Summary of session 3

-Chapter 4. Solving Systems of Linear Equations

- 4.3. Pivoting and Constructing Algorithm:

  Basic Gaussian Elimination

  Pivoting

  Guassian Elimination with Scaled Pivoting

  Diagonal Dominant Matrix

  Triangular System

4.4 Norms and the Analysis of Errors

- Vector Norms

Vector Norms: On a vector space V, a norm is a function $\|\cdot\|$ from V to set of nonnegative reals that obeys these three postulates:

1. $\|\mathbf{x}\| > 0$ if $\mathbf{x} \neq 0,\ \mathbf{x} \in V.$
2. $\|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\|$ if $\lambda \in R, \mathbf{x} \in V.$
3. $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ if $\mathbf{x},\ \mathbf{y} \in V$ (triangle inequality)

We can think of $\|\mathbf{x}\|$ as the `length` or `magnitude` of the vector $\mathbf{x}$. A norm on a vector space generalizes the notion of absolute value, $|\mathrm{r}|$, for a real or complex number. The most familar norm is the `Euclidean` l_2-`norm` defined

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} x_i^{\,2}\right)^{1/2} \qquad \text{where } \mathbf{x} = (x_1, x_2, \mathrm{K}\ , x_n)^T.$$

other most familiar norm is the $l_\infty-$`norm`: $\|\mathbf{x}\|_\infty = \max_{1\leq i\leq n} |x_i|.$

Also l_1-`norm` is important one. $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$

4.4 Norms and the Analysis of Errors

- Vector Norms

Example

Using the norm $\|.\|_1$, compare the lenghts of the following three vectors in R^n. Then repeat the calculation for the norms $\|.\|_2$, and $\|.\|_\infty$.

$$\mathbf{x} = (4,-4,4,4)^T, \quad \mathbf{v} = (0,5,5,5)^T, \quad \mathbf{w} = (6,0,0,0)^T$$

Solution:

M	$\|.\|_1$	$\|.\|_2$	$\|.\|_\infty$
x M	16.	8.	4.
v M	15.	8.66.	5.
w M	6.	6.	6.

4.4 Norms and the Analysis of Errors

- Matrix Norms

Matrix Norms: On a matrix space $R^{n\times n}$, a norm is a function $\|\cdot\|$ from $R^{n\times n}$ to set of nonnegative reals that obeys these postulates:

1. $\|\mathbf{A}\| > 0$ if $\mathbf{A} \neq 0$, $\mathbf{A} \in R^{n\times n}$.
2. $\|\lambda\mathbf{A}\| = |\lambda|\|\mathbf{A}\|$ if $\lambda \in R$, $\mathbf{A} \in R^{n\times n}$.
3. $\|\mathbf{A}+\mathbf{B}\| \le \|\mathbf{A}\| + \|\mathbf{B}\|$ if $\mathbf{A}, \mathbf{B} \in R^{n\times n}$ (triangle inequility)

$\|\mathbf{A}\|$ is said to be a norm of matrix $\mathbf{A}$.

There are many types of matrix norms which satisfy the above conditions. For studying the linear equation system, the matrix norms satisfying the following extra conditions are important

4. $\|\mathbf{AB}\| \le \|\mathbf{A}\| \cdot \|\mathbf{B}\|$, $\mathbf{A}, \mathbf{B} \in R^{n\times n}$
5. $\|\mathbf{Ax}\| \le \|\mathbf{A}\| \cdot \|\mathbf{x}\|$, $\mathbf{A} \in R^{n\times n}$, $\mathbf{x} \in R^{n}$

4.4 Norms and the Analysis of Errors

- Matrix Norms

The matrix norms subordinate to l_∞ - *norm*, l_1 - *norm*, l_2 - *norm* are defined as

$$\|\mathbf{A}\|_\infty = \max_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}| \qquad (\|\mathbf{x}\|_\infty = \max_{1\le i\le n}|x_i|)$$

$$\|\mathbf{A}\|_1 = \max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}| \qquad (\|\mathbf{x}\|_1 = \sum_{i=1}^{n}|x_i|)$$

$$\|\mathbf{A}\|_2 = \sqrt{\rho(\mathbf{A}^T\mathbf{A})} \qquad (\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} x_i^{\,2}\right)^{1/2})$$

where $\rho(\mathbf{A}^T\mathbf{A})$ is **spectral radius** of $\mathbf{A}^T\mathbf{A}$, which is defined as

$$\rho(\mathbf{A}^T\mathbf{A}) = \max\{|\lambda_i| : \lambda_i \text{ is the eigenvalue of } \mathbf{A}^T\mathbf{A}\}$$

4.4 Norms and the Analysis of Errors

- Matrix Norms

It can be proved that they satisfy the conditions to a matrix norm.

Example: $\|\mathbf{A}\|_1 = \max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|$

1. it is obvious that $\|\mathbf{A}\|_1 \ge 0$, and $\|\mathbf{A}\|_1 = 0$, i.e.

$$\max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}| = 0 \text{ when but only when } j = 1,2,\mathsf{L}\ ,n$$

$$\sum_{i=1}^{n}|a_{ij}| = 0$$

so that for any $i\ (1\le i\le n)$, $|a_{ij}| = 0$, and therefore $\mathbf{A} = 0$.

2. for any real a, $a\mathbf{A} = (aa_{ij})$, so

$$\|a\mathbf{A}\|_1 = \max_{1\le j\le n}\sum_{i=1}^{n}|aa_{ij}| = \max_{1\le j\le n}|a|\sum_{i=1}^{n}|a_{ij}| = |a|\max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}| = |a|\ \|\mathbf{A}\|_1$$

4.4 Norms and the Analysis of Errors

- Matrix Norms

3. for any two matrix $\mathbf{A}=(a_{ij})$ and $\mathbf{B}=(b_{ij})$

$$\|\mathbf{A}+\mathbf{B}\|_1 = \max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}+b_{ij}| \le \max_{1\le j\le n}\sum_{i=1}^{n}(|a_{ij}|+|b_{ij}|)$$

$$\le \max_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}| + \max_{1\le j\le n}\sum_{i=1}^{n}|b_{ij}| = \|\mathbf{A}\|_1 + \|\mathbf{B}\|_1$$

