LECTURE 30

Least Squares Solution

A) Two solution methods

B) The Normal Equations

C) The Subspace Matrix

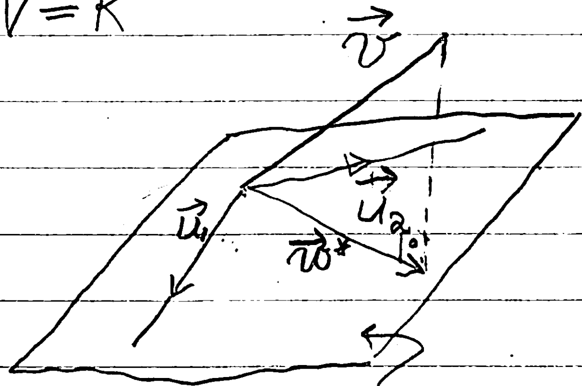
D) Solution via Subspace Matrix

THE LEAST SQUARES SOLUTION

A) Two Methods

There are two ways of solving the least squares problem.

$V = R^m$



$W_N = Sp(\underbrace{\{u_1, u_2, \ldots, u_N\}}_{B})$

1. From the given basis for W_N, using the G-S process, construct an orthonormal basis, say, $\{e_1, \ldots, e_N\} \equiv C$. Then determine the orthogonal projection

The optimal least squares vector →

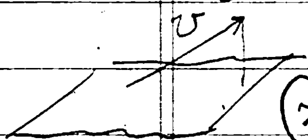
$$\boxed{\vec{w}^* = e_1 \langle \vec{e}_1 | \vec{v} \rangle + \ldots + e_N \langle \vec{e}_N | v \rangle,} \quad (*)$$

which according to the Fundamental Theorem is the least squares approximation.

2. Knowing that the least squares solution $w^* \in W_N$, (i) use the given basis $\{u_1, \ldots, u_N\}$ for W_N to represent w^* as a to-be-determined linear combination,

$$w^* = \overset{?}{b_1^*}\vec{u}_1 + \cdots + \overset{?}{b_N^*} u_N$$

and (ii) apply to the error vector $v - w^*$ the orthogonality constraint



$$(*) \quad \boxed{\langle \underbrace{v - w^*}_{h^*} \mid w \rangle = 0} \qquad w = \vec{u}_1, \vec{u}_2, \ldots, \vec{u}_N$$

N different times by letting w equal each of the N different basis vectors $u_1, \ldots, u_N$. Thus one has N different equations for the N to-be-determined unknowns $b_1^*, \ldots, b_N^*$.

B) Normal Equations

Comment:

It therefore is not necessary to use an O.N. basis for W_N inorder to find the optimal $\vec{w}^*$. This fact makes computations easy and is expressed by the following

Theorem ("Normal" Equations)

Let $\{u_1, \ldots, u_N\}$ be any basis, orthogonal or nonorthogonal, for $W_N \subset V$

Let $\vec{v}$ be any vector in V. Then

$$\langle w \mid \underbrace{v - w^*}_{h^*} \rangle = 0 \quad \forall\, w \in W_N \quad \text{i.e. } h^* \perp W_N$$

if and only if ($\Updownarrow$)

$$\boxed{\begin{array}{c} \langle u_1 \mid v - w^* \rangle = 0 \\ \vdots \\ \langle u_N \mid v - w^* \rangle = 0 \end{array}} \qquad (**)$$

error is handout

Proof: Use $w = c_1 u_1 + \cdots + c_N u_N$ to show $\Uparrow$. Use $w = u_1, \ldots, u_N$ to show $\Downarrow$.

not u

Comment 1

The boxed Eqs(**) at the bottom of page 30.3 are known as the normal equations for the least squares problem.

Introducing Eq. (*) on page 30.2 into the Eqs (**) one obtains N equations in the N unknowns $b_1^*, \dots, b_N^*$:

$$\boxed{\begin{array}{c} \langle u_1|u_1\rangle b_1^* + \cdots + \langle u_1|u_N\rangle b_N^* = \langle u_1|v\rangle \\ \vdots \\ \langle u_N|u_1\rangle b_1^* + \cdots + \langle u_N|u_N\rangle b_N^* = \langle u_N|v\rangle \end{array}} \quad (***)$$

Comment 2

If $\{u_1, \dots, u_N\}$ is an orthogonal basis one recovers

$$w^* = \sum_{k=1}^{N} \vec{u}_k b_k^* = \sum_{k=1}^{N} \vec{u}_k \frac{\langle u_k|v\rangle}{\langle u_k|u_k\rangle} = \sum_{k=1}^{N} e_k \langle e_k|v\rangle$$

which is the equation at the very bottom of page 29.10 of Lecture 29.

C) THE SUBSPACE MATRIX

Comment 3

One obtains additional knowledge about the geometrical and algebraic properties of the least squares problem by introducing the subspace matrix

$$A : R^N \to W_N : A = \begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_N \\ \vdots & \vdots & & \vdots \end{bmatrix} \qquad (\star)$$

whose N columns are the basis vectors which span the subspace W_N.

Every aspect of the least squares problem, including the normal equations, Eqs (***) on page 30.4, can be reexpressed more economically in terms of that matrix A.

