lecture 16. Orthogonality and least Squares Problems

$u, v \in \mathbb{R}^m$, $u \perp v$ if $u^T v = 0$.

In this case, $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.

Two Subspaces S and T: $S \perp T$ if

$\forall s \in S$, and $\forall t \in T$, $s^T t = 0$, i.e., $s \perp t$.

Four Subspaces associated with $A \in \mathbb{R}^{m \times n}$

1. If $S \perp T$, then $S \cap T = \{0\}$

The formula $\dim(S+T) + \dim(S \cap T) = \dim(S) + \dim(T)$

implies $\dim(S) + \dim(T) = \dim(S+T) \leq m$

So two planes in $\mathbb{R}^3$ can't be orth. to each other

2. when $S \perp T$, and $\dim(S) + \dim(T) = m$.

then T is called the orthogonal complement of S

denoted by $S^{\perp} = T$, versiversa, $T^{\perp} = S$.

and $(S^{\perp})^{\perp} = S$.

Example
$A^T \underline{Ax} = 0 \Rightarrow Ax =$
$Ax \in N(A^T) \cap C(A) \Rightarrow Ax = 0$

$A \in R^{m \times n}$
3. $C(A)^{\perp} = N(A^T)$, $\forall x \in R^m$.

$x = x_c + x_n$, with $x_c \in C(A)$, and $x_n \in N(A^T)$

and $x_c \perp x_n$.

$$Ax = 0 \Leftrightarrow \begin{pmatrix} A_1^T x \\ \vdots \\ A_m^T x \end{pmatrix} = 0$$

$C(A^T)^{\perp} = N(A)$, $\forall y = R^n$

$y = y_c + y_n$. $y_c \in C(A^T)$, $y_n \in N(A)$. $y_c \perp y_n$

Example. Assume $Ax = b$ has solution, find the solution with minimal length $\|\cdot\|$. any solution

$x = a + b$, with $a \in C(A^T)$, $b \in N(A)$

$\|x\|^2 = \|a\|^2 + \|b\|^2$, set $b = 0$.

Example $S \subset \mathbb{R}^3$, and $S = \{0\}$, $S^\perp = ?$

If $S = \text{Span}\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}$, $S^\perp = \text{Span}\left\{\begin{pmatrix}1\\-1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\-1\end{pmatrix}\right\}$

$S = \text{Span}\left\{\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\1\\-1\end{pmatrix}\right\}$, $S^\perp = \text{Span}\left\{\begin{pmatrix}1\\-1\\0\end{pmatrix}\right\}$

Example. $a \perp \text{Span}\{v_1, \cdots, v_k\}$ iff. $a \perp v_i$, $i = 1, \cdots, k$

Proof. "$\Rightarrow$" Since $v_i \in \text{Span}\{v_1, \cdots v_k\}$. $\Rightarrow a \perp v_i$.

"$\Leftarrow$" $\forall x \in \text{Span}\{v_1, \cdots v_k\}$, $x = \sum_{i=1}^{k} \alpha_i v_i$

$$a^T x = \sum_{i=1}^{k} \alpha_i \underbrace{a^T v_i}_{=0} = 0$$

least squares prolems. $A \in \mathbb{R}^{m \times n}$

We have consider $Ax = b$, when $b \in C(A)$, there might be one or many solusions. But if $b \notin C(A)$, we can ask for an approximate solusion in the sense of making $r = b - Ax$ small but not exactly zero, r is the so-called residual of an approximate sol x.

least squares in one approach of quantifying the size of "smallness" of r, in terms of its

2-norm $\quad \|r\|^2 = r_1^2 + r_2^2 + \cdots + r_m^2$, $r = \begin{pmatrix} r_1 \\ \vdots \\ r_m \end{pmatrix}$

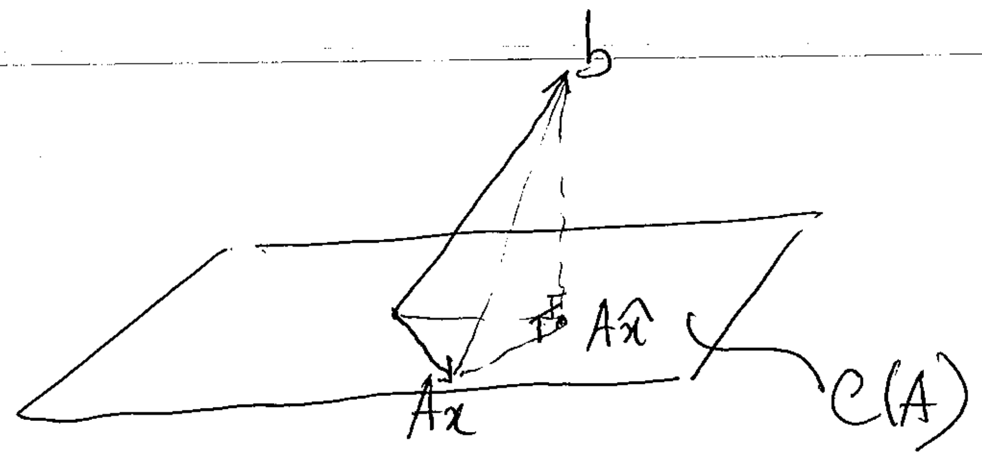
Other possibilities exist, e.g. $\quad r_i = (b_i - A_i x)^2$

$$\|r\|_1 = |r_1| + |r_2| + \cdots + |r_m| \qquad A = \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix}$$

So we want find an $x \in \mathbb{R}^n$, s.t.

$$\min \|r\|^2 = \min_{x \in \mathbb{R}^n} \|b - Ax\|^2$$

This is to say we find $A\hat{x} \in C(A)$ s.t. the residual $b - A\hat{x}$ is the smallest in $\|\cdot\|_2$



Orthogonality principle: $\|r\|^2$ is minimized iff.

