Numerical Analysis and Computing

Lecture Notes #16
— Matrix Algebra —
Norms of Vectors and Matrices
Eigenvalues and Eigenvectors
Iterative Techniques

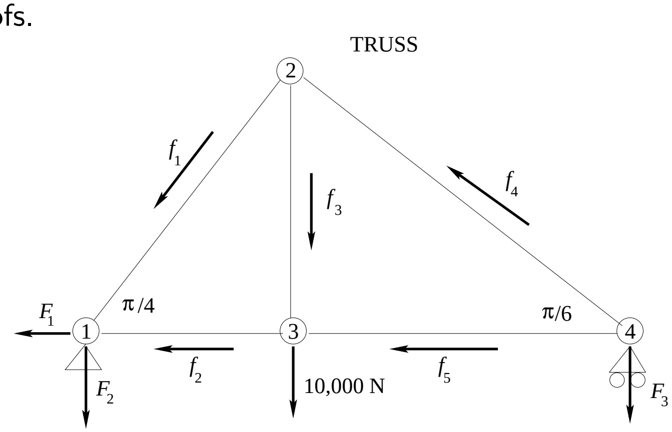
Joe Mahaffy,
⟨mahaffy@math.sdsu.edu⟩

Department of Mathematics
Dynamical Systems Group
Computational Sciences Research Center
San Diego State University
San Diego, CA 92182-7720

Outline

Matrix Application - Truss

Trusses are lightweight structures capable of carrying heavy loads, e.g., roofs.



Physics of Trusses

The truss on the previous slide has the following properties:

1. Fixed at Joint 1
2. Slides at Joint 4
3. Holds a mass of 10,000 N at Joint 3
4. All the Joints are pin joints
5. The forces of tension are indicated on the diagram

Static Equilibrium

At each joint the forces must add to the zero vector.

Joint	Horizontal Force	Vertical Force
1	$-F_1 + \frac{\sqrt{2}}{2} f_1 + f_2 = 0$	$\frac{\sqrt{2}}{2} f_1 - F_2 = 0$
2	$-\frac{\sqrt{2}}{2} f_1 + \frac{\sqrt{3}}{2} f_4 = 0$	$-\frac{\sqrt{2}}{2} f_1 - f_3 - \frac{1}{2} f_4 = 0$
3	$-f_2 + f_5 = 0$	$f_3 - 10,000 = 0$
4	$-\frac{\sqrt{3}}{2} f_4 - f_5 = 0$	$\frac{1}{2} f_4 - F_3 = 0$

This creates an 8×8 linear system with 47 zero entries and 17 nonzero entries.

Sparse matrix – Solve by iterative methods

Earlier Iterative Schemes

Earlier we used iterative methods to find roots of equations

$$f(x) = 0$$

or fixed points of

$$x = g(x)$$

The latter requires $|g'(x)| < 1$ for convergence.

Want to extend to n-dimensional linear systems.

Basic Definitions

Definition

A *Vector norm* on $\mathbb{R}^n$ is a function $||\cdot||$ mapping $\mathbb{R}^n \to \mathbb{R}$ with the following properties:
(i) $||\mathbf{x}|| \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$
(ii) $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
(iii) $||\alpha\mathbf{x}|| = |\alpha|\,||\mathbf{x}||$ for all $\alpha \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$ (scalar multiplication)
(iv) $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (triangle inequality)

Common Norms

The l_1 norm is given by

$$||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i|$$

The l_2 norm or **Euclidean norm** is given by

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

The l_∞ norm or **Max norm** is given by

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

The Euclidean norm represents the usual notion of distance (Pythagorean theorem for distance).

Triangle Inequality

We need to show the triangle inequality for $||\cdot||_2$.

Theorem (Cauchy-Schwarz)

For each $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\mathbf{x}^t\mathbf{y} = \sum_{i=1}^n x_i y_i \le \left(\sum_{i=1}^n x_i^2\right)^{1/2} \left(\sum_{i=1}^n y_i^2\right)^{1/2} = ||\mathbf{x}||_2 \cdot ||\mathbf{y}||_2$$

Triangle Inequality - Proof

Cauchy-Schwarz.

