Linear Systems. Gauss Elimination

Ned Nedialkov

McMaster University

2, 7 March 2023

Outline

Introduction

- Given an $n \times n$ nonsingular matrix A and an n-vector b solve

$$Ax = b$$

- The following are equivalent
  - A is nonsingular
  - The determinant of A is nonzero, $\det(A) \neq 0$
  - There exists A^{-1} such that $A^{-1}A = AA^{-1} = I$, where I is the $n \times n$ identity matrix
- Dense system: A may have a small number of nonzeros
- Sparse system: most of the elements are zeros
  See Florida Sparse Matrix Collection
- Direct methods: based on Gauss elimination
- Iterative methods: conjugate gradient, GMRES, ...

Gauss elimination (GE)

Example 1 (Forward elimination).

$$Ax = \begin{bmatrix} 1 & -1 & 3 \\ 1 & 1 & 0 \\ 3 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 11 \\ 3 \\ 3 \end{bmatrix} = b$$

Multiply: first row by 1 and subtract from second row; first row by 3 and subtract from third row

$$A|b = \left[\begin{array}{ccc|c} 1 & -1 & 3 & 11 \\ 1 & 1 & 0 & 3 \\ 3 & -2 & 1 & 3 \end{array}\right] \begin{array}{cc} \times 1 & \times 3 \\ \downarrow & \\ & \downarrow \end{array}$$

$$A|b \leftarrow \left[\begin{array}{ccc|c} 1 & -1 & 3 & 11 \\ 0 & 2 & -3 & -8 \\ 0 & 1 & -8 & -30 \end{array}\right]$$

Example 1. cont.

Multiply second row by $\frac{1}{2}$ and subtract from third row

$$A|b \leftarrow \left[\begin{array}{rrr|r} 1 & -1 & 3 & 11 \\ 0 & 2 & -3 & -8 \\ 0 & 1 & -8 & -30 \end{array}\right] \quad \begin{array}{c} \\ \times\frac{1}{2} \\ \downarrow \end{array}$$

$$A|b \leftarrow \left[\begin{array}{rrr|r} 1 & -1 & 3 & 11 \\ 0 & 2 & -3 & -8 \\ 0 & 0 & -6.5 & -26 \end{array}\right]$$

Example 2 (Backward substitution).

$$\begin{bmatrix} 1 & -1 & 3 \\ 0 & 2 & -3 \\ 0 & 0 & -6.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \begin{bmatrix} 11 \\ -8 \\ -26 \end{bmatrix}$$

$$\begin{aligned} x_3 &= b_3/a_{33} && = -26/(-6.5) && = 4 \\ x_2 &= (b_2 - a_{23}x_3)/a_{22} && = (-8 - (-3) \times 4)/2 && = 2 \\ x_1 &= (b_1 - a_{12}x_2 - a_{13}x_3)/a_{11} && = (11 - (-1) \times 2 - 3 \times 4)/1 && = 1 \end{aligned}$$

LU decomposition

- Decompose A as $A = LU$, where
  - L is unit lower-triangular
    1's on the main diagonal, 0's above it
  - U is upper-triangular
    0's below the main diagonal
- Consider solving $Ax = b$. From

$$\begin{aligned} Ax = LUx &= b \\ L\underbrace{(Ux)}_{y} &= b \end{aligned}$$

  - solve $Ly = b$ for y
  - solve $Ux = y$ for x

A is $n \times n$

- Solving by GE takes $O(n^3)$ arithmetic operations.
- LU decomposition takes $O(n^3)$ arithmetic operations.
- Solving each of $Ly = b$ and $Ux = y$ takes $O(n^2)$ arithmetic operations.
- Suppose we need to solve m systems $Ax = b^{(i)}$, $i = 1, \ldots, m$. A is the same, the right-hand side changes.
- If we solve them with GE $O(mn^3)$
- Do LU decomposition first $O(n^3)$
- Solve $Ly = b^{(i)},\ Ux = y$, for $i = 1 : m$ $O(mn^2)$
  Total LU+triangular solves $O(n^3 + mn^2)$

Example 3 (LU decomposition).

$$A = \begin{bmatrix} 1 & -1 & 3 \\ 1 & 1 & 0 \\ 3 & -2 & 1 \end{bmatrix} \quad \begin{matrix} \times 1 & \times 3 \\ \downarrow & \\ & \downarrow \end{matrix}$$

- multipliers $l_{2,1} = 1$, $l_{3,1} = 3$

$$L_1 A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 3 \\ 1 & 1 & 0 \\ 3 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 3 \\ 0 & 2 & -3 \\ 0 & 1 & -8 \end{bmatrix} = A^{(1)}$$

- multiplier $l_{3,2} = \frac{1}{2}$

Example 3. cont.

$$L_2A^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 3 \\ 0 & 2 & -3 \\ 0 & 1 & -8 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & -1 & 3 \\ 0 & 2 & -3 \\ 0 & 0 & -6.5 \end{bmatrix} = A^{(2)} = U$$

We have

$$L_2A^{(1)} = (L_2L_1)A = U$$
$$A = \underbrace{(L_1^{-1}L_2^{-1})}_{L}U$$

To compute L_1^{-1}, L_2^{-1} flip the signs of nonzero entries below the main diagonal.

