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 $YA = I$,

$$X = (YA)X = Y(AX) = Y$$

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

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 $Q^T Q = I$, then $\quad Q^{-1} = Q^T$
  - for a 2×2 matrix A with $\det(A) = A_{11}A_{22} - A_{21}A_{12} \neq 0$,

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

Inverse via QR factorization

- 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

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

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

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

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

write as $Ac = b$ with

Ex: polynomial interpolation

Vandermonde matrix:

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

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

- $Ay = 0$ means $p(t_1) = \ldots = p(t_n) = 0$ where $p(t)$ is polynomial of degree $n-1$ or less:

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

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

Ex: polynomial interpolation

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

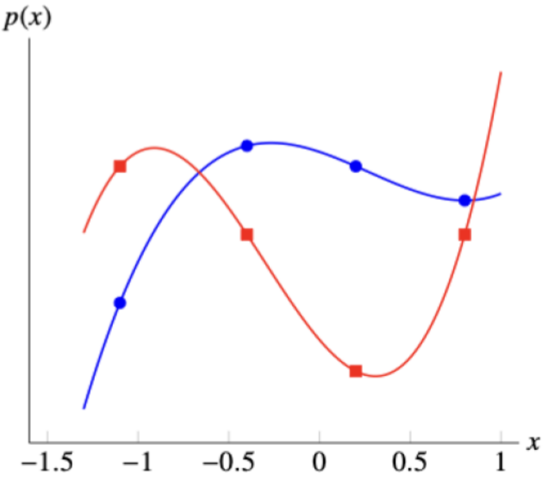
$$A^{-1} = \begin{bmatrix} -0.0370 & 0.3492 & 0.7521 & -0.0643 \\ 0.1388 & -1.8651 & 1.6239 & 0.1023 \\ 0.3470 & 0.1984 & -1.4957 & 0.9503 \\ -0.5784 & 1.9841 & -2.1368 & 0.7310 \end{bmatrix}$$

- observe, e.g., c_1 is not very sensitive to b_1 or b_4
- first col gives coeffs of polynomial that satisfies

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

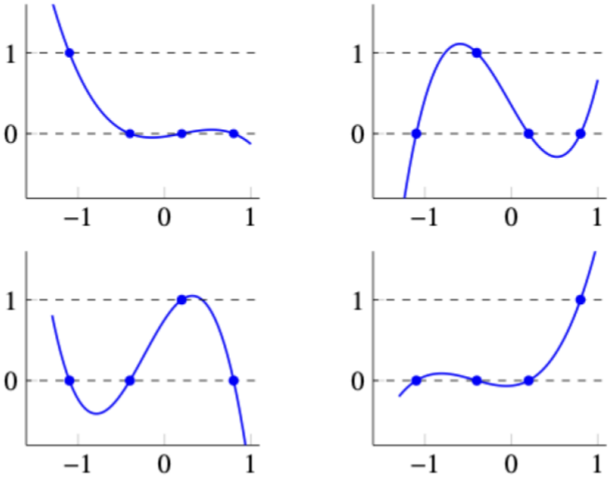
called (first) *Lagrange polynomial*

Example



Lagrange polynomials

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



Invertibility of Gram matrix

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

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

$$0 = x^T(A^TA)x = (Ax)^T(Ax) = \|Ax\|^2 = 0$$

Pseudo-inverse of tall matrix

- for A with linearly independent cols, the psuedo-inverse is

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

- 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
- in terms of QR factorization: $A^\dagger = R^{-1} Q^T$

Projection on range

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

$$AA^\dagger = QRR^{-1}Q^T = QQ^T$$

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

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:

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

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

- 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: pseudo-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:

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

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

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

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

$$AA^\dagger = BCC^\dagger B^\dagger = BB^\dagger = B(B^TB)^{-1}B^T$$

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

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

- 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 λ, if

$$Ax = \lambda x$$

- 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

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

- similarity transforms preserve eigenvalues:

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

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

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

special interest will be *orthogonal* similarity transforms

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 Λ 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

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

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

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 $\{v_1, \ldots, v_n\}$:

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

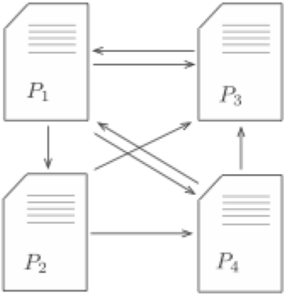
- by superposition for $f(x) = Ax$, $Ax = (w_1^T x)\lambda_1 v_1 + \ldots + (w_n^T x)\lambda_n v_n = T\Lambda T^{-1}x$

Ex: Google PageRank

[OM, example 3.5]

- say web is composed of n pages, labeled with $j = 1, \ldots, n$, 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 k: x_j/n_j
- let B_k be set of “backlinks” for page k (pages point to k). score of page k is:

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



Ex: Google PageRank

- 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