Announcement

- Wed (4/13) lecture will be on Zoom (by Prof. Ratliff)
- will start Module 2

# EE445 Mod1-Lec2: Linear Algebra V

References:

- [VMLS]: Chapter 11
- [OM] by Calafiore & El-Ghaoui: 3.3

# Inverse of $A$

- if matrix $A$ has both a left-inverse and a right-inverse, they are unique and equal
  - $A$ must be square
  - we say $A$ is *invertible* or non-singular $(\det(A) \neq 0)$
- to see this: if $AX = I$ and $YA = I$,

$$X = (\underbrace{YA}_{I})X = Y(\underbrace{AX}_{I}) = Y$$

- inverse of product: $(AB)^{-1} = B^{-1}A^{-1}$

(intuitively, order is reversed since we're reversing the role of input & output, or row & column)

# Inverse of $A$

- for a square matrix $A$, the following are equivalent:
  - $A$ is invertible
  - columns of $A$ are linearly independent
  - rows of $A$ are linearly independent
- examples:
  - if $Q$ is square with $Q^TQ = I$, then $\underline{Q^{-1} = Q^T}$
  - for a $2 \times 2$ matrix $A$ with $\det(A) = \underline{A_{11}A_{22} - A_{21}A_{12} \neq} 0$,

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}$$

# Inverse via QR factorization

$$\underset{n\times n}{A} = QR = \underset{n\times n}{\begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}} \underset{n\times n}{\begin{bmatrix} \ddots & & \\ & \ddots & \\ 0 & & \ddots \end{bmatrix}}$$

- if $A$ is invertible, $Ax = b$ has the unique solution $x = A^{-1}b$ for any $b$
- if $A = QR$, the inverse is given by

$$A^{-1} = (QR)^{-1} = R^{-1}Q^{-1} = R^{-1}Q^T$$

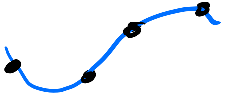
$$Ax = b$$
$$QRx = b$$
$$Rx = Q^T b$$

- easy way to solve for $x$:
  1. compute the QR factorization $A = QR$
  2. compute $Q^T b$
  3. solve the triangular equation $Rx = Q^T b$ using back-substitution

$$\begin{bmatrix} \ddots & & & \\ & \ddots & & \\ & & \circ & \circ \\ 0 & & & \circ \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = Q^T b$$

- solve starting from last row $\Rightarrow x_n$
- then move up to solve for $x_{n-1}, \dots, x_1$

# Ex: polynomial interpolation



let's find coefficients of polynomial $p(x) = c_1 + c_2 x + c_3 x^2 + c_4 x^3$ that satisfies

$x = -1.1$

$$p(-1.1) = b_1, \quad p(-0.4) = b_2, \quad p(0.1) = b_3, \quad p(0.8) = b_4$$

write as $Ac = b$ with

$$\underbrace{\begin{bmatrix} 1 & (-1.1) & (-1.1)^2 & (-1.1)^3 \\ 1 & (-0.4) & (-0.4)^2 & (-0.4)^3 \\ 1 & (0.1) & (\ \ )^2 & (\ \ )^3 \\ 1 & (0.8) & (\ \ )^2 & (\ \ )^3 \end{bmatrix}}_{A} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

# Ex: polynomial interpolation

Vandermonde matrix:

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 & \dots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \dots & t_n^{n-1} \\ \vdots & \vdots & & \vdots & \\ 1 & t_n & t_n^2 & \dots & t_n^{n-1} \end{bmatrix}_{n \times n}$$

assume $t_i \neq t_j$ for $i \neq j$

we show $A$ is invertible, by showing if $Ay = 0$ then $y = 0$

- $\underset{n \times n}{A}y = 0$ means $p(t_1) = \ldots = p(t_n) = 0$ where $p(t)$ is polynomial of degree $n-1$ or less:

$$p(t) = \underline{y_1} + \underline{y_2}t + \underline{y_3}t^2 + \ldots + \underline{y_n}t^{n-1}$$