This result gives for each $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\begin{aligned} ||\mathbf{x}+\mathbf{y}||^2 &= \sum_{i=1}^n (x_i + y_i)^2 \\ &= \sum_{i=1}^n x_i^2 + 2\sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 \\ &\le ||\mathbf{x}||^2 + 2||\mathbf{x}||||\mathbf{y}|| + ||\mathbf{y}||^2 \end{aligned}$$

Taking the square root of the above gives the **Triangle Inequality** □

Distance

Definition

If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the l_2 and l_∞ distances between $\mathbf{x}$ and $\mathbf{y}$ is a function $||\cdot||$ mapping $\mathbb{R}^n \to \mathbb{R}$ with the following properties:are defined by

$$\begin{aligned} ||\mathbf{x}-\mathbf{y}||_2 &= \left(\sum_{i=1}^n (x_i - y_i)^2\right)^{1/2} \\ ||\mathbf{x}-\mathbf{y}||_\infty &= max_{1\le i\le n}|x_i - y_i| \end{aligned}$$

Convergence

Also, we need the concept of **convergence** in n-dimensions.

Definition

A sequence of vectors $\{\mathbf{x}^{(k)}\}_{k=1}^\infty$ in $\mathbb{R}^n$ is said to **converge** to $\mathbf{x}$ with respect to norm $||\cdot||$ if given any $\epsilon > 0$ there exists an integer $N(\epsilon)$ such that

$$||\mathbf{x}^{(k)} - \mathbf{x}|| < \epsilon \quad \text{for all} \quad k \ge N(\epsilon).$$

Basic Theorems

Theorem

The sequence of vectors $\{\mathbf{x}^{(k)}\}_{k=1}^{\infty} \to \mathbf{x}$ *in* $\mathbb{R}^n$ *with respect to* $||\cdot||_{\infty}$ *if and only if*

$$\lim_{k\to\infty} x_i^{(k)} = x_i \quad \text{for each} \quad i = 1, 2, ..., n.$$

Theorem

For each $\mathbf{x} \in \mathbb{R}^n$

$$||\mathbf{x}||_{\infty} \leq ||\mathbf{x}||_2 \leq \sqrt{n}||\mathbf{x}||_{\infty}.$$

It can be shown that all norms on $\mathbb{R}^n$ are equivalent.

Matrix Norm

We need to extend our definitions to include matrices.

Definition

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.
(i) $||A|| \geq 0$
(ii) $||A|| = 0$ if and only if A is 0 (all zero entries)
(iii) $||\alpha A|| = |\alpha|\,||A||$ (scalar multiplication)
(iv) $||A + B|| \leq ||A|| + ||B||$ (triangle inequality)
(v) $||AB|| \leq ||A||\,||B||$

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

Natural Matrix Norm

Theorem

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

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

is a matrix norm.

This is the **natural** or **induced matrix norm** associated with the vector norm.

For any $\mathbf{z} \neq \mathbf{0}$, $\mathbf{x} = \frac{\mathbf{z}}{||\mathbf{z}||}$ is a unit vector

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

Matrix Action

The natural norm describes how a matrix stretches unit vectors relative to that norm. (Think eigenvalues!)

Theorem

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

$$||A||_{\infty} = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}| \qquad \text{(largest row sum)}$$

Matrix Mapping

An $n \times m$ matrix is a function that takes m-dimensional vectors into n-dimensional vectors.

For square matrices A, we have $A : \mathbb{R}^n \to \mathbb{R}^n$.

Certain vectors are parallel to $A\mathbf{x}$, so $A\mathbf{x} = \lambda\mathbf{x}$ or $(A - \lambda I)\mathbf{x} = \mathbf{0}$.

These values λ, the **eigenvalues**, are significant for convergence of iterative methods.

Eigenvalues and Eigenvectors

Definition

If A is an $n \times n$ matrix, the characteristic polynomial of A is defined by

$$p(\lambda) = \det(A - \lambda I)$$

Definition

If p is the characteristic polynomial of the matrix A, the zeroes of p are **eigenvalues** (or *characteristic values*) of A. If λ is an eigenvalue of A and $\mathbf{x} \neq \mathbf{0}$ satisfies $(A - \lambda I)\mathbf{x} = \mathbf{0}$, then $\mathbf{x}$ is an **eigenvector** (or *characteristic vector*) of A corresponding to the eigenvalue λ.

Geometry of Eigenvalues and Eigenvectors

If $\mathbf{x}$ is an eigenvector associated with λ, then $A\mathbf{x} = \lambda\mathbf{x}$, so the matrix A takes the vector $\mathbf{x}$ into a scalar multiple of itself.

If λ is real and $\lambda > 1$, then A has the effect of stretching $\mathbf{x}$ by a factor of λ.

If λ is real and $0 < \lambda < 1$, then A has the effect of shrinking $\mathbf{x}$ by a factor of λ.

If $\lambda < 0$, the effects are similar, but the direction of $A\mathbf{x}$ is reversed.

