z]$ will next be solved, but this time starting from the last unknown towards the first one, a process known as backward substitution,

$$\psi_4 = z_4$$

$$\psi_3 + u_3\psi_4 = z_3 \Rightarrow \psi_3 = z_3 - u_3\psi_4$$

$$\psi_2 + u_2\psi_3 = z_2 \Rightarrow \psi_2 = z_2 - u_2\psi_3$$

$$\psi_1 + u_1\psi_2 = z_1 \Rightarrow \psi_1 = z_1 - u_1\psi_2$$

This particular sequence simply takes into account the special structures of the $[L]$ and $[U]$ matrices.The generalization to larger size matrix is a very straightforward exercise.

In one dimension, this is very efficient to solve the static equations, but it becomes too expensive in three dimensions. This is because of the large number of operations required in performing the LU decomposition itself, which is equivalent to solving the matrix system by Gaussian elimination, which takes too many steps, on the order of N^3 operations. Iterative methods of solving the linear system will be used in this case.

Iterative Methods

We will consider here only a very few of the standard iterative methods. Many other methods can be found in the literature, but most of them are more or less complex variations of the basic methods.

Iterative methods become necessary in two and three dimensional problems, which involve a very large number of mesh points. When they are well optimized, they can be very efficient. The matrix notation used to describe the methods are only used for numerical analysis purposes. In practice, their implementation in computer codes does not use the iterative matrix formulation, since only the relationships for the individual coefficients are necessary. The extra storage for the iterative matrix is sufficiently high to preclude their use.

Stationary Iterations

Let us consider a linear system of the form

$$[A][\psi] = [b] \qquad \text{(EQ 62)}$$

Stationary iterative methods are of the form

$$[\psi]^{k+1} = [M][\psi]^{k} + [g] \qquad \text{(EQ 63)}$$

They are said to be stationary because the matrix $[M]$ does not depend on the iteration index k. The same iterative matrix is used throughout the iterative process.

The consistency condition wants that the true solution of system (62) must be a stationary point of the iterative process (63). In other words, if the true solution $[\psi]$ of the linear system is used in place of $[\psi]^k$, then $[\psi]^{k+1}$ must return $[\psi]$ also.

Convergence

A matrix $[M]$ is said to be convergent if

$$\lim_{k \to \infty} [M]^k = 0$$

where this time the index k is an exponent

- $[M]$ is convergent if and only if $\rho([M]) < 1$
- $[M]$ is convergent if and only if $\|[M]\| < 1$.

To see this, let us define the error at iteration k in the following way,

$$[e]^k = [\psi]^k - [\psi]$$

where $[\psi]$ is the exact solution of e (62). Thus

$$[\psi]^k = [e]^k + [\psi] \qquad \text{(EQ 64)}$$

and also

$$[\psi]^{k+1} = [e]^{k+1} + [\psi] \qquad \text{(EQ 65)}$$

Substituting (64) and (65) into (63),

$$\begin{aligned} [\psi]^{k+1} &= [M][\psi]^k + [g] \\ [e]^{k+1} + [\psi] &= [M]([e]^k + [\psi]) + [g] \\ [e]^{k+1} + [\psi] &= [M][e]^k + [M][\psi] + [g] \\ [e]^{k+1} + [\psi] &= [M][e]^k + [\psi] \\ [e]^{k+1} &= [M][e]^k \end{aligned}$$

Consider now the behavior of the error in the iterative process,

$$
\begin{aligned}
[e]^0 &= [\psi]^0 - [\psi] \\
[e]^1 &= [M][\psi]^0 \\
[e]^2 &= [M][\psi]^1 = [M][M][\psi]^0 = [M]^2[\psi]^0 \\
[e]^k &= [M]^k[\psi]^0
\end{aligned}
$$

So that finally,

$$[e]^k \to 0 \Leftrightarrow [M]^k \to 0 \Leftrightarrow \rho([M]) < 1$$

The iterative matrix must be convergent if the error on the solution vector is to eventually disappear as the number of iterations increases. To estimate if an iterative method is convergent, the spectral radius of the iterative matrix will have to be determined. This is a non trivial task, and the large body of literature on the subject can be consulted for details.

Jacobi Method

The method of Jacobi only offers a theoretical interest, because the spectral radius of the associated iterative matrix is simple to calculate. The matrix $[A]$ is split in the following way,

$$[A] = [L] + [D] + [U]$$

where $[D]$ is a diagonal matrix, containing only the diagonal elements of $[A]$, $[L]$ is a matrix containing only those elements of $[A]$ that are below the diagonal (lower triangular matrix), and $[U]$ contains those elements of $[A]$ that are above the diagonal (upper triangular matrix).

In practice, the Jacobi method is obtained by sending to the right hand side all the terms that are not on the diagonal. The vector $[\psi]^{k+1}$ does not replace the vector $[\psi]^{k}$ until the end of iteration k. The two vectors $[\psi]^{k}$ and $[\psi]^{k+1}$ must therefore be held in memory. This is one of the reasons for which the Jacobi method is not used in practice; another reason being poor rate of convergence, as compared to other methods.

