Numerical Methods

Solution of Linear Systems

Iterative Methods

2013-2014

[DΣIM]

Iterative Methods

- Using an iterative method we will obtain the solution of a linear system by a progressive refinement.
- It will be necessary to have a way to measure the distance between two vectors from to consecutive iterations.
- The concept of vectorial distance is associated to the concept of norm

Vectorial Norms

- A vectorial norm in $\mathbf{R^n}$ is a function $\|\cdot\|$ from $\mathbf{R^n} \rightarrow \mathbf{R}$ with the following properties:

i) $\|\mathbf{x}\| \geq 0$ per a tot $\mathbf{x} \in \mathbf{R^n}$

ii) $\|\mathbf{x}\| = 0$ si i només si $\mathbf{x} = \mathbf{0}$

iii) $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ per a tot $\alpha \in \mathbf{R}$ i $\mathbf{x} \in \mathbf{R^n}$

iv) $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ per a tot $\mathbf{x},\mathbf{y} \in \mathbf{R^n}$

Vectorial Norms

- There are many possibilities to define a norm. The more frequently used are:
- The euclidean norm or l_2

$$\|\mathbf{x}\|_2 = \left\{ \sum_{i=1}^{n} x_i^2 \right\}^{1/2}$$

- The infinite norm or l_∞

$$\|\mathbf{x}\|_\infty = \max_{1 \le i \le n} |x_i|$$

Distance

- When we have defined a norm, we can introduce the distance between to vectors:
- This distance will be the norm of the difference vector. Then, we have:

$$\|\mathbf{x}-\mathbf{y}\|_2 = \left\{ \sum_{i=1}^{n} |x_i - y_i|^2 \right\}^{1/2}$$

- Or alternatively:

$$\|\mathbf{x}-\mathbf{y}\|_\infty = \max_{1\le i\le n} |x_i - y_i|$$

Convergence

- Given a distance, we can decide if a succession of vectors is reaching certain limit, that we will take as our solution vector.
- A succession of vectors $\{\boldsymbol{x}^{(k)}\}$ is convergent to a vector $\mathbf{x}$ in $\mathbf{R}^{\mathbf{n}}$ with respect to certain norm $\|\cdot\|$, if given a value $\varepsilon>0$ there is an integer $N(\varepsilon)$ such as:

$$\left\|\mathbf{x}^{\mathbf{k}}-\mathbf{x}\right\|<\varepsilon \quad \text{for all} \quad k \geq N(\varepsilon)$$

Convergence

- The following theorem gives a way to recognize a convergent succession.
- ***Theorem:*** The vector succession $\{\mathbf{x}^{(k)}\}$ is convergent to a vector $\mathbf{x}$ in $\mathbf{R}^n$ with respect of some norm $\|\cdot\|$, if and only if:

$$\lim_{k\to\infty} x_i^{(k)} = x_i \quad \text{for each } i = 1,\ldots,n$$

Matrix Norms

- A matrix norm defined in the matrix set of real coefficient matrices of dimension n x n is a function $\|\cdot\|$ from $\mathbf{R}^{nxn} \rightarrow \mathbf{R}$ with the following properties:

i) $\|\mathbf{A}\| \geq 0$ per a tot $\mathbf{A} \in \mathbf{R}^{nxn}$

ii) $\|\mathbf{A}\| = 0$ si i només si $\mathbf{A} = \mathbf{0}$

iii) $\|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\|$ per a tot $\alpha \in \mathbf{R}$ i $\mathbf{A} \in \mathbf{R}^{nxn}$

iv) $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$ per a tot $\mathbf{A}, \mathbf{B} \in \mathbf{R}^{nxn}$

v) $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{B}\|$

Matrix Norms

- If $\|\cdot\|$ is a vectorial norm in $\mathbf{R^n}$, then the relation:

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\|$$

- Defines a matrix norm in the set of real matrices of dimension *n* x *n* that we will call the ***natural norm***.
- We are going to use only natural norms.