Spectral Radius

The **spectral radius**, $\rho(A)$, provides a valuable measure of the eigenvalues, which helps determine if a numerical scheme will converge.

Definition

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

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

where λ is an eigenvalue of A.

Theorem for $\rho(A)$

Theorem

If A is an $n \times n$ matrix,
(i) $||A||_2 = (\rho(A^t A))^{1/2}$.
(ii) $\rho(A) \leq ||A||$ for any natural norm $||\cdot||$.

Proof of (ii): Let $||\mathbf{x}||$ be a unit eigenvector or A with respect to the eigenvalue λ

$$|\lambda| = |\lambda|\,||\mathbf{x}|| = ||\lambda\mathbf{x}|| = ||A\mathbf{x}|| \leq ||A||\,||\mathbf{x}|| = ||A||.$$

Thus,

$$\rho(A) = \max|\lambda| \leq ||A||.$$

If A is symmetric, then $\rho(A) = ||A||_2$.

Interesting Result for $\rho(A)$

An interesting and useful result: For any matrix A and any $\epsilon > 0$, there exists a natural norm $||\cdot||$ with the property that

$$\rho(A) \leq ||A|| < \rho(A) + \epsilon.$$

So $\rho(A)$ is the greatest lower bound for the natural norms on A.

Convergence of Matrix

Definition

An $n \times n$ matrix A is **convergent** if

$$\lim_{k\to\infty} (A^k)_{ij} = 0, \quad \text{for each} \quad i = 1, ..., n \text{ and } j = 1, ..., n.$$

Example: Consider

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

It is easy to see that

$$A^k = \begin{pmatrix} \frac{1}{2^k} & 0 \\ \frac{k}{2^{k+1}} & \frac{1}{2^k} \end{pmatrix} \to 0.$$

Convergence Theorem for Matrices

Theorem

The following statements are equivalent,
(i) A is a convergent matrix.
(ii) $\lim_{n\to\infty} ||A^n|| = 0$ for some natural norm.
(iii) $\lim_{n\to\infty} ||A^n|| = 0$ for all natural norms.
(iv) $\rho(A) < 1$.
(v) $\lim_{n\to\infty} A^n\mathbf{x} = \mathbf{0}$ for every $\mathbf{x}$.

Introduction – Iterative Methods

Gaussian elimination and other **direct methods** are best for small dimensional systems.

Jacobi and Gauss-Seidel iterative methods were developed in late 18^{th} century to solve

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

by iteration.

Iterative methods are more efficient for large sparse matrix systems, both in computer storage and computation.

Common examples include electric circuits, structural mechanics, and partial differential equations.

Basic Idea – Iterative Scheme

The iterative scheme starts with an initial guess, $\mathbf{x}^{(0)}$ to the linear system

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

Transform this system into the form

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

The iterative scheme becomes

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

Illustrative Example (1 of 4)

Consider the following linear system $A\mathbf{x} = \mathbf{b}$

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

This has the unique solution $\mathbf{x} = (1, 2, -1, 1)^T$.

Illustrative Example (2 of 4)

The previous system is easily converted to the form

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

by solving for each x_i.

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

Illustrative Example (3 of 4)

Thus, the system $A\mathbf{x} = \mathbf{b}$ becomes

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

with

$$T = \begin{bmatrix} 0 & \frac{1}{10} & -\frac{1}{5} & 0 \\ \frac{1}{11} & 0 & \frac{1}{11} & -\frac{3}{11} \\ -\frac{1}{5} & \frac{1}{10} & 0 & \frac{1}{10} \\ 0 & -\frac{3}{8} & \frac{1}{8} & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} \frac{3}{5} \\ \frac{25}{11} \\ -\frac{11}{10} \\ \frac{15}{8} \end{bmatrix}$$

Illustrative Example (4 of 4)

The iterative scheme becomes

$$\begin{aligned} x_1^{(k)} &= \phantom{-\tfrac{1}{5}x_1^{(k-1)}} \quad \tfrac{1}{10}x_2^{(k-1)} - \tfrac{1}{5}x_3^{(k-1)} + \tfrac{3}{5} \\ x_2^{(k)} &= \tfrac{1}{11}x_1^{(k-1)} + \tfrac{1}{11}x_3^{(k-1)} - \tfrac{3}{11}x_4^{(k-1)} + \tfrac{25}{11} \\ x_3^{(k)} &= -\tfrac{1}{5}x_1^{(k-1)} + \tfrac{1}{10}x_2^{(k-1)} + \tfrac{1}{10}x_4^{(k-1)} - \tfrac{11}{10} \\ x_4^{(k)} &= -\tfrac{3}{8}x_2^{(k-1)} + \tfrac{1}{8}x_3^{(k-1)} + \tfrac{15}{8} \end{aligned}$$