4. $$\|\mathbf{AB}\|_1 = \max_{1\le j\le n}\sum_{i=1}^{n}\left|\sum_{k=1}^{n}a_{ik}b_{kj}\right| \le \max_{1\le j\le n}\sum_{i=1}^{n}(\sum_{k=1}^{n}|a_{ik}||b_{kj}|)$$

$$= \max_{1\le j\le n}\sum_{k=1}^{n}(\sum_{i=1}^{n}|a_{ik}|)|b_{kj}| \le \max_{1\le j\le n}\sum_{k=1}^{n}(\max_{1\le k\le n}\sum_{i=1}^{n}|a_{ik}|)|b_{kj}|$$

$$= (\max_{1\le k\le n}\sum_{i=1}^{n}|a_{ik}|)\max_{1\le j\le n}\sum_{k=1}^{n}|b_{kj}| = \|\mathbf{A}\|_1\,\|\mathbf{B}\|_1$$

5. Similarly, $\|\mathbf{A}x\|_1 \le \|\mathbf{A}\|_1\,\|x\|_1$

4.4 Norms and the Analysis of Errors

- Matrix Norms

Example : $\mathbf{A} = \begin{bmatrix} 1.1 & -2 \\ 2.5 & 3.5 \end{bmatrix}$

calculate norms $\|\mathbf{A}\|_1$, $\|\mathbf{A}\|_\infty$, $\|\mathbf{A}\|_2$.

Solution: $\|\mathbf{A}\|_1 = \max\{1.1+2.5,\ |-2|+3.5\} = 5.5$

$$\|\mathbf{A}\|_\infty = \max\{1.1+\ |-2|,\ 2.5+3.5\} = 6$$

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 7.46 & -10.95 \\ -10.95 & 16.25 \end{bmatrix}$$

Its eigenvalues are $\lambda_1 = 23.6541,\ \lambda_2 = 0.05591$

$$\|\mathbf{A}\|_2 = \sqrt{\lambda_1} = 4.86355$$

4.4 Norms and the Analysis of Errors

- Condition Number

Let's look at how these norm concepts work. Considering a system

$$\mathbf{Ax} = \mathbf{b}$$

where $\mathbf{A}$ is supposed to be invertible.

Example: If $\mathbf{A}$ is perturbed to obtain a new matrix $\mathbf{B}$, then the solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ to become a new vector $\tilde{\mathbf{x}} = \mathbf{B}^{-1}\mathbf{b}$. How large is this latter perturbation in absolute and relative terms?

Solution: using any vector norm and its subordinate matrix norm we have

$$\|\mathbf{x} - \tilde{\mathbf{x}}\| = \|\mathbf{x} - \mathbf{B}^{-1}\mathbf{b}\| = \|\mathbf{x} - \mathbf{B}^{-1}\mathbf{Ax}\| = \|(\mathbf{I} - \mathbf{B}^{-1}\mathbf{A})\mathbf{x}\| \le \|\mathbf{I} - \mathbf{B}^{-1}\mathbf{A}\|\,\|\mathbf{x}\|$$

This gives the magnitude of the perturbation in $\mathbf{x}$.

4.4 Norms and the Analysis of Errors

- Condition Number

If the relative perturbation is being measured, we can write

$$\frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|}{\|\mathbf{x}\|} \le \|\mathbf{I}-\mathbf{B}^{-1}\mathbf{A}\|$$

This inequality gives an upper bound on $\|\mathbf{x}-\tilde{\mathbf{x}}\|/\|\mathbf{x}\|$, and it is taken as a measure of the relative error between $\mathbf{x}$ and $\tilde{\mathbf{x}}$.

Example Suppose that the vector $\mathbf{b}$ is perturbed to obtain a vector $\tilde{\mathbf{b}}$. If $\mathbf{x}$ and $\tilde{\mathbf{x}}$ satisfy $\mathbf{A}\mathbf{x}=\mathbf{b}$ and $\mathbf{A}\tilde{\mathbf{x}}=\tilde{\mathbf{b}}$, how much do $\mathbf{x}$ and $\tilde{\mathbf{x}}$ differ, in absolute and relative terms?

4.4 Norms and the Analysis of Errors

- Condition Number

Solution : Assuming that $\mathbf{A}$ is invertible, we have

$$\|\mathbf{x}-\tilde{\mathbf{x}}\| = \|\mathbf{A}^{-1}\mathbf{b}-\mathbf{A}^{-1}\tilde{\mathbf{b}}\| = \|\mathbf{A}^{-1}(\mathbf{b}-\tilde{\mathbf{b}})\| \le \|\mathbf{A}^{-1}\|\,\|\mathbf{b}-\tilde{\mathbf{b}}\|$$

This gives a measure of the perturbation in $\mathbf{x}$. To estimate the relative perturbation, we write

$$\|\mathbf{x}-\tilde{\mathbf{x}}\| \le \|\mathbf{A}^{-1}\|\,\|\mathbf{b}-\tilde{\mathbf{b}}\| = \|\mathbf{A}^{-1}\|\,\|\mathbf{A}\mathbf{x}\|\frac{\|\mathbf{b}-\tilde{\mathbf{b}}\|}{\|\mathbf{b}\|} \le \|\mathbf{A}^{-1}\|\,\|\mathbf{A}\|\,\|\mathbf{x}\|\frac{\|\mathbf{b}-\tilde{\mathbf{b}}\|}{\|\mathbf{b}\|}$$

Hence

$$\frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|}{\|\mathbf{x}\|} \le \kappa(\mathbf{A})\frac{\|\mathbf{b}-\tilde{\mathbf{b}}\|}{\|\mathbf{b}\|}$$

where $\kappa(\mathbf{A}) \equiv \|\mathbf{A}\|\,\|\mathbf{A}^{-1}\|$, and the number $\kappa(\mathbf{A})$ is called a *condition number* of the matrix $\mathbf{A}$.

4.4 Norms and the Analysis of Errors

- Condition Number

The inequality tells us that the relative error in $\mathbf{x}$ is no greater than $\kappa(\mathbf{A})$ times the relative error in $\mathbf{b}$. The condition number depends on the vector norm chosen at the beginning of the analysis. From inequality we see that if the condition number is small, then small perturbations in $\mathbf{b}$ lead to small perturbations in $\mathbf{x}$. The inequality $\kappa(\mathbf{A}) \geq 1$ is always true.