Matrix Norms

- Some natural matrix forms are the norm l_∞

$$\|\mathbf{A}\|_\infty = \max_{\|\mathbf{x}\|_\infty = 1} \|\mathbf{A}\mathbf{x}\|_\infty$$

- And the euclidean norm l_2

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|_2 = 1} \|\mathbf{A}\mathbf{x}\|_2$$

- It is true that

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

Spectral Radius

- The spectral radius of a matrix is defined by:

$$\rho(\mathbf{A}) = \max_{1 \le i \le n} \left| \lambda_i \right| \quad i = 1, \ldots, n$$

- Where λ_i is a proper value of the matrix. The following theorem is true:
- ***Theorem:*** If **A** is a *n* x *n* real matrix, then for any natural norm:

$$\text{i)} \quad \left[\rho(\mathbf{A}^{\mathbf{T}} \cdot \mathbf{A}) \right]^{1/2} = \left\| \mathbf{A} \right\|_2$$

$$\text{ii)} \quad \rho(\mathbf{A}) \le \left\| \mathbf{A} \right\|$$

Iterative Methods

- When using an iterative method to solve a linear system of the form:

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

- We want to fin a succession of vectors $\{\mathbf{x}^{(k)}\}$, $k \geq 0$ verifying:

$$\lim_{k \to \infty} \mathbf{x}^{(k)} = \mathbf{x}$$

- If this is true we will say that the iterative method is convergent.

Iterative Methods

- The linear system

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

- Can be rewritten as:

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

- This allows to write the following recurrent rule:

$$\mathbf{x}^{(k+1)} = \mathbf{B}\mathbf{x}^{(k)} + \mathbf{c} \quad k \geq 0$$

Iterative Methods

- Even if $\mathbf{x}^{(0)}$ can be an arbitrary vector, the near to the final solution the better, as then the convergence will be faster.
- The matrix **B** is called the iterative matrix.
- For an iterative method to be convergent independently of the initial vector the following theorem must be fulfilled:

Iterative Methods

- ***Theorem:*** Convergence of iterative methods:
  - a) An iterative method $\mathbf{x}^{(k+1)} = \mathbf{B}\mathbf{x}^{(k)} + \mathbf{c}$ is convergent for any $\mathbf{x}^{(0)}$ if and only if $\rho(B)<1$
  - b) An iterative method $\mathbf{x}^{(k+1)} = \mathbf{B}\mathbf{x}^{(k)} + \mathbf{c}$ is convergent for any $\mathbf{x}^{(0)}$ if and only if there is some vectorial norm $\|\cdot\|$ associated to a matrix norm, such that $\|\mathbf{B}\|<1$

Iterative Methods

- For any solution vector, it will be true that $\mathbf{x}=\mathbf{B}\mathbf{x}+\mathbf{c}$. Then we can write

$$\mathbf{x}^{(k+1)}-\mathbf{x}=\mathbf{B}\left(\mathbf{x}^{(k)}-\mathbf{x}\right)=\cdots=\mathbf{B}^{k+1}\left(\mathbf{x}^{(0)}-\mathbf{x}\right)$$

- Suppose now that we have a basis of proper vectors of $\mathbf{B}$, $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$. We have:

$$\mathbf{x}^{(0)}-\mathbf{x}=\alpha_1\mathbf{u}_1+\cdots+\alpha_n\mathbf{u}_n$$

Iterative Methods

- If we apply $k+1$ times matrix **B**, we have

$$\mathbf{x}^{(k+1)} - \mathbf{x} = \alpha_1 \lambda_1^k \mathbf{u_1} + \cdots + \alpha_n \lambda_n^k \mathbf{u_n}$$

- And if it is true that $|\lambda_i| < 1$ for $i=1,\ldots,n$, i.e., if $\rho(B) < 1$ it will be true that:

$$\lim_{k \to \infty} \mathbf{x}^{(k)} = \mathbf{x}$$

Iterative Methods

- However, the proper values are not normally know beforehand. In this case we can use the norm:

$$\left\|\mathbf{x}^{(k+1)}-\mathbf{x}\right\|=\left\|\mathbf{B}^{k+1}(\mathbf{x}^{(0)}-\mathbf{x})\right\|\leq\left\|\mathbf{B}^{k+1}\right\|\cdot\left\|\mathbf{x}^{(0)}-\mathbf{x}\right\|$$

- As the norms have the multiplicative property:

$$\left\|\mathbf{B}^{k+1}\right\|\leq\left\|\mathbf{B}\right\|\cdot\left\|\mathbf{B}^{k}\right\|\leq\cdots\leq\left\|\mathbf{B}\right\|^{k+1}$$

- If $\|\mathbf{B}\|<1$ we have:

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

Jacobi's Method

- Given the linear system

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

- We can rewrite matrix **A** as the sum of three new matrices. A lower triangular with zeros in the main diagonal **L**, an upper triangular matrix with zeros in the main diagonal **U** and a diagonal matrix **D**

