Numerical Analysis

Lecture 5: The Solution of Linear Systems

(Iterative Solver)

Instructor: Prof. Qiwei Zhan
Zhejiang University

Outline

Vector Norms

- Let $\mathbb{R}^n$ denote the set of all n-dimensional column vectors with real-number components.
- To define a distance in $\mathbb{R}^n$ we use the notion of a norm, which is the generalization of the absolute value on $\mathbb{R}$, the set of real numbers.

Definition: Vector Norm

A vector norm on $\mathbb{R}^n$ is a function, $||\cdot||$, from $\mathbb{R}^n$ into $\mathbb{R}$ with the following properties:

1. $||\mathbf{x}|| \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$
2. $||\mathbf{x}|| = 0$ iff $\mathbf{x} = 0$
3. $||\alpha\mathbf{x}|| = |\alpha|||\mathbf{x}||$ for all $\alpha \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$
4. $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ for all $||\mathbf{x}||, ||\mathbf{y}|| \in \mathbb{R}^n$.

Matrix Norms

Definition: Matrix Norm

A matrix norm on the set of all $n \times n$ matrices is a real-valued function, $||\cdot||$, defined on this set, satisfying for all $n \times n$ matrices A and B and all real numbers α:

1. $||A|| \geq 0$
2. $||A|| = 0$, iff A is O, the matrix with all 0 entries
3. $||\alpha A|| = |\alpha|||A||$
4. $||A + B|| \leq ||A|| + ||B||$
5. $||AB|| \leq ||A||||B||$

The distance between $n \times n$ matrices A and B with respect to this matrix norm is $||A - B||$.

Matrix Norms

Frobenius Norm

$$||A||_F = \sqrt{\sum_{i=1}^{n}\sum_{i=1}^{n}|a_{ij}|^2}$$

The p-norms

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

Matrix Norms

Frobenius Norm

$$||A||_F = \sqrt{\sum_{i=1}^{n}\sum_{i=1}^{n}|a_{ij}|^2}$$

The p-norms

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

- It is clear that $||A||_p$ is the p-norm of the largest vector obtained by applying A to a unit p-norm vector:

$$||A||_p = \sup_{\mathbf{x}\neq 0} \left|\left| A\left(\frac{\mathbf{x}}{||\mathbf{x}||_p}\right)\right|\right|_p = \max_{||\mathbf{x}||_p=1} ||A\mathbf{x}||_p$$

Matrix Norms

Theorem: Matrix Norm

If $||\cdot||$ is a vector norm on $\mathbb{R}^n$, then

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

is a matrix norm.

- Matrix norms defined by vector norms are called the natural, or induced, matrix norm associated with the vector norm.
- In this course, all matrix norms will be assumed to be natural matrix norms unless specified otherwise.

Matrix Norms

Corollary

For any vector $\mathbf{x} \neq \mathbf{0}$, matrix A, and any natural norm $||\cdot||$, we have

$$||A\mathbf{x}|| \leq ||A|| \cdot ||\mathbf{x}||$$

- The measure given to a matrix under a natural norm describes how the matrix stretches unit vectors relative to that norm. The maximum stretch is the norm of the matrix. The matrix norms we will consider have the forms

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

and

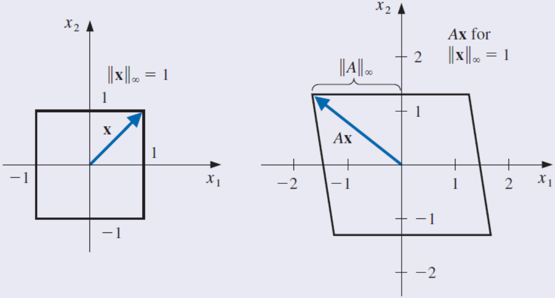
$$||A||_2 = \max_{||\mathbf{x}||_2 = 1} ||A\mathbf{x}||_2$$

Matrix Norms

Illustration

An illustration of $||A||_\infty = \max_{||\mathbf{x}||_\infty = 1} ||A\mathbf{x}||_\infty$ when $n = 2$ for the matrix

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

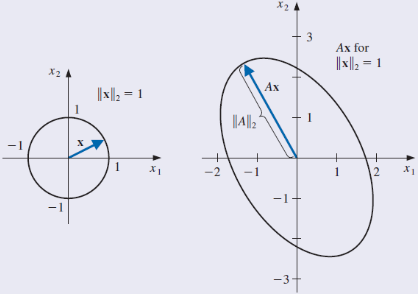


Matrix Norms

Illustration

An illustration of $||A||_2 = \max_{||\mathbf{x}||_2=1} ||A\mathbf{x}||_2$ when $n = 2$ for the matrix

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



Matrix Norms

Theorem

If $A = (a_{ij})$ *is an* $n \times n$ *matrix, then*

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

Matrix Norms

Proof (1/3)

- First we show that $||A||_\infty \leq \max\limits_{1 \leq i \leq n} \sum\limits_{j=1}^{n} |a_{ij}|$.