We look at an example (where $\varepsilon > 0$)

$$\mathbf{A} = \begin{bmatrix} 1 & 1+\varepsilon \\ 1-\varepsilon & 1 \end{bmatrix} \qquad \mathbf{A}^{-1} = \varepsilon^{-2} \begin{bmatrix} 1 & -1-\varepsilon \\ -1+\varepsilon & 1 \end{bmatrix}$$

If the ∞-norm is used, we have

$$\|\mathbf{A}\|_\infty = 2+\varepsilon \quad \text{and} \quad \|\mathbf{A}^{-1}\|_\infty = \varepsilon^{-2}(2+\varepsilon).$$

Hence, $\kappa(\mathbf{A}) = [(2+\varepsilon)/\varepsilon]^2 > 4/\varepsilon^2$. If $\varepsilon \leq 0.01$, then $\kappa(\mathbf{A}) \geq 40{,}000$. In such a case, a small relative perturbation in $\mathbf{b}$ may induce a relative perturbation 40,000 times greater than the solution of the system $\mathbf{Ax} = \mathbf{b}$.

4.4 Norms and the Analysis of Errors

- Condition Number

If we solve a system of equations

$$\mathbf{Ax} = \mathbf{b}$$

numerically, we obtain not the exact solution $\mathbf{x}$ but an approximate solution $\tilde{\mathbf{x}}$.

One can test $\tilde{\mathbf{x}}$ by forming $\mathbf{A}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}$ to see if it is close to $\mathbf{b}$. Thus we obtain the **residual vector**

$$\mathbf{r} = \mathbf{b}\text{-}\tilde{\mathbf{b}}$$

The difference between the exact solution $\mathbf{x}$ and the approximation $\tilde{\mathbf{x}}$ is called the **error vector**

$$\mathbf{e} = \mathbf{x}\text{-}\tilde{\mathbf{x}}$$

4.4 Norms and the Analysis of Errors

- Condition Number

The following relationship

$$Ae = r$$

between the error vector and the residual vector is of fundamental importance.

We now establish relationships between the relative errors in $\mathbf{x}$ and $\mathbf{b}$. In other words, we want to relate

$$\|\mathbf{x} - \tilde{\mathbf{x}}\| / \|\mathbf{x}\| \quad \text{and} \quad \|\mathbf{b} - \tilde{\mathbf{b}}\| / \|\mathbf{b}\| = \|\mathbf{r}\| / \|\mathbf{b}\|.$$

The following theorem shows that the condition number $\kappa(\mathbf{A})$ plays an important role.

4.4 Norms and the Analysis of Errors

- Condition Number

Theorem 3 – on Bounds Involving Condition Number

Statement:

When solving systems of equations $\mathbf{Ax} = \mathbf{b}$, the condition number $\kappa(\mathbf{A})$, the residual vector $\mathbf{r}$, and the error vector $\mathbf{e}$ satisfy the the following inequality:

$$\frac{1}{\kappa(\mathbf{A})}\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|} \le \frac{\|\mathbf{e}\|}{\|\mathbf{x}\|} \le \kappa(\mathbf{A})\frac{\|\mathbf{r}\|}{\|\mathbf{b}\|}$$

4.4 Norms and the Analysis of Errors

- Condition Number

Proof:

The equality on the right can be written as

$$\|\mathbf{e}\| \, \|\mathbf{b}\| \leq \|\mathbf{A}\| \, \|\mathbf{A}^{-1}\| \, \|\mathbf{r}\| \, \|\mathbf{x}\|$$

and this is true since

$$\|\mathbf{e}\| \, \|\mathbf{b}\| = \|\mathbf{A}^{-1}\mathbf{r}\| \, \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}^{-1}\| \, \|\mathbf{r}\| \, \|\mathbf{A}\| \, \|\mathbf{x}\|$$

In fact, the inequality on the right in the theorem is

$$\frac{\|\mathbf{x} - \tilde{\mathbf{x}}\|}{\|\mathbf{x}\|} \leq \kappa(\mathbf{A}) \frac{\|\mathbf{b} - \tilde{\mathbf{b}}\|}{\|\mathbf{b}\|}$$

The inequality on the left can be written as

$$\|\mathbf{r}\| \, \|\mathbf{x}\| \leq \|\mathbf{A}\| \, \|\mathbf{A}^{-1}\| \, \|\mathbf{b}\| \, \|\mathbf{e}\|$$

and this follows at once from

$$\|\mathbf{r}\| \, \|\mathbf{x}\| = \|\mathbf{A}\mathbf{e}\| \, \|\mathbf{A}^{-1}\mathbf{b}\| \leq \|\mathbf{A}\| \, \|\mathbf{e}\| \, \|\mathbf{A}^{-1}\| \, \|\mathbf{b}\|$$

4.4 Norms and the Analysis of Errors

- Condition Number

A matrix with a large condition number is said to be **ill conditioned**. For an ill conditioned matrix **A**, there will be cases in which the solution of a system $\mathbf{Ax} = \mathbf{b}$ will be very sensitive to small changes in the vector **b**. In other words, to attain a certain precision in the determination of **x**, we shall require significantly higher precision in **b**. If the condition number of **A** is of moderate size, the matrix is said to be **well conditioned**.

4.4 Norms and the Analysis of Errors

- Condition Number

The following is an example of this.

For equations

$$\begin{cases} x_1 + x_2 = 2 \\ x_1 + 1.00001x_2 = 2 \end{cases}$$

the solution is $x_1 = 2,\ x_2 = 0.$

For equations

$$\begin{cases} x_1 + x_2 = 2 \\ x_1 + 1.00001x_2 = 2.00001 \end{cases}$$

the solution is $x_1 = 1,\ x_2 = 1.$

Comparing these two systems, there is only a small difference on the right side of the equations, but the solutions are so different.

4.4 Norms and the Analysis of Errors

- Condition Number

The condition number is calculated as following

$$\kappa(\mathbf{A}) = \|\mathbf{A}\|\|\mathbf{A}^{-1}\|$$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1+10^{-5} \end{bmatrix}$$

$$\mathbf{A}^{-1} = \begin{bmatrix} 1+10^{5} & -10^{5} \\ -10^{5} & 10^{5} \end{bmatrix}$$

therefore condition number $\kappa(\mathbf{A})$

$$\kappa(\mathbf{A})_{\infty} = \|\mathbf{A}\|_{\infty}\|\mathbf{A}^{-1}\|_{\infty} = (2+10^{-5})(2\times 10^{5}+1) > 4\times 10^{5}$$

The condition number is very big, so the system is ill conditioned.

4.4 Norms and the Analysis of Errors

- Condition Number

Questions

1. $A \in R^{n\times n}$ is symmetric and positive definite, define

$$\|x\|_A = (Ax, x)^{\frac{1}{2}},$$

prove that $\|x\|_A$ is a vector norm on $R^{n\times n}$.