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

Mètode de Jacobi

- Les components de les matrius són

$$l_{ij} = \begin{cases} a_{ij} & \text{si} \quad i > j \\ 0 & \text{si} \quad i \leq j \end{cases}$$

$$d_{ij} = \begin{cases} a_{ii} & \text{si} \quad i = j \\ 0 & \text{si} \quad i \neq j \end{cases}$$

$$u_{ij} = \begin{cases} a_{ij} & \text{si} \quad i < j \\ 0 & \text{si} \quad i \geq j \end{cases}$$

Jacobi's Method

- The linear system can be rewritten as:

$$\mathbf{Ax} = \mathbf{b} \Leftrightarrow \left(\mathbf{L} + \mathbf{D} + \mathbf{U}\right)\mathbf{x} = \mathbf{b}$$

- If all the diagonal elements are different from zero $a_{ii} \neq 0$, we can make:

$$(\mathbf{L} + \mathbf{U})\mathbf{x} + \mathbf{Dx} = \mathbf{b} \Leftrightarrow \mathbf{Dx} = -(\mathbf{L} + \mathbf{U}) + \mathbf{b}$$

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

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

Jacobi's Method

- From the last relation we can obtain the following recursive formula:

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

- Which written in components is:

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

Gauss-Seidel Method

- Another possibility is to decompose the matrix **A** in the following way:

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

- Then we have:

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

Gauss Seidel Method

- And we obtain the recursive relation:

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

- The iteration matrix is now: $\mathbf{B}=-(\mathbf{L}+\mathbf{D})^{-1}\mathbf{U}$.
- Normally, however, we use the expression:

$$\begin{aligned}
\mathbf{x}^{(k+1)} &= -(\mathbf{L}+\mathbf{D})^{-1}\mathbf{U}\mathbf{x}^{(k)} + (\mathbf{L}+\mathbf{D})^{-1}\mathbf{b} \Leftrightarrow \\
(\mathbf{L}+\mathbf{D})\mathbf{x}^{(k+1)} &= -\mathbf{U}\mathbf{x}^{(k)} + \mathbf{b} \Leftrightarrow \\
\mathbf{D}\mathbf{x}^{(k+1)} + \mathbf{L}\mathbf{x}^{(k+1)} &= -\mathbf{U}\mathbf{x}^{(k)} + \mathbf{b} \Leftrightarrow \\
\mathbf{x}^{(k+1)} &= \mathbf{D}^{-1}\left(-\mathbf{L}\mathbf{x}^{(k+1)} - \mathbf{U}\mathbf{x}^{(k)} + \mathbf{b}\right)
\end{aligned}$$

Gauss Seidel Method

- This relation in components is written as:

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

- Both methods are similar. In the Jacobi's method each component is calculated as a function of the components in the previous step. In Gauss-Seidel method we use the new calculated components

[DΣIM]

Diagonal Dominant Matrices

- Many numerical problems need to solve linear systems where the matrix is a diagonal dominant matrix. In this situation the following theorem is satisfied
- ***Theorem:*** If **A** is an strict diagonal dominant matrix the Jacobi and Gauss-Seidel methods are convergent.

Stopping iterative methods

- When stopping iterative methods we must take into account that the exact solution is unknown and we will be forced to use the calculated values.
- First of all, we have to choose a tolerance $\varepsilon>0$ which will be a measure of the precision that we want to reach.

Stopping iterative methods

- Next, we need to fix a criterion to stop our computations. For instance:

$$\left\| \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \right\| < \varepsilon$$

- O alternatively:

$$\frac{\left\| \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \right\|}{\left\| \mathbf{x}^{(k)} \right\|} < \varepsilon$$

Error Analysis

- Even that we do not know the exact solution, we can place an upper limit to the error after k iterations. As:

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| \leq \left\|\mathbf{B}\right\|^{k} \cdot \left\|\mathbf{x}^{(0)} - \mathbf{x}\right\|$$

- An we start from the null vector $\mathbf{x^{(0)}=0}$, we have the constraint:

$$\frac{\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\|}{\left\|\mathbf{x}\right\|} < \left\|\mathbf{B}\right\|^{(k)}$$

Error Analysis

- As:

$$\rho(B) = \max_{1 \le i \le n} |\lambda_i| \le \|B\|$$

- We can write:

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}\right\| \le \rho(\mathbf{B})^k \cdot \left\|\mathbf{x}^{(0)} - \mathbf{x}\right\|$$

- Suppose that we know the spectral radius and we want an estimation of the number of iterations needed to obtain a certain relative error of the order of 10^{-t}

[DΣIM]

Error Analysis

- We can write the relation:

$$\frac{\left\|\mathbf{x}^{(k)}-\mathbf{x}\right\|}{\left\|\mathbf{x}\right\|}<\rho(\mathbf{B})^{k}\leq 10^{-t}$$

- Or alternatively:

$$k\geq\frac{t}{-\log_{10}\rho(\mathbf{B})}$$

- The smaller the spectral radius, the faster will be our method.