Proof (1/3)

- First we show that $||A||_\infty \leq \max\limits_{1\leq i\leq n} \sum\limits_{j=1}^{n} |a_{ij}|$.
- Let $\mathbf{x}$ be an n-dimensional vector with $1 = ||\mathbf{x}||_\infty = \max\limits_{1\leq i\leq n} |x_i|$. Since $A\mathbf{x}$ is also an n-dimensional vector,

$$||A\mathbf{x}||_\infty = \max_{1\leq i\leq n} |(A\mathbf{x})_i| = \max_{1\leq i\leq n} \left| \sum_{j=1}^{n} a_{ij}x_j \right| \leq \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}| \max_{1\leq i\leq n} |x_j|.$$

But $\max\limits_{1\leq i\leq n} |x_j| = ||\mathbf{x}||_\infty = 1$, so

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

Matrix Norms

Proof (2/3)

- and consequently,

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

Matrix Norms

Proof (2/3)

- and consequently,

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

- Now we will show the opposite inequality. Let p be an integer with

$$\sum_{j=1}^{n} |a_{pj}| = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|,$$

  and $\mathbf{x}$ be the vector with components

$$x_j = \begin{cases} 1, & \text{if } a_{pj} \ge 0 \\ -1, & \text{if } a_{pj} < 0 \end{cases}$$

Matrix Norms

Proof (3/3)

Then $||\mathbf{x}||_\infty = 1$ and $a_{pj}x_j = |a_{pj}|$, for all $j = 1, 2, \cdots, n$, so

$$||A\mathbf{x}||_\infty = \max_{1\le i\le n} \left|\sum_{j=1}^{n} a_{ij}x_j\right| \ge \left|\sum_{j=1}^{n} a_{pj}x_j\right| = \left|\sum_{j=1}^{n} a_{pj}\right| = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|$$

This result implies that

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

Putting together, we get

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

Matrix Norms

Example

Determine $||A||_\infty$ for the matrix

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & -1 \\ 5 & -1 & 1 \end{bmatrix}$$

Outline

Eigenvalues and Eigenvectors

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

Eigenvalues and Eigenvectors

We have

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$$

Definition (Characteristic Polynomial)

If $\mathbf{A}$ is a square matrix, the characteristic polynomial of $\mathbf{A}$ is defined by

$$p(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$$

Eigenvalues and Eigenvectors

Definition (Characteristic Polynomial)

If $\mathbf{A}$ is a square matrix, the characteristic polynomial of $\mathbf{A}$ is defined by

$$p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$$

Comments

- p is an nth-degree polynomial and, consequently, has at most n distinct zeros, some of which might be complex.
- If λ is a zero of p, then, since $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$, we can prove that the linear system defined by

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}$$

  has a solution with $\mathbf{x} \neq \mathbf{0}$.

Eigenvalues and Eigenvectors

Finding the Eigenvalues & Eigenvectors

- To determine the eigenvalues of a matrix, we can use the fact that λ is an eigenvalue of A if and only if

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0$$

- Once an eigenvalue λ has been found, a corresponding eigenvector $\mathbf{x} \neq \mathbf{0}$ is determined by solving the system

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$$

Eigenvalues and Eigenvectors

Example

Show that there are no nonzero vectors $\mathbf{x}$ in $\mathbb{R}^2$ with $\mathbf{Ax}$ parallel to $\mathbf{x}$ if

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

Solution (1/2)

The eigenvalues of A are the solutions to the characteristic polynomial

$$0 = \det(\mathbf{A} - \lambda\mathbf{I}) = \det \begin{bmatrix} -\lambda & 1 \\ -1 & -\lambda \end{bmatrix} = \lambda^2 + 1$$

so the eigenvalues of A are the complex numbers $\lambda_1 = i$ and $\lambda_2 = -i$.

Eigenvalues and Eigenvectors

Solution (2/2)

- A corresponding eigenvector **x** for λ_1 needs to satisfy

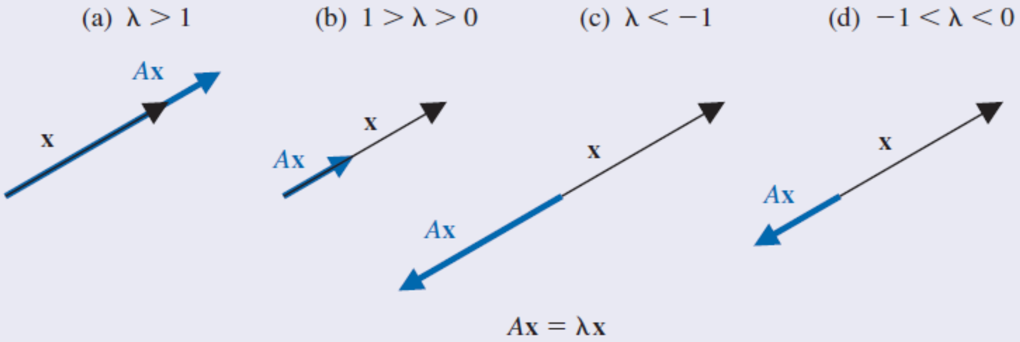
$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -ix_1 + x_2 \\ -x_1 - ix_2, \end{bmatrix}$$

  this is, $0 = -ix_1 + x_2$, so $x_2 = ix_1$, and $0 = -x_1 - ix_2$.
- Hence if x is an eigenvector of A, then exactly one of its components is real and the other is complex.

