Optimization

Minimize $f(x)$ subjected to $x \in S$

x : variable or parameters

$f(\,) \rightarrow$ objective function / cost function

$S \rightarrow$ feasible region / search space

Solution

$x^* \in S$ such that

$$f(x^*) \leq f(x) \quad \forall x \in S$$

$x^* \rightarrow$ solution and $f(x^*)$ is optimal objective function value

Note : x^* may not be unique and may not even exist.

$\Rightarrow$ Maximize $f(x) \equiv -$ Minimize $-f(x)$

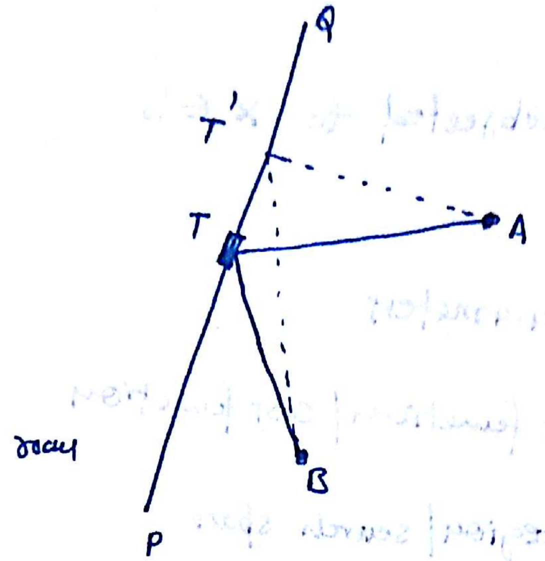
Minimize $f(x)$ subjected to $x \in S$

may be written as

$$\min_{x} f(x)$$

$$\text{st } x \in S$$

Problem formulation: Bus terminus location problem



Coordinats of A & B

$x_A = (x_{A_1}, x_{A_2})$ & $x_B = (x_{B_1}, x_{B_2})$

Equation of line (Road)

$ax_1 + bx_2 + c$

Location of bus terminus

$x_T = (x_{T_1}, x_{T_2})$ [Variable]

Objective is to minimize [using Euclidean distance (d)]

$d(x_A, x_T) + d(x_T, x_B)$

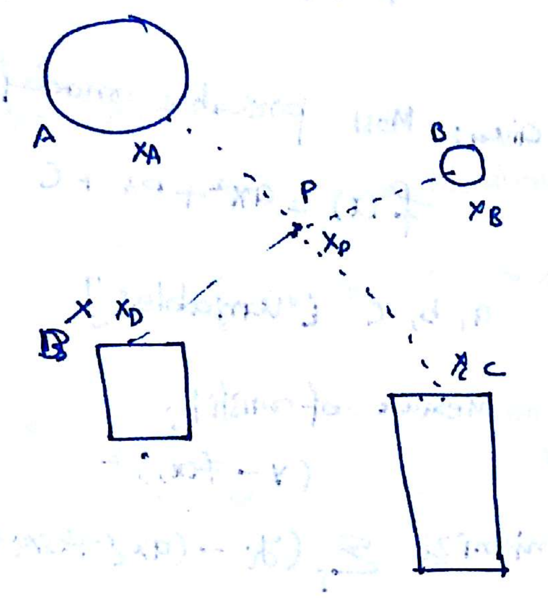
~~but~~ with constraint : T lies on PQ

$$\min_{x_{T_1}, x_{T_2}} d(x_A, x_T) + d(x_T, x_B)$$

$$\text{s.t.} \quad ax_{T_1} + bx_{T_2} + c = 0$$

→ The problem may be varied ~~to~~ like: suppose road is not a straight line, rather it is a curve then, then if we know the equation of that curve, condition in the constraint part will change accordingly.

Facility location Problem



There are four population areas (sectors or colonies) and aim is locate a facility centre somewhere so that each population area can be served optimally.