Error Analysis

- It is not possible to decide a priory if the Jacobi method will be faster than the method of Gauss-Seidel.
- Only when the conditions of the theorem Stein-Rosenberg are satisfied it is possible to decide which one will be faster. This theorem is satisfied for the matrices used in the numerical solution of elliptical PDEs

Error Analysis

- ***Stein-Rosenberg theorem:*** Let $\mathbf{B_J}$ be the matrix in the Jacobi iterations and $\mathbf{B_G}$ the matrix in Gauss-Seidel iterations. IF $a_{ij} \leq 0$ for $i \neq j$ and $a_{ii} > 0$ for each $i = 1,2,...,n$ only one of the assertions is true:
  - *a*) $0 < \rho(\mathbf{B_G}) < \rho(\mathbf{B_J}) < 1$
  - *b*) $1 < \rho(\mathbf{B_J}) < \rho(\mathbf{B_G})$
  - *c*) $\rho(\mathbf{B_J}) = \rho(\mathbf{B_G}) = 0$
  - *d*) $\rho(\mathbf{B_J}) = \rho(\mathbf{B_G}) = 1$

Error Analysis

- If one of the methods is convergent the other will be convergent also. In this case is better to use the Gauss Seidel method, as it has a lower spectral radius.
- This is a general result. The faster method will be always the one with smaller spectral radius or with smaller matrix norm for the iteration matrix.

Error Analysis

- The numerical solution of a linear system will not always give exact solutions. This may be due to error in the initial data, both in the matrix or in the independent vector and also due to the rounding errors in the computations.
- We will consider now, how these errors are propagated to the final solution.

Error Analysis

- We will never be able to eliminate the errors due to rounding in the computations but we can consider the errors in the initial data.
- Suppose that we have certain error in the independent term **b**: **δb**, which means that in reality we are using the vector **b+δb**. We will find the solution **x*=x+δx**, satisfying

$$\mathbf{A}\mathbf{x}^{*} = \mathbf{A}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$$

Error Analysis

- This means:

$$\mathbf{Ax} + \mathbf{A}\delta\mathbf{x} = \mathbf{b} + \delta\mathbf{b}$$

- And as **Ax=b** we have

$$\|\delta\mathbf{x}\| = \left\|\mathbf{A}^{-1}\delta\mathbf{b}\right\| \leq \left\|\mathbf{A}^{-1}\right\| \cdot \|\delta\mathbf{b}\|$$

- But as they are consistent norms we can write:

$$\delta\mathbf{x} = \mathbf{A}^{-1}\delta\mathbf{b}$$

Error Analysis

- Now, dividing by $\|\mathbf{x}\|$ we have

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|\mathbf{A}^{-1}\|\|\delta\mathbf{b}\|}{\|\mathbf{x}\|} = \frac{\|\mathbf{A}\|\|\mathbf{A}^{-1}\|\|\delta\mathbf{b}\|}{\|\mathbf{A}\|\|\mathbf{x}\|}$$

$$\leq \|\mathbf{A}\|\|\mathbf{A}^{-1}\|\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}$$

- As $\|\mathbf{b}\| = \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \ \|\mathbf{x}\|$

Error Analysis

- Note that the error in the initial values is amplified by the factor $\mu(\mathbf{A}) = \|\mathbf{A}\| \, \|\mathbf{A}^{-1}\|$ which is called the ***matrix condition number***.
- This factor is always greater or equal to the unit

$$\boldsymbol{\mu(A) = \|A\|\|A^{-1}\| \geq \|AA^{-1}\| = \|I\| = 1}$$

Error Analysis

- This is a very important factor in the solution of linear systems, not only in the solution using iterative methods. It will always determine the precision that we can reach when solving the system.
- Linear problems with a matrix with a high condition number, $\mu(\mathbf{A})>>1$, will be difficult to resolve numerically. We say that these are ***ill conditioned systems.***

Error Analysis

- A great condition number indicates that some of the straight lines that define the system are nearly parallel. This makes difficult to find a good numerical solution.
- Even as the lines will cut in a single point in the space, numerically we do not have this infinite precision and we can consider the lines as bands with some thickness determined by the rounding numerical precision.