As a consequence, there are no nonzero vectors **x** in $\mathbb{R}^2$ with **Ax** parallel to **x**.

Eigenvalues and Eigenvectors

Geometric Interpretation of λ



Definition (Spectral Radius)

The spectral radius $\rho(A)$ of a matrix A is defined by

$$\rho(A) = \max |\lambda|,$$

where λ is an eigenvalue of A. For complex $\lambda = \alpha + \beta i$, we define $|\lambda| = (\alpha^2 + \beta^2)^{1/2}$.

Example

For the matrix

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 2 \\ 1 & -1 & 4 \end{bmatrix}$$

note that

$$\rho(A) = \max\{2, 3\} = 3$$

Eigenvalues and Eigenvectors

Theorem

If $\mathbf{A}$ *is a* $n \times n$ *matrix, then*

1. $||\mathbf{A}||_2 = [\rho(\mathbf{A}^T\mathbf{A})]^{1/2}$
2. $\rho(\mathbf{A}) \leq ||\mathbf{A}||$, *for any natural norm* $||.||$

Eigenvalues and Eigenvectors

Proof (i)

Step 1: Let $\mu = [\rho(\mathbf{A}^T\mathbf{A})]^{1/2}$,

$$||A\mathbf{x}||_2^2 = \mathbf{x}^T A^T A\mathbf{x} \leq \mu^2 \mathbf{x}^T\mathbf{x}$$

Thus,

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

Step 2: If $\mathbf{u}$ is an eigenvector of $A^T A$ corresponding to μ^2, then

$$\mathbf{u}^T A^T A\mathbf{u} = \mu^2 \mathbf{u}^T\mathbf{u},$$

which shows that equality holds.

Eigenvalues and Eigenvectors

Proof (ii)

Suppose λ is an eigenvalue of A with eigenvector $\mathbf{x}$ and $||\mathbf{x}|| = 1$. Then $A\mathbf{x} = \lambda\mathbf{x}$ and

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

Thus

$$\rho(\mathbf{A}) = max|\lambda| \leq ||A||$$

Eigenvalues and Eigenvectors

Example

Determine the L_2 norm of

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

Eigenvalues and Eigenvectors

Solution (1/3)

We first need the eigenvalues of A^tA, where

$$A^tA = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ -1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 6 & 4 \\ -1 & 4 & 5 \end{bmatrix}$$

Eigenvalues and Eigenvectors

Solution (2/3)

If

$$\begin{aligned} 0 = \det(A^tA - \lambda I) &= \det \begin{bmatrix} 3-\lambda & 2 & -1 \\ 2 & 6-\lambda & 4 \\ -1 & 4 & 5-\lambda \end{bmatrix} \\ &= -\lambda^3 + 14\lambda^2 - 42\lambda \\ &= -\lambda(\lambda^2 - 14\lambda + 42) \end{aligned}$$

then $\lambda = 0$ or $\lambda = 7 \pm \sqrt{7}$.

Solution (3/3)

By part (i) of the theorem, we have

$$
\begin{aligned}
||A||_2 &= \sqrt{\rho(A^t A)} \\
&= \sqrt{\max\{0, 7-\sqrt{7}, 7+\sqrt{7}\}} \\
&= \sqrt{7+\sqrt{7}} \\
\approx 3.106
\end{aligned}
$$

Outline

Convergent Matrix

Convergent Matrix

We call an $n \times n$ matrix A convergent if

$$\lim_{k\to\infty} (A^k)_{ij} = 0$$

for each $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, n$.

Example

$$A = \begin{bmatrix} \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

Convergent Matrix

Theorem

The following statements are equivalent

1. *A is a convergent matrix.*
2. $\lim_{n\to\infty} ||A^n|| = 0$, *for some natural norm.*
3. $\lim_{n\to\infty} ||A^n|| = 0$, *for all natural norms.*
4. $\rho(A) < 1$.
5. $\lim_{n\to\infty} A^n\mathbf{x} = \mathbf{0}$, *for every* $\mathbf{x}$.

Outline

Iterative Technique

An iterative technique to solve the $n \times n$ linear system $A\mathbf{x} = \mathbf{b}$ starts with an initial approximation $\mathbf{x}^{(0)}$ to the solution $\mathbf{x}$ and generates a sequence of vectors $\{\mathbf{x}^{(k)}\}$ that converges to $\mathbf{x}$.

Iterative Technique

An iterative technique to solve the $n \times n$ linear system $A\mathbf{x} = \mathbf{b}$ starts with an initial approximation $\mathbf{x}^{(0)}$ to the solution $\mathbf{x}$ and generates a sequence of vectors $\{\mathbf{x}^{(k)}\}$ that converges to $\mathbf{x}$.

