× Friday – no lecture.

× HW1 & specify

CS450: Numerical Analysis

Howard Hao Yang

Assistant Professor, ZJU-UIUC Institute

18/10/2023

Linear Least Square

$a \cdot x = b \quad \in \mathbb{R}$

$a \neq 0 \rightarrow x = \frac{b}{a}$

- Suppose we want to solve a linear system $\boldsymbol{Ax} = \boldsymbol{b}$, with $\boldsymbol{A}$ and $\boldsymbol{b}$ being known, what does this mean?
  - What does $\boldsymbol{Ax}$ mean?
  - What does $\boldsymbol{Ax} = \boldsymbol{b}$ mean?

$\begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix} \cdot x = b$

Combination of $a_1 \cdots a_n \rightarrow b$

- What happens if $\boldsymbol{A}$ is not a square matrix?
- Instead of solving $\boldsymbol{Ax} = \boldsymbol{b}$, we aim to find $\boldsymbol{x}$ to minimize $\|\boldsymbol{Ax} - \boldsymbol{b}\|_2$
  - Generally, a perfect fitting may not be possible, and we look for an approximation
  - In data science, this is called Linear Regression

Solving the Linear Least Square

- How to find $\boldsymbol{x}^* = \text{argmin}_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$ where $\boldsymbol{A} \in R^{m \times n}$?
  - The residual error can be written as

$$Err(\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

  - Let $\frac{\partial Err(\boldsymbol{x})}{\partial \boldsymbol{x}} = 2A^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{0}$, which is equivalent to $\boxed{(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}}$
  - If $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is nonsingular, we obtain the solution as

$$\boldsymbol{x} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

  - This is known as the normal equation

assess the error

condition #

Recall: Matrix Condition Number

$A^{-1}A = AA^{-1} = I$

- Given a square matrix $\boldsymbol{A}$, the condition number is

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

- The usefulness of the condition number is in assessing the accuracy of solutions to linear systems
- In other words, it measures the danger of unacceptable round off error in solving linear systems (i.e. $\boldsymbol{Ax} = \boldsymbol{b}$)
- How can we define condition number for rectangular matrices?

Conditioning of Rectangular Matrix

$A^{-1}A = AA^{-1} = I$ $\quad I \leftarrow (A^{-1})Ax = A^{-1}b$

- Given an overdetermined system $\boldsymbol{Ax} = \boldsymbol{b}$ where $\boldsymbol{A} \in R^{m\times n}, m > n$ and $\boxed{\text{rank}(\boldsymbol{A}) = n}$,
  - recall that $\kappa(\boldsymbol{A}) = \|\boldsymbol{A}\|_2 \cdot \|\boldsymbol{A}^{-1}\|_2$
  - we can calculate $\|\boldsymbol{A}\|_2$, but $\boldsymbol{A}$ is not invertable, therefore, …
  - we define a pseudoinverse of $\boldsymbol{A}$ by $\boldsymbol{A}^+ = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$, note that $\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{I}$

$(A^TA)x = A^Tb$

$\square\, I = I$

$x = (A^TA)^{-1}A^Tb$

$\boxed{A} \to \|A\|_2\,?$

$\sigma_1, \sigma_2 \neq 0$

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

$$= \max\{\sigma_1, \sigma_2\} \to \sigma_{max}$$

$$\underset{3\times 2}{A} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{bmatrix} \xrightarrow{SVD} \underset{3\times 3}{\boxed{U}} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{bmatrix}_{3\times 2} \underset{2\times 2}{\square\, V^T}$$

$$A^+ = (\underset{2\times 2}{A^TA})^{-1}\underset{2\times 3}{A^T} \to 2\times 3$$

$$A^+ = (A^TA)^{-1}A^T$$

$$A^+A = (A^TA)^{-1}A^TA = I$$

$$AA^+ = ?\, I$$

Conditioning of Rectangular Matrix

$A^+A = I$ $\quad \begin{bmatrix} 1 & \\ & 0 \end{bmatrix}$

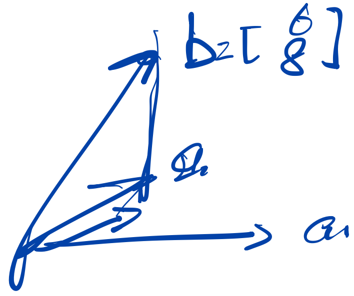
- What does $\boldsymbol{A}^+$ imply?
- For linear system $\boldsymbol{Ax} = \boldsymbol{b}$, a solution given by $\boxed{\boldsymbol{x}^+ = \boldsymbol{A}^+\boldsymbol{b}}$ is the minimum norm least square solution