2. Given

$$\mathbf{A} = \begin{bmatrix} 100 & 99 \\ 99 & 98 \end{bmatrix},$$

calculate the condition number $\kappa(\mathbf{A})_v$ $(v = 2, \infty)$.

4.6 Solution of Equations by Iterative Methods

The Gaussian algorithm and its variants are termed **direct methods** for solving the matrix problem $\mathbf{Ax} = \mathbf{b}$, which proceed through a finite number of steps and produce a solution $\mathbf{x}$ that would be completely accurate if not consider the roundoff errors.

An **indirect method**, by contrast, produces a sequence of vectors that ideally *converges* to the solution. The computation is halted when an approximate solution is obtained having some specified accuracy or after a certain number of iterations. Indirect methods are almost always **iterative** in nature.

4.6 Solution of Equations by Iterative Methods

For large linear systems containing thousands of equations, iterative methods often have decisive advantages over direct methods in terms of speed and demands on computer memory. Sometimes, if the accuracy requirements are not stringent, a modest number of iterations will suffice to produce an acceptable solution. For sparse systems, iterative methods are often very efficient. In sparse problems, the nonzero elements of **A** are sometimes stored in sparse-storage format; in other cases, it is not necessary to store **A** at all. The latter situation is common in the numerical solution of partial differential equations. In this case, each row of **A** might be generated as needed but not retained after use. Another advantage of iterative methods is that they are usually stable, and they will accually dampen errors (due to roundoff) as the process continues.

4.6 Solution of Equations by Iterative Methods

- Jacobi method

To convey the general idea, we describe two fundamental iterative methods. We start from considering the linear system

$$\begin{bmatrix} 7 & -6 \\ -8 & 9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$$

How can it be solved by an iterative process?

Solution : A straightforward procedure would be to solve the *i*th equation for the *i*th unknown as follows :

$$x_1^{(k)} = \frac{6}{7} x_2^{(k-1)} + \frac{3}{7}, \quad x_2^{(k)} = \frac{8}{9} x_1^{(k-1)} - \frac{4}{9}$$

This is known as the **Jacobi method** or **iteration**. Initially, we select for $x_1^{(0)}$ and $x_2^{(0)}$ the best available guess for the solution, or simply set them to 0. The equations above then generate what we hope are improved values $x_1^{(1)}$ and $x_2^{(1)}$.

4.6 Solution of Equations by Iterative Methods

- Jacobi method

The process is repeated a prescribed number of times or until a certain precision appears to have been achieved in the vectors $\left(x_1^{(k)}, x_2^{(k)}\right)^T$. Here are some selected values of the iterates of the Jacobi method for this example

k	$x_1^{(k)}$	$x_2^{(k)}$
0	0.00000	0.00000
10	0.14865	–0.19820
20	0.18682	–0.24909
30	0.19662	–0.26215
40	0.19913	–0.26551
50	0.19978	–0.26637

4.6 Solution of Equations by Iterative Methods

- Gauss-Seidel Method

It is apparent that this iterative process could be modefied so the newest value for $x_1^{(k)}$ is used immediately in the second equation. The resulting method is called the **Gauss-Seidel method** or **iteration**. Its equations are

$$x_1^{(k)} = \frac{6}{7}x_2^{(k-1)} + \frac{3}{7}, \quad x_2^{(k)} = \frac{8}{9}x_1^{(k)} - \frac{4}{9}$$

Some of the output from the Gauss-Seidel method follows:

k	$x_1^{(k)}$	$x_2^{(k)}$
0	0.00000	0.00000
10	0.21978	–0.24909
20	0.20130	–0.26531
30	0.20009	–0.26659
40	0.20001	–0.26666
50	0.20000	–0.26667

The Gauss-Seidal iteration is seen faster than Jacobi.

4.6 Solution of Equations by Iterative Methods

- Gauss-Seidel Method

Both the Jacobi and the Gauss-Seidel iterations seems to be converging to the same limit, and the latter is converging faster. Also, notice that, in contrast to a direct method, the precision we obtain in the solution depends on when the iterative process is halted.

The exact solution is $x_1 = \frac{1}{5}, \quad x_2 = -\frac{8}{30}.$

4.6 Solution of Equations by Iterative Methods

- Basic Concepts

We now consider iterative methods in a more general mathematical setting. A general type of iterative process for solving the system

$$\mathbf{Ax} = \mathbf{b} \qquad (1)$$

can be described as follows:

A certain matrix **Q**, called the **splitting matrix**, is prescribed, and the origional problem is rewritten in the equivalent form

$$\mathbf{Qx} = (\mathbf{Q} - \mathbf{A})\mathbf{x} + \mathbf{b} \qquad (2)$$

Equation (2) suggests an iterative process, defined by writting

$$\mathbf{Qx}^{(k)} = (\mathbf{Q} - \mathbf{A})\mathbf{x}^{(k-1)} + \mathbf{b} \quad (k \geq 1) \qquad (3)$$

4.6 Solution of Equations by Iterative Methods

- Basic Concepts

The initial vector $\mathbf{x}^{(0)}$ can be arbitrary; if a good guess of the solution is available, it should be used for $\mathbf{x}^{(0)}$. We shall say that the iterative in (3) converges for any initial vector $\mathbf{x}^{(0)}$. A sequence of vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$ can be computed from (3), and our objective is to choose $\mathbf{Q}$ so that these two conditions are met.

1. The sequence $\left[\mathbf{x}^{(k)}\right]$ is easily computed.

2. The sequence $\left[\mathbf{x}^{(k)}\right]$ converges rapidly to a solution.

4.6 Solution of Equations by Iterative Methods

- Basic Concepts

$$\mathbf{Ax} = \mathbf{b} \qquad (1)$$

If the sequence $\left[\mathbf{x}^{(k)}\right]$ converges, say to a vector $\mathbf{x}$, then $\mathbf{x}$ is automatically a solution. Indeed, if we simply take the limit in (3) and use the continuity of the algebraic operations, the result is

$$\mathbf{Qx} = (\mathbf{Q} - \mathbf{A})\mathbf{x} + \mathbf{b} \qquad (4)$$

which means that $\mathbf{Ax} = \mathbf{b}$.

$$\mathbf{Qx}^{(k)} = (\mathbf{Q} - \mathbf{A})\mathbf{x}^{(k-1)} + \mathbf{b} \quad (k \geq 1) \qquad (3)$$

To assure that (1) has a solution for any vector $\mathbf{b}$, we shall assume that $\mathbf{A}$ and $\mathbf{Q}$ are nonsingular, so (3) can be solved for the unknown vectors $\mathbf{x}^{(k)}$. Havig made these assumptions, we can use the following equation for the *theoretical* analysis:

$$\mathbf{x}^{(k)} = (\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A})\mathbf{x}^{(k-1)} + \mathbf{Q}^{-1}\mathbf{b} \qquad (5)$$

Equation (5) is convenient for the analysis, but in numerical work $\mathbf{x}^{(k)}$ is almost always obtained by solving (3) with out the use of $\mathbf{Q}^{-1}$.