With an initial guess of $\mathbf{x} = (0,0,0,0)^T$, we have

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

It takes 10 iterations to converge to a tolerance of 10^{-3}. Error is given by $\frac{||\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}||_\infty}{||\mathbf{x}^{(k)}||_\infty}$

Jacobi Iteration

The example above illustrates the **Jacobi iterative method**.
To solve the linear system

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

Find x_i (for $a_{ii} \neq 0$) by iterating

$$x_i^{(k)} = \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\frac{-a_{ij}x_j^{(k-1)}}{a_{ii}} \right) + \frac{b_i}{a_{ii}} \qquad \text{for } i = 1, ..., n$$

Jacobi Iteration – Matrix Form (1 of 2)

If A is given by

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

Split this into

$$\begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_{nn} \end{bmatrix} - \begin{bmatrix} 0 & \dots & \dots & 0 \\ -a_{21} & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ -a_{n1} & \dots & -a_{n,n-1} & 0 \end{bmatrix} - \begin{bmatrix} 0 & -a_{12} & \dots & -a_{1n} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -a_{n-1,n} \\ 0 & \dots & \dots & 0 \end{bmatrix}$$

or

$$A = D - L - U$$

Jacobi Iteration – Matrix Form (2 of 2)

We are solving $A\mathbf{x} = \mathbf{b}$ with $A = D - L - U$ from above.
It follows that:

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

or

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

The **Jacobi iteration method** becomes

$$\mathbf{x} = T_j\mathbf{x} + \mathbf{c}_j$$

where $T_j = D^{-1}(L + U)$ and $\mathbf{c}_j = D^{-1}\mathbf{b}$.

Notes on Solving $A\mathbf{x} = \mathbf{b}$

If any of the $a_{ii} = 0$ and the matrix A is nonsingular, then the equations can be reordered so that all $a_{ii} \neq 0$.

Convergence (if possible) is accelerated by taking the a_{ii} as large as possible.

Gauss-Seidel Iteration

One possible improvement is that $\mathbf{x}^{(k-1)}$ are used to compute $x_i^{(k)}$.
However, for $i > 1$, the values of $x_1^{(k)}, \dots x_{i-1}^{(k)}$ are already computed and should be improved values.
If we use these updated values in the algorithm we obtain:

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

This modification is called the **Gauss-Seidel iterative method**.

Return to Illustrative Example

The Gauss-Seidel iterative scheme becomes

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

With an initial guess of $\mathbf{x} = (0, 0, 0, 0)^T$, it takes 5 iterations to converge to a tolerance of 10^{-3}.
Again the error is given by

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

Gauss-Seidel Iteration – Matrix Form

With the same definitions as before, $A = D - L - U$, we can write the equation $A\mathbf{x} = \mathbf{b}$ as

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

The **Gauss-Seidel iterative method** becomes

$$\mathbf{x}^{(k)} = \underbrace{(D-L)^{-1}U}_{T_g}\mathbf{x}^{(k-1)} + \underbrace{(D-L)^{-1}\mathbf{b}}_{\mathbf{c}_g}$$

or

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

The matrix $D - L$ is nonsingular if and only if $a_{ii} \neq 0$ for each $i = 1, ..., n$.

Convergence

Usually the Gauss-Seidel iterative method converges faster than the Jacobi method.

Examples do exist where the Jacobi method converges and the Gauss-Seidel method fails to converge.

Also, examples exist where the Gauss-Seidel method converges and the Jacobi method fails to converge.

Convergence Theorems (1 of 3)

We want convergence criterion for the general iteration scheme of the form

$$\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}, \qquad k = 1, 2, ...$$

Lemma

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

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

The previous lemma is important in proving the main convergence theorem.

Convergence Theorems (2 of 3)

Theorem

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

$$\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}, \qquad k = 1, 2, ...$$

converges to the unique solution of

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

if and only if $\rho(T) < 1$.

The proof of the theorem helps establish error bounds from the iterative methods.

