Ferhat Abbas University-Sétif 1
Faculty of Sciences
Department of Chemistry

Chapter 4 : Numerical Solution of Systems of Linear Equations

In practice, engineers are often faced with problems whose solution involves solving a system of equations that models the various elements under consideration. For example, determining currents and voltages in electrical networks requires solving a system of linear equations.

We seek the vector $X \in \mathbb{R}^n, X = (x_1, x_2, \ldots, x_n)$, which is the solution of the following linear system :

$$AX = b \Longleftrightarrow \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases} \tag{1}$$

This system admits a unique solution when the determinant of A is non-zero, which we will assume from now on. Solving this system using direct methods becomes impractical when n is relatively large. Therefore, it is preferable to use iterative methods based on constructing a sequence that converges to the solution of the system.

In this chapter, we will introduce two iterative methods that provide an approximate solution to the system of linear equations using a linear function f such that $X^{k+1} = f(X^k)$, $k \in \mathbb{N}$. These methods are very easy to implement and program, require minimal memory, and produce results as accurate as desired.

Given an arbitrary initial vector X^0, we construct a sequence of vectors

$$X^0, X^1, \cdots, X^k, \cdots$$

which converges to the solution X^* of the linear system $AX = b$. We consider the linear system (1) with A being an invertible square matrix of order n and b a vector in $\mathbb{R}^n$. For any invertible square matrix M of order n, the system (1) is equivalent to

$$MX - (M - A)X = b$$

or, by setting $N = M - A$, $B = M^{-1}N$, and $c = M^{-1}b$, we obtain

$$X = BX + c\,.$$

This allows us to define the following iterative formula :

$$\begin{cases} X^0 \in \mathbb{R}^n \text{ vecteur initiale} \\ X^{k+1} = BX^k + c. \end{cases} \tag{2}$$

Let X^* be the exact solution of (1). If we denote $e^k = \|X^k - X^*\|$ as the k-th error vector, we obtain

$$\begin{aligned} e_k &= \|X^k - X^*\| = \|(BX^{k-1} + c) - (BX^* + c)\| = B\|X^{k-1} - X^*\| \\ &= Be_{k-1} = B^k e_0 \end{aligned}$$

Remark 1. In practice, if we impose a precision ε, we can estimate the error by :

$$\|X^k - X^{k-1}\| \leq \varepsilon$$

This means that, for all $i \in \{1, \ldots, n\}$, we have :

$$|x_i^k - x_i^{k-1}| \leq \varepsilon.$$

Theorem 1. *The iterative method* (2) *converges if the sequence of vectors* $\{e^k\}_{k\in\mathbb{N}}$ *converges to zero regardless of the initial vector* X^0*, provided that one of the three norms is less than 1 :*

- $\|B\|_1 = \max_j(\sum_{i=1}^{n} |B_{ij}|)$
- $\|B\|_\infty = \max_i(\sum_{j=1}^{n} |B_{ij}|)$
- $\|B\|_2 = \sqrt{\rho(BB^t)}$.

Depending on the choices of the matrices M and N, we have different iterative methods. Let D be the matrix formed by the diagonal elements of A, E be the matrix formed by the $-a_{ij}$ when $i > j$, and F be the matrix formed by the $-a_{ij}$ when $i < j$, so that $A = D - (E + F)$.

- The matrix D is a diagonal matrix of A, given by :

$$D = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

- The matrix E is a lower triangular matrix of A with a zero diagonal.

$$E = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ -a_{21} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & 0 \end{pmatrix}$$

- The matrix F is an upper triangular matrix of A with a zero diagonal.

$$F = \begin{pmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ 0 & 0 & \cdots & a_{2n} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

1. Jacobi Method

In the Jacobi iterative method, the matrix A of the system $AX = b$ is decomposed as $A = M - N$. The matrix M corresponds to the diagonal of A (with zeros outside the diagonal), so $M = D$, and the matrix N is the matrix A in which the diagonal elements have been replaced by zeros, i.e., $N = E + F$. The matrix $J = M^{-1}N = D^{-1}(E + F) = I - D^{-1}A$ is called the Jacobi matrix. Starting from an initial vector $X^0 = (x_1^0, x_2^0, \ldots, x_n^0)^t$, at each step, we compute X^k using the following formula :

$$x_i^{k+1} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1, j\neq i}^{n} a_{ij} x_j^k \right), i = 1, 2 \ldots, n\,. \tag{3}$$

Remark 2. The Jacobi iterative method does not always converge. If A is a positive definite matrix, the iterative method converges. Similarly, if A is a strictly diagonally dominant matrix, i.e., $|a_{ii}| > \sum_{j\neq i} |a_{ij}|$, then the Jacobi method is convergent.

Exemple 1. Consider the following system

$$\begin{cases} 4x_1 + 2x_2 + x_3 &= 4 \\ -x_1 + 2x_2 &= 2 \\ 2x_1 + x_2 + 4x_3 &= 9\,. \end{cases}$$

Let $X^0 = (0, 0, 0)^t$ be the initial vector. By calculating the first five iterations, we obtain :

$$X^1 = \begin{pmatrix} 1 \\ 1 \\ 9/4 \end{pmatrix}, X^2 = \begin{pmatrix} -1/16 \\ 3/2 \\ 3/2 \end{pmatrix}, X^3 = \begin{pmatrix} -1/8 \\ -1/32 \\ 61/32 \end{pmatrix}, X^4 = \begin{pmatrix} 5/128 \\ 15/16 \\ 265/128 \end{pmatrix}, \text{ et } X^5 = \begin{pmatrix} 7/512 \\ 261/256 \\ 511/256 \end{pmatrix}$$