4.6 Solution of Equations by Iterative Methods

- Basic Concepts

The actual solution $\mathbf{x}$ satisfies the equation

$$\mathbf{x} = (\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A})\mathbf{x} + \mathbf{Q}^{-1}\mathbf{b} \qquad (6)$$

By subtracting the terms in (6) from those in (5), we obtain

$$\mathbf{x}^{(k)} - \mathbf{x} = (\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A})(\mathbf{x}^{(k-1)} - \mathbf{x}) \qquad (7)$$

Now select any convenient vector norm and its subroutine matrix norm. We obtain

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| \leq \left\|\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A}\right\| \left\|\mathbf{x}^{(k-1)} - \mathbf{x}\right\| \qquad (8)$$

By repeating this step, we eventually arrive at the inequality

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| \leq \left\|\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A}\right\|^{k} \left\|\mathbf{x}^{(0)} - \mathbf{x}\right\| \qquad (10)$$

Thus, if $\left\|\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A}\right\| < 1$, we can conclude at once that

$$\lim_{k\to\infty} \left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| = 0 \qquad (11)$$

for any $\mathbf{x}^{(0)}$.

4.6 Solution of Equations by Iterative Methods

- Basic Concepts

Thus, we have the following theorem.

Theorem 1 Theorem on Iterative Method Convergence

If $\left\|\mathbf{I}-\mathbf{Q}^{-1}\mathbf{A}\right\|<1$ for some subordinate matrix norm, then the sequence produced by equation (3) converges to the solution of $\mathbf{Ax}=\mathbf{b}$ for any initial vector $\mathbf{x}^{(0)}$.

$$(\mathbf{Qx}^{(k)}=(\mathbf{Q}-\mathbf{A})\mathbf{x}^{(k-1)}+\mathbf{b} \quad (k\geq 1) \quad (3))$$

4.6 Solution of Equations by Iterative Methods

- Richardson Method

As an illustration of these concepts, we consider the **Richardson method**, in which **Q** is chosen to be identity matrix. Equation (3) in this case reads as follows:

$$\mathbf{x}^{(k)} = (\mathbf{I} - \mathbf{A})\mathbf{x}^{(k-1)} + \mathbf{b} = \mathbf{x}^{(k-1)} + \mathbf{r}^{(k-1)} \qquad (12)$$

where $\mathbf{r}^{(k-1)}$ is the residual vector, defined by $\mathbf{r}^{(k-1)} = \mathbf{b} - \mathbf{A}\mathbf{x}^{(k-1)}$. According to Thoerem 1, the Richardson iteration will produce a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ (in the limit) if $\|\mathbf{I} - \mathbf{A}\| < 1$ for some subordinate matrix norm.

$$(\mathbf{Q}\mathbf{x}^{(k)} = (\mathbf{Q} - \mathbf{A})\mathbf{x}^{(k-1)} + \mathbf{b} \quad (k \geq 1) \quad (3)\ )$$

4.6 Solution of Equations by Iterative Methods

- Jacobi Method

We consider the linear equation system

$$\mathbf{Ax} = \mathbf{b}$$

We can write matrix A in the following form

$$\mathbf{A} = \mathbf{D} - \mathbf{L} - \mathbf{U}$$

where **D** is a diagonal matrix, and $d_{ii} = a_{ii}$, **L** a lower matrix and **U** an upper matrix, $l_{ii} = u_{ii} = 0$. When **D** is nonsingular, i.e. $a_{ii} \neq 0$, we have

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

We then have the iterative formula

$$\mathbf{x}^{(k)} = \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})\mathbf{x}^{(k-1)} + \mathbf{D}^{-1}\mathbf{b}, \qquad k = 1,2,...$$

This is the **Jacobi iteration.**

4.6 Solution of Equations by Iterative Methods

- Jacobi Method

$$(\mathbf{x} = (\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A})\mathbf{x} + \mathbf{Q}^{-1}\mathbf{b} \qquad (6)\ )$$

In Jacobi method, the splitting matrix is diagonal matrix.

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

$$= (\mathbf{I} - \mathbf{D}^{-1}\mathbf{A})\mathbf{x} + \mathbf{D}^{-1}\mathbf{b}$$

In **Jacobi** iteration $\mathbf{Q}$ is the diagonal matrix $\mathbf{D}$ whose diagonal entries are the same as those in the matrix $\mathbf{A} = (a_{ij})$. In this case, the generic element of $\mathbf{Q}^{-1}\mathbf{A}$ is a_{ij}/a_{ii}. The diagonal elements of this matrix are all 1, and hence,

$$\left\|\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A}\right\|_{\infty} = \max_{1 \le i \le n} \sum_{\substack{j=1 \\ j \ne i}}^{n} \left| a_{ij}/a_{ii} \right| \qquad (13)$$

4.6 Solution of Equations by Iterative Methods

- Jacobi Method

Theorem 2 Theorem on Convergence of Jacobi Method

If **A** is **strictly** diagonally dominant, then the sequence produced by the Jacobi iteration converges to the solution of $\mathbf{Ax} = \mathbf{b}$ for any starting vector.

Proof : Diagonal dominance means that

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \qquad (1 \le i \le n)$$

From (13), we then conclude that

$$\left\| \mathbf{I} - \mathbf{Q}^{-1}\mathbf{A} \right\|_{\infty} < 1$$

By Theorem 1, the Jacobi iteration converges.

4.6 Solution of Equations by Iterative Methods

- Gauss-Seidel Method

In theoretical analysis, the matrix form is used. In calculation the following component form is often used:

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

when calculating $x_i^{(k)}$, $x_1^{(k)}$, $x_2^{(k)}$,...,$x_{i-1}^{(k)}$ have been calculated.

If we use the latest values of them, the formula becomes

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

This is the **Gauss - Seidel** iteration.

Its matrix form is

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

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

4.6 Solution of Equations by Iterative Methods

- Gauss-Seidel Method

Therefore in the Gauss-Seidel matrix, **Q** is taken to be the lower triangular part of **A**, i.e. $\mathbf{Q} = \mathbf{(D - L)}$.

