2.29 Numerical Fluid Mechanics
Spring 2015 – **Lecture 6**

REVIEW Lecture 5:

Systems of Linear Equations

- Direct Methods for solving Linear Equation Systems
  - Determinants and Cramer's Rule
  - Gauss Elimination
    - Algorithm
      - Forward Elimination/Reduction to Upper Triangular System
      - Back-Substitution
      - Number of Operations: $O(\frac{2}{3}n^3 + n^2) + O(n^2)$
    - Numerical implementation and stability
      - Partial Pivoting
      - Equilibration
      - Full pivoting
      - Well suited for dense matrices
      - Issues: round-off, cost, does not vectorize/parallelize well
    - Special cases, Multiple RHSs, Operation count $O(n^3 + pn^2) + O(pn^2)$

TODAY's Lecture: Systems of Linear Equations II

- Direct Methods
  - Cramer's Rule
  - Gauss Elimination
    - Algorithm
    - Numerical implementation and stability
      - Partial Pivoting
      - Equilibration
      - Full Pivoting
      - Well suited for dense matrices
      - Issues: round-off, cost, does not vectorize/parallelize well
    - Special cases, Multiple right hand sides, Operation count
  - LU decomposition/factorization
  - Error Analysis for Linear Systems
      - Condition Number
  - Special Matrices: Tri-diagonal systems
- Iterative Methods
  - Jacobi's method
  - Gauss-Seidel iteration
  - Convergence

Reading Assignment

- Chapters 9 and 10 of "Chapra and Canale, Numerical Methods for Engineers, 2006/2010/204."
  - Any chapter on "Solving linear systems of equations" in references on CFD that we provided. For example: chapter 5 of "J. H. Ferziger and M. Peric, Computational Methods for Fluid Dynamics. Springer, NY, 3rd edition, 2002"

LU Decomposition/Factorization:

LU Decomposition: Separates time-consuming elimination for A from that for b / B

The coefficient Matrix $\overline{\overline{\mathbf{A}}}$ is decomposed as

$$\overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{L}}} \cdot \overline{\overline{\mathbf{U}}}$$

where $\overline{\overline{\mathbf{L}}}$ is a lower triangular matrix
and $\overline{\overline{\mathbf{U}}}$ is an upper triangular matrix

$$\overline{\overline{\mathbf{L}}} = [l_{ij}] = \begin{bmatrix} l_{11} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ l_{21} & l_{22} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & & \cdot & l_{kk} & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot & 0 \\ l_{n1} & \cdot & \cdot & \cdot & \cdot & l_{n,n-1} & l_{nn} \end{bmatrix}$$

Then the solution is performed in two simple steps

1. $\overline{\overline{\mathbf{L}}}\vec{y} = \vec{b}$ Forward substitution
2. $\overline{\overline{\mathbf{U}}}\vec{x} = \vec{y}$ Back substitution

$$\overline{\overline{\mathbf{U}}} = [u_{ij}] = \begin{bmatrix} u_{11} & u_{12} & \cdot & \cdot & \cdot & \cdot & u_{1n} \\ 0 & u_{22} & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & & & & \\ \cdot & & \cdot & u_{kk} & \cdot & & \\ \cdot & & & & \cdot & \cdot & u_{n-1,n} \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & u_{nn} \end{bmatrix}$$

$\boxed{\text{How to determine } \overline{\overline{\mathbf{L}}} \text{ and } \overline{\overline{\mathbf{U}}} \text{ ?}}$

LU Decomposition / Factorization

via Gauss Elimination, assuming no pivoting needed

$$\overline{\overline{\mathbf{A}}}^{(i)} = \left[a_{ij}^{(i)}\right] = \begin{array}{cc} & j \longrightarrow \\ i \downarrow & \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdot & \cdot & \cdot & \cdot & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdot & \cdot & \cdot & \cdot & a_{2n}^{(2)} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & a_{ii}^{(i)} & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & a_{ni}^{(i)} & \cdot & \cdot & a_{nn}^{(i)} \end{bmatrix} \end{array}$$

After reduction step *i-1*:

Above and on diagonal: $i \leq j$

Unchanged after step *i-1*: $a_{ij}^{(n)} = \cdots a_{ij}^{(i)}$

Below diagonal: $j < i$

Become and remain *0* in step *j*: $a_{ij}^{(n)} = \cdots a_{ij}^{(j+1)} = 0$

Gauss Elimination (GE): iteration eqns. for the reduction at step k are

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)}, \quad m_{ik} = a_{ik}^{(k)} / a_{kk}^{(k)}$$