$x_A = (x_{A_1}, x_{A_2})$, $x_B = (x_{B_1}, x_{B_2})$, $x_C = \cdots$

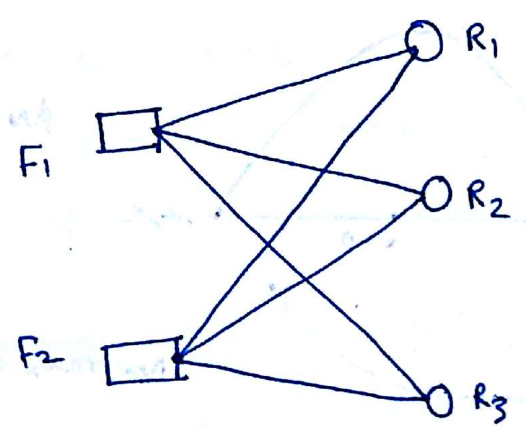
$x_P = (x_{P_1}, x_{P_2})$

$$\min_{x_{P_1}, x_{P_2}} d(x_A, x_P) + d(x_B, x_P) + d(x_C, x_P) + d(x_D, x_P)$$

s.t. $x_A \in A$, $x_B \in B$, $x_C \in C$, $x_D \in D$

Transportation problem:

There are two unit of fertilizer production F_1 & F_2, and ~~There~~ three retailers R_1, R_2 and R_3 to distribute fertilizer for farmers. The objective is to ~~ap~~ find the number of units transported to each retailer ~~such that~~ to minimize the cost.



$a_i \rightarrow$ Capacity (production) of the plant F_i (per week)

$b_j \rightarrow$ Demand of the outlet R_j

$c_{ij} \rightarrow$ cost of shipping one unit of product from F_i to R_j

$x_{ij} \rightarrow$ Number of units of the product shipped from F_i to R_j (VARIABLES) (to be decided)

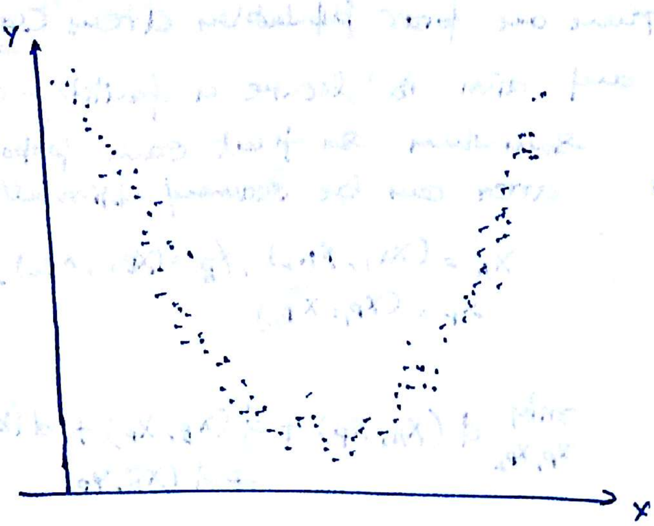
$$\min_{x} \sum_{ij} c_{ij} x_{ij}$$

s.t. $\sum_{j=1}^{3} x_{ij} \leq a_i$, $i = 1, 2$ ~~$j=1,2,3$~~ (limited by production capacity)

$\sum_{i=1}^{2} x_{ij} \geq b_j$ $j = 1, 2, 3$ (Demand of each retailer must be fulfilled)

$x_{ij} \geq 0$ (as number of unit shipped cannot be negative)

Data fitting problem



Given: $\{x_i, y_i\}_{i=1}^{n}$, n data points

Given: Most probable model type

$$f(x) = ax^2 + bx + c$$

a, b, c [variables]