In Gauss-Seidel method, the updated values of x_i replaced the old values immediately, whereas in the Jacobi method, all new components of the x-vector are computed before the replacement takes place. In the Jacobi method, the new components of x can be calculated simutaneously, whereeas in Gauss-Seidel method they are computed serially since the computation of the new x_i requires all the new values of x_1, x_2, ...,x_{i-1}. Because of this differences, the Jacobi iteration may be preferable on computers that allow vector or parallel processing.

4.6 Solution of Equations by Iterative Methods

- SOR Method

If the splitting matrix $\mathbf{Q}$ is chosen to be the lower triangular part with a parameter of $\mathbf{A}$:

$$\mathbf{Q} = \frac{1}{\omega}(\mathbf{D} - \omega\mathbf{L}),$$

where $\omega>0$. In this case,

$$\mathbf{x} = (\mathbf{I} - \mathbf{Q}^{-1}\mathbf{A})\mathbf{x} + \mathbf{Q}^{-1}\mathbf{b} = (\mathbf{I} - \omega(\mathbf{D} - \omega\mathbf{L})^{-1}\mathbf{A})\mathbf{x} + \omega(\mathbf{D} - \omega\mathbf{L})^{-1}\mathbf{b},$$

thus we can obtain another iterative method, and the iterative matrix is

$$\mathbf{L}_{\omega} \equiv \mathbf{I} - \omega(\mathbf{D} - \omega\mathbf{L})^{-1}\mathbf{A} = (\mathbf{D} - \omega\mathbf{L})^{-1}((1-\omega)\mathbf{D} + \omega\mathbf{U}).$$

This method is known as the **successive over relaxation** method, commonly abbreviated as **SOR** method.

The matrix form of the **SOR** method for solving $\mathbf{Ax} = \mathbf{b}$ is

$$\mathbf{x}^{(k)} = (\mathbf{D} - \omega\mathbf{L})^{-1}((1-\omega)\mathbf{D} + \omega\mathbf{U})\mathbf{x}^{(k-1)} + \omega(\mathbf{D} - \omega\mathbf{L})^{-1}\mathbf{b}, \ k = 1,2,\cdots$$

from which we can get

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

or

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

4.6 Solution of Equations by Iterative Methods

- SOR Method

Then we give the component form of the **SOR** method:

$$x_i^{(k)} = x_i^{(k-1)} + \omega(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i}^{n} a_{ij}x_j^{(k-1)}) / a_{ii}, \quad i = 1,2,\cdots,n, \quad k = 1,2,\cdots$$

or

$$\begin{cases} x_i^{(k)} = x_i^{(k-1)} + \Delta x_i \\ \Delta x_i = \omega(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i}^{n} a_{ij}x_j^{(k-1)}) / a_{ii} \end{cases} \quad i = 1,2,\cdots,n, \quad k = 1,2,\cdots$$

Recall the component form of the **Gauss - Seidel** method:

$$x_i^{(k)} = (b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k-1)}) / a_{ii}, \quad i = 1,2,\cdots,n, \quad k = 1,2,\cdots$$

which can be written as

$$\begin{cases} x_i^{(k)} = x_i^{(k-1)} + \Delta x_i \\ \Delta x_i = (b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i}^{n} a_{ij}x_j^{(k-1)}) / a_{ii} \end{cases} \quad i = 1,2,\cdots,n, \quad k = 1,2,\cdots$$

Thus, the **SOR** method becomes the **Gauss - Seidel** mehod when $\omega = 1$.

4.6 Solution of Equations by Iterative Methods

- SOR Method

The **SOR** method can be seen as a revision of the **Gauss - Seidel** mehod, which is explained as follows.

- 由Gauss-Seidel迭代法得

$$\bar{x}_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right)$$

$$x_i^{(k+1)} = \omega \bar{x}_i^{(k+1)} + (1-\omega)x_i^{(k)}$$

即

$$x_i^{(k+1)} = x_i^{(k)} + \frac{\omega}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i}^{n} a_{ij}x_j^{(k)}\right)$$

称为超松弛迭代法(SOR方法),其中ω为松弛因子.

4.6 Solution of Equations by Iterative Methods

- Convergence Analysis of Iterative Method

Theorem 5 Theorem on Necessary and Suffient Conditions for Iterative Method Convergence

Suppose $\mathbf{A}$ is invertible, the linear system

$$\mathbf{Ax} = \mathbf{b}$$

can be writen as the following system

$$\mathbf{x} = \mathbf{Gx} + \mathbf{c}$$

For the iteration formula

$$\mathbf{x}^{(k)} = \mathbf{Gx}^{(k-1)} + \mathbf{c}$$

to produce a sequence converging to $(\mathbf{I} - \mathbf{G})^{-1}\mathbf{c}$, for any starting vector $\mathbf{x}^{(0)}$, it is necessary and suffient that $\rho(\mathbf{G}) < 1$.

4.6 Solution of Equations by Iterative Methods

- Convergence Analysis of Iterative Method

Proof:

Suffient Suppose $\rho(\mathbf{G}) < 1$, it is easy to know that $\mathbf{Ax} = \mathbf{b}$ has a unique solution $\mathbf{x}^*$,

$$\mathbf{x}^* = \mathbf{G}\mathbf{x}^* + \mathbf{c},$$

the error vector

$$\mathbf{e}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}^* = \mathbf{G}(\mathbf{x}^{(k-1)} - \mathbf{x}^*) = \ldots = \mathbf{G}^k\mathbf{e}^{(0)}, \quad \mathbf{e}^{(0)} = \mathbf{x}^{(0)} - \mathbf{x}^*$$

since $\rho(\mathbf{G}) < 1$, then $\lim \mathbf{G}^k = \mathbf{0}$. Therefore, for any $\mathbf{x}^{(0)}$, we have $\lim \mathbf{e}^{(k)} = \mathbf{0}$, that is, $\lim \mathbf{x}^{(k)} = \mathbf{x}^*$.

Necessary If for any $\mathbf{x}^{(0)}$, $\lim \mathbf{x}^{(k)} = \mathbf{x}^*$, since

$$\mathbf{x}^{(k+1)} = \mathbf{G}\mathbf{x}^{(k)} + \mathbf{c},$$

it is obvious that $\mathbf{x}^*$ is the solution of $\mathbf{Ax} = \mathbf{b}$, and for any $\mathbf{x}^{(0)}$, we have

$$\lim \mathbf{e}^{(k)} = \lim(\mathbf{x}^{(k)} - \mathbf{x}^*) = \lim \mathbf{G}^k\mathbf{e}^{(0)} = 0$$

then $\lim \mathbf{G}^k = \mathbf{0}$, therefore, we have $\rho(\mathbf{G}) < 1$.