- Iterative techniques are seldom used for solving linear systems of small dimension since the time required for sufficient accuracy exceeds that required for direct techniques such as Gaussian elimination.

Iterative Technique

Iterative Technique

An iterative technique to solve the $n \times n$ linear system $A\mathbf{x} = \mathbf{b}$ starts with an initial approximation $\mathbf{x}^{(0)}$ to the solution $\mathbf{x}$ and generates a sequence of vectors $\{\mathbf{x}^{(k)}\}$ that converges to $\mathbf{x}$.

- Iterative techniques are seldom used for solving linear systems of small dimension since the time required for sufficient accuracy exceeds that required for direct techniques such as Gaussian elimination.
- For large systems with a high percentage of 0 entries, however, these techniques are efficient in terms of both computer storage and computation.

The Jacobi Iterative Method

Basic Idea

Convert

$$A\mathbf{x} = \mathbf{b}$$

into an equivalent system of the form

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

and approximate solution by computing

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

The Jacobi Iterative Method

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$$\begin{aligned} A &= \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ddots & 0 & a_{nn} \end{bmatrix} - \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ -a_{21} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ -a_{n1} & \cdots & -a_{n,n1} & 0 \end{bmatrix} - \begin{bmatrix} 0 & -a_{12} & \cdots & -a_{1n} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -a_{n-1,n} \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \\ &= D - L - U \end{aligned}$$

Formalize it!!

The Jacobi Iterative Method

The equation

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

is then transformed into

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

and, if D^{-1} exists, that is, if $a_{ii} \neq 0$ for each i, then

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

This results in the matrix form of the Jacobi iterative technique:

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

The Jacobi Iterative Method

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

The Jacobi Iterative Method

For each $k \geq 1$, generate the components $x_i^{(k)}$ of $\mathbf{x}^{(k)}$ from the components of $\mathbf{x}^{(k-1)}$ by

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

for $i = 1, 2, \cdots, n$.

The Jacobi Iterative Method

Example

The linear system $A\mathbf{x} = \mathbf{b}$ given by

$$\begin{array}{rrcrcrcrcr}
E_1: & 10x_1 & - & x_2 & + & 2x_3 & & & = & 6 \\
E_2: & -x_1 & + & 11x_2 & - & x_3 & + & 3x_4 & = & 25 \\
E_3: & 2x_1 & - & x_2 & + & 10x_3 & - & x_4 & = & -11 \\
E_4: & & & 3x_2 & - & x_3 & + & 8x_4 & = & 15
\end{array}$$

has the unique solution $\mathbf{x} = (1, 2, -1, 1)^t$. Use Jacobi's iterative technique to find approximations $\mathbf{x}^{(k)}$ to $\mathbf{x}$ starting with $\mathbf{x}^{(0)} = (0, 0, 0, 0)^t$ until

$$\frac{||\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}||_\infty}{||\mathbf{x}^{(k)}||_\infty} < 10^{-3}$$

The Jacobi Iterative Method

Example: Solution (1/4)

We first solve equation E_i for x_i, for each $i = 1, 2, 3, 4$, to obtain

$$
\begin{aligned}
x_1 &= \phantom{-\frac{1}{5}x_1 +} \frac{1}{10}x_2 - \frac{1}{5}x_3 \phantom{+ \frac{1}{10}x_4} + \frac{3}{5} \\
x_2 &= \frac{1}{11}x_1 \phantom{+ \frac{1}{10}x_2} + \frac{1}{11}x_3 - \frac{3}{11}x_4 + \frac{25}{11} \\
x_3 &= -\frac{1}{5}x_1 + \frac{1}{10}x_2 \phantom{+ \frac{1}{8}x_3} + \frac{1}{10}x_4 - \frac{11}{10} \\
x_4 &= \phantom{-\frac{1}{5}x_1} - \frac{3}{8}x_2 + \frac{1}{8}x_3 \phantom{+ \frac{1}{10}x_4} + \frac{15}{8}
\end{aligned}
$$

The Jacobi Iterative Method

Example: Solution (2/4)

From the initial approximation $\mathbf{x}^{(0)} = (0, 0, 0, 0)^t$ we have $\mathbf{x}^{(1)}$ given by

$$
\begin{aligned}
x_1^{(1)} &= \frac{1}{10}x_2^{(0)} - \frac{1}{5}x_3^{(0)} + \frac{3}{5} = 0.6000 \\
x_2^{(1)} &= \frac{1}{11}x_1^{(0)} + \frac{1}{11}x_3^{(0)} - \frac{3}{11}x_4^{(0)} + \frac{25}{11} = 2.2727 \\
x_3^{(1)} &= -\frac{1}{5}x_1^{(0)} + \frac{1}{10}x_2^{(0)} + \frac{1}{10}x_4^{(0)} - \frac{11}{10} = -1.1000 \\
x_4^{(1)} &= -\frac{3}{8}x_2^{(0)} + \frac{1}{8}x_3^{(0)} + \frac{15}{8} = 1.8750
\end{aligned}
$$