Measure of misfit:

$$(y - f(x))^2$$

The objective is to minimize $\sum_i (y_i - (ax_i^2 + bx_i + c))^2$

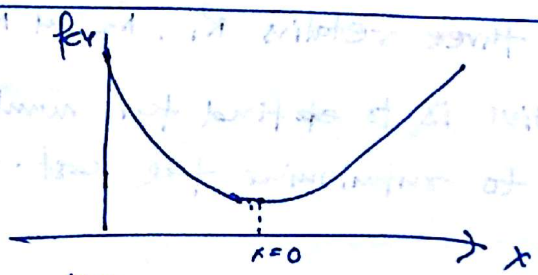
- No constraints

$$\min_{a,b,c} \sum_{i=1}^{n} (y_i - (ax_i^2 + bx_i + c))^2$$

---

① $f(x) = (x-2)^2 + 3$

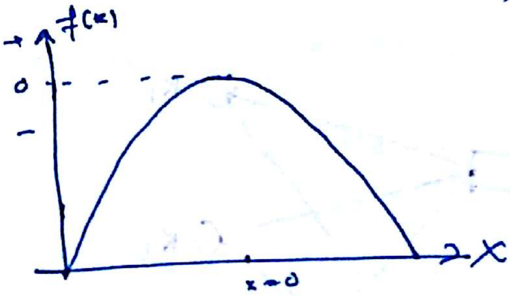
$x^* = 2$, $f(x^*) = 3$



No maxima

② $f(x) = -x^2$

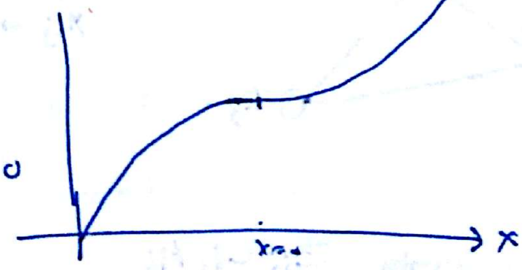
$x^* = 0$, $f(x^*) = 0$



No minima

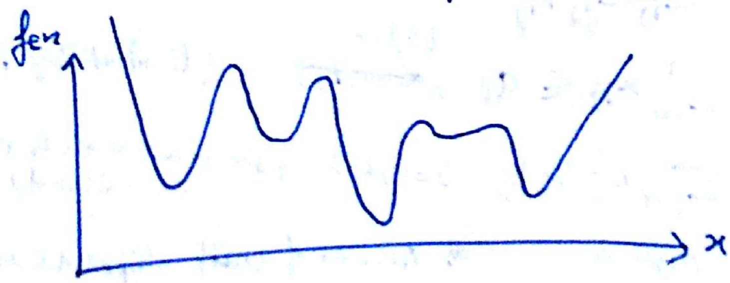
③ $f(x) = x^3$

~~f(x)~~ saddle point $x^* = 0$



No max or min

④



Many local min

one global min

sets

A set is collection of objects satisfying certain property P

Examples:

- A set of natural number $\{1, 2, 3, \dots\}$
- $\{x \in \mathbb{R} : 1 \leq x \leq 3\}$

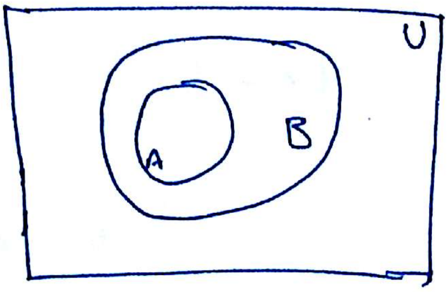
empty set → contains no elements denoted by φ

Let set A and B

Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$

Intersection: $A \cap B = \{x : x \in A \text{ and } x \in B\}$

Difference: $A \setminus B = \{x : x \in A \text{ and } x \notin B\}$



Subset $A \subseteq B$

if there are some elem of B $y \in B$ but they are not in A $y \notin A$

then A is strict subset of B $A \subset B$

Supremum and Infimum of a set

A set A of real numbers is said to be bounded above, if there is a real number y such that $x \leq y$ $\forall x \in A$. The smallest possible real number y satisfying $x \leq y$ $\forall x \in A$ is called the least upper bound or supremum of A and denoted by $\sup\{x : x \in A\}$

- Similarly one can define greatest lower bound or infimum $\inf\{x : x \in A\}$

Example: consider A set $A = \{x : 1 \leq x < 3\}$

$\sup\{x : x \in A\} = 3$ (Note $3 \notin A$)

$\inf\{x : x \in A\} = 1 \ (\in A)$

i.e Sup or inf may or may not belong to set c4 d Linear Independence

A set of vectors $\bar{x}_1, \bar{x}_2, \dots \bar{x}_k$ is said to be linearly independent if

$$\sum_{i=1}^{k} \alpha_i \bar{x}_i = 0 \quad \Rightarrow \alpha_i = 0 \ \forall i$$

otherwise, they are linearly dependent and one of them is linear combination of the others.