$A^+ = (A^TA)^{-1}A^T$

- Example: The shortest least squares solution to the following

$$\begin{cases} 3\cdot x_1 + 0\cdot x_2 = 6 \\ 0\cdot x_1 + 0\cdot x_2 = 8 \end{cases} \rightarrow \underbrace{\overset{A}{\begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 8 \end{bmatrix}$$

$\mathrm{col}\left(\begin{bmatrix} 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right)$



is given by

$\|x^+\| \leq \|x'\|$

$$\boldsymbol{x}^+ = A^+\boldsymbol{b} = \overset{A^+}{\begin{bmatrix} 1/3 & 0 \\ 0 & 0 \end{bmatrix}} \begin{bmatrix} 6 \\ 8 \end{bmatrix} = \begin{bmatrix} \mathbf{2} \\ \mathbf{0} \end{bmatrix} \rightarrow \begin{bmatrix} 6 \\ 0 \end{bmatrix} \leftarrow \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Conditioning of Rectangular Matrix (cont'd)

ZJU-UIUC INSTITUTE

$\square \Rightarrow \kappa(A) = \frac{\lambda_{max}}{\lambda_{min}}$

$A \in \mathbb{R}^{m \times n}$ $\quad$ $\mathrm{rank}(A) = n$

- The condition number of a rectangular matrix $\boldsymbol{A}$ is defined as

$A^+ = (A^TA)^{-1}A^T$

$$\kappa(\boldsymbol{A}) = \|\boldsymbol{A}\|_2 \cdot \|\boldsymbol{A}^+\|_2$$

Δ σ_{max}

$A = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{bmatrix}$ $\quad$ 3×2, $\mathrm{rank}(A) = 2$

  - What if $\mathrm{rank}(\boldsymbol{A}) < n$? $\kappa(\boldsymbol{A}) = \infty$

?

  - We also define $\kappa(\boldsymbol{A}) = \sigma_{max}/\sigma_{min}$, is it still true here?

SVD: $A = U \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{bmatrix} V^T$ $\quad$ $U \in \mathbb{R}^{3\times 3}$, $V \in \mathbb{R}^{2\times 2}$ $\Rightarrow$ $A^TA = V \begin{bmatrix} \sigma_1^2 & \\ & \sigma_2^2 \end{bmatrix} V^T$

$(A^TA)^{-1} = V^T \begin{bmatrix} \frac{1}{\sigma_1^2} & \\ & \frac{1}{\sigma_2^2} \end{bmatrix} V$

$(A^TA)^{-1}A^T = V^T \begin{bmatrix} \frac{1}{\sigma_1} & \\ & \frac{1}{\sigma_2} \end{bmatrix} U$ $\Rightarrow$ $\|(A^TA)^{-1}A^T\|_2 = \max\left\{\frac{1}{\sigma_1}, \frac{1}{\sigma_2}\right\}$

$= \frac{1}{\sigma_{min}}$

$\Rightarrow \kappa(A) = \frac{\sigma_{max}}{\sigma_{min}}$

Example (1)

- Calculate the condition number of the following rectangular matrix $\boldsymbol{A} \in \mathbb{R}^{6\times 3}$

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cong \begin{bmatrix} 1237 \\ 1941 \\ 2417 \\ 711 \\ 1177 \\ 475 \end{bmatrix} = \boldsymbol{b}.$$

$$\Rightarrow \boldsymbol{A}^+ = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T = \frac{1}{4}\begin{bmatrix} 2 & 1 & 1 & -1 & -1 & 0 \\ 1 & 2 & 1 & 1 & 0 & -1 \\ 1 & 1 & 2 & 0 & 1 & 1 \end{bmatrix}$$