This gives the final changes occurring in reduction steps $k = 1$ to $k = i\text{-}1$

LU Decomposition / Factorization
via Gauss Elimination, assuming no pivoting needed

$$
\overline{\overline{\mathbf{A}}}^{(i)} = \left[a_{ij}^{(i)}\right] = \begin{array}{cc} & \xrightarrow{\quad j \quad} \\ i\downarrow & \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdot & \cdot & \cdot & \cdot & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdot & \cdot & \cdot & \cdot & a_{2n}^{(2)} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & a_{ii}^{(i)} & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & a_{ni}^{(i)} & \cdot & \cdot & a_{nn}^{(i)} \end{bmatrix} \end{array}
$$

After reduction step *i-1*:

Above and on diagonal: $i \le j$

Unchanged after step *i-1*: $a_{ij}^{(n)} = \cdots a_{ij}^{(i)}$

Below diagonal: $j < i$

Become and remain *0* in step *j*: $a_{ij}^{(n)} = \cdots a_{ij}^{(j+1)} = 0$

Gauss Elimination (GE): iteration eqns. for the reduction at step k are

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)}, \quad m_{ik} = a_{ik}^{(k)} / a_{kk}^{(k)}$$

This gives the final changes occurring in reduction steps $k = 1$ to $k = i\text{-}1$

Now, to evaluate the changes that accumulated from when one started the elimination, let's try to sum this iteration equation, from:

- 1 to i-1 for above and on diagonal
- 1 to j for below diagonal

As done in class, you can also sum up to an arbitrary r and see which terms remain.

LU Decomposition / Factorization
via Gauss Elimination, assuming no pivoting needed

$j \longrightarrow$

$i \downarrow$

$$\overline{\overline{\mathbf{A}}}^{(i)} = \left[a_{ij}^{(i)}\right] = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdot & \cdot & \cdot & \cdot & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdot & \cdot & \cdot & \cdot & a_{2n}^{(2)} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & a_{ii}^{(i)} & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & a_{ni}^{(i)} & \cdot & \cdot & a_{nn}^{(i)} \end{bmatrix}$$

After reduction step *i-1*:

Above and on diagonal: $i \leq j$

Unchanged after step *i-1*: $a_{ij}^{(n)} = \cdots a_{ij}^{(i)}$

Below diagonal: $j < i$

Become and remain *0* in step *j*: $a_{ij}^{(n)} = \cdots a_{ij}^{(j+1)} = 0$

Gauss Elimination (GE): iteration eqns. for the reduction at step k are

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)}, \quad m_{ik} = a_{ik}^{(k)} / a_{kk}^{(k)}$$

This gives the final changes occurring in reduction steps $k = 1$ to $k = i\text{-}1$

∑ these step-k eqns. from (k=1 to i-1) =>
Gives the total change above diagonal:

$$i \leq j \ : \quad a_{ij}^{(i)} = a_{ij} - \sum_{k=1}^{i-1} m_{ik} a_{kj}^{(k)}$$

∑ this step-k eqns. from (k=1 to j) =>
Gives the total change below diagonal:

$$i > j \ : \quad a_{ij}^{(i)} = 0 = a_{ij} - \sum_{k=1}^{j} m_{ik} a_{kj}^{(k)}$$

LU Decomposition / Factorization
via Gauss Elimination, <u>assuming no pivoting needed</u>

$$\overline{\overline{\mathbf{A}}}^{(i)} = \left[a_{ij}^{(i)}\right] = \begin{array}{c|c} & j \rightarrow \\ \hline i \downarrow & \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdot & \cdot & \cdot & \cdot & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \cdot & \cdot & \cdot & \cdot & a_{2n}^{(2)} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & a_{ii}^{(i)} & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 0 & a_{ni}^{(i)} & \cdot & \cdot & a_{nn}^{(i)} \end{bmatrix} \end{array}$$

After reduction step *i-1*:

Above and on diagonal: $i \le j$

Unchanged after step *i-1*: $a_{ij}^{(n)} = \cdots a_{ij}^{(i)}$

Below diagonal: $j < i$

Become and remain *0* in step *j*: $a_{ij}^{(n)} = \cdots a_{ij}^{(j+1)} = 0$

Summary: summing the changes in reduction steps $k = 1$ to $k = i\text{-}1$:

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)}, \quad m_{ik} = a_{ik}^{(k)} / a_{kk}^{(k)}$$

We obtained: Total change above diagonal

$$i \le j \;:\; a_{ij}^{(i)} = a_{ij} - \sum_{k=1}^{i-1} m_{ik} a_{kj}^{(k)} \qquad (1)$$

We obtained: Total change below diagonal

$$i > j \;:\; a_{ij}^{(i)} = 0 = a_{ij} - \sum_{k=1}^{j} m_{ik} a_{kj}^{(k)} \qquad (2)$$

→ Now, if we define:

$$m_{ii} = 1, \quad i = 1, \ldots n$$

and use them in equations (1) and (2) =>

$$\begin{cases} i \le j \;:\; a_{ij} = \sum_{k=1}^{i} m_{ik} a_{kj}^{(k)} \\ i > j \;:\; a_{ij} = \sum_{k=1}^{j} m_{ik} a_{kj}^{(k)} \end{cases}$$

$$\boxed{\Rightarrow a_{ij} = \sum_{k=1}^{\min(i,j)} m_{ik} a_{kj}^{(k)}}$$

LU Decomposition / Factorization
via Gauss Elimination, assuming no pivoting needed

Result seems to be a 'Matrix product': $a_{ij} = \sum_{k=1}^{\min(i,j)} m_{ik} a_{kj}^{(k)}$

Sum stops at diagonal

Below diagonal

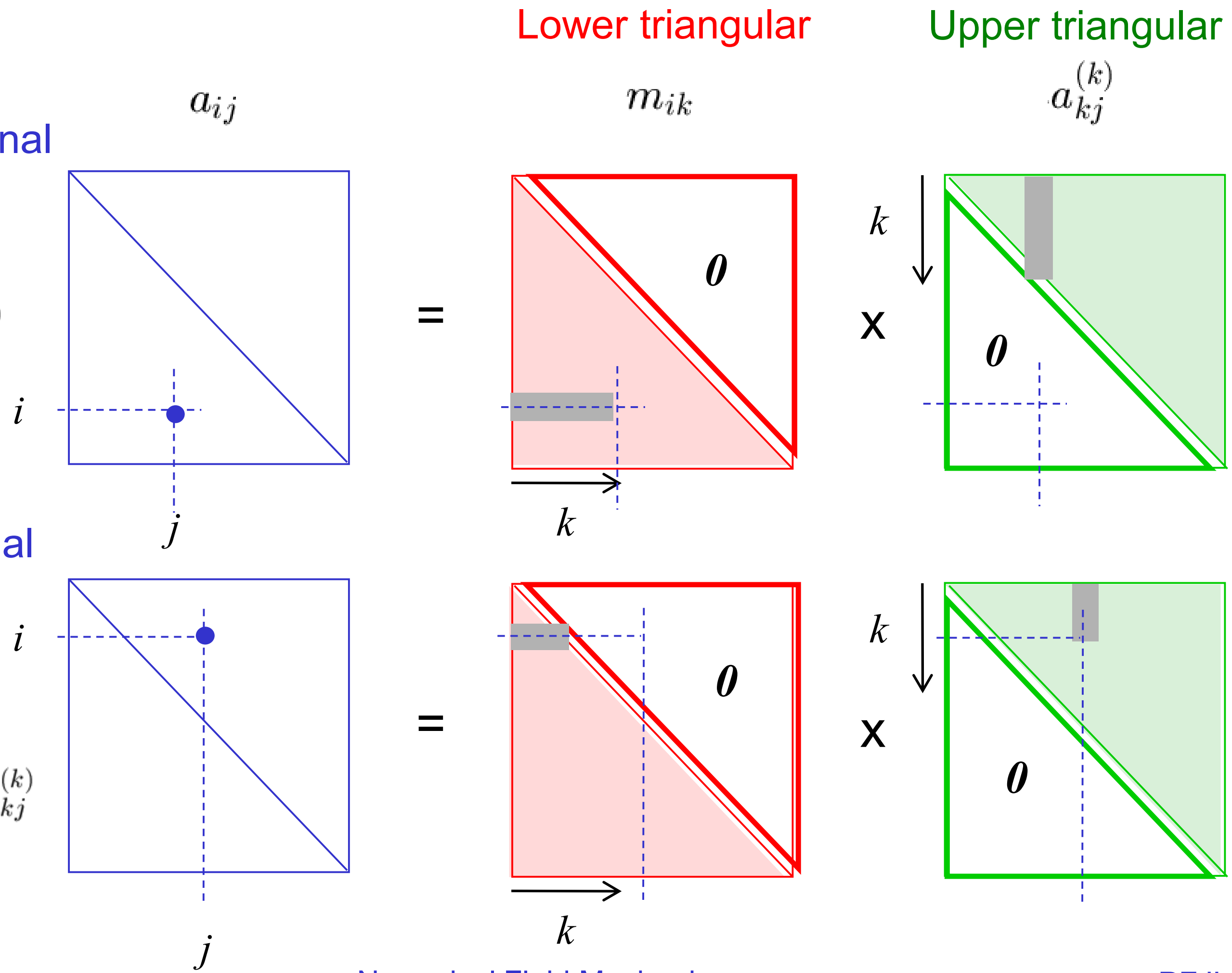
$i > j$:

$i > j \; : \quad a_{ij} = \sum_{k=1}^{j} m_{ik} a_{kj}^{(k)}$

Above diagonal

$i \le j$:

$i \le j \; : \quad a_{ij} = \sum_{k=1}^{i} m_{ik} a_{kj}^{(k)}$



LU Decomposition / Factorization
via Gauss Elimination, <u>assuming no pivoting needed</u>

GE reduction directly yields LU factorization

$$\overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{L}}} \cdot \overline{\overline{\mathbf{U}}}$$

Lower triangular

$$\overline{\overline{\mathbf{L}}} = l_{ij} = \begin{cases} 0 & i < j \\ 1 & i = j \\ m_{ij} & i > j \end{cases}$$

Upper triangular

$$\overline{\overline{\mathbf{U}}} = u_{ij} = \begin{cases} a_{ij}^{(i)} & i \leq j \\ 0 & i > j \end{cases}$$

Number of Operations for LU?

Compact storage:
no need for additional memory (the unitary diagonal of L does not need to be stored)

$$\begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdot & \cdot & \cdot & \cdot & a_{1n}^{(1)} \\ m_{21} & a_{22}^{(2)} & \cdot & \cdot & \cdot & \cdot & a_{2n}^{(2)} \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & & \cdot & \cdot & & & \cdot \\ \cdot & & & \cdot & \cdot & & \cdot \\ m_{n1} & \cdot & \cdot & \cdot & \cdot & m_{n,n-1} & a_{nn}^{(n)} \end{bmatrix}$$

Lower diagonal implied

$$m_{ii} = 1, \quad i = 1, \ldots n$$

(referred to as the Doolittle decomposition)

LU Decomposition / Factorization

via Gauss Elimination, assuming no pivoting needed

GE Reduction directly yields LU factorization

$$\overline{\overline{\mathbf{A}}} = \overline{\overline{\mathbf{L}}} \cdot \overline{\overline{\mathbf{U}}}$$

Lower triangular

$$\overline{\overline{\mathbf{L}}} = l_{ij} = \begin{cases} 0 & i < j \\ 1 & i = j \\ m_{ij} & i > j \end{cases}$$

Upper triangular

$$\overline{\overline{\mathbf{U}}} = u_{ij} = \begin{cases} a_{ij}^{(i)} & i \leq j \\ 0 & i > j \end{cases}$$

Compact storage

$$\begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdot & \cdot & \cdot & \cdot & a_{1n}^{(1)} \\ m_{21} & a_{22}^{(2)} & \cdot & \cdot & \cdot & \cdot & a_{2n}^{(2)} \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & & \cdot & \cdot & & & \cdot \\ \cdot & & & \cdot & \cdot & & \cdot \\ m_{n1} & \cdot & \cdot & \cdot & \cdot & m_{n,n-1} & a_{nn}^{(n)} \end{bmatrix}$$

Lower diagonal implied

$$m_{ii} = 1, \quad i = 1, \ldots n$$

Number of Operations for LU?