Functions

A function f from a set A to a set B is a rule that assigns to each x in A a unique element $f(x)$ in B.

$$f : A \to B$$

- A : Domain of f
- $\{y \in B : (\exists x)[y = f(x)]\}$: Range of f
- Range of $f \subseteq B$

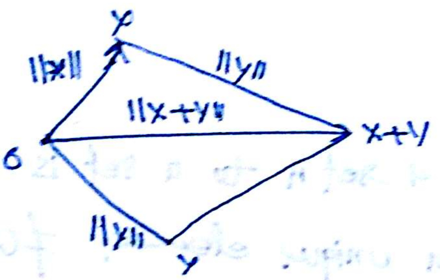
Examples

- $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = x^2$
- $f : (-1, 1) \to \mathbb{R}$ defined as $f(x) = \frac{1}{|x| - 1}$

Norm

A norm on $\mathbb{R}^n$ is a real-valued-value function $\|\cdot\|: \mathbb{R}^n \to \mathbb{R}^n$ which obeys ($x \to$ vector x)

- $\|x\| \geq 0$ for every $x \in \mathbb{R}^n$, and $\|x\| = 0$ if and only if $x = 0$
- $\|\alpha x\| = |\alpha| \, \|x\|$ for every $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ and
- $\|x+y\| \leq \|x\| + \|y\|$ $\forall x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$



Some popular Norms

Let $x \in \mathbb{R}^n$

- L_2 or Euclidean norm

$$\|x\|_2 = \left[\sum_{i=1}^{n} (x_i)^2\right]^{1/2}$$

- L_1 norm

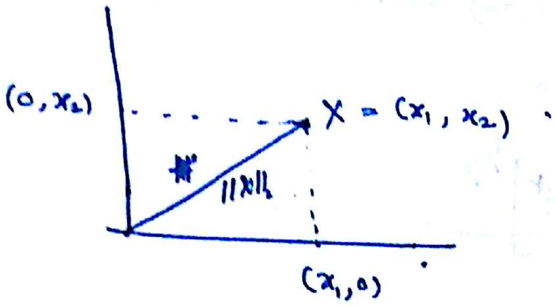
$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