Then

$$L = L_1^{-1} L_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & \frac{1}{2} & 1 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & \frac{1}{2} & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} 1 & -1 & 3 \\ 0 & 2 & -3 \\ 0 & 0 & -6.5 \end{bmatrix}}_{U} = \underbrace{\begin{bmatrix} 1 & -1 & 3 \\ 1 & 1 & 0 \\ 3 & -2 & 1 \end{bmatrix}}_{A}$$

Partial pivoting

Example 4. Consider solving

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

Multiply the first row by $2/\epsilon$ and subtract from the second

$$\begin{bmatrix} \epsilon & 1 \\ 0 & 1 - \frac{2}{\epsilon} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 - \frac{2}{\epsilon} \end{bmatrix}$$

$$x_2 = \frac{3 - \frac{2}{\epsilon}}{1 - \frac{2}{\epsilon}} = 1 + \frac{2\epsilon}{\epsilon - 2}$$

If $|\epsilon| \ll 2$, in FP arithmetic we compute $\widetilde{x}_2 = 1$.

Then we compute $\widetilde{x}_1 = (1 - \widetilde{x}_2)/\epsilon = 0$.

Example 4. cont.

But

$$x_1 = \frac{1 - x_2}{\epsilon} = \frac{-\frac{2\epsilon}{\epsilon - 2}}{\epsilon} = \frac{2}{2 - \epsilon} \approx 1.$$

Write $x_2 = \widetilde{x}_2 + \delta$, $-\widetilde{x}_2 = -x_2 - \delta$.

What we compute is

$$\begin{aligned} \widetilde{x}_1 = \frac{1 - \widetilde{x}_2}{\epsilon} = \frac{1 - x_2}{\epsilon} + \frac{\delta}{\epsilon} = x_1 + \frac{\delta}{\epsilon} \\ = x_1 + \frac{2}{\epsilon - 2} \approx x_1 - 1 \end{aligned}$$

The error in x_2 is multiplied by $1/\epsilon$, which is large if $|\epsilon|$ is small.

Example 4. cont.

Swap the rows

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

Multiply first rows by $\epsilon/2$ and subtract from second row

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

Then

$$x_2 = \frac{1 - 3\frac{\epsilon}{2}}{1 - \frac{\epsilon}{2}} \approx 1.$$

We compute $\widetilde{x}_2 = 1$ and $\widetilde{x}_1 = (3 - \widetilde{x}_2)/2 = 1$.

Example 4. cont.

Using $x_2 = \widetilde{x}_2 + \delta$,

$$\widetilde{x}_1 = \frac{3 - \widetilde{x}_2}{2} = \frac{3 - x_2}{2} + \frac{\delta}{2}.$$

The error in $\widetilde{x}_2$ is divided by 2.

Partial pivoting

Partial pivoting

- at step $k = 1 : n - 1$ chose the row q for which $|a_{qk}|$ is the largest
- eliminate with row q
  now we divide by the largest element in column k

Permutations

- A permutation matrix is a square matrix that contains
  - 1 in each row and column
  - 0 everywhere else
- A permutation matrix P is orthogonal: $PP^T = I$, $\det P = \pm 1$. I is the identity matrix.
- PA reorders the rows of A, AP reorders the columns of A.
- An elementary permutation matrix P differs from the identity matrix in two rows and columns.
  Such a P is symmetric, $P = P^T$ and $PP = I$, $P = P^{-1}$.
- A general permutation matrix can be written as the product of elementary permutation matrices.
- Swapping rows i and j in matrix A can be written as PA, where P is I with rows i and j swapped.
  Such a P is an elementary permutation matrix.

In practice, represent a permutation matrix P as an integer vector p where

- if row i has 1 in column j then
- $p_i = j$

Example 5.

$$P = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$p = [3, 2, 1, 4]$$

Example 6.

Consider LU factorization with partial pivoting of a 3×3 matrix A.

On the first step, assume we swap rows 1 and 2. Denote

$$P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Write the elimination in the first column as $L_1 P_1 A$.

One the second step, assume we swap rows 2 and 3 of $L_1 P_1 A$. Denoting

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

write the elimination in the second columns as $L_2 P_2 L_1 P_1 A = U$.