- if $y \neq 0$, $p(t)$ cannot have more than $n-1$ distinct real roots
- so $\underline{p(t_1) = \ldots = p(t_n) = 0}$ only possible if $\underline{y = 0}$

# Ex: polynomial interpolation

$i^{th}$ row coeff's show how $c_i$ depends on $b_1, \ldots, b_n$ – small coeff means $c_i$ is not very sensitive

- coefficients given by $c = A^{-1}b$ with

$$A^{-1} = \begin{bmatrix} -0.0370 & 0.3492 & 0.7521 & -0.0643 \\ 0.1388 & -1.8651 & 1.6239 & 0.1023 \\ 0.3470 & 0.1984 & -1.4957 & 0.9503 \\ -0.5784 & 1.9841 & -2.1368 & 0.7310 \end{bmatrix}$$

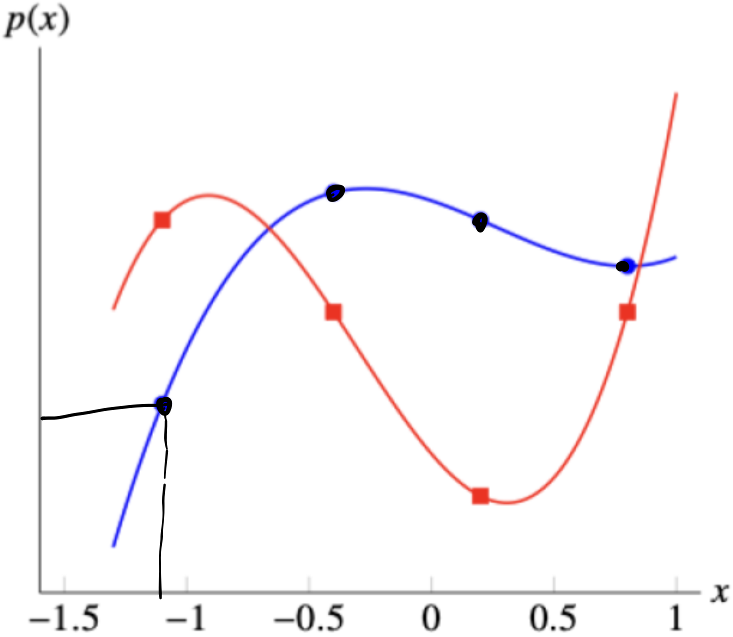
- observe, e.g., $c_1$ is not very sensitive to $b_1$ or $b_4$ $\rightsquigarrow$ because $(A^{-1})_{11}$ and $(A^{-1})_{14}$ are small
- first col gives coeffs of polynomial that satisfies

$$p(-1.1) = 1, \quad p(-0.4) = 0, \quad p(0.1) = 0, \quad p(0.8) = 0$$

called (first) *Lagrange polynomial*

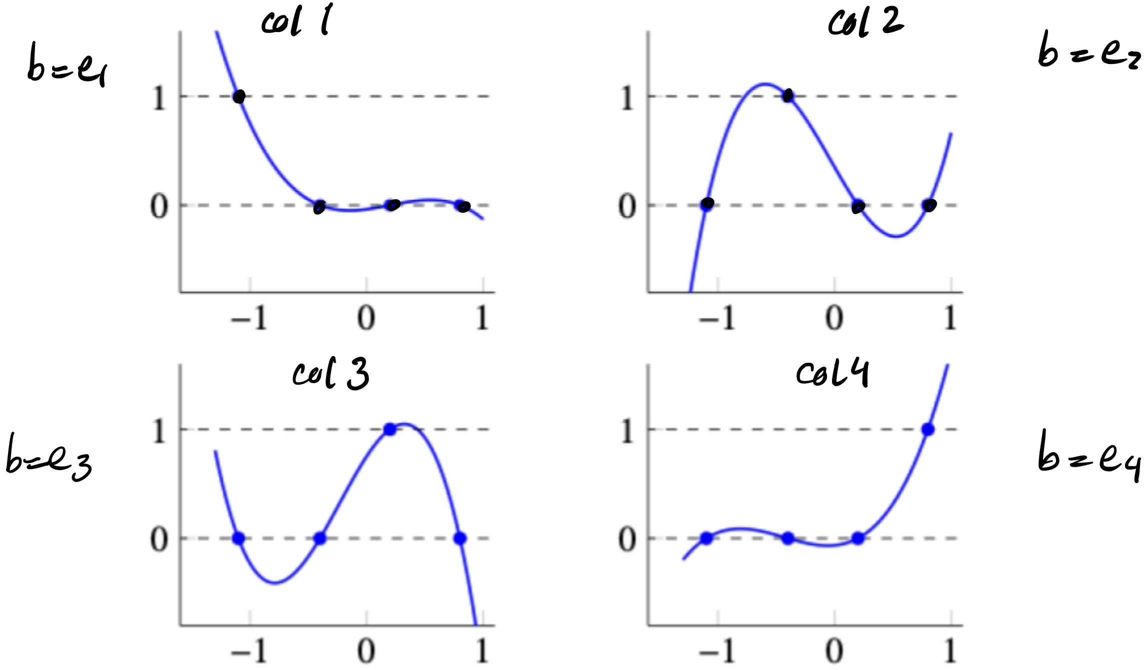
$$A^{-1}e_1 = a_1 \Rightarrow Aa_1 = e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