Same as Gauss Elimination:
less in Elimination phase (no RHS operations), but more in double back-substitution phase

Pivoting in LU Decomposition / Factorization

Before reduction, step k

$$
\begin{bmatrix}
a_{11}^{(1)} & a_{12}^{(1)} & \cdot & \cdot & \cdot & \cdot & a_{1n}^{(1)} \\
m_{21} & a_{22}^{(2)} & \cdot & \cdot & \cdot & \cdot & a_{2n}^{(2)} \\
\cdot & \cdot & \cdot & & & & \cdot \\
\cdot & & m_{k,k-1} & a_{kk}^{(k)} & \cdot & & \cdot \\
\cdot & & \cdot & \cdot & \cdot & \cdot & \cdot \\
m_{n1} & \cdot & m_{n,k-1} & a_{nk}^{(k)} & \cdot & \cdots & a_{n,n}^{(k)}
\end{bmatrix}
$$

Pivoting if

$$\left|a_{ik}^{(k)}\right| \gg \left|a_{kk}^{(k)}\right| \;, \quad i > k$$

To do this interchange of rows i and k,

use a pivot vector: $\begin{cases} p_k = i \\ \text{or else} \\ p_k = k \end{cases}$

Pivot element vector

$$p_i, \quad i = 1, \ldots n$$

Forward substitution, step k

$$\overline{\overline{\mathbf{L}}}\vec{y} = \vec{b}$$

Interchange rows i and k

$$\begin{Bmatrix} b_1 \\ \cdot \\ \cdot \\ b_k \\ \cdot \\ b_i \\ \cdot \\ b_n \end{Bmatrix}$$

Dummy var.

In code, use b(p(k)), which amounts to:

If $p_k = i \Rightarrow \begin{cases} b_i^{(k)} = b_k \\ b_k = b_i \\ b_i = b_i^{(k)} \end{cases}$

LU Decomposition / Factorization: Variations

- Doolittle decomposition:
  - $m_{ii}=1$ (implied but could be stored in **L**)
- Crout decomposition:
  - Directly impose diagonal of **U** equal to 1's (instead of L)
  - Sweeps both by columns and rows (columns for L and rows for U)
  - Reduce storage needs
  - Each element of **A** only employed once
- Matrix inverse: $\mathbf{AX=I} \Rightarrow \mathbf{(LU)X=I}$
  - Numbers of ops: $O\left(\underbrace{\frac{2n^3}{3}}_{\text{LU Decomp.}} + \underbrace{pn^2}_{\text{Forward Substitution}} + \underbrace{pn^2}_{\text{Backward Substitution}}\right)$ for $p=n$, $\Rightarrow \frac{2n^3}{3} + 2n^3 = \frac{8n^3}{3}$

Recall Lecture 3: The Condition Number

- The *condition* of a mathematical problem relates to its sensitivity to changes in its input values
- A computation is *numerically unstable* if the uncertainty of the input values are magnified by the numerical method
- Considering x and $f(x)$, the *condition number* is the ratio of the relative error in $f(x)$ to that in x.
- Using first-order Taylor series $f(\bar{x}) = f(\bar{x}) + f'(\bar{x})(\bar{x} - \bar{x})$
- Relative error in $f(x)$: $\dfrac{f(x) - f(\bar{x})}{f(\bar{x})} \cong \dfrac{f'(\bar{x})(x - \bar{x})}{f(\bar{x})}$
- Relative error in x: $\dfrac{(x - \bar{x})}{\bar{x}}$
- Condition Nb = Ratio of relative errors: $\boxed{K_p = \left| \dfrac{\bar{x}\, f'(\bar{x})}{f(\bar{x})} \right|}$

Linear Systems of Equations
Error Analysis

Function of one variable

$$y = f(x)$$

Condition number

$$\left|\frac{f(\overline{x}) - f(x)}{f(x)}\right| = K \left|\frac{\overline{x} - x}{x}\right| \;, \quad \overline{x} = x + \delta x$$

$$\left|\frac{\delta y}{y}\right| = K \left|\frac{\delta x}{x}\right|$$

The condition number K is a measure of the amplification of the relative error by the function f(x)

Linear systems

How is the relative error of $\overline{\mathbf{x}}$ dependent on errors in $\overline{\mathbf{b}}$?

