Goal

1) Notation
2) Linear Algebra

I. Notation

Remark: 1) Let $\mathbb{N}$ denote the set of natural numbers including 0.
2) Let $\mathbb{Z}$ denote the set of integers
3) Let $\mathbb{R}$ denote the set of reals
4) Let $\mathbb{R}^n$ denote the set of ntuples of real numbers
5) Let $\mathbb{C}$ denote the complex numbers.

Def: Given two sets A and B we say that

a) $f: A \to B$ is a function if $\forall a \in A$ (for all) (in) f assigns one and only one element in B denoted $f(a) \in B$ called the value of f at a

b) A is called the domain of f

c) B is called the codomain of f

Ex: $f: \mathbb{N} \to \mathbb{R}$ $\quad f(k) = (1/2)^k$

domain is $\mathbb{N}$

Codomain is $\mathbb{R}$

f only takes values between $(0,1]$

II. Linear Algebra

Remark: a) Given $P \in \mathbb{R}^{m \times n}$, we say that $[P]_{ij}$ is the entry in row i and column j.

b) We will denote matrices in capital letters, vectors in lower case, and scalars in greek letters.

Def: a) The identity matrix is a square matrix with all diagonal entries equal to 1 and all off-diagonal entries equal to 0.

Given $P, Q \in \mathbb{R}^{m \times n}$, $R \in \mathbb{R}^{n \times p}$ and $S \in \mathbb{R}^{n \times n}$, then

b) $P + Q \in \mathbb{R}^{m \times n}$ where $[P+Q]_{ij} = [P]_{ij} + [Q]_{ij}$

c) The transpose of P or $P^T \in \mathbb{R}^{n \times m}$ where $[P^T]_{ji} \doteq [P]_{ij}$

d) The trace of S or $tr(S) \in \mathbb{R}$ where $tr(S) = \sum_{i=1}^{n} [S]_{ii}$

e) $PR \in \mathbb{R}^{m \times p}$ where $[PR]_{ik} = \sum_{j=1}^{n} [P]_{ij} [R]_{jk}$

Note: $(PR)^T = R^T P^T$

A. Linear Independence

Def: A set of vectors $\{v_i \in \mathbb{R}^n\}_{i=1}^m$ is called **linearly independent** if and only if (iff) $\sum_{i=1}^n \alpha_i v_i = 0$ implies ($\Rightarrow$) $\alpha_i = 0 \ \forall i$. Otherwise the set of vectors is called **linearly dependent**.

Ex: (a) $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ is linearly independent

(b) $\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ is linearly dependent

(c) $\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ -3 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \\ 0 \end{pmatrix} \right\}$ is linearly dependent

Def: Let $A \in \mathbb{R}^{m \times p}$. The **rank** of A denoted $rk(A)$ is the maximum number of linearly independent columns or rows of A.

Ex: (a)
$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ rk(A) = 3$$

(b) $A = \left[\begin{array}{c|c|c} 1 & 2 & 1 \\ -2 & -3 & 1 \end{array} \right.$, $rk(A) = 2$

$$\left[\begin{array}{c|c|c} 2 & 3 & 1 \\ 3 & 5 & 0 \end{array}\right]$$

Note: MATLAB command rank

Ex: Let $A = \begin{bmatrix} | & | & | & & | \\ a_1 & a_2 & a_3 & \cdots & a_n \\ | & | & | & & | \end{bmatrix}$ be rank deficient then $\exists \{\alpha_i\}$ such that $\sum \alpha_i a_i = 0$ with $\alpha_i \neq 0$. This can be written down as

$$A \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = 0$$

same

Theorem: A is full rank iff $Av = 0 \Rightarrow v = 0$

Ex: $A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$

$Av = 0$ if $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ then $Av = 0$

so A is not full rank.

What does it mean to be rank deficient?