The Jacobi Iterative Method

Example: Solution (3/4)

Additional iterates, $\mathbf{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, x_3^{(k)}, x_4^{(k)})^t$, are generated in a similar manner and are summarized as follows:

k	0	1	2	3	4		10
$x_1^{(k)}$	0.0	0.6000	1.0473	0.9326	1.0152	$\cdots$	1.0001
$x_2^{(k)}$	0.0	2.2727	1.7159	2.053	1.9537	$\cdots$	1.9998
$x_3^{(k)}$	0.0	−1.1000	−0.8052	−1.0493	−0.9681	$\cdots$	−0.9998
$x_4^{(k)}$	0.0	1.8750	0.8852	1.1309	0.9739	$\cdots$	0.9998

The Jacobi Iterative Method

Example: Solution (4/4)

The process was stopped after 10 iterations because

$$\frac{\|\mathbf{x}^{(10)} - \mathbf{x}^{(9)}\|_\infty}{\|\mathbf{x}^{(10)}\|_\infty} = \frac{8.0 \times 10^{-4}}{1.9998} < 10^{-3}$$

In fact, $\|\mathbf{x}^{(10)} - \mathbf{x}\|_\infty = 0.0002$.

The Jacobi Iterative Algorithm

INPUT the number of equations and unknowns n; the entries a_{ij}, $1 \leq i, j \leq n$ of the matrix A; the entries b_i, $1 \leq i \leq n$ of b; the entries XO_i, $1 \leq i \leq n$ of $\mathbf{XO} = \mathbf{x}^{(0)}$; tolerance TOL; maximum number of iterations N.

OUTPUT the approximate solution $x_1, \ldots, x_n$ or a message that the number of iterations was exceeded.

Step 1 Set $k = 1$.

Step 2 While ($k \leq N$) do Steps 3–6.

Step 3 For $i = 1, \ldots, n$

$$\text{set } x_i = \frac{1}{a_{ii}}\left[-\sum_{\substack{j=1 \\ j \neq i}}^{n} (a_{ij} XO_j) + b_i\right].$$

Step 4 If $||\mathbf{x} - \mathbf{XO}|| < TOL$ then OUTPUT $(x_1, \ldots, x_n)$;
(The procedure was successful.)
STOP.

Step 5 Set $k = k + 1$.

The Jacobi Iterative Algorithm

Comments

- Step 3 of the algorithm requires that $a_{ii} \neq 0$, for each $i = 1, 2, \cdots, n$. If one of the a_{ii} entries is 0 and the system is nonsingular, a reordering of the equations can be performed so that no $a_{ii} = 0$.
- To speed convergence, the equations should be arranged so that a_{ii} is as large as possible.
- Another possible stopping criterion in Step 4 is to iterate until

$$\frac{||\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}||}{||\mathbf{x}^{(k)}||}$$

  is smaller than some prescribed tolerance. For this purpose, any convenient norm can be used, the usual choice is the l_∞ norm.

The Gauss-Seidel Method

The Jacobi Iterative Method

For each $k \geq 1$, generate the components $x_i^{(k)}$ of $\mathbf{x}^{(k)}$ from the components of $\mathbf{x}^{(k-1)}$ by

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

for $i = 1, 2, \cdots, n$.

The Gauss-Seidel Method

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

for $i = 1, 2, \cdots, n$. Recall Fixed Points Acceleration (Lecture 3, P 29).

The Gauss-Seidel Method

The Gauss-Seidel Method

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

for $i = 1, 2, \cdots, n$.

With the definitions of D, L, and U, we have the Gauss-Seidel method represented by

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

and

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

The Gauss-Seidel Method

Example

Use the Gauss-Seidel iterative technique to find approximate solutions to

$$\begin{aligned}
10x_1 - x_2 + 2x_3 &= 6, \\
-x_1 + 11x_2 - x_3 + 3x_4 &= 25, \\
2x_1 - x_2 + 10x_3 - x_4 &= -11, \\
3x_2 - x_3 + 8x_4 &= 15
\end{aligned}$$

starting with $\mathbf{x} = (0, 0, 0, 0)^t$ and iterating until

$$\frac{\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_\infty}{\|\mathbf{x}^{(k)}\|_\infty} < 10^{-3}.$$

The Gauss-Seidel Method

Example

For the Gauss-Seidel method we write the system, for each $k = 1, 2, \ldots$ as

$$
\begin{aligned}
x_1^{(k)} &= \qquad\qquad \frac{1}{10}x_2^{(k-1)} - \frac{1}{5}x_3^{(k-1)} \qquad\qquad + \frac{3}{5}, \\
x_2^{(k)} &= \frac{1}{11}x_1^{(k)} \qquad\qquad + \frac{1}{11}x_3^{(k-1)} - \frac{3}{11}x_4^{(k-1)} + \frac{25}{11}, \\
x_3^{(k)} &= -\frac{1}{5}x_1^{(k)} + \frac{1}{10}x_2^{(k)} \qquad\qquad + \frac{1}{10}x_4^{(k-1)} - \frac{11}{10}, \\
x_4^{(k)} &= \qquad\qquad - \frac{3}{8}x_2^{(k)} + \frac{1}{8}x_3^{(k)} \qquad\qquad + \frac{15}{8}.
\end{aligned}
$$