- L_∞ norm

$$\|x\|_\infty = \max_{i=1,2,\dots n} |x_i|$$

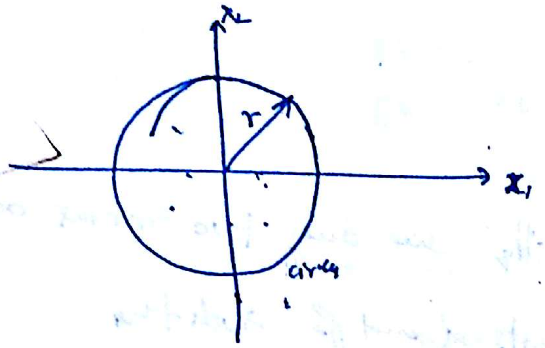
ation

fer

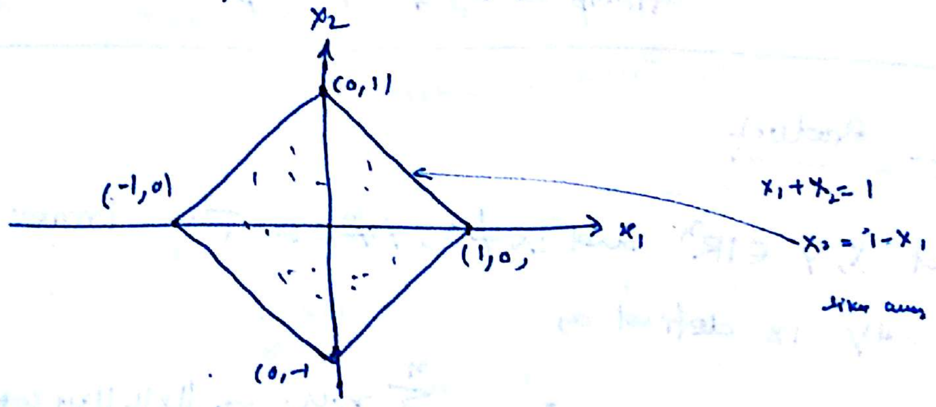


consider a set $S = \{ x \in \mathbb{R}^2 : \|x\|_2 \le r \}$

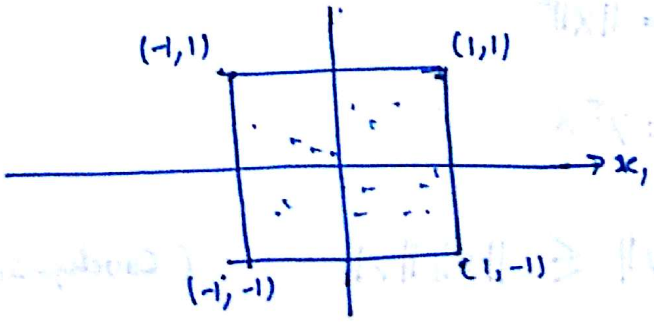
it is shown



$S = \{ x \in \mathbb{R}^2 : \|x\|_1 \le 1 \}$



$S = \{ x \in \mathbb{R}^2 : \|x\|_\infty \le 1 \}$



- In general, the L_p $(1 \le p < \infty)$ vector norms is defined as

$$\|x\|_p = \left[\sum_{i=1}^{n} |x_i|^p\right]^{1/p}$$

- Does the convergence of a particular optimization algorithm depends on what norm its stopping criterion used?

Ans: no

using the Result

If $\|\cdot\|_p$ and $\|\cdot\|_q$ are any two norms on $\mathbb{R}^n$, then there exist positive constants α and β such that

$$\alpha\|x\|_p \le \|x\|_q \le \beta\|x\|_p \qquad \forall x \in \mathbb{R}^n$$

Inner Product

Let $x, y \in \mathbb{R}^n$ and $x \ne 0$, $y \ne 0$. The inner or dot product of x and y is defined as

$$x \cdot y \equiv x^T y = \sum_{i=1}^{n} x_i y_i = \|x\| \cdot \|y\| \cos\theta$$

where θ is the angle between x and y

Note

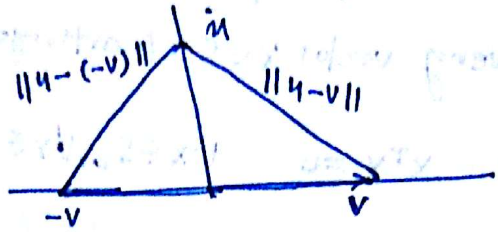
① $x^T x = \|x\|^2$

② $x^T y = y^T x$

③ $\|x \cdot y\| \le \|x\|\,\|y\|$ (Cauchy-Schwartz inequality)

ombir
utir

Orthogonality



Geometry

$$[\text{dist}(u,-v)]^2 = [\text{dist}(u,v)]^2 \quad -(1)$$

$$\begin{aligned}[\text{dist}(u,-v)]^2 &= \|u-(-v)\|^2 \\ &= \|u+v\|^2 \\ &= (u+v)\cdot(u+v) \\ &= u\cdot(u+v)+v\cdot(u+v) \\ &= u\cdot u + u\cdot v + v\cdot u + v\cdot v \\ &= \|u\|^2+\|v\|^2+2u\cdot v\end{aligned}$$