4.6 Solution of Equations by Iterative Methods

- Convergence Analysis of Iterative Method

Corollary Corollary on Necessary and Sufficient Conditions for **Jacobi** and **Gauss - Seidel** and **SOR** Methods Convergence

For the linear system $\mathbf{Ax} = \mathbf{b}$, suppose $\mathbf{A} = \mathbf{D} - \mathbf{L} - \mathbf{U}$ is invertible, where $\mathbf{D}$ is also invertible, then

(1) **Jacobi iteration** method is convergent, it is necessary and sufficient that $\rho(\mathbf{G}_J) < 1$, where $\mathbf{G}_J = \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})$.

(2) **Gauss - Seidel** iteration method is convergent, it is necessary and sufficient that $\rho(\mathbf{G}_{G\text{-}S}) < 1$, where $\mathbf{G}_{G\text{-}S} = (\mathbf{D} - \mathbf{L})^{-1}\mathbf{U}$.

(3) **SOR** iteration method is convergent, it is necessary and sufficient that $\rho(\mathbf{L}_\omega) < 1$, where $\mathbf{L}_\omega = (\mathbf{D} - \omega\mathbf{L})^{-1}((1-\omega)\mathbf{D} + \omega\mathbf{U})$.

4.6 Solution of Equations by Iterative Methods

- Gauss-Seidel Method - examples

Using Jacobi and Gauss-Seidel methods to solve

$$\begin{cases} 8x_1 - x_2 + x_3 = 1 \\ 2x_1 + 10x_2 - x_3 = 4 \\ x_1 + x_2 - 5x_3 = 3 \end{cases}$$

By Jacobi method:

$$\begin{cases} x_1^{(k)} = \frac{1}{8}(1 + x_2^{(k-1)} - x_3^{(k-1)}) = 0.125(1 + x_2^{(k-1)} - x_3^{(k-1)}) \\ x_2^{(k)} = \frac{1}{10}(4 - 2x_1^{(k-1)} + x_3^{(k-1)}) = 0.1(4 - 2x_1^{(k-1)} + x_3^{(k-1)}) \\ x_3^{(k)} = \frac{1}{-5}(3 - x_1^{(k-1)} - x_2^{(k-1)}) = 0.2(-3 + x_1^{(k-1)} + x_2^{(k-1)}) \end{cases}$$

By Gauss-Seidel method:

$$\begin{cases} x_1^{(k)} = 0.125(1 + x_2^{(k-1)} - x_3^{(k-1)}) \\ x_2^{(k)} = 0.100(4 - 2x_1^{(k)} + x_3^{(k-1)}) \\ x_3^{(k)} = 0.200(-3 + x_1^{(k)} + x_2^{(k)}) \end{cases}$$

4.6 Solution of Equations by Iterative Methods

- Gauss-Seidel Method - examples

The results are

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$
0	0.1250	0.4000	−0.6000
1	0.2500	0.3150	−0.4950
2	0.2263	0.3005	−0.4870
3	0.2235	0.3060	−0.4946
4	0.2251	0.3058	−0.4941
5	0.2250	0.3056	−.4938
6	0.2249	0.3056	−0.4939
7	0.2249	0.3056	−0.4939

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$
0	0.1250	0.4000	−0.6000
1	0.2500	0.2900	−0.4920
2	0.2228	0.3062	−0.4942
3	0.2251	0.3056	−0.4939
4	0.2249	0.3056	−0.4939
5	0.2249	0.3056	−0.4939

Several Examples

- Indirect method

Example 1 Using iterative methods to solve

$$\begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\begin{pmatrix} 5 \\ 5 \end{pmatrix}$$

Solution: Using an iterative method:

$$\begin{cases} x_1^{(k)} = -\dfrac{1}{3}x_2^{(k-1)} + \dfrac{5}{3} \\ x_2^{(k)} = -\dfrac{1}{2}x_1^{(k-1)} + \dfrac{5}{2} \end{cases}$$

what we construct is

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^{(\mathrm{k})} = \begin{pmatrix} 0 & -\dfrac{1}{3} \\ -\dfrac{1}{2} & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^{(k-1)} + \begin{pmatrix} \dfrac{5}{3} \\ \dfrac{5}{2} \end{pmatrix} \qquad (5)$$

we start from $x^{(0)} = [0,0]^{\mathrm{T}}$. The exact solution is $x_1 = 1,\ x_2 = 2.$

Several Examples

- Indirect method

If we obtain x_1 from the second equation and x_2 from the first

$$\begin{cases} x_1^{(k)} = -2x_2^{(k-1)} + 5 \\ x_2^{(k)} = -3x_1^{(k-1)} + 5 \end{cases}$$

what we construct is

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^{(k)} = \begin{pmatrix} 0 & -2 \\ -3 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^{(k-1)} + \begin{pmatrix} 5 \\ 5 \end{pmatrix} \qquad (6)$$

We find that (5) convergent and (6) is divergent.

Several Examples

- Indirect method

In (5), $\mathbf{G} = \begin{pmatrix} 0 & -\frac{1}{3} \\ -\frac{1}{2} & 0 \end{pmatrix}$ and its eigenvalues may be obtained

$$\lambda^2 - \frac{1}{6} = 0 \quad \Rightarrow \quad \lambda_1 = -\frac{1}{\sqrt{6}}, \ \lambda_2 = \frac{1}{\sqrt{6}} \quad \Rightarrow \quad \rho(\mathrm{G}) = \frac{1}{\sqrt{6}} < 1$$

Therefore (5) is convergent.

But, in (6), $\mathbf{G} = \begin{pmatrix} 0 & -2 \\ -3 & 0 \end{pmatrix}$ and its eigenvalues are

$$\lambda^2 - 6 = 0 \quad \Rightarrow \quad \lambda_1 = -\sqrt{6}, \ \lambda_2 = \sqrt{6} \quad \Rightarrow \quad \rho(\mathrm{G}) = \sqrt{6} > 1$$

Therefore (6) is divergent.

Several Examples

- Indirect method

Example 2: Consider a system

$$\begin{pmatrix} 2 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

examine its convergence when using the Jacobi and Gauss-Seidel methods.