$S \in \mathbb{R}^{n \times n}$ $\quad rk(S) < n$ then it is rank deficient

Def: The kernel of $A \in \mathbb{R}^{n \times p}$ denoted $\ker(A) =$

$\{x \in \mathbb{R}^p \mid Ax \overset{?}{=} 0\}$

Ex: a) $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\ker(A) = \{x \in \mathbb{R}^2 \mid \begin{pmatrix} 0 \\ \alpha \end{pmatrix}, \alpha \in \mathbb{R}\}$

Note the dimension of the kernel of A is 1 since one linearly independent vector describes all elements of ker(A).

b) $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $\ker(A) = \{0\}$

which has dimension 0.

Theorem: (Rank-Nullity Theorem) Let $A \in \mathbb{R}^{n \times n}$ then $rk(A) + \dim(\ker(A)) = n$

Ex: (a) $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ $rk(A) = 2 - \dim(\ker(A))$
$= 1$

(b) $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ $rk(A) = 3 - \dim(\ker(A))$
$= 3$

B. Determinant

Def: The determinant of $S \in \mathbb{R}^{n \times n}$ denoted $\det(S)$

or $|S|$ is defined as

$$\det(S) = \sum_{j=1}^{n} (-1)^{j+1} [S]_{1j} \det(S_{1j})$$

where $S_{1j} \in \mathbb{R}^{(n-1)\times(n-1)}$ is obtained by deleting the first row and j^{th} column of S, $\det(\alpha) = \alpha$, for $\alpha \in \mathbb{R}$.

Ex: $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 4 & 9 \end{bmatrix}$ $\det(A) = 2 * \det\left(\begin{bmatrix} 3 & 4 \\ 4 & 9 \end{bmatrix}\right)$

$= 2 * (3 * 9 - 4 * 4)$

$= 22$

Note: Use MATLAB det feature

Lemma: $\det(A) \neq 0$ iff A is full rank.

C. Eigenvalues

Def: The eigenvalues of $S \in \mathbb{R}^{n \times n}$ are scalars $\lambda \in \mathbb{C}$ s.t. $Sv = \lambda v$ for some nonzero $v \in \mathbb{C}^n$ which is called an eigenvector.

Note: The eigenvectors correspond to the preferred directions of the matrix.

Ex: Find the eigenvalues of $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 4 & 9 \end{bmatrix}$

$$Av = \lambda v \Rightarrow (A - \lambda I)v = 0$$

$$A(v - \lambda v) = 0$$

is nonzero

$$\det(A - \lambda I) = 0$$

$$\begin{vmatrix} 2-\lambda & 0 & 0 \\ 0 & 3-\lambda & 4 \\ 0 & 4 & 9-\lambda \end{vmatrix} = -\lambda^3 + 14\lambda^2 - 35\lambda + 2$$

$$\Downarrow$$

$$\lambda = 1, 2, 11$$

$$(A - \lambda I)v = 0 \text{ and } v \neq 0$$

$$\dim(\ker(A - \lambda I)) \neq 0$$

$$\mathrm{rk}(A - \lambda I) + \underline{\dim(\ker(A - \lambda I))} = n$$

Note: There are elegant algorithms to compute eigenvalues (eig in MATLAB)

Def: Given a matrix $S \in \mathbb{R}^{n \times n}$, the degree n polynomial $\det(S - \lambda I)$ is called the characteristic polynomial of S.

Theorem: (a) If $A \in \mathbb{R}^{n \times n}$ then it has n-eigenvalues

Let $A, B \in \mathbb{R}^{n \times n}$, then

(b) If $\{\lambda_i\}_{i=1}^n$ are the eigenvalues of A then

(i) $|A| = \prod_{i=1}^n \lambda_i = \lambda_1 \cdots \lambda_n$

(ii) $tr(A) = \sum_{i=1}^n \lambda_i$

(c) $|AB| = |A||B|$

D. Inverses

<u>Def</u>: Let $A \in \mathbb{R}^{n\times n}$. Suppose $\exists B \in \mathbb{R}^{n\times n}$ s.t. $AB = I$, then B is called the inverse of A, denoted A^{-1}.