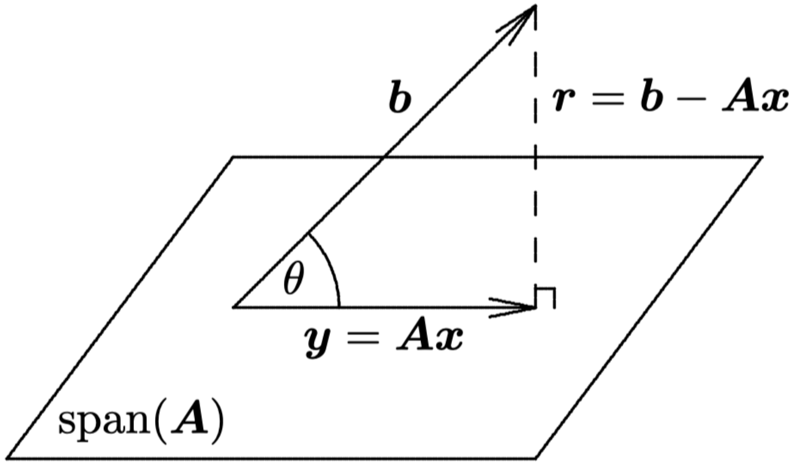
- Consequently, $\|\boldsymbol{A}\|_2 = 2$ and $\|\boldsymbol{A}^+\|_2 = 1$, which gives $\kappa(\boldsymbol{A}) = \|\boldsymbol{A}\|_2 \cdot \|\boldsymbol{A}^+\|_2 = 2$

Conditioning of Linear Least Square

$\boxed{A}x=\boxed{b} \rightarrow A(x+\Delta x) = b+\Delta b$

$\frac{\|\Delta x\|}{\|x\|} \le \frac{\|\Delta b\|}{\|b\|} \cdot \kappa(A)$

- Back to the problem of finding $\boldsymbol{x}^* = \text{argmin}_{\boldsymbol{x}} \|\boldsymbol{Ax} - \boldsymbol{b}\|_2$ where $\boldsymbol{A} \in R^{m \times n}$
- Unlike the conditioning of a linear system $\boldsymbol{Ax} = \boldsymbol{b}$, which only depends on the condition of the $\boldsymbol{A}$; the conditioning/sensitivity of the solution to an LSQ depends on both $\boldsymbol{A}$ and $\boldsymbol{b}$
  - It is more stable when $\boldsymbol{b}$ lies near span($\boldsymbol{A}$)
  - It is sensitive if $\boldsymbol{b}$ lies near orthogonal to span($\boldsymbol{A}$)



Conditioning of LSQ (cont'd)

$x = \arg\min \|Ax-b\|_2$

$\operatorname{rank}(A) = n$

$Ax = b$

$a = b \cdot \cos\theta$

$\frac{\|\Delta x\|}{\|x\|} \le \square \frac{\|\Delta b\|}{\|b\|}$?

- Recall that solution of the normal equation $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b}$ solves the LSQ
- For a perturbation on the RHS, it gives $\boldsymbol{A}^T\boldsymbol{A}(\boldsymbol{x}+\Delta\boldsymbol{x}) = \boldsymbol{A}^T(\boldsymbol{b}+\Delta\boldsymbol{b})$, i.e., $\boldsymbol{A}^T\boldsymbol{A}\Delta\boldsymbol{x} = \boldsymbol{A}^T\Delta\boldsymbol{b}$. As such, $\Delta\boldsymbol{x} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\Delta\boldsymbol{b} = \boldsymbol{A}^+\Delta\boldsymbol{b}$, and this leads to:

$$\left\{\begin{aligned} (A^TA)x &= A^Tb \\ (A^TA)(x+\Delta x) &= A^T(b+\Delta b) \end{aligned}\right.$$