Solution: decomposing the coefficient matrix such that $\mathbf{A} = \mathbf{D} - \mathbf{L} - \mathbf{U}$.

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

Several Examples

- Indirect method

The iteration matrix in Jacobi method is

$$\mathbf{G}_{\mathrm{J}} = \mathbf{D}^{-1}(\mathbf{L}+\mathbf{U}) = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1/2 \end{pmatrix} \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 & -1/2 \\ -1 & 0 & -1 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

Its characteristic equation is

$$\left|\lambda\mathbf{I}-\mathbf{G}_{\mathrm{J}}\right| = \begin{vmatrix} \lambda & -1/2 & 1/2 \\ 1 & \lambda & 1 \\ -1/2 & -1/2 & \lambda \end{vmatrix} = \lambda^3 + \frac{5}{4}\lambda = 0 \quad \Rightarrow \quad \lambda_1 = 0,\ \lambda_{2,3} = \pm\frac{\sqrt{5}}{2}i$$

Since $\rho(\mathbf{G}_{\mathrm{J}}) = \dfrac{\sqrt{5}}{2} > 1$, Jacobi iteration is divergent.

Several Examples

- Indirect method

But for Gauss-Seidel method, the iterative matrix is

$$\mathbf{G}_{\text{G-S}} = (\mathbf{D}-\mathbf{L})^{-1}\mathbf{U} = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & -2 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 & -1/2 \\ 0 & -1/2 & -1/2 \\ 0 & 0 & -1/2 \end{pmatrix}$$

Obviously, its eigenvalues $\lambda_1 = 0,\ \lambda_{2,3} = -\frac{1}{2},\ \rho(\mathbf{G}_{\text{G-S}}) = \frac{1}{2} < 1,$ so the Gauss-Seidel method is convergent.

Several Examples

- Indirect method

Example 3: Consider a system

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

discuss its convergence when using Jacobi and G-S methods.

Solutions: decomposing the coefficient matrix such that $\mathbf{A} = \mathbf{D} - \mathbf{L} - \mathbf{U}$.

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

$$\mathbf{G}_{\mathrm{J}} = \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{pmatrix}$$

Several Examples

- Indirect method

Its characteristic equation is

$$\left|\lambda \mathbf{I}-\mathbf{G}_{\mathrm{J}}\right|=\begin{vmatrix}\lambda & 2 & -2\\ 1 & \lambda & 1\\ 2 & 2 & \lambda\end{vmatrix}=\lambda^{3}=0 \;\Rightarrow\; \lambda_1=\lambda_2=\lambda_3=0,$$

Since $\rho(\mathbf{G}_{\mathrm{J}})=0<1$, Jacobi iteration is convergent.

For Gauss-Seidel method, the iterative matrix is

$$\mathbf{G}_{\text{G-S}}=(\mathbf{D}-\mathbf{L})^{-1}\mathbf{U}=\begin{pmatrix}1 & 0 & 0\\ 1 & 1 & 0\\ 2 & 2 & 1\end{pmatrix}^{-1}\begin{pmatrix}0 & -2 & 2\\ 0 & 0 & -1\\ 0 & 0 & 0\end{pmatrix}=\begin{pmatrix}0 & -2 & 2\\ 0 & 2 & -3\\ 0 & 0 & 2\end{pmatrix}$$

Obviously, its eigenvalues $\lambda_1=0,\ \lambda_{2,3}=2,\ \rho(\mathbf{G}_{\text{G-S}})=2>1$, so the Gauss-Seidel method is divergent.

Several Examples

- Indirect method

Example 4: Consider a system

$$\begin{bmatrix} -4 & 1 & 1 & 1 \\ 1 & -4 & 1 & 1 \\ 1 & 1 & -4 & 1 \\ 1 & 1 & 1 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

use the **SOR** method to solve it, given the true solution $\mathbf{x}^* = (-1,-1,-1,-1)^{\mathrm{T}}$.

Solutions: Let $\mathbf{x}^{(0)} = \mathbf{0}$, the iterative formula is

$$\begin{cases} x_1^{(k)} = x_1^{(k-1)} - \omega(1 + 4x_1^{(k-1)} - x_2^{(k-1)} - x_3^{(k-1)} - x_4^{(k-1)})/4 \\ x_2^{(k)} = x_2^{(k-1)} - \omega(1 - x_1^{(k)} + 4x_2^{(k-1)} - x_3^{(k-1)} - x_4^{(k-1)})/4 \\ x_3^{(k)} = x_3^{(k-1)} - \omega(1 - x_1^{(k)} - x_2^{(k)} + 4x_3^{(k-1)} - x_4^{(k-1)})/4 \\ x_4^{(k)} = x_4^{(k-1)} - \omega(1 - x_1^{(k)} - x_2^{(k)} - x_3^{(k)} + 4x_4^{(k-1)})/4 \end{cases}$$

choose $\omega = 1.3$, the result of 11 iterations is

$$\mathbf{x}^{(11)} = (-0.99999646, -1.00000310, -0.99999953, -0.99999912)^{\mathrm{T}},$$

the error is

$$\left\| \varepsilon^{(11)} \right\|_2 \le 0.46 \times 10^{-5}.$$

Several Examples

- Indirect method

When ω is chosen to be other values, the iteration numbers are given in the table below. It can be seen that a good choice for ω can accelerate the convergence of the iteration. For this example, the best choice is $\omega = 1.3$.

ω	iteration number $\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|_2 < 10^{-5}$	ω	iteration number $\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|_2 < 10^{-5}$
1.0	22	1.5	17
1.1	17	1.6	23
1.2	12	1.7	33
1.3	11	1.8	53
1.4	14	1.9	109

Exercises

*Ex*1. Suppose that $\mathrm{x} = [2, -3, 4]^{\mathrm{T}}$, $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 4 \\ 0 & -2 & 4 \end{bmatrix}$, calculate $\|\mathrm{x}\|_\infty, \| x \|_1, \| x \|_2$, $\| A \|_\infty, \| A \|_1, \| A \|_2$ *and* $\rho(\mathrm{A})$.

*Ex*2. Prove that for the following given equation system, the Jacobi iteration diverges and Gauss-Seidel iteration converges.

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Exercises

*Ex*3. Use the SOR method (ω=0.9) to solve

$$\begin{bmatrix} 5 & 2 & 1 \\ -1 & 4 & 2 \\ 2 & -3 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -12 \\ 20 \\ 3 \end{bmatrix}.$$

When $\left\| \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \right\|_\infty < 10^{-4}$, stop the iteration.