Iterative matrix. In terms of the decomposition of the $[A]$ matrix, the Jacobi method is given by

$$[D][\psi]^{k+1} = -([L]+[U])[\psi]^{k}+[b]$$

and consequently,

$$[\psi]^{k+1} = -[D]^{-1}([L]+[U])[\psi]^{k}+[D]^{-1}[b] \qquad \text{(EQ 66)}$$

which gives for the iterative process (63),

$$\begin{aligned}[M] &= -[D]^{-1}([L]+[U]) \\ [g] &= [D]^{-1}[b]\end{aligned}$$

The matrix$[M]$ can be written in another way, by replacing the sum $[L]+[U]$ in terms of $[A]$,

$$\begin{aligned}[M] &= -[D]^{-1}([L]+[U]) \\ &= -[D]^{-1}([A]-[D])\end{aligned}$$

and then

$$[M] = ([I] - [D]^{-1}[A])$$

it is easy to show that the stationary point of (66) is the actual solution of $[A][\psi] = [b]$. Suppose that for a given iteration index K we find that $[\psi]^{K+1} = [\psi]^{K}$. Then

$$[\psi]^K = -[D]^{-1}([L]+[U])[\psi]^K + [D]^{-1}[b]$$
$$[D][\psi]^K = -([L]+[U])[\psi]^K + [b]$$
$$([D]+[L]+[U])[\psi]^K = [b]$$

and therefore

$$[A][\psi]^K = [b]$$

By virtue of the unique solution of non-singular linear systems, we can only conclude that $[\psi]^K$ is the exact solution sought for.

Gauss-Seidel Method

The Gauss-Seidel method resembles much to the Jacobi method. However, it uses in the calculation sequence the new elements of $[\psi]^{k+1}$ as soon as they are available. Only one vector has to be kept in memory, and it contains both new and old elements.

In terms of the decomposition of the matrix $[A]$, we have

$$[D][\psi]^{k+1} = -[L][\psi]^{k+1} - [U][\psi]^k + b$$

As is the case, the elements of the matrix $[A]$ under the diagonal multiply the most recent calculated values, those at iteration $k+1$, of the vector $[\psi]^{k+1}$, while the elements of $[A]$ above the diagonal multiply the components of $[\psi]^k$, that have not been evaluated yet. This gives

$$([D]+[L])[\psi]^{k+1} = -[U][\psi]^k + [b]$$
$$[\psi]^{k+1} = -([D]+[L])^{-1}[U][\psi]^k + ([D]+[L])^{-1}[b]$$

and therefore,

$$\begin{aligned}[M] &= -([D]+[L])^{-1}[U]\\ [g] &= ([D]+[L])^{-1}[b]\end{aligned} \qquad \text{(EQ 67)}$$

or, in another form,

$$\begin{aligned}[M] &= -([D]+[L])^{-1}[U]\\ &= -([D]([I]+[D]^{-1}[L]))^{-1}[U]\end{aligned}$$

which gives

$$[M] = -([I]+[D]^{-1}[L])^{-1}[D]^{-1}[U] \qquad \text{(EQ 68)}$$

while the $[g]$ vector becomes

$$[g] = ([D]([I]+[D]^{-1}[L]))^{-1}[b]$$

so that

$$[g] = ([I]+[D]^{-1}[L])^{-1}[D]^{-1}[b] \qquad \text{(EQ 69)}$$

In practice, the Gauss-Seidel method converges quite well. It is easy to program on a computer, and for this reason is quite popular.

It can be shown that when the Jacobi method converges, then the Gauss-Seidel method is also convergent, and that the spectral radius of the iterative matrix is smaller than that of the Jacobi method. in the case where the Jacobi method diverges, then the Gauss-Seidel method will diverge even faster.

SOR Method

The *Successive* Over Relaxation method (SOR), is based on the Gauss-Seidel method, but replaces the new component of the solution vector by a linear combination of the previous one and the new one, in a two step calculation.

It is given by,

$$\begin{aligned}[L][\psi]^{k+1} + [D][\psi]^{k+1/2} + [U][\psi]^{k} &= [b] \\ [\psi]^{k+1} &= (1-\omega)[\psi]^{k} + \omega[\psi]^{k+1/2}\end{aligned} \quad \text{(EQ 70)}$$

It can be shown that

$$[M] = ([I] + \omega[D]^{-1}[L])^{-1}\{[I](1-\omega) - \omega[D]^{-1}[U]\} \quad \text{(EQ 71)}$$

$$[g] = ([I] + \omega[D]^{-1}[L])^{-1}\omega[D]^{-1}[b] \quad \text{(EQ 72)}$$

In practice, the SOR method is probably used more than the Gauss-Seidel method (which is a special case with $\omega = 1$) for most realistic situations. It can also be shown that the method is consistent only for

$0 \leq \omega \leq 2$. There is also a very strong dependence of the spectral radius on the value of ω. There exists an optimal value ω_{opt} of the relaxation parameter, and there is no easy way to calculate it apriori. Also, the convergence rate will be much slower if the value of ω used in a given computation is under-estimated in relation to ω_{opt}, than if we over-estimate it. If we have on hand a good value of ω, the convergence rate will be much higher than that of the Gauss-Seidel method.