$$(A^TA)\cdot\Delta x = A^T\Delta b$$

$$\Delta x = (A^TA)^{-1}A^T\Delta b = A^+\Delta b$$

$$\begin{aligned}
\frac{\|\Delta \boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} &\le \|\boldsymbol{A}^+\|_2 \frac{\|\Delta \boldsymbol{b}\|_2}{\|\boldsymbol{x}\|_2} \\
&= \operatorname{cond}(\boldsymbol{A}) \frac{\|\boldsymbol{b}\|_2}{\|\boldsymbol{A}\|_2 \cdot \|\boldsymbol{x}\|_2} \frac{\|\Delta \boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2} \\
&\le \operatorname{cond}(\boldsymbol{A}) \boxed{\frac{\|\boldsymbol{b}\|_2}{\|\boldsymbol{A}\boldsymbol{x}\|_2}} \frac{\|\Delta \boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2} \\
&= \operatorname{cond}(\boldsymbol{A}) \frac{1}{\cos(\theta)} \frac{\|\Delta \boldsymbol{b}\|_2}{\|\boldsymbol{b}\|_2}
\end{aligned}$$

$(\|\Delta x\|_2 \le \|A^+\|_2 \cdot \|\Delta b\|_2)$

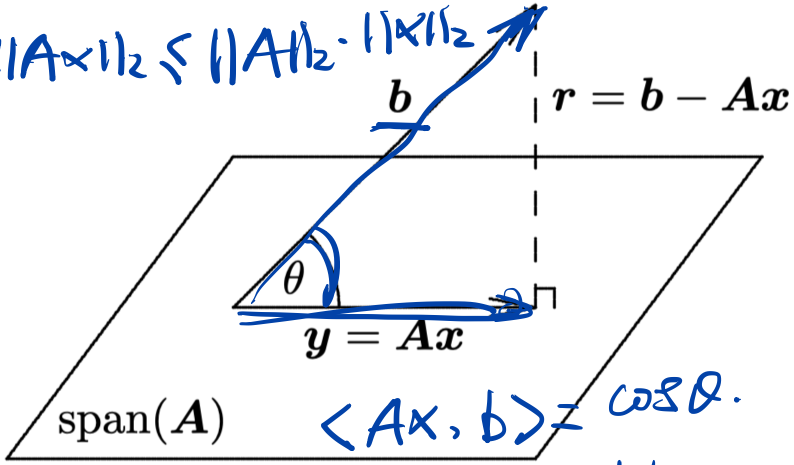
$\|A^+\|_2 = \frac{\kappa(A)}{\|A\|_2}$

$\|Ax\|_2 \le \|A\|_2 \cdot \|x\|_2$

Fix

large

$\frac{\|b\|_2}{\|Ax\|_2}$



$\frac{\langle Ax, b\rangle}{\|Ax\|_2 \cdot \|b\|_2} = \cos\theta$

- What determines θ?
- What values of θ are bad?

Solving Linear Systems

- For a square matrix $\boldsymbol{A}$, how to (systematically) solve for $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$?
  - Transform it into one whose solution is the same but easier to compute
- Example: solve for x_1, x_2, x_3, x_4 for the following

$$
\begin{aligned}
E_1 &: \quad x_1 + x_2 + 3x_4 = 4,\\
E_2 &: \quad 2x_1 + x_2 - x_3 + x_4 = 1,\\
E_3 &: \quad 3x_1 - x_2 - x_3 + 2x_4 = -3,\\
E_4 &: \quad -x_1 + 2x_2 + 3x_3 - x_4 = 4,
\end{aligned}
$$

$E_2 - 2\times E_1$

$E_3 - 3\times E_1$

$E_4 + E_1$

$E_2: \cancel{2x_1} \; -x_2 - x_3 - 5x_4 = -7$

$n\times n$

$a_{11}\cdot x_1 + a_{12}x_2 + \cdots + a_{1n}\cdot x_n = b_1$

$Ax = b$

$$
\left[\begin{array}{c|ccc}
a_{11} & a_{12} & \cdots & a_{1n}\\
\hline
0 & & & \\
0 & & A_{n-1} & \\
\vdots & & & \\
0 & & &
\end{array}\right] x = b'
$$