# Example



# Lagrange polynomials

Lagrange polynomials corresponding to points -1.1, -0.4, 0.2, 0.8



# Invertibility of Gram matrix

$Ax=0$ only if $x=0$

$A$ has linearly independent columns if and only if $A^TA$ (Gram matrix of $A$) is invertible

- to see this, we'll show $\quad Ax = 0 \Leftrightarrow A^TAx = 0$
- $\Rightarrow$: if $Ax = 0$ then $(A^TA)x = A^T(Ax) = A^T0 = 0$
- $\Leftarrow$: if $(A^TA)x = 0$ then

$$0 = \underbrace{x^T\underbrace{(A^TA)x}_{=0}}_{\underbrace{(x^TA^T)}_{(Ax)^T}(Ax)} = (Ax)^T(Ax) = \|Ax\|^2 = 0 \quad \Rightarrow Ax=0$$

# Pseudo-inverse of tall matrix

$$A = \begin{bmatrix} | & | & | \\ | & | & | \end{bmatrix}_{m\times n} = \underset{m\times n}{Q}\, R_{n\times n}$$

- for $A$ with linearly independent cols, the pseudo-inverse is

$$\underset{\text{dagger}}{A^\dagger} = (A^T A)^{-1} A^T$$

- it is a left-inverse of $A$:

$$A^\dagger A = (A^T A)^{-1} A^T A = I$$

- reduces to $A^{-1}$ when $A$ is square $\longrightarrow$ $A^\dagger = A^{-1} \underbrace{A^{-T} A^T}_{=I} = A^{-1}$
- in terms of QR factorization: $\quad A^\dagger = R^{-1} Q^T$

$$A = \left[\quad\right]_{m\times n} = \overset{Q}{\left[\quad\right]}\overset{R}{\left[\quad\right]}$$

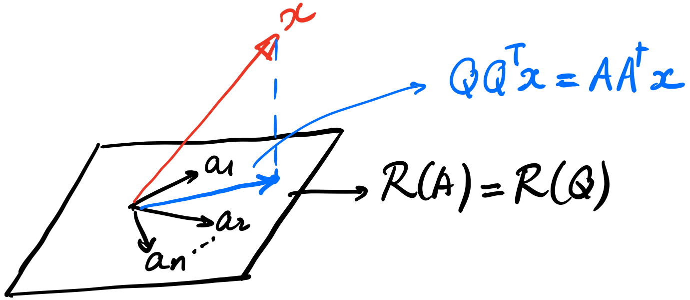
# Projection on range

- for $A$ with linearly independent cols, combining $A = QR$ and $A^\dagger = R^{-1}Q^T$ gives

$$AA^\dagger = QRR^{-1}Q^T = QQ^T = \left[\;\right]\left[\quad\right]$$

(note order of product in $AA^\dagger$ and difference with $A^\dagger A = I$)