Exemple 2. Let us solve the following system using the Jacobi method :

$$\begin{cases} 3x_1 + x_2 - x_3 = 2 \\ x_1 + 5x_2 + 2x_3 = 17 \\ 2x_1 - x_2 - 6x_3 = -18 \end{cases}$$

We have,

$$\begin{cases} i = 1, & x_1^{k+1} = \frac{1}{3}\left(2 - x_2^k + x_3^k\right) \\ i = 2, & x_2^{k+1} = \frac{1}{5}\left(17 - x_1^k - 2x_3^k\right) \\ i = 3, & x_3^{k+1} = \frac{-1}{6}\left(-18 - 2x_1^k + x_2^k\right) \end{cases}$$

Let $X^0 = (0,0,0)^t$ be the initial vector, we obtain : $X_1 = \begin{pmatrix} 2/3 \\ 17/5 \\ 3 \end{pmatrix}$, $X_2 = \begin{pmatrix} 8/15 \\ 31/15 \\ 2.6555 \end{pmatrix}$.

After 10 iterations, we obtain the following table of results :

k	x_1^k	x_2^k	x_3^k
0	0	0	0
1	0,666666	3,4	3
2	0,533333	2,066667	2,655556
3	0,862963	2,231111	2,833333
4	0,867407	2,094074	2,915802
5	0,940576	2,0601198	2,970123
6	0,959975	2,035835	2,970159
7	0,978108	2,019941	2,980686
8	0,986915	2,012104	2,989379
9	0,992425	2,006865	2,993621
10	0,995585	2,004067	2,996331

From this table, we notice that the values converge towards the solution $X = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

2. Méthode de Gauss-Seidel

The Gauss-Seidel method is an improvement over the Jacobi method, as it accelerates the iterative process. Starting from the Jacobi method, the computation of the vectors $X^1, X^2, \ldots, X^k, \ldots$ leads to convergence. This means that each new vector is better than the previous one.

In the Jacobi method, we notice that to compute the component x_2^2 of the vector X^2, we use the components of X^1, even though x_1^2 is already computed and is better than x_1^1. This is the principle behind the Gauss-Seidel method : each component is used as soon as it is computed.

Thus, to calculate the component x_i^{k+1}, we use all the components from x_1^{k+1} to x_{i-1}^{k+1} already computed in iteration $(k+1)$, as well as the components x_{i+1}^k to x_n^k that are still at iteration k.

Given that the matrix A is decomposed as :

$$A = M - N,$$

we take :

$$M = D - E, \quad N = F.$$

This modifies equation (3) as follows : for $k \geq 0$ (assuming again that $a_{ii} \neq 0$ for $i = 1, \ldots, n$).

$$x_i^{k+1} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{k+1} - \sum_{j=i+1}^{n} a_{ij} x_j^k \right), i = 1, 2 \ldots, n\,. \tag{4}$$

Remark 3. The Gauss-Seidel method does not always converge. If A is a positive definite matrix, the iterative method converges. Similarly, if A is a diagonally dominant matrix, i.e., if

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|,$$

then the Gauss-Seidel method converges.

Example 1. Solve the following system using the Gauss-Seidel method with 3 iterations and an initial vector $X^0 = (0, 0, 0)^t$.

$$\begin{cases} -x_1 + x_2 + 3x_3 & = -1 \\ x_1 + 2x_2 & = 2 \\ 3x_1 + x_2 - x_3 & = 1 \end{cases}$$

This system can be written in reduced form :

$$\begin{cases} i = 1, \quad x_{k+1}^1 & = 1 + x_k^2 + 3x_k^3 \\ i = 2, \quad x_{k+1}^2 & = 1 - \frac{1}{2} x_{k+1}^1 \\ i = 3, \quad x_{k+1}^3 & = -1 + 3x_{k+1}^1 - x_{k+1}^2 \end{cases}$$

- First iteration, we obtain $X^1 = \begin{pmatrix} 1 \\ 0.5 \\ 1.5 \end{pmatrix}$

- Second iteration, we obtain $X^2 = \begin{pmatrix} 6 \\ -2 \\ 19 \end{pmatrix}$

- Third iteration, we obtain $X^3 = \begin{pmatrix} 56 \\ -27 \\ 194 \end{pmatrix}$.

Example 2. Solve the same linear system from Example 2 using the Gauss-Seidel method.

For each iteration k, the iterative scheme of the Gauss-Seidel method is written in this case :

$$\begin{cases} i = 1, \quad x_1^{k+1} & = \frac{1}{3}\left(2 - x_2^k + x_3^k\right) \\ i = 2, \quad x_2^{k+1} & = \frac{1}{5}\left(17 - x_1^{k+1} - 2x_3^k\right) \\ i = 3, \quad x_3^{k+1} & = \frac{-1}{6}\left(-18 - 2x_1^{k+1} + x_2^{k+1}\right) \end{cases}$$

Starting from $X^0 = (0, 0, 0)^t$, we find $X^1 = \left(\frac{2}{3}, \frac{49}{15}, \frac{241}{90}\right)^t$. After 10 iterations, we obtain the following table of results :

k	x_k^1	x_k^2	x_k^3
0	0	0	0
1	0.6666667	3.266667	2.677778
2	0.4703704	2,234815	2.784321
3	0.8498354	2.116305	2.930561
4	0.9380855	2.040158	2.972669
5	0.9775034	2.015432	2.989929
6	0.9914991	2.005729	2.996212
7	0.9968271	2.002150	2.998584
8	0.9988115	2.000804	2.999470
9	0.9995553	2.000301	2.999802
10	0.9998335	2.000113	2.999926

It can be observed that for the same number of iterations, the approximate solution obtained by the Gauss-Seidel method is more accurate. The Gauss-Seidel method generally converges more quickly than the Jacobi method, but not always.