Example 6. cont.

$$
\begin{aligned}
L_2 P_2 L_1 P_1 A &= L_2 (P_2 L_1 P_2^{-1})(P_2 P_1) A \\
&= L_2 \underbrace{(P_2 L_1 P_2)}_{L_1'} \underbrace{(P_2 P_1)}_{P} A
\end{aligned}
$$

L_1' has the same structure as L_1:

$$
\begin{aligned}
L_1' = P_2 L_1 P_2 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ l_{31} & 0 & 1 \\ l_{21} & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ l_{31} & 1 & 0 \\ l_{21} & 0 & 1 \end{bmatrix}
\end{aligned}
$$

$P = P_2 P_1$ is a permutation matrix, product of elementary P_2 and P_1.

Example 6. cont.

We have

$$\begin{aligned} L_2 L_1' P A &= U \\ P A &= \underbrace{(L_2 L_1')^{-1}}_{L} U \\ &= L U \end{aligned}$$

Matlab's lu

```
>> A = gallery('frank',3), [L,U,P] = lu(A), PA=P*A
A =
     3     2     1
     2     2     1
     0     1     1
L =
    1.0000         0         0
         0    1.0000         0
    0.6667    0.6667    1.0000
U =
    3.0000    2.0000    1.0000
         0    1.0000    1.0000
         0         0   -0.3333
P =
     1     0     0
     0     0     1
     0     1     0
PA =
     3     2     1
     0     1     1
     2     2     1
```

```
>> A = gallery('frank',3), [L,U,p] = lu(A,'vector'), A(p,:)
A =
     3     2     1
     2     2     1
     0     1     1
L =
    1.0000         0         0
         0    1.0000         0
    0.6667    0.6667    1.0000
U =
    3.0000    2.0000    1.0000
         0    1.0000    1.0000
         0         0   -0.3333
p =
     1     3     2
ans =
     3     2     1
     0     1     1
     2     2     1
```

```
>> A = gallery('frank',3), [L,U] = lu(A)
A =
     3     2     1
     2     2     1
     0     1     1
L =
    1.0000         0         0
    0.6667    0.6667    1.0000
         0    1.0000         0
U =
    3.0000    2.0000    1.0000
         0    1.0000    1.0000
         0         0   -0.3333
```

Solution error

Consider $Ax = b$

- Let $\widetilde{x}$ be the computed solution, and let x be the exact solution
- Relative error in the solution is

$$\frac{\|x - \widetilde{x}\|}{\|x\|}$$

- Residual is

$$r = b - A\widetilde{x}$$

$$r = 0 \iff b - A\widetilde{x} = 0 \iff \widetilde{x} = x$$

- In practice $r \neq 0$

Given $\widetilde{x}$, how large is

$$\frac{\|x - \widetilde{x}\|}{\|x\|}$$

Using $r = b - A\widetilde{x} = Ax - A\widetilde{x} = A(x - \widetilde{x})$,

$$x - \widetilde{x} = A^{-1}r$$
$$\|x - \widetilde{x}\| = \|A^{-1}r\| \leq \|A^{-1}\|\|r\|$$

Using $b = Ax$, $\|b\| = \|Ax\| \leq \|A\|\|x\|$, and

$$\|x\| \geq \frac{\|b\|}{\|A\|}$$

The condition number of A is

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

We have

$$\frac{\|x - \widetilde{x}\|}{\|x\|} \leq \frac{\|A^{-1}\| \|r\|}{\frac{\|b\|}{\|A\|}} = \|A^{-1}\| \|A\| \frac{\|r\|}{\|b\|} = \text{cond}(A) \frac{\|r\|}{\|b\|}$$

$$\frac{\|x - \widetilde{x}\|}{\|x\|} \leq \text{cond}(A) \frac{\|r\|}{\|b\|}$$

- If $\text{cond}(A)$ is not large and $\|r\|/\|b\|$ is small then small relative error.
- GE with partial pivoting generally produces small $\|r\|/\|b\|$.
- As a rule of thumb, if $\text{cond}(A) \approx 10^k$, then about k decimal digits are lost when solving $Ax = b$.

Vector norms

lp norms

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \qquad 1 \le p \le \infty$$

- $p = 1$, one norm $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
- $p = \infty$, infinity or max norm $\|x\|_\infty = \max_{i=1,\dots,n} |x_i|$
- $p = 2$, two or Euclidean norm $\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$

Matrix norms

- $A \in \mathbb{R}^{m \times n}$, $\| \cdot \|$ is a vector norm
- Matrix norm induced by this vector norm

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|$$

- Properties
  1. $\|A\| \geq 0$, and $\|A\| = 0$ iff $A = 0$, the zero matrix
  2. $\|\alpha A\| = |\alpha| \|A\|$, $\alpha \in \mathbb{R}$
  3. $\|A + B\| = \|A\| + \|B\|$, for any $A, B \in \mathbb{R}^{m \times n}$
  4. $\|AB\| \leq \|A\| \cdot \|B\|$, for any $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$

- Infinity norm, max row sum

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

- One norm, max column sum

$$\|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|$$

- Two norm

$$\|A\|_2 = \max_i \sqrt{\lambda_i(A^T A)},$$

where $\lambda_i(A^T A)$ is the ith eigenvalue of $A^T A$