Thus one has

$$w^* = \vec{u}_1 b_1^* + \cdots + \vec{u}_N b_N^*$$

$$= A \begin{bmatrix} b_1^* \\ \vdots \\ b_N^* \end{bmatrix} \equiv A\, \vec{b}^*$$

For illustrative purposes let us assume that $W_N \subset R^m$ is a subspace of R^m and that $v \in R^m$. Then the inner products in Eqs (***) on page 30.4 can be written in terms of row vectors and column vectors:

$$\langle \vec{u}_i \,|\, \vec{u}_j \rangle = [\cdots \vec{u}_i^T \cdots] \begin{bmatrix} \vdots \\ \vec{u}_j \\ \vdots \end{bmatrix} = u_i^T u_j$$

Thus Eq. (***) becomes

$$\underbrace{\begin{bmatrix} \vec{u}_1^T \vec{u}_1 & \vec{u}_1^T \vec{u}_2 & \cdots & \vec{u}_1^T \vec{u}_N \\ \vdots & & & \vdots \\ \vec{u}_N^T \vec{u}_1 & \vec{u}_N^T \vec{u}_2 & \cdots & \vec{u}_N^T \vec{u}_N \end{bmatrix}}_{N\times m \begin{bmatrix} - \vec{u}_1^T - \\ \vdots \\ - \vec{u}_N^T - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_N \\ | & | & & | \end{bmatrix}_{m\times N} = A^T A \quad N\times N} \begin{bmatrix} b_1^* \\ b_2^* \\ \vdots \\ b_N^* \end{bmatrix} = \begin{bmatrix} \vec{u}_1^T \vec{v} \\ \vdots \\ \vec{u}_N^T \vec{v} \end{bmatrix} \quad (\star\star)$$

The matrix $A^T A$ on the l.h.s. is the product of the transpose A^T of matrix A, Eq. (★) on page 30.6, and A itself:

$$A^T A = \underbrace{\begin{bmatrix} \cdots u_1^T \cdots \\ \vdots \\ \cdots u_N^T \cdots \end{bmatrix}}_{A^T} \underbrace{\begin{bmatrix} \vdots & \vdots & & \vdots \\ u_1 & u_2 & \cdots & u_N \\ \vdots & \vdots & & \vdots \end{bmatrix}}_{A} = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & \cdots & u_1^T u_N \\ \vdots \\ u_N^T u_1 & u_N^T u_1 & & u_N^T u_N \end{bmatrix}$$

It follows that Eq. (∗∗∗) on page 30.4, or equivalently, Eq. (★★) on page 30.6, have the form

$$A^T A \begin{bmatrix} b_1^* \\ \vdots \\ b_N^* \end{bmatrix} = A^T v$$

or

$$A^T A\, b^* = A^T v$$

where $b^* \in R^N$ and $v \in R^m$

OPTIMIZATION

D) SOLUTION VIA SUBSPACE MATRIX,

Summary

For the least squares problem, one has

(which is expressed by orthogonality statement $\langle w | w^* - v\rangle = 0$ $\forall w \in W_N$)

(i) the normal equation

$\langle w | w^* - v\rangle = 0 \iff A^T A\, b^* = A^T v$

for $w = u_1, \ldots, u_N$

where $A = \begin{bmatrix} \vdots & \vdots & & \vdots \\ u_1 & u_2 & & u_N \\ \vdots & \vdots & & \vdots \end{bmatrix} : R^N \to R^m$ (= # of data points)

$$\boxed{\mathcal{R}(A) = W_N}$$

(ii) $\boxed{A^T A \text{ is invertible}}$

Proof:

a) $\dim \mathcal{N}(A) + \dim \mathcal{R}(A) = \dim \text{Domain } A$

$\boxed{\text{Exercise}} \to \dim \mathcal{N}(A) = \mathcal{N}(A^T A)$, $\quad \dim \mathcal{R}(A) = \dim W_N = N$, $\quad \dim \text{Domain } A = N$

$\therefore \dim \mathcal{N}(A^T A) = 0$ i.e. $A^T A$ is one-to-one,

b) $\dim \mathcal{N}(A^T A) + \dim \mathcal{R}(A^T A) = \dim \text{domain } A^T A$

$0 + \dim \mathcal{R}(A^T A) = N$

$\therefore \mathcal{R}(A^T A) = R^N$ i.e. $A^T A$ is onto

c) Hence $A^T A : R^N \to R^N$ is invertible

(iii) The normal equation $A^TAb^* = A^Tv$ has a unique solution

$$\boxed{b^* = (A^TA)^{-1}A^Tv}$$

$$\begin{bmatrix} — u_1 — \\ \vdots \\ — u_N — \end{bmatrix}$$

(iv) The least squares solution is

$$\boxed{\begin{aligned} w^* &= Ab^* \\ w^* &= A(A^TA)^{-1}A^Tv \end{aligned}}$$

Comment ("Left Inverse")

The eq'n

$$b^* = (A^TA)^{-1}A^Tv$$

illustrates the existence of a left inverse for a transformation

$$b^* \in R^N \mapsto Ab^* \in R^m \qquad N < m$$

which is not onto, but is one-to-one, i.e. A has lin. indep. columns.

One can find the <u>best</u> sol'n to

$$Ab^* = v$$

even though $v \in R(A)$. This best sol'n is obtained by multiplying by A^T

$$A^TAb^* = A^Tv.$$

The left inverse of A is therefore

$(A^TA)A^T$: $b^* = (A^TA)^{-1}A^Tv$.