<u>Ex</u>: (a) $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, then $A^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

(b) $A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, then A^{-1} does not exist

<u>Theorem</u>: Let $A, B \in \mathbb{R}^{n\times n}$

a) If A^{-1} exists it is unique.

b) $\det(A) \neq 0$ iff A^{-1} exists

c) If $\det(A) \neq 0$, then $\det(A^{-1}) = (\det(A))^{-1}$

d) If A, B are invertible $(AB)^{-1} = B^{-1}A^{-1}$

Note: MATLAB function inv

$$Ax = b$$
$$x = A^{-1}b$$

MATLAB backslash

E. Singular Value Decomposition (SVD)

Theorem: Suppose $M \in \mathbb{R}^{m \times n}$ then $\exists\, U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$ and $V \in \mathbb{R}^{n \times n}$ s.t.

MM^T or M^TM

(a) $M = U \Sigma V^T$ (called the svd of M)

(b) the columns of U are the set of orthonormal eigenvectors of MM^T s.t. $U^T U = I$.

$$U = \begin{bmatrix} | & | & | & \\ u_1 & u_2 & u_3 & \cdots \\ | & | & | & \end{bmatrix}$$

$u_i^T \cdot u_j \in \mathbb{R}$

$$u_i^T \cdot u_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

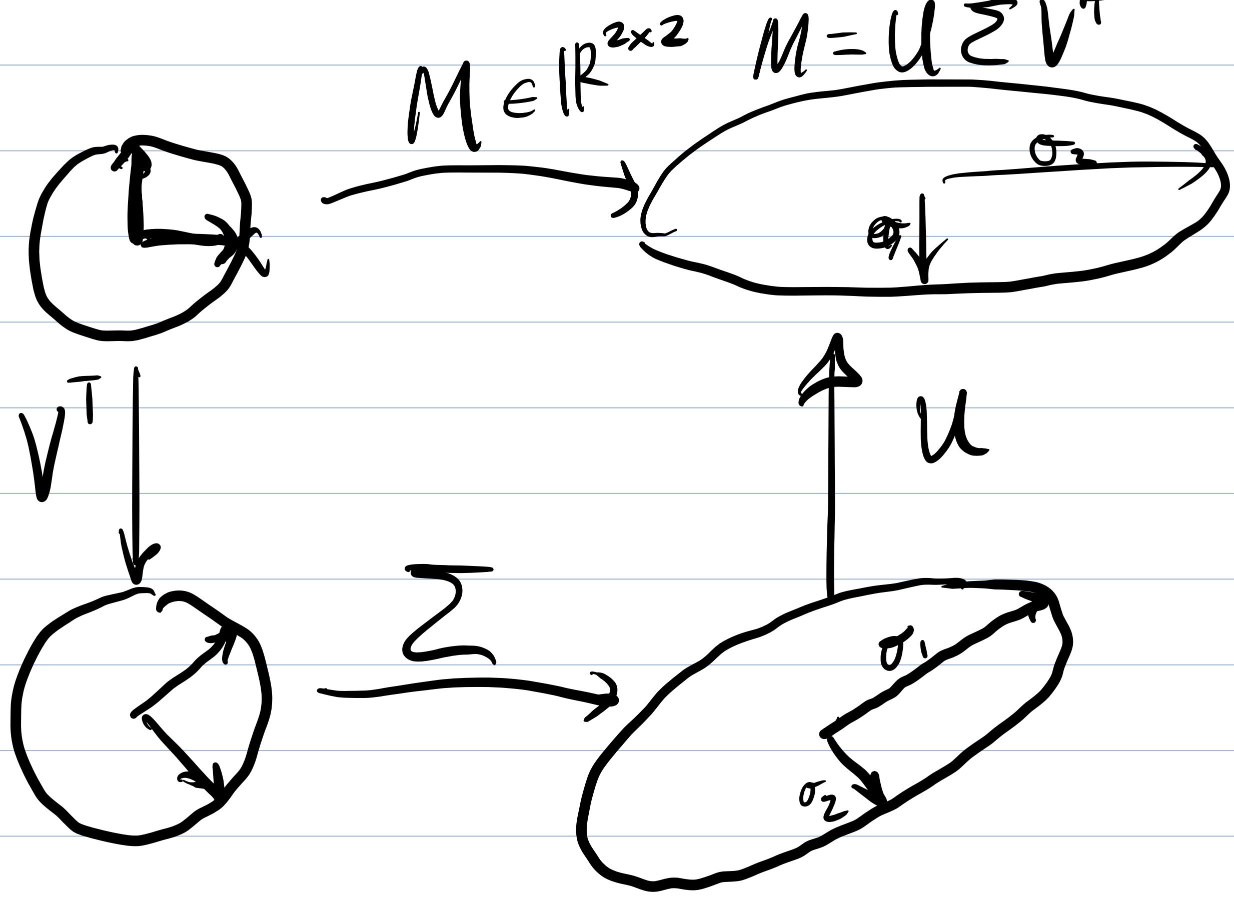
(c) the columns of V are the set of orthonormal eigenvectors of M^TM s.t. $VV^T = I$

