2.5 Least Squares Estimation and the Kalman Filter

Weighted Least Squares

To minimize $w_1^2(A\vec{x}-\vec{b})_1^2 + w_2^2(A\vec{x}-\vec{b})_2^2 + \cdots$

$$= \|W\vec{e}\|^2 = w_1^2 e_1^2 + w_2^2 e_2^2 + \cdots$$

To find the best least squares solution to

$$WA\vec{x} = W\vec{b}$$

$$\Rightarrow \quad (WA)^T WA\vec{x} = (WA)^T W\vec{b}$$

$$A^T W^T W A\vec{x} = A^T W^T W\vec{b}$$

$$\vec{x} = (A^T C A)^{-1} A^T C \vec{b} = L\vec{b}, \text{ where } C = W^T W$$

Example:

Minimize $(x-70)^2 + (x-80)^2 + (x-120)^2$

$$\Rightarrow \quad (x-70) + (x-80) + (x-120) = 0$$

or $$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 70 \\ 80 \\ 120 \end{bmatrix}$$

or $$3x = 270$$

Minimize $w_1^2(x-70)^2 + w_2^2(x-80)^2 + w_3^2(x-120)^2$

$$\Rightarrow \quad w_1^2(x-70) + w_2^2(x-80) + w_3^2(x-120) = 0$$

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} w_1^2 & & \\ & w_2^2 & \\ & & w_3^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} w_1^2 & & \\ & w_2^2 & \\ & & w_3^2 \end{bmatrix} \begin{bmatrix} 70 \\ 80 \\ 120 \end{bmatrix}$$

or $$x = \frac{70 w_1^2 + 80 w_2^2 + 120 w_3^2}{w_1^2 + w_2^2 + w_3^2}$$

Review of Probability Theory

S : a sample space

$E, F (\subset S)$: events

$p : S \to [0,1]$: probability s.t $\sum_{s \in S} p(s) = 1$

$$p(E) = \sum_{s \in E} p(s)$$

E, F : independent $\Leftrightarrow p(E \cap F) = p(E) \cdot p(F)$

$X, Y : S \to \mathbb{R}$: random variables

$E(X) = \sum_{s \in S} p(s) X(s)$: expected value

$$E(X+Y) = E(X) + E(Y)$$
$$E(aX+b) = aE(X) + b$$

∘ X, Y : independent
$\Leftrightarrow p(X=r_1 \text{ and } Y=r_2) = p(X=r_1) \cdot p(Y=r_2)$

∘ X, Y : independent random variables
$\Rightarrow E(XY) = E(X) \cdot E(Y)$

$V(X) = \sum (X(s) - E(X))^2 p(s)$: variance

$\sqrt{V(X)}$: standard deviation

$\Rightarrow V(X) = E(X^2) - E(X)^2$