$r \perp C(A) \iff a_i \perp r. \ \forall i = 1, \cdots, n.$

where $A = (a_1, \cdots, a_n) \in \mathbb{R}^{m \times n}$

From $\begin{pmatrix} a_1^T r \\ a_2^T r \\ \vdots \\ a_n^T r \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ we obtain

$$A^T r = 0 \qquad A^T(b - A\hat{x}) = 0$$

<u>Normal Eqn</u> $A^T A \hat{x} = A^T b$, this is an $n \times n$ linear system. Since $\operatorname{rank}[A^T A, A^T b] = \operatorname{rank}(A^T A)$ ~~it always has a solution, no matter what b is.~~

It has a unique solution $\hat{x} = (A^T A)^{-1} A^T b$ when A has full col rank, i.e., $r(A) = n \leq m$.

Now $p = A\hat{x} = A(A^T A)^{-1} A^T b = P_A b$

P_A is a <u>projection</u> matrix: $P = P^T$, $P^2 = P$.

Will discuss $r(A) < n$ <u>later</u>!

Now $b = P_A b + (I - P_A) b$

Since $P_A b \perp (I - P_A) b$. for arbitrary $x \in \mathbb{R}^n$

$$\|b - Ax\|^2 = \|(P_A b - Ax) + (I - P_A) b\|^2$$

Now $Ax \in C(A)$, $Ax \perp (I - P_A) b$.

$$= \|P_A b - Ax\|^2 + \|(I - P_A) b\|^2$$

The residual is minimized iff $Ax = P_A b$.

Given two lines:

$$P = (x, x, x), \quad x \in \mathbb{R}.$$

$$Q = (y, 3y, -1), \quad y \in \mathbb{R}.$$

Do the two lines meet at some point?

We can formulate the above as a LS problem

$$\min_{P,Q} \|P - Q\|^2 = \min_{x,y} \left\| \begin{pmatrix}1\\1\\1\end{pmatrix} x - \begin{pmatrix}1\\3\\0\end{pmatrix} y - \begin{pmatrix}0\\0\\-1\end{pmatrix} \right\|^2$$

$$= \left\| \begin{pmatrix}1 & 1\\1 & 3\\1 & 0\end{pmatrix} \begin{pmatrix}x\\y\end{pmatrix} - \begin{pmatrix}0\\0\\-1\end{pmatrix} \right\|^2.$$

Example. Fitting a straight line

$$f(t) = c + dt$$

to a set of points (t_i, b_i), $i = 1, \cdots, m$.

The linear system is

$$c + dt_1 = b_1$$
$$c + dt_2 = b_2$$
$$\vdots$$
$$c + dt_m = b_m$$

$Ax = b$, with $A = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix}$

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

$$A^T A = \begin{pmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{pmatrix}, \quad A^T b = \begin{pmatrix} \sum b_i \\ \sum t_i b_i \end{pmatrix}$$

$$\hat{x} = (A^T A)^{-1} A^T b = \frac{1}{m \sum t_i^2 - (\sum t_i)^2} \begin{pmatrix} \sum t_i^2 & -\sum t_i \\ -\sum t_i & m \end{pmatrix} \begin{pmatrix} \sum b_i \\ \sum t_i b_i \end{pmatrix}$$

$$= \begin{pmatrix} (\sum t_i^2)(\sum b_i) - (\sum t_i)(\sum t_i b_i) \\ m(\sum t_i b_i) - (\sum t_i)(\sum b_i) \end{pmatrix} \Big/ \left(m \sum t_i^2 - (\sum t_i)^2\right)$$

This can be easily generalized to fitting a polynomial to a set of data, or supervised learning.

Supervised learning.

let. $g_1(x), \dots, g_n(x)$ be a set of basis functions, e.g., $1, x, \dots, x^{n-1}$, $x \in \mathbb{R}$, but in general $x \in \mathbb{R}^p$, multi-dimensional.

~~We are given a set of data points (x_i, y_i)~~

We want find a linear combination of the g_i's

$$g(x) = \alpha_1 g_1(x) + \dots + \alpha_n g_n(x)$$

s.t. $g(x_i) \approx y_i$, $i = 1, \dots, m$. If we minimize $\sum_{i=1}^{m} \left(y_i - g(x_i)\right)^2$, we have the LS problem $\min \|Ax - b\|^2$, where

$$A = \begin{pmatrix} g_1(x_1) \cdots g_n(x_1) \\ \vdots \\ g_1(x_m) \cdots g_n(x_m) \end{pmatrix}, \quad x = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}, \quad b = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}.$$

Th. Let $C(A)$ denote the col space of A.

If $C(A)$ is a subspace of $C(B)$, then $r(A) \leqslant r(B)$.

Let's prove the following $r(A+B) \leqslant r([A, B])$

Proof. $A = [a_1, \dots, a_n]$, $B = [b_1, \dots, b_n]$.

$a_i \in C([A, B])$, $b_i \in C([A, B])$

---

hence $a_i + b_i \in C([A, B])$, which implies

$$C(A+B) \subseteq C([A, B]) \Rightarrow r(A+B) \leqslant r([A, B])$$

Corollary. $r(A+B) \leqslant r(A) + r(B)$.

Corollary $|r(A) - r(B)| \leqslant r(A-B)$

$A \in R^{m\times n}$, $B \in R^{n\times p}$ if $AB = 0$, then

$r(A) + r(B) \leqslant n$.

Proof. $AB = 0 \Rightarrow C(B) \subseteq N(A) \Rightarrow r(B) \leqslant \dim(N(A))$

$= n - r(A)$