The Gauss-Seidel Method

Example

When $\mathbf{x}^{(0)} = (0, 0, 0, 0)^t$, we have $\mathbf{x}^{(1)} = (0.6000, 2.3272, -0.9873, 0.8789)^t$.

k	0	1	2	3	4	5
$x_1^{(k)}$	0.0000	0.6000	1.030	1.0065	1.0009	1.0001
$x_2^{(k)}$	0.0000	2.3272	2.037	2.0036	2.0003	2.0000
$x_3^{(k)}$	0.0000	−0.9873	−1.014	−1.0025	−1.0003	−1.0000
$x_4^{(k)}$	0.0000	0.8789	0.9844	0.9983	0.9999	1.0000

The Gauss-Seidel Method

Remarks

- It is almost always true that the Gauss-Seidel method is superior to the Jacobi method.
- But there are linear systems for which the Jacobi method converges and the Gauss-Seidel method does not. See Chapter 7, Problem 9 & 10.
- um... We need a further discussion!!

General Iteration Methods

Convergence Issue

To study the convergence of general iteration techniques, we need to analyze the formula

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

for each $k = 1, 2, \cdots$, where $\mathbf{x}^{(0)}$ is arbitrary.

General Iteration Methods

Lemma

If the spectral radius satisfies $\rho(T) < 1$, then $(I - T)^{-1}$ exists, and

$$(I - T)^{-1} = I + T + T^2 + \cdots = \sum_{j=0}^{\infty} T^j.$$

General Iteration Methods

Lemma

If the spectral radius satisfies $\rho(T) < 1$, then $(I - T)^{-1}$ exists, and

$$(I - T)^{-1} = I + T + T^2 + \cdots = \sum_{j=0}^{\infty} T^j.$$

Proof (1/2)

- Because $T\mathbf{x} = \lambda\mathbf{x}$ is true precisely when $(I - T)\mathbf{x} = (1 - \lambda)\mathbf{x}$, we have λ as an eigenvalue of T precisely when $1 - \lambda$ is an eigenvalue of $I - T$.
- But $|\lambda| \leq \rho(T) < 1$, so $\lambda = 1$ is not an eigenvalue of T, and 0 cannot be an eigenvalue of $I - T$.
- Hence, $(I - T)^{-1}$ exists.

General Iteration Methods

Proof (2/2)

Let

$$S_m = I + T + T^2 + \cdots + T^m$$

then

$$(I - T)S_m = (I + T + T^2 + \cdots + T^m) - (T + T^2 + \cdots + T^m + T^{m+1}) = I - T^{m+1}$$

and, since T is convergent, which implies

$$(I - T) \lim_{m\to\infty} S_m = \lim_{m\to\infty} (I - T)S_m = \lim_{m\to\infty} (I - T^{m+1}) = I$$

Thus,

$$(I - T)^{-1} = \lim_{m\to\infty} S_m = I + T + T^2 + \cdots = \sum_{j=0}^{\infty} T^j.$$

Convergent Matrix

Theorem

The following statements are equivalent

1. *A is a convergent matrix.*
2. $\lim_{n\to\infty} ||A^n|| = 0$, *for some natural norm.*
3. $\lim_{n\to\infty} ||A^n|| = 0$, *for all natural norms.*
4. $\rho(A) < 1$.
5. $\lim_{n\to\infty} A^n\mathbf{x} = \mathbf{0}$, *for every* $\mathbf{x}$.

General Iteration Methods

Theorem

For any $\mathbf{x}^{(0)} \in \mathbb{R}^n$, *the sequence* $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ *defined by*

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

for each $k \geq 1$, *converges to the unique solution of* $\mathbf{x} = T\mathbf{x} + \mathbf{c}$ *iff* $\rho(T) < 1$.

Proof(1/5)

First assume that $\rho(T) < 1$. Then,

$$\begin{aligned} \mathbf{x}^{(k)} &= T\mathbf{x}^{(k-1)} + \mathbf{c} \\ &= T(T\mathbf{x}^{(k-2)} + \mathbf{c}) + \mathbf{c} \\ &= T^2\mathbf{x}^{(k-2)} + (T + I)\mathbf{c} \\ &\vdots \\ &= T^k\mathbf{x}^{(0)} + (T^{k-1} + \cdots + T + I)\mathbf{c}. \end{aligned}$$

Because $\rho(T) < 1$, which implies that T is convergent, and

$$\lim_{k\to\infty} T^k\mathbf{x}^{(0)} = \mathbf{0}.$$

Proof(2/5)

The previous lemma implies that

$$\begin{aligned}
\lim_{k\to\infty} \mathbf{x}^{(k)} &= \lim_{k\to\infty} T^k \mathbf{x}^{(0)} + \left(\sum_{j=0}^{\infty} T^j\right) \mathbf{c} \\
&= \mathbf{0} + (I - T)^{-1}\mathbf{c} \\
&= (I - T)^{-1}\mathbf{c}
\end{aligned}$$

Hence, the sequence $\{\mathbf{x}^{(k)}\}$ converges to the vector $\mathbf{x} = (I - T)^{-1}\mathbf{c}$ and $\mathbf{x} = T\mathbf{x} + \mathbf{c}$.