- $QQ^Tx$ gives the *orthogonal projection* of $x$ on the range of $Q$ (we'll see more in Module 3):



# Pseudo-inverse of wide matrix

- similarly, if $A$ is wide, with linearly independent rows, $AA^T$ is invertible. pseudo-inverse is defined as

$$A^\dagger = A^T(AA^T)^{-1}$$

- it is a right-inverse of $A$ (can check)
- reduces to $A^{-1}$ when $A$ is square

# Psuedo-inverse of a general matrix

$A_{m\times n} = B_{m\times r} \, C_{r\times n}$

suppose $A$ is $m \times n$ with rank $r$ ($r < \min\{m, n\}$), so has a factorization $A = BC$

- $B$ is $m \times r$ with linearly independent columns, its psuedo-inverse is:

$$B^\dagger = (B^T B)^{-1} B^T$$

- $C$ is $r \times n$ with linearly indepedent rows, its psuedo-inverse is:

$$C^\dagger = C^T (CC^T)^{-1}$$

we define the psuedo-inverse of $A$ as $\quad A^\dagger = C^\dagger B^\dagger \qquad A^\dagger = (BC)^\dagger = C^\dagger B^\dagger$

- extends def of psuedo-inverse to non-full-rank matrices. also known as *Moore-Penrose (generalized) inverse*.
- (later, we'll also give expression in terms of SVD. . . )

# Ex: psuedo-inverse of diagonal matrix

- rank of diagonal matrix= # of nonzero diagonal entries
- $A^\dagger$ is a diagonal matrix with

$$(A^\dagger)_{ii} = \begin{cases} 1/A_{ii} & \text{if } A_{ii} \neq 0 \\ 0 & \text{if } A_{ii} = 0 \end{cases}$$

example:

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix} \qquad A^\dagger = \begin{bmatrix} \underline{-1} & 0 & 0 & 0 \\ 0 & \underline{1/2} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \underline{-1/3} \end{bmatrix}$$

# Meaning of $AA^\dagger$ and $A^\dagger A$

$A_{m\times n} = B_{m\times r} C_{r\times n}$

$r = \text{rank}(A)$

if $A$ does not have full rank, $A^\dagger$ is not a left or right inverse

- **interpretation of $AA^\dagger$:**

$$AA^\dagger = \underbrace{B}\overbrace{C C^\dagger}^{=I}\underbrace{B^\dagger} = BB^\dagger = B(B^TB)^{-1}B^T$$

- $BB^\dagger$ gives the orthogonal projection on $\mathcal{R}(B)$ (and $\mathcal{R}(A) = \mathcal{R}(B)$)
- **interpretation of $A^\dagger A$:**

$$A^\dagger A = C^\dagger \overbrace{B^\dagger B}^{=I} C = C^\dagger C = C^T(CC^T)^{-1}C$$

- orthogonal projection onto $\mathcal{R}(A^T) = \mathcal{R}(C^T)$

# Eigenvalues & eigenvectors

a nonzero vector $x$ is an eigenvector of the $n \times n$ matrix $A$, with eigenvalue $\lambda$, if

$$Ax = \lambda x \qquad (A - \lambda I)x = 0$$

- the matrix $\lambda I - A$ is singular, $x$ is a (nonzero) vector in $\mathcal{N}(\lambda I - A)$
- the eigenvalues of $A$ are the roots of the characteristic polynomial:

$$\det(\lambda I - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + (-1)^n \det(A) = 0$$

- the polynomial roots (and eigenvectors) may be complex
- there are exactly $n$ eigenvalues (counted with their multiplicity)
- set of eignevalues called the *spectrum* of $A$

# Similarity transform

also known as 'coordinate change' matrix

two matrices $A$ and $B$ are *similar* if $\quad B = T^{-1}AT \quad$ for some nonsingular matrix $T$

det(ABC) = det(CAB)
↪ a property of det

- similarity transforms preserve eigenvalues:

$$\det(\lambda I - B) = \det(\lambda \overbrace{I}^{T^{-1}T} - T^{-1}AT) = \det(T^{-1}(\lambda I - A)T) = \det(\lambda I - A)$$