Similarly $[\text{dist}(u,v)]^2 = \|u\|^2 + \|-v\|^2 + 2u\cdot(-v)$

$= \|u\|^2 + \|v\|^2 - 2u\cdot v$

Equation (1) is true iff $2u\cdot v = -2u\cdot v$

which happen if and only if

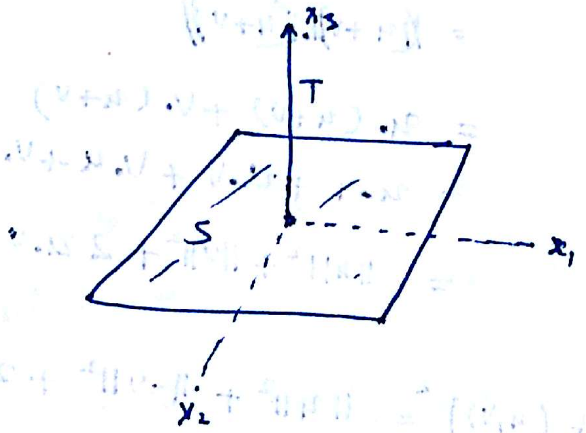
$u\cdot v = 0$

> Two vector u and v in $\mathbb{R}^n$ are orthogonal (to each other)
>
> if $u\cdot v = 0$ or $u^Tv = 0$

Two subspaces S and T of the same vector space $\mathbb{R}^n$ are orthogonal if every vector $x \in S$ is orthogonal to every vector $y \in T$

i.e. $x^T y = 0 \quad \forall x \in S$ and $\forall y \in T$

→ Given a subspace S of $\mathbb{R}^n$, the space of all vectors orthogonal to S is called the orthogonal complement of S



S and T are orthogonal complement of each other

Mutual orthogonality

Vectors $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ are said to be mutually orthogonal if $x_i^T x_j = 0 \quad \forall\ i \neq j$

In addition, if $\|x_i\| = 1$ for every i

the set $\{x_1, x_2, \ldots, x_k\}$ is said to be orthonormal

$\vec{w} = \vec{p}$

If $X_1, X_2, \dots, X_k$ are mutually orthogonal nonzero vectors, they are linearly independent

To prove it we need to show that

$$\sum_{i=1}^{k} \alpha_i X_i = 0 \Rightarrow \alpha_i = 0 \ \forall i$$

Proof:

Let $\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k = 0$

Therefore, $(\alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k)^T X_1 = 0$

or $$\sum_{i=1}^{k} \alpha_i X_i^T X_1 = 0$$

This gives $\alpha_1 X_1^T X_1 + 0 + 0 + \dots + 0 = 0$

$\Rightarrow \alpha_1 X_1^T X_1 = 0$

$\Rightarrow \alpha_1 \|X_1\|^2 = 0$ but $\|X_1\|^2 \neq 0$

Thus $\alpha_1 = 0$

Similarly we can show $\alpha_i = 0 \ \forall i$

Let $\{\bar{u}_1, \bar{u}_2, \dots \bar{u}_p\}$ be an orthogonal basis for a subspace S of $\mathbb{R}^n$. For each Y in S, the weights in the linear combination $\bar{Y} = c_1 \bar{u}_1 + \dots + c_p \bar{u}_p$

and given by $c_j = \dfrac{\bar{Y} \cdot \bar{u}_j}{\bar{u}_j \cdot \bar{u}_j}$ $(j = 1, 2, \dots, p)$

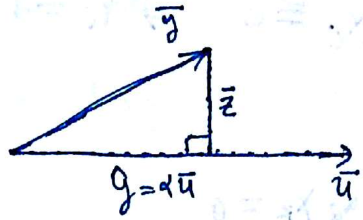
Proof: take dot product with $\bar{u}_1$ $\bar{Y}^T \bar{u}_1 = (c_1 \bar{u}_1 + c_2 \bar{u}_2 + \dots + c_p \bar{u}_p)^T \bar{u}_1$