General Iteration Methods

Proof(3/5)

To prove the converse, we will show that for any $\mathbf{z} \in \mathbb{R}^n$, we have $\lim_{k\to\infty} T^k \mathbf{z} = \mathbf{0}$, which is equivalent to $\rho(T) < 1$.

- Let $\mathbf{z}$ be an arbitrary vector, and $\mathbf{x}$ be the unique solution to $\mathbf{x} = T\mathbf{x} + \mathbf{c}$.
- Define $\mathbf{x}^{(0)} = \mathbf{x} - \mathbf{z}$, and for $k \geq 1$, $\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}$.
- Then $\{\mathbf{x}^{(k)}\}$ converges to $\mathbf{x}$.

General Iteration Methods

Proof(4/5)

Also,

$$\mathbf{x} - \mathbf{x}^{(k)} = (T\mathbf{x} + \mathbf{c}) - (T\mathbf{x}^{(k-1)} + \mathbf{c}) = T\left(\mathbf{x} - \mathbf{x}^{(k-1)}\right)$$

so

$$\begin{aligned}
\mathbf{x} - \mathbf{x}^{(k)} &= T\left(\mathbf{x} - \mathbf{x}^{(k-1)}\right) \\
&= T^2\left(\mathbf{x} - \mathbf{x}^{(k-2)}\right) \\
&= \vdots \\
&= T^k\left(\mathbf{x} - \mathbf{x}^{(0)}\right) \\
&= T^k\mathbf{z}
\end{aligned}$$

Proof(5/5)

Hence

$$\begin{aligned}\lim_{k\to\infty} T^k \mathbf{z} &= \lim_{k\to\infty} T^k \left(\mathbf{x} - \mathbf{x}^{(0)}\right) \\ &= \lim_{k\to\infty} \left(\mathbf{x} - \mathbf{x}^{(k)}\right) \\ &= \mathbf{0}\end{aligned}$$

But $\mathbf{z} \in \mathbb{R}^n$ was arbitrary, so T is convergent and $\rho(T) < 1$.

General Iteration Methods

Corollary

If $||T|| < 1$ *for any natural matrix norm and* $\mathbf{c}$ *is a given vector, then the sequence* $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ *defined by* $\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}$ *converges, for any* $\mathbf{x}^{(0)} \in \mathbb{R}^n$*, to a vector* $\mathbf{x} \in \mathbb{R}^n$*, with* $\mathbf{x} = T\mathbf{x} + \mathbf{c}$*, and the following error bounds hold:*

1. $||\mathbf{x} - \mathbf{x}^{(k)}|| \leq ||T||^k ||\mathbf{x}^{(0)} - \mathbf{x}||$
2. $||\mathbf{x} - \mathbf{x}^{(k)}|| \leq \frac{||T||^k}{1-||T||} ||\mathbf{x}^{(1)} - \mathbf{x}^{(0)}||$.

Refer to Page 60

Convergence of Jacobi Methods

We have seen that the Jacobi and Gauss-Seidel iterative techniques can be written

$$\mathbf{x}^{(k)} = T_j\mathbf{x}^{(k-1)} + \mathbf{c}_j \ \text{ and } \ \mathbf{x}^{(k)} = T_g\mathbf{x}^{(k-1)} + \mathbf{c}_g$$

using the matrices

$$T_j = D^{-1}(L+U) \ \text{ and } \ T_g = (D-L)^{-1}U$$

If $\rho(T_j)$ or $\rho(T_g)$ is less than 1, then the corresponding sequence $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ will converge to the solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$.

General Iteration Methods

Convergence of Jacobi Methods

For example, the Jacobi scheme has

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

and, if $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ converges to $\mathbf{x}$, then

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

This implies that

$$D\mathbf{x} = (L+U)\mathbf{x} + \mathbf{b} \;\text{ and }\; (D-L-U)\mathbf{x} = \mathbf{b}.$$

Since $D-L-U = A$, the solution $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{b}$.

General Iteration Methods

Theorem

If A is strictly diagonally dominant, then for any choice of $\mathbf{x}^{(0)}$, both the Jacobi and Gauss-Seidel methods give sequences $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ that converge to the unique solution of $A\mathbf{x} = \mathbf{b}$.

Remarks

- No general results exist to tell which of the two techniques, Jacobi or Gauss-Seidel, will be most successful for an arbitrary linear system.
- In special cases, however, the answer is known, as is demonstrated in the following theorem.

