3

Lecture #1

Goal
1) Notation
2) Linear Algebra

I Notation

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

Def: Given two sets A and B we say that
a) $f: A \to B$ is a function if $\forall a \in A$ 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: Consider the function $f: \mathbb{N} \to \mathbb{R}$ where $f(k) = (1/2)^k$. It has domain $\mathbb{N}$ and codomain $\mathbb{R}$, but it only takes values that are in $(0, 1]$.

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 upper case and vectors in lower case. We will use greek letters for scalars.

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 $[P^T]_{ji} = [P]_{ij}$

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) $\sum \alpha_i v_i = 0$ implies ($\Rightarrow$) $\alpha_i = 0 \ \forall i$.

Otherwise the set of vectors is called 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

first vector minus second vector is equal to the third vector

Def: (a) 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 (this will be the same...)

(b)
$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ has rank } 3$$

(c)
$$A = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} \text{ has rank } 2$$

Note: There are efficient algorithms to compute the rank of matrices (use the rank command in MATLAB)

Ex: Let $A = [a_1 \; a_2 \cdots a_n]$ be rank deficient then $\exists \{\alpha_i\}$ such that $\sum \alpha_i a_i = 0$, with some $\alpha_i \neq 0$. This can be written as $A(\alpha_1 \cdots \alpha_N)^T = 0$.

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

Ex: $A=\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ Suppose we want to determine if A is full rank.
$Av=0$ if $v=(1\ 0)^T$ then $Av=0$,
So A is not full rank.
Note any scaled version of v would have worked.

Def: The kernel of $A\in\mathbb{R}^{n\times p}$ denoted $\ker(A)=\{x\in\mathbb{R}^n \mid Ax=0\}$.

Ex: a) $A=\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\ker(A)=\left\{x\in\mathbb{R}^2 \middle| \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \alpha\in\mathbb{R}\right\}$
Note the dimension of the kernel of A is $\underline{1}$ since one linearly independent vector describes all elements of $\ker(A)$.
b) $A=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ $\ker(A)=\{0\}$
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) $A=\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $rk(A)=2-\dim(\ker(A))=1$
(b) $A=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ $rk(A)=3-\dim(\ker(A))$
$=3$

Scanned by CamScanner

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 column j in the matrix S, where $\det(\alpha) = \alpha$ for $\alpha \in \mathbb{R}$.

Ex:

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

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

$$= 22$$

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 $\lambda \in \mathbb{C}$ such that $Sv = \lambda v$ for some nonzero $v \in \mathbb{C}^n$ which is called an eigenvector.

Note: The eigenvectors correspond to the "directions" of the matrix along which it acts only by scaling.

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

We want λ such that $Av=\lambda v \Rightarrow$
$(A-I\lambda)v=0$ so we want to
find when $(A-I\lambda)$ is rank deficient $\Leftrightarrow$
$\det(A-\lambda I)=0$

$$\begin{vmatrix} 2-\lambda & 0 & 0 \\ 0 & 3-\lambda & 4 \\ 0 & 4 & 9-\lambda \end{vmatrix} = -\lambda^3+14\lambda^2-35\lambda+22$$
$$\Rightarrow \lambda=1,2,11.$$

Note: There are numerically efficient tools to compute det and eig using 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) $|A| = \prod_{i=1}^n \lambda_i$
ii) $tr(A) = \sum_{i=1}^n \lambda_i$
(c) $|AB|=|A||B|$

D. Inverses

Def: 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}.

Ex: (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.

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

a) If A 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 invertible then $(AB)^{-1} = B^{-1}A^{-1}$

Note: There are algorithms to compute A^{-1} using MATLAB function inv.

E. Singular Value Decomposition (SVD)

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

(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 such that $U^T U = I$

(c) the columns of V are the set of orthonormal eigenvectors of $M^T M$ such that $VV^T = I$.

(d) Σ is a diagonal matrix whose diagonal elements are the square root of the eigenvalues of MM^T and $M^T M$.