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

Example

$$\overline{\overline{\mathbf{A}}} = \begin{bmatrix} 1.0 & 1.0 \\ 1.0 & 1.0001 \end{bmatrix} \;, \quad \det(\overline{\overline{\mathbf{A}}}) = 0.0001$$

Using MATLAB with different $\overline{\mathbf{b}}$'s (see tbt8.m):

$$\overline{\mathbf{b}} = \begin{Bmatrix} 2 \\ 2 \end{Bmatrix} \Rightarrow \overline{\mathbf{x}} = \begin{Bmatrix} 2 \\ 0 \end{Bmatrix}$$

$$\overline{\mathbf{b}} = \begin{Bmatrix} 2 \\ 2.0001 \end{Bmatrix} \Rightarrow \overline{\mathbf{x}} = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$$

Small changes in $\overline{\mathbf{b}}$ give large changes in $\overline{\mathbf{x}}$

The system is ill-Conditioned

Linear Systems of Equations: **Norms**

Vector and Matrix Norms:

$$\begin{cases} \|\overline{\mathbf{x}}\|_\infty = \max_i |x_i| \\ \left\|\overline{\overline{\mathbf{A}}}\right\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}| \end{cases}$$

Evaluation of Condition Numbers requires use of Norms

Properties:

$$\overline{\overline{\mathbf{A}}} \neq \overline{\overline{\mathbf{0}}} \Rightarrow \left\|\overline{\overline{\mathbf{A}}}\right\| > 0$$

$$\left\|\alpha\overline{\overline{\mathbf{A}}}\right\| = |\alpha| \left\|\overline{\overline{\mathbf{A}}}\right\|$$

$$\left\|\overline{\overline{\mathbf{A}}} + \overline{\overline{\mathbf{B}}}\right\| \leq \left\|\overline{\overline{\mathbf{A}}}\right\| + \left\|\overline{\overline{\mathbf{B}}}\right\|$$

Sub-multiplicative / Associative Norms (n-by-n matrices with such norms form a Banach Algebra/space)

$$\begin{cases} \left\|\overline{\overline{\mathbf{A}\mathbf{B}}}\right\| \leq \left\|\overline{\overline{\mathbf{A}}}\right\| \left\|\overline{\overline{\mathbf{B}}}\right\| \\ \boxed{\left\|\overline{\overline{\mathbf{A}}}\overline{\mathbf{x}}\right\| \leq \left\|\overline{\overline{\mathbf{A}}}\right\| \|\overline{\mathbf{x}}\|} \end{cases}$$

Examples of Matrix Norms

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

"Maximum Column Sum"

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

"Maximum Row Sum"

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{1/2}$$

"The Frobenius norm" (also called Euclidean norm)", which for matrices differs from:

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^* A)}$$

"The l-2 norm" (also called spectral norm)

Linear Systems of Equations
Error Analysis: Perturbed Right-hand Side

Vector and Matrix Norms

$$||\bar{\mathbf{x}}||_\infty = \max_i |x_i|$$

$$\left\|\bar{\bar{\mathbf{A}}}\right\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

Properties

$$\bar{\bar{\mathbf{A}}} \neq \bar{\bar{\mathbf{0}}} \Rightarrow \left\|\bar{\bar{\mathbf{A}}}\right\| > 0$$

$$\left\|\alpha\bar{\bar{\mathbf{A}}}\right\| = |\alpha| \left\|\bar{\bar{\mathbf{A}}}\right\|$$

$$\left\|\bar{\bar{\mathbf{A}}} + \bar{\bar{\mathbf{B}}}\right\| \leq \left\|\bar{\bar{\mathbf{A}}}\right\| + \left\|\bar{\bar{\mathbf{B}}}\right\|$$

$$\left\|\overline{\overline{\mathbf{AB}}}\right\| \leq \left\|\bar{\bar{\mathbf{A}}}\right\| \left\|\bar{\bar{\mathbf{B}}}\right\|$$

$$\boxed{\left\|\bar{\bar{\mathbf{A}}}\bar{\mathbf{x}}\right\| \leq \left\|\bar{\bar{\mathbf{A}}}\right\| ||\bar{\mathbf{x}}||}$$

Perturbed Right-hand Side implies

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

$$\Downarrow$$

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

Subtract original equation

$$\begin{aligned} \bar{\bar{\mathbf{A}}}\delta\bar{\mathbf{x}} &= \delta\bar{\mathbf{b}} \\ \delta\bar{\mathbf{x}} &= \bar{\bar{\mathbf{A}}}^{-1}\delta\bar{\mathbf{b}} \end{aligned}$$

$$\Downarrow$$

$$\left\{ \begin{aligned} ||\delta\bar{\mathbf{x}}|| &\leq \left\|\bar{\bar{\mathbf{A}}}^{-1}\right\| ||\delta\bar{\mathbf{b}}|| \\ ||\bar{\mathbf{b}}|| = \left\|\bar{\bar{\mathbf{A}}}\bar{\mathbf{x}}\right\| &\leq \left\|\bar{\bar{\mathbf{A}}}\right\| ||\bar{\mathbf{x}}|| \end{aligned} \right\} \Rightarrow$$