General Iteration Methods

Theorem (Stein-Rosenberg)

If $a_{ij} \leq 0$, for each $i \neq j$ and $a_{ii} > 0$, for each $i = 1, 2, \cdots, n$, then one and only one of the following statements holds:

1. $0 \leq \rho(T_g) < \rho(T_j) < 1$;
2. $1 < \rho(T_j) < \rho(T_g)$;
3. $\rho(T_j) = \rho(T_g) = 0$;
4. $\rho(T_j) = \rho(T_g) = 1$;

Error Bounds

Residual Vector

Suppose $\tilde{\mathbf{x}} \in \mathbb{R}^n$ is an approximation to the solution of the linear system defined by $A\mathbf{x} = \mathbf{b}$. The residual vector for $\tilde{\mathbf{x}}$ with respect to this system is $\mathbf{r} = \mathbf{b} - A\tilde{\mathbf{x}}$.

Error Bounds

Residual Vector

Suppose $\tilde{\mathbf{x}} \in \mathbb{R}^n$ is an approximation to the solution of the linear system defined by $A\mathbf{x} = \mathbf{b}$. The residual vector for $\tilde{\mathbf{x}}$ with respect to this system is $\mathbf{r} = \mathbf{b} - A\tilde{\mathbf{x}}$.

Condition Number

The condition number of a nonsingular matrix A w.r.t a norm $||\cdot||$ is

$$K(A) = ||A|| \cdot ||A^{-1}||$$

Error Bounds

Conditioning

A matrix A is well-conditioned if $K(A)$ is close to 1, and is ill-conditioned when $K(A)$ is significantly greater than 1.

Remarks

- For any nonsingular matrix A and natural norm $||\cdot||$,

$$1 = ||I|| = ||A \cdot A^{-1}|| \leq ||A|| \cdot ||A^{-1}|| = K(A)$$

- Conditioning refers to the relative security that a small residual vector implies a correspondingly accurate approximate solution.

Error Bounds

Theorem

Suppose that $\tilde{\mathbf{x}}$ *is an approximation to the solution of* $A\mathbf{x} = \mathbf{b}$, A *is a nonsingular matrix, and* $\mathbf{r}$ *is the residual vector for* $\tilde{\mathbf{x}}$. *Then for any natural norm,*

$$||\mathbf{x} - \tilde{\mathbf{x}}|| \leq ||\mathbf{r}|| \cdot ||A^{-1}||$$

and if $\mathbf{x} \neq \mathbf{0}$ *and* $\mathbf{b} \neq \mathbf{0}$,

$$\frac{||\mathbf{x} - \tilde{\mathbf{x}}||}{||\mathbf{x}||} \leq K(A)\frac{||\mathbf{r}||}{||\mathbf{b}||}.$$

Error Bounds

Proof

Since $\mathbf{r} = \mathbf{b} - A\tilde{\mathbf{x}} = A\mathbf{x} - A\tilde{\mathbf{x}}$ and A is nonsingular, we have $x - \tilde{\mathbf{x}} = A^{-1}\mathbf{r}$. Thus,

$$||\mathbf{x} - \tilde{\mathbf{x}}|| = ||A^{-1}\mathbf{r}|| \leq ||A^{-1}|| \cdot ||\mathbf{r}||.$$

Moreover, since $\mathbf{b} = A\mathbf{x}$, we have $||\mathbf{b}|| \leq ||A|| \cdot ||\mathbf{x}||$. So $1/||\mathbf{x}|| \leq ||A||/||\mathbf{b}||$ and

$$\frac{||\mathbf{x} - \tilde{\mathbf{x}}||}{||\mathbf{x}||} \leq ||A|| \cdot ||A^{-1}|| \frac{||\mathbf{r}||}{||\mathbf{b}||} = K(A) \frac{||\mathbf{r}||}{||\mathbf{b}||}.$$

Error Bounds

Theorem

Suppose A is nonsingular and

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

The solution $\tilde{\mathbf{x}}$ *to* $(A + \delta A)\tilde{\mathbf{x}} = \mathbf{b} + \delta\mathbf{b}$ *approximates the solution* $\mathbf{x}$ *of* $A\mathbf{x} = \mathbf{b}$ *with the error estimate*

$$\frac{||\mathbf{x} - \tilde{\mathbf{x}}||}{||\mathbf{x}||} \leq \frac{K(A)||A||}{||A|| - K(A)||\delta A||}\left(\frac{||\delta\mathbf{b}||}{||\mathbf{b}||} + \frac{||\delta A||}{||A||}\right)$$

Error Bounds

Remarks

- If the matrix A is well-conditioned, then small changes in A and $\mathbf{b}$ produce correspondingly small changes in the solution $\mathbf{x}$.
- If, on the other hand, A is ill-conditioned, then small changes in A and $\mathbf{b}$ may produce large changes in $\mathbf{x}$.

Assignments

- Reading Assignment for this class: Chap 7
- Reading Assignment for next class: Chapter 9
- Homework 3.