d) Σ is a diagonal matrix whose diagonal elements are the square root of MM^T and M^TM

sometimes called the singular values of M.

<u>Remark</u>: 1) Matrices whose transpose are their inverse sometimes called orthogonal matrices or unitary matrices.

2) Unitary/Orthogonal matrices describe transformations between coordinate systems.



$M \in \mathbb{R}^{p \times r} \longrightarrow \mathbb{R}^{p \times p}$

$V^T \in \mathbb{R}^{r \times r}$ | $V^T \in \mathbb{R}^{r \times r}$

$\Sigma \in \mathbb{R}^{p \times r}$ | $\Sigma \in \mathbb{R}^{r \times r}$

$U \in \mathbb{R}^{p \times p}$ | $U \in \mathbb{R}^{r \times r}$

$\underline{AB \neq BA}$

$U \Sigma V^T$

$V^T \Sigma U$

Ma $\qquad U\Sigma V^T a$

$a = \sum \alpha_i v_i$

$V^T a$

$Ma = b$ $\quad M \in \mathbb{R}^{n\times p}$

$a = M^{-1} b$

$$\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_s \end{bmatrix}$$

eig(MM^T)

eig$(M^T M)$

Ex: Suppose $(1,6)$, $(2,3)$, $(3,7)$, and $(4,10) \in \mathbb{R}^2$ we want to fit a function $y = \beta_1 + \beta_2 x$ to the data

$$\min_{\beta_1, \beta_2 \in \mathbb{R}} \left\| \begin{bmatrix} 6 \\ 3 \\ 7 \\ 10 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \right\|_2^2$$

$$\left\| \begin{bmatrix} 6 - \beta_1 - \beta_2 \\ 3 - \beta_1 - 2\beta_2 \\ 7 - \beta_1 - 3\beta_2 \\ 10 - \beta_1 - 4\beta_2 \end{bmatrix} \right\|_2^2$$

$$= \left(\sqrt{(6-\beta_1-\beta_2)^2 + (3-\beta_1-2\beta_2)^2 + \dots} \right)^2$$

$$\min_{\beta \in \mathbb{R}^2} \|Y - X\beta\|_2^2 = \min_{\beta \in \mathbb{R}^2} \sum_{i=1}^{4} \left(y_i - \sum_{j=1}^{2} x_{ij}\beta_j \right)^2$$

$r_i = y_i - \sum_{j=1}^{2} x_{ij}\beta_j$ and let $S = \sum_{i=1}^{4} r_i^2$

then $\frac{dS}{d\beta_i} = 2\sum_{i=1}^{4} r_i \frac{dr_i}{d\beta_j}$, $\frac{dr_i}{d\beta_j} = -x_{ij}$