$= c_1 \bar{u}_1^T \bar{u}_1 + \underbrace{c_2 \bar{u}_2^T \bar{u}_1}_{=0} + \dots + \underbrace{c_p \bar{u}_p^T u_1}_{=0} = c_1 \bar{u}_1^T u_1$

$\Rightarrow c_1 = \dfrac{\bar{Y}^T \bar{u}_1}{\bar{u}_1^T u_1}$ or $c_1 = \dfrac{\bar{Y} \cdot \bar{u}_1}{\bar{u}_1 \cdot \bar{u}_1}$ Similarly for other coefficients

Orthogonal Projection

Consider a nonzero vector $\bar{u}$ in $\mathbb{R}^n$. Now consider the decomposition of a vector $\bar{y}$ in $\mathbb{R}^n$ as follow



Thus $\quad \bar{y} = \hat{y} + \bar{z} \quad$ when $\hat{y} = \alpha\bar{u}$ (same multiple of $\bar{u}$)

$\bar{y} = \alpha\bar{u} + \bar{z} \qquad \Rightarrow z = \bar{y} - \alpha\bar{u}$

$\hat{y}$ is orthogonal projection of $\bar{y}$ on $\bar{u}$

and $\bar{z}$ is component of $\bar{y}$ orthogonal to $\bar{u}$

Note that $\hat{y}$ and $\bar{z}$ are ~~ot~~ orthogonal to each other,

thus $\quad \hat{y}\cdot\bar{z} = 0 \quad$ also $\bar{u}\cdot\bar{z} = 0$ as $\bar{z}$ is $\perp$ $\bar{u}$ as well

$= \cancel{(\alpha\bar{u})\cdot(\bar{y}-\alpha\bar{u})}$

$\bar{u}\cdot(\bar{y}-\alpha\bar{u}) \underset{or}{=} \bar{u}^T(\bar{y}-\alpha\bar{u}) = 0$

$\Rightarrow \bar{u}^T\bar{y} - \alpha\bar{u}^T\bar{u} = 0$

$\Rightarrow \alpha = \dfrac{\bar{u}^T\bar{y}}{\bar{u}^T\bar{u}} = \dfrac{y^Tu}{u^Tu}$

Orthogonal projection of $\bar{y}$ onto $\bar{u}$

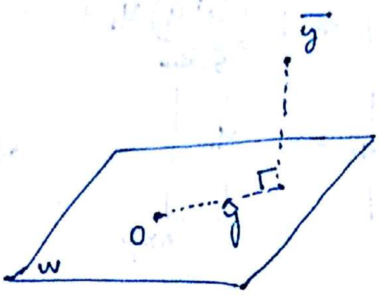
$$\boxed{\hat{y} = \operatorname{Proj}\bar{y} = \frac{\bar{y}^T\bar{u}}{\bar{u}^Tu} = \frac{\bar{y}\cdot\bar{u}}{\bar{u}\cdot\bar{u}}}$$

Given a vector $\bar{y}$ and a subspace W in $\mathbb{R}^n$

There is a vector $\hat{y}$ in W such that

(1) $\hat{y}$ is unique vector in W for which $(\bar{y} - \hat{y}) \perp W$

and (2) $\hat{y}$ is the unique vector in W closest to $\bar{y}$.



---

let W be a subspace of $\mathbb{R}^n$. Then each $\bar{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\bar{y} = \hat{y} + z$$

where $\hat{y}$ is in W and $\bar{z}$ is in $W^{\perp}$.

In fact, if $\{\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_p\}$ is any orthogonal basis of W, then

$$\hat{y} = \frac{\bar{y}\cdot\bar{u}_1}{\bar{u}_1\cdot\bar{u}_1}\bar{u}_1 + \frac{\bar{y}\cdot\bar{u}_2}{\bar{u}_2\cdot\bar{u}_2}\bar{u}_2 + \cdots + \frac{\bar{y}\cdot\bar{u}_p}{\bar{u}_p\cdot\bar{u}_p}\bar{u}_p$$