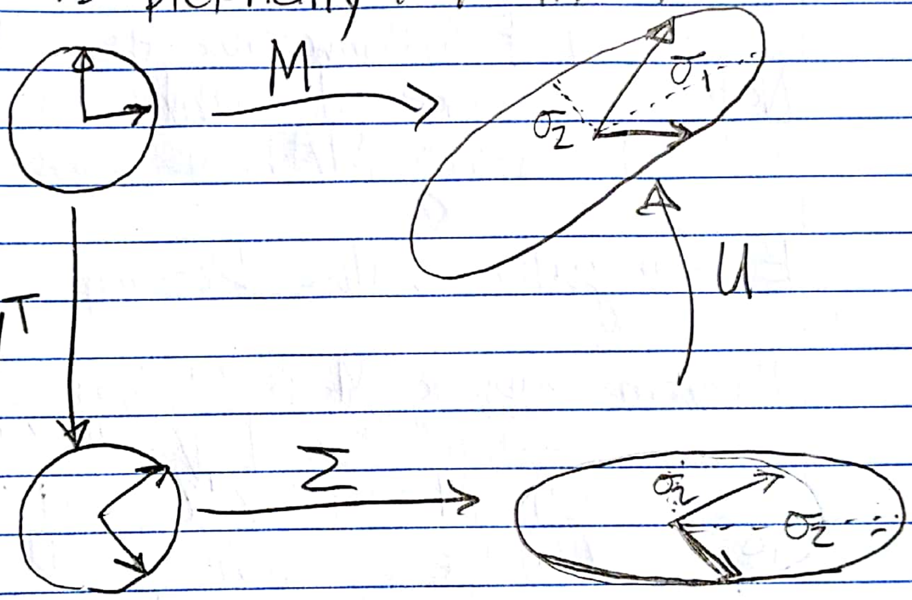
Def: The diagonal entries σ_i of Σ are called the singular values of M.

Remark: 1) Matrices whose transpose are their inverse are called unitary matrices.

2) Unitary matrices describe transformations between different coordinate systems.

More on this later

3) Σ scales the coordinate axes, so SVD pictorially looks like this:



Ex: Suppose $(1,6)$, $(2,3)$, $(3,7)$, and $(4,10)$ are given and we want to fit a function $y = \beta_1 + \beta_2 x$ to the data. We want to find the best β_1 and β_2:

$$\min_{\beta_1, \beta_2} \left\| \begin{bmatrix} 6 \\ 5 \\ 7 \\ 10 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \right\|_2^2$$

that is we want to solve:

$$\min_{\beta\in\mathbb{R}^n} \|y - X\beta\|_2^2 = \min_{\beta\in\mathbb{R}^n} \sum_{i=1}^{m} \left(y_i - \sum_j x_{ij}\beta_j\right)^2$$

How do we solve this problem?
Use Calculus, take gradient and set equal to zero.

Let $r_i = y_i - \sum x_{ij}\beta_j$ and let $S = \sum_{i=1}^{m} r_i^2$

then $\frac{dS}{dr_i} = 2\sum r_i \frac{dr_i}{d\beta_j}$, $\frac{dr_i}{d\beta_j} = -X_{ij}$

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

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

which is equal to:

$$\sum_{i=1}^{m}\sum_{k=1}^{n} X_{ij}X_{ik}\hat{\beta}_k = \sum_{i=1}^{m} X_{ij}y_i$$

$$(X^TX)\hat{\beta} = X^Ty$$

but how do we solve for $\hat{\beta}$?

<u>Note:</u> a) If X's columns are linearly independent then X^TX is invertible

<u>Pf:</u> We show $X^TXv = 0 \Rightarrow v = 0$.

$$0 = v^TX^TXv = (Xv)^T(Xv) = \|Xv\|_2^2$$

This implies $v = 0$ since X's columns are linearly independent.

b) We refer to $(X^TX)^{-1}X^T$ the pseudoinverse of X:

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

c) $\hat{\beta} = (X^TX)^{-1}X^TY.$