Thus if $\hat{\beta}$ minimizes S, then

$$2\left(\sum_{i=1}^{4} \left(y_i - \sum_{k=1}^{2} x_{ik}\hat{\beta}_k \right)(-x_{ij}) \right) = 0$$

which is equal to

$$\sum_{i=1}^{4} \sum_{k=1}^{2} x_{ij} x_{ik} \hat{\beta}_k = \sum_{i=1}^{4} x_{ij} y_i$$

$$(X^TX)\beta = X^Ty$$

why?

$$\hat{\beta} = \underbrace{(X^TX)^{-1}X^T}_{\text{the pseudoinverse of } X}\, y$$

$$(X^TX)^{-1}X^T = (V\Sigma^T U^T U\Sigma V^T)^{-1} V\Sigma^T U^T$$

$$= (V\Sigma^T\Sigma V^T)^{-1} V\Sigma^T U^T$$

$X = U\Sigma V^T$

$X^T = V\Sigma^T U^T$

$$= (V^T)^{-1}(\Sigma^T\Sigma)^{-1}\cancel{V^{-1}V}\Sigma^T U^T$$

$$= (V^T)^{-1}(\Sigma^T\Sigma)^{-1}\Sigma^T U^T$$

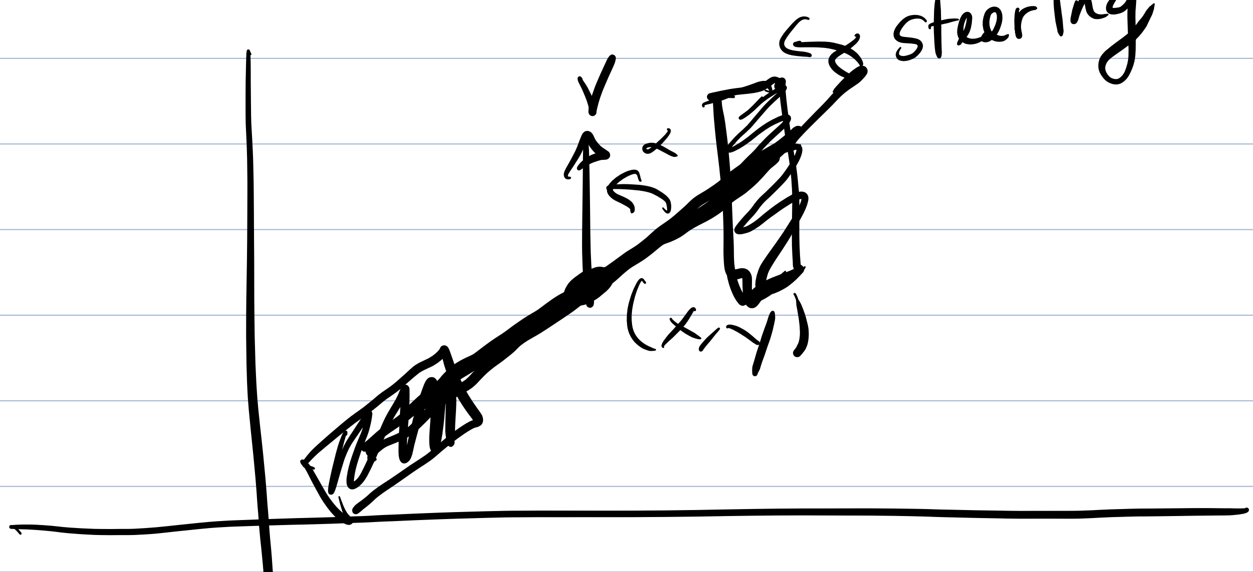
$(AB)^{-1} = B^{-1}A^{-1}$

$$= V(\Sigma^T\Sigma)^{-1}\Sigma^T U^T$$

if Σ was invertible, then

$$= V\Sigma^{-1}\cancel{(\Sigma^T)^{-1}\Sigma^T}U^T$$

$$= V\Sigma^{-1}U^T$$



[illegible]