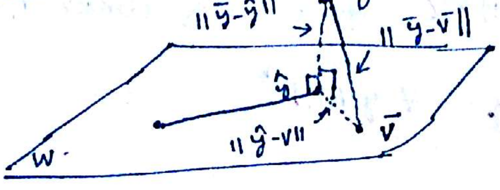
and $\bar{z} = \bar{y} - \hat{y}$

The Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, $\bar{y}$ an vector in $\mathbb{R}^n$, and $\hat{y}$ the orthogonal projection of $\bar{y}$ onto W. Then $\hat{y}$ is the closest point in W to $\bar{y}$

i.e. $\quad \|\bar{y} - \hat{y}\| < \|\bar{y} - \bar{v}\|$

for all $\bar{v}$ in W distinct from $\hat{y}$



$$\|\bar{y}-\bar{v}\|^2 = \|\bar{y}-\hat{y}\|^2 + \|\hat{y}-\bar{v}\|^2$$

If $\{u_1, \ldots, u_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$,

then

as $\bar{u}_j \cdot \bar{u}_j = 1$

$$\text{Proj}_W \bar{y} = (\bar{y}\cdot\bar{u}_1)\bar{u}_1 + (\bar{y}\cdot\bar{u}_2)\bar{u}_2 + (\bar{y}\cdot\bar{u}_3)\bar{u}_3 + \cdots + (\bar{y}\cdot\bar{u}_p)\bar{u}_p$$

$$= (\bar{y}^T\bar{u}_1)\bar{u}_1 + (\bar{y}^T\bar{u}_2)\bar{u}_2 + \cdots + (\bar{y}^T\bar{u}_p)\bar{u}_p$$

$$\text{Proj}_W \bar{y} = (\bar{u}_1^T\bar{y})\bar{u}_1 + \underbrace{(\bar{u}_2^T\bar{y})}_{\text{scalar}}\bar{u}_2 + \cdots + (\bar{u}_p^T\bar{y})\bar{u}_p$$

let

$$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_p \\ | & | & & | \end{bmatrix}_{n\times p}$$

u_i is vector of dimension n

as u_i is basis vector of $\mathbb{R}^n$

$\text{Proj}_W \bar{y}$ is linear combination of column vector $\{u_1, u_2, \ldots u_p$

with weights $\bar{u}_1^T\bar{y}, u_2^T\bar{y}, \ldots, u_p^T y$

$$\text{Proj}_W \bar{y} = (\bar{u}_1^T y)u_1 + (u_2^T y)u_2 + \cdots + (u_p^T y)u_p = \underbrace{\begin{bmatrix} u_1 & u_2 & u_p \end{bmatrix}_{n\times p}}_{U} \underbrace{\begin{bmatrix} u_1^T y \\ u_2^T y \\ \vdots \\ u_p^T y \end{bmatrix}_{p\times 1}}_{U^T y}$$

scalar

Now
$$U^T = \begin{bmatrix} - u_1^T - \\ - u_2^T - \\ \vdots \\ u_p^T \end{bmatrix}_{p\times n}$$

$$U^T y = \begin{bmatrix} - u_1^T - \\ - u_2^T - \\ \vdots \\ - u_p^T - \end{bmatrix}_{p\times n} \begin{bmatrix} | \\ y \\ | \end{bmatrix}_{n\times 1} = \begin{bmatrix} u_1^T y \\ u_2^T y \\ u_p^T y \end{bmatrix}$$

therefore
$$\text{Proj}_W \bar{y} = \bar{U}\bar{U}^T\bar{y} \quad \forall \bar{y} \text{ in } \mathbb{R}^n$$

$U^T U\bar{y} = I_p \bar{y} = y$ for all y in $\mathbb{R}^p$ where U is $n\times p$ matrix with orthonormal columns

$UU^T\bar{y} = \text{Proj}_W y$ for $\forall y$ in $\mathbb{R}^n$

If U is $n\times n$ orthonormal matrix then W is whole $\mathbb{R}^n$ space

& $UU^T y = I_n y = y \quad \forall y$ in $\mathbb{R}^n$