Relative Error Magnification

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

Condition Number

$$K(\bar{\bar{\mathbf{A}}}) = \left\|\bar{\bar{\mathbf{A}}}^{-1}\right\| \left\|\bar{\bar{\mathbf{A}}}\right\|$$

Linear Systems of Equations
Error Analysis: Perturbed Coefficient Matrix

Vector and Matrix Norms

$$||\overline{\mathbf{x}}||_\infty = \max_i |x_i|$$

$$\left\|\overline{\overline{\mathbf{A}}}\right\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

Properties

$$\overline{\overline{\mathbf{A}}} \neq \overline{\overline{\mathbf{0}}} \Rightarrow \left\|\overline{\overline{\mathbf{A}}}\right\| > 0$$

$$\left\|\alpha\overline{\overline{\mathbf{A}}}\right\| = |\alpha| \left\|\overline{\overline{\mathbf{A}}}\right\|$$

$$\left\|\overline{\overline{\mathbf{A}}} + \overline{\overline{\mathbf{B}}}\right\| \leq \left\|\overline{\overline{\mathbf{A}}}\right\| + \left\|\overline{\overline{\mathbf{B}}}\right\|$$

$$\boxed{\begin{gathered}\left\|\overline{\overline{\mathbf{A}}}\overline{\overline{\mathbf{B}}}\right\| \leq \left\|\overline{\overline{\mathbf{A}}}\right\| \left\|\overline{\overline{\mathbf{B}}}\right\| \\ \left\|\overline{\overline{\mathbf{A}}}\overline{\mathbf{x}}\right\| \leq \left\|\overline{\overline{\mathbf{A}}}\right\| ||\overline{\mathbf{x}}||\end{gathered}}$$

Perturbed Coefficient Matrix implies

$$\left(\overline{\overline{\mathbf{A}}} + \delta\overline{\overline{\mathbf{A}}}\right)(\overline{\mathbf{x}} + \delta\overline{\mathbf{x}}) = \overline{\mathbf{b}}$$

Subtract unperturbed equation

$$\overline{\overline{\mathbf{A}}}\delta\overline{\mathbf{x}} + \delta\overline{\overline{\mathbf{A}}}(\overline{\mathbf{x}} + \delta\overline{\mathbf{x}}) = \overline{\mathbf{0}}$$

(Neglect 2nd order)

$$\delta\overline{\mathbf{x}} = -\overline{\overline{\mathbf{A}}}^{-1}\delta\overline{\overline{\mathbf{A}}}(\overline{\mathbf{x}} + \delta\overline{\mathbf{x}}) \simeq -\overline{\overline{\mathbf{A}}}^{-1}\delta\overline{\overline{\mathbf{A}}}\overline{\mathbf{x}}$$

$$||\delta\overline{\mathbf{x}}|| \leq \left\|\overline{\overline{\mathbf{A}}}^{-1}\right\| \left\|\delta\overline{\overline{\mathbf{A}}}\right\| ||\overline{\mathbf{x}}||$$

Relative Error Magnification

$$\frac{||\delta\overline{\mathbf{x}}||}{||\overline{\mathbf{x}}||} \leq \left\|\overline{\overline{\mathbf{A}}}^{-1}\right\| \left\|\overline{\overline{\mathbf{A}}}\right\| \frac{\left\|\delta\overline{\overline{\mathbf{A}}}\right\|}{\left\|\overline{\overline{\mathbf{A}}}\right\|}$$

Condition Number

$$K(\overline{\overline{\mathbf{A}}}) = \left\|\overline{\overline{\mathbf{A}}}^{-1}\right\| \left\|\overline{\overline{\mathbf{A}}}\right\|$$

Example: Ill-Conditioned System

$$\begin{bmatrix} 1.0 & 1.0 \\ 1.0 & 1.0001 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\det(\overline{\overline{\mathbf{A}}}) = 0.0001$$