- if $x$ is an eigenvector of $A$ then $y = T^{-1}x$ is an eigenvector og $B$:

scalar

$$By = (T^{-1}AT)(T^{-1}x) = T^{-1}Ax = T^{-1}(\lambda x) = \lambda y$$

special interest will be *orthogonal* similarity transforms

$T^{-1} = (T)^T$

# Diagonalizable matrices

- a matrix is diagonalizable if it is similar to a diagonal matrix:

$$T^{-1}AT = \Lambda$$

  for some nonsingular matrix $T$

- diagonal entries of $\Lambda$ are the eigenvalues of $A$
- cols of $T$ are eigenvectors of $A$:

$$A(Te_i) = T\Lambda e_i = \Lambda_{ii}(Te_i)$$

- cols of $T$ give $n$ linearly independent eigenvectors
- (not all square matrices are diagonalizable)

# Spectral decomposition

suppose $A$ is diagonalizable, with

$$A = T^{-1}\Lambda T \quad = \quad \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} \begin{bmatrix} w_1^T \\ w_2^T \\ \vdots \\ w_n^T \end{bmatrix}$$

$$= \quad \lambda_1 v_1 w_1^T + \lambda_2 v_2 w_2^T + \ldots + \lambda_n v_n w_n^T$$

this is a spectral decomposition of the linear function $f(x) = Ax$

- entries of $T^{-1}x$ are coeffs of $x$ in the eigenvector basis $\{v_1, \ldots, v_n\}$:

$$x = TT^{-1} = (w_i^T x)v_1 + \ldots + (w_n^T x)v_n$$

- by superposition for $f(x) = Ax$, $Ax = (w_1^T x)\lambda_1 v_1 + \ldots + (w_n^T x)\lambda_n v_n = T\Lambda T^{-1}x$

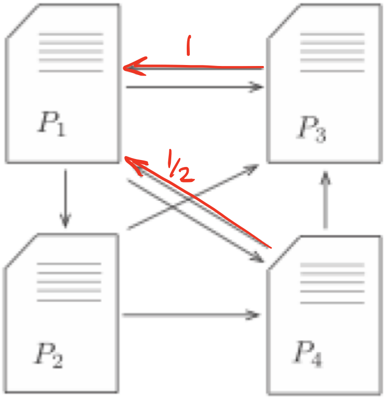
# Ex: Google PageRank

- ranks page importance (for display order)
  - nodes: pages
  - edges: directed link from $k_1$ to $k_2$, if page $k_1$ contains a link to $k_2$

[OM, example 3.5]

- say web is composed of $n$ pages, labeled with $j = 1, \dots, n$, and model as a directed graph.
- denote by $x_j$ the importance score (or "voting power") of page $j$, to be evenly divided among outgoing links from node $j$: $x_j / n_j$
- let $B_k$ be set of "backlinks" for page $k$ (pages point to $k$). score of page $k$ is:

$$x_k = \sum_{j_k \in B_k} \frac{x_j}{n_j}, \quad k = 1, \dots, n$$




$$\begin{cases} x_1 = x_3 + \frac{1}{2} x_4 \\ x_2 = \frac{1}{3} x_1 \\ x_3 = \frac{1}{3} x_1 + \frac{1}{2} x_2 + \frac{1}{2} x_4 \\ x_4 = \frac{1}{3} x_1 + \frac{1}{2} x_2 \end{cases}$$

# Ex: Google PageRank

# of outgoing edges from page

- in figure, we have $n_1 = 3,\ n_2 = 2,\ n_3 = 1,\ n_4 = 2$, hence we get system of linear eq's:

$$x = Ax, \quad A \begin{bmatrix} 0 & 0 & 1 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

- $A$ is called *link matrix*. $x$ is eignevector of $A$ associated with $\lambda = 1$
- here we get

$$x = v_1 = (12,\ 4,\ 9,\ 6)$$

- thus page 1 appears to be the most relevant according to PageRank scoring (larger score)
- real-world challenge: computing $v_1$ for a HUGE matrix...