$A_{n-1}\cdot x_{n-1} = b_{n-1}$

$n-1 < n-1$

ZJU-UIUC INSTITUTE

Solving Linear Systems (cont'd)

$Ax = b \Rightarrow MAx = Mb$

- Example: solve for x_1, x_2, x_3, x_4 for the following

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

$$\begin{aligned}
E_1 &: \quad x_1 + x_2 + 3x_4 = 4, \\
E_2 &: \quad 2x_1 + x_2 - x_3 + x_4 = 1, \\
E_3 &: \quad 3x_1 - x_2 - x_3 + 2x_4 = -3, \\
E_4 &: \quad -x_1 + 2x_2 + 3x_3 - x_4 = 4,
\end{aligned}$$

$$x = (MA)^{-1} Mb = A^{-1} M^{-1} M b = A^{-1} b$$

- Step #1: use E_1 to eliminate the unknown x_1 from E_2, E_3, and E_4

$$\begin{aligned}
E_1 &: \quad x_1 + x_2 + 3x_4 = 4, \\
E_2 &: \quad -x_2 - x_3 - 5x_4 = -7, \\
E_3 &: \quad -4x_2 - x_3 - 7x_4 = -15, \\
E_4 &: \quad 3x_2 + 3x_3 + 2x_4 = 8.
\end{aligned}$$

- Q: why the solutions remained the same after this process?

Solving Linear Systems

- Given a square matrix $\boldsymbol{A}$, if we want to solve for $\boldsymbol{Ax} = \boldsymbol{b}$, then premultiply both sides with any non-singular matrix $\boldsymbol{M}$ would not affect the solution
- Why?: solution to $\boldsymbol{MAx} = \boldsymbol{Mb}$ is given by

$$\boldsymbol{x} = (\boldsymbol{MA})^{-1}\boldsymbol{Mb} = \boldsymbol{A}^{-1}\boldsymbol{M}^{-1}\boldsymbol{Mb} = \boldsymbol{A}^{-1}\boldsymbol{b}$$

Solving Linear Systems (cont'd)

- Step #1: use E_1 to eliminate the unknown x_1 from E_2, E_3, and E_4

$$\begin{aligned}
E_1 &: \quad x_1 + x_2 \qquad\quad + 3x_4 = 4,\\
E_2 &: \qquad\ - x_2 - x_3 - 5x_4 = -7,\\
E_3 &: \qquad - 4x_2 - x_3 - 7x_4 = -15,\\
E_4 &: \qquad\quad 3x_2 + 3x_3 + 2x_4 = 8.
\end{aligned}$$

- Continue the process, and finally reduce to the following

$$\begin{aligned}
E_1 &: \quad x_1 + x_2 \qquad + 3x_4 = 4,\\
E_2 &: \qquad - x_2 - x_3 - 5x_4 = -7,\\
E_3 &: \qquad\qquad 3x_3 + 13x_4 = 13,\\
E_4 &: \qquad\qquad\qquad - 13x_4 = -13.
\end{aligned}$$

$x_4 = \frac{-13}{-13} = 1$

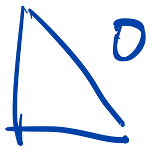
- The matrix is in triangular form, can be solved via backward-substitution

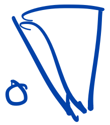
Triangular Linear Systems

- What type of linear system is easy to solve?
  - If one equation in the system involves only one component of solution, then that component can be computed by division
  - If another equation in system involves only one additional solution component, then by substituting one known component into it, we can solve for other component
  - If this pattern continues, all components of solution can be computed in succession
  - System with this property is called triangular

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= a_{1,n+1} \\
a_{22}x_2 + \cdots + a_{2n}x_n &= a_{2,n+1} \\
\ddots \qquad \vdots \quad & \quad\ \vdots \\
a_{nn}x_n &= a_{n,n+1}
\end{aligned}
$$