Using Cramer's rule:

$$\left\{ \begin{aligned} a_{11} &= \frac{1.0001}{0.0001} = 10,001 \\ a_{12} &= \frac{-1}{0.0001} = -10,000 \\ a_{21} &= \frac{-1}{0.0001} = -10,000 \\ a_{11} &= \frac{1.0}{0.0001} = 10,000 \end{aligned} \right.$$

$$\left. \begin{aligned} \left\|\overline{\overline{\mathbf{A}}}\right\|_\infty &= 2.0001 \\ \left\|\overline{\overline{\mathbf{A}}}^{-1}\right\|_\infty &= 20,001 \end{aligned} \right\} \Rightarrow K(\overline{\overline{\mathbf{A}}}) \simeq \boxed{40,000}$$

Ill-conditioned system

4-digit Arithmetic

tbt6.m

```
n=4
a = [ [1.0 1.0]' [1.0 1.0001]']
b= [1 2]'

ai=inv(a);
a_nrm=max( abs(a(1,1)) + abs(a(1,2)) ,
           abs(a(2,1)) + abs(a(2,2)) )
ai_nrm=max( abs(ai(1,1)) + abs(ai(1,2)) ,
            abs(ai(2,1)) + abs(ai(2,2)) )
k=a_nrm*ai_nrm

r=ai * b

x=[0 0];
m21=a(2,1)/a(1,1);
a(2,1)=0;
a(2,2)  = radd(a(2,2),-m21*a(1,2),n);
b(2)    = radd(b(2),-m21*b(1),n);

x(2)    = b(2)/a(2,2);
x(1)    = (radd(b(1), -a(1,2)*x(2),n))/a(1,1);
x'
```

Example: Better-Conditioned System

$$\begin{bmatrix} 0.0001 & 1.0 \\ 1.0 & 1.0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

4-digit Arithmetic

$$\det(\overline{\overline{\mathbf{A}}}) = 0.9999$$

Using Cramer's rule:

$$\left\{ \begin{aligned} a_{11} &= \frac{-1}{0.9999} = -1,0001 \\ a_{12} &= \frac{1}{0.9999} = 1.0001 \\ a_{21} &= \frac{1}{0.9999} = 1.0001 \\ a_{11} &= \frac{-0.0001}{0.9999} = -0.0001 \end{aligned} \right.$$

tbt7.m

```
n=4
a = [ [0.0001 1.0]' [1.0 1.0]']
b= [1 2]'

ai=inv(a);
a_nrm=max( abs(a(1,1)) + abs(a(1,2)) ,
           abs(a(2,1)) + abs(a(2,2)) )
ai_nrm=max( abs(ai(1,1)) + abs(ai(1,2)) ,
            abs(ai(2,1)) + abs(ai(2,2)) )
k=a_nrm*ai_nrm

r=ai * b

x=[0 0];
m21=a(2,1)/a(1,1);
a(2,1)=0;
a(2,2)  = radd(a(2,2),-m21*a(1,2),n);
b(2)    = radd(b(2),-m21*b(1),n);

x(2)    = b(2)/a(2,2);
x(1)    = (radd(b(1), -a(1,2)*x(2),n))/a(1,1);
x'
```

$$\left. \begin{aligned} \left\|\overline{\overline{\mathbf{A}}}\right\|_\infty &= 2.0 \\ \left\|\overline{\overline{\mathbf{A}}}^{-1}\right\|_\infty &= 2.0002 \end{aligned} \right\} \Rightarrow K(\overline{\overline{\mathbf{A}}}) \simeq \boxed{4}$$

Relatively Well-conditioned system

Recall Lecture 3: The Condition Number

- The *condition* of a mathematical problem relates to its sensitivity to changes in its input values
- A computation is *numerically unstable* if the uncertainty of the input values are magnified by the numerical method
- Considering x and $f(x)$, the *condition number* is the ratio of the relative error in $f(x)$ to that in x.
- Using first-order Taylor series f
- Relative error in $f(x)$: $\dfrac{f(x)-f(\bar{x})}{f(\bar{x})} \cong \dfrac{f'(\bar{x})(x-\bar{x})}{f(\bar{x})}$
- Relative error in x: $\dfrac{(x-\bar{x})}{\bar{x}}$
- Condition Nb = Ratio of relative errors: $\boxed{K_p = \left|\dfrac{\bar{x}\; f'(\bar{x})}{f(\bar{x})}\right|}$

2.29 Numerical Fluid Mechanics
Spring 2015