Convergence Theorems (3 of 3)

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

$$\mathbf{x}^{(k)} = T\mathbf{x}^{(k-1)} + \mathbf{c}, \qquad k = 1, 2, \dots$$

coverges for any $\mathbf{x}^{(0)} \in \mathbb{R}^n$ to a vector $\mathbf{x} \in \mathbb{R}^n$ and the following error bounds hold:

(i) $||\mathbf{x} - \mathbf{x}^{(k)}|| \leq ||T||^k ||\mathbf{x} - \mathbf{x}^{(0)}||$

(ii) $||\mathbf{x} - \mathbf{x}^{(k)}|| \leq \dfrac{||T||^k}{1 - ||T||^k} ||\mathbf{x}^{(1)} - \mathbf{x}^{(0)}||$

Convergence of Jacobi and Gauss-Seidel

The Jacobi method is given by:

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

where $T_j = D^{-1}(L + U)$.
The Gauss-Seidel method is given by:

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

where $T_g = (D - L)^{-1}U$.

These iterative schemes converge if

$$\rho(T_j) < 1 \qquad \text{or} \qquad \rho(T_g) < 1.$$

More on Convergence of Jacobi and Gauss-Seidel

Definition

The $n \times n$ matrix A is said to be **strictly diagonally dominant** when

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

holds for each $i = 1, 2, \dots n$.

Rate of Convergence

Theorem

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

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

The rapidity of convergence is seen from previous Corollary:

$$||\mathbf{x}^{(k)} - \mathbf{x}|| \approx \rho(T)^k ||\mathbf{x}^{(0)} - \mathbf{x}||$$

Theorem for Some Matrices

Theorem (Stein-Rosenberg)

If $a_{ik} < 0$ for each $i \neq k$ and $a_{ii} > 0$ for each $i = 1, ...n$, then one and only one of the following hold:
(a) $0 \leq \rho(T_g) < \rho(T_j) < 1$,
(b) $1 < \rho(T_j) < \rho(T_g)$,
(c) $\rho(T_j) = \rho(T_g) = 0$,
(d) $\rho(T_j) = \rho(T_g) = 1$.

Part a implies that when one method converges, then both converge with the Gauss-Seidel method converging faster.
Part b implies that when one method diverges, then both diverge with the Gauss-Seidel divergence being more pronounced.

Residuals

Definition

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

We want residuals to converge as rapidly as possible to $\mathbf{0}$.
The Gauss-Seidel method chooses $\mathbf{x}_{i+1}^{(k)}$ so that the i^{th} component of $\mathbf{r}_{i+1}^{(k)}$ is zero.
Making one coordinate zero is often not the optimal way to reduce the norm of the residual, $\mathbf{r}_{i+1}^{(k)}$.

Modify Gauss-Seidel Iteration

The Gauss-Seidel method satisfies:

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

which can be written:

$$x_i^{(k)} = x_i^{(k-1)} + \frac{r_{ii}}{a_{ii}}$$

We modify this to

$$x_i^{(k)} = x_i^{(k-1)} + \omega\frac{r_{ii}}{a_{ii}}$$

where certain choices of $\omega > 0$ reduce the norm of the residual vector and consequently improve the rate of convergence.

SOR Method

The method from previous slide are called **relaxation methods**.
When $0 < \omega < 1$, the procedures are called **under-relaxation methods** and can be used to obtain convergence of systems that fail to converge by the Gauss-Seidel method.
For choices of $\omega > 1$, the procedures are called **over-relaxation methods**, abbreviated **SOR** for **Successive Over-Relaxation** methods, which can accelerate convergence.
The SOR Method is given by:

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

Matrix Form of SOR

Rearranging the SOR Method:

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

In vector form this is

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

or

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

Let $T_\omega = (D - \omega L)^{-1}[(1-\omega)D + \omega U]$ and $\mathbf{c}_\omega = \omega(D - \omega L)^{-1}\mathbf{b}$, then

$$\mathbf{x}^{(k)} = T_\omega \mathbf{x}^{(k-1)} + \mathbf{c}_\omega.$$

SOR Theorems

Theorem (Kahan)

If $a_{ii} \neq 0$ for each $i = 1, ..., n$, then $\rho(T_\omega) \geq |\omega - 1|$. This implies that the SOR method can converge only if $0 < \omega < 2$.

Theorem (Ostrowski-Reich)

If A is a positive definite matrix and $0 < \omega < 2$, then the SOR method converges for any choice of initial approximate vector, $\mathbf{x}^{(0)}$

SOR Theorems

Theorem

If A is positive definite and tridiagonal, then $\rho(T_g) = [\rho(T_j)]^2 < 1$ and the optimal choice of ω for the SOR method is

$$\omega = \frac{2}{1 + \sqrt{1 - [\rho(T_j)]^2}}.$$

with this choice of ω, we have $\rho(T_\omega) = \omega - 1$.