Triangular Matrices





$Ax = b \Rightarrow Lx = b'$

$Ux = \bar{b}$

- Two specific triangular forms are of particular interest
  - Lower triangular: all entries above main diagonal are zero, $a_{ij} = 0$ for $i < j$
  - Upper triangular: all entries below main diagonal are zero, $a_{ij} = 0$ for $i > j$
- Successive substitution process described earlier is especially easy to formulate for lower or upper triangular systems
- Any triangular matrix can be permuted into upper or lower triangular form by suitable row permutation

Solving Linear Systems (cont'd)

- Example: Solve the following triangular system

$$\begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 8 \end{bmatrix}$$

  - Using back-substitution for this upper triangular system, last equation, $4x_3 = 8$, is solved directly to obtain $x_3 = 2$
  - Next, x_3 is substituted into second equation to obtain $x_2 = 2$
  - Finally, both x_3 and x_2 are substituted into first equation to obtain $x_1 = -1$

Elimination

- To transform general linear system into triangular form, we need to replace selected nonzero entries of matrix by zeros
- This can be accomplished by taking linear combinations of rows
- Consider a two-dimensional vector $\boldsymbol{a} = [a_1\ a_2]$
- If $a_1 \neq 0$, then

$$\begin{bmatrix} 1 & 0 \\ -a_2/a_1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ 0 \end{bmatrix}$$

Elementary Elimination Matrices

- More generally, we can annihilate all entries below k-th in a n-dimensional vector $\boldsymbol{a}$ by transformation

$$\boldsymbol{M}_k \boldsymbol{a} = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & -m_{k+1} & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -m_n & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_k \\ a_{k+1} \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $m_i = \frac{a_i}{a_k}$, $i = k+1, \dots, n$

- Divisor a_k, called pivot, must be nonzero

Elementary Elimination Matrices (cont'd)

Example: For a vector $\boldsymbol{a} = [2, 4, -2]^T$, the elementary elimination matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are (recall: $m_i = \frac{a_i}{a_k}$, $i = k + 1, \ldots, n$)

$$\boldsymbol{M}_1\boldsymbol{a} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

$$\boldsymbol{M}_2\boldsymbol{a} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix}$$

Elementary Elimination Matrices (cont'd)

- Matrix $\boldsymbol{M}_k$, called elementary elimination matrix, adds multiple of row k to each subsequent row, with multipliers m_i chosen so that result is zero
- $\boldsymbol{M}_k$ is unit lower triangular and nonsingular
- $\boldsymbol{M}_k = \boldsymbol{I} - \boldsymbol{m}_k \boldsymbol{e}_k^T$, where $\boldsymbol{m}_k = [0, \ldots, 0, m_{k+1}, \ldots, m_n]^T$ and $\boldsymbol{e}_k$ is the k-th column of identity matrix
- $\boldsymbol{M}_k^{-1} = \boldsymbol{I} + \boldsymbol{m}_k \boldsymbol{e}_k^T$, which means $\boldsymbol{M}_k^{-1} = \boldsymbol{L}_k$ is the same as $\boldsymbol{M}_k$ except signs of multipliers are reversed

Elementary Elimination Matrices (cont'd)

- If $\boldsymbol{M}_j$, $j > k$, is another elementary elimination matrix, with vectors of multi pliers $\boldsymbol{m}_j$, then

$$\boldsymbol{M}_k\boldsymbol{M}_j = \left(\boldsymbol{I} - \boldsymbol{m}_k\boldsymbol{e}_k^T\right)\left(\boldsymbol{I} - \boldsymbol{m}_j\boldsymbol{e}_j^T\right)$$

$$= \boldsymbol{I} - \boldsymbol{m}_k\boldsymbol{e}_k^T - \boldsymbol{m}_j\boldsymbol{e}_j^T + \boldsymbol{m}_k\boldsymbol{e}_k^T\boldsymbol{m}_j\boldsymbol{e}_j^T$$

$$= \boldsymbol{I} - \boldsymbol{m}_k\boldsymbol{e}_k^T - \boldsymbol{m}_j\boldsymbol{e}_j^T$$

where means product is essentially "union", and similarly for product of inverses, $\boldsymbol{L}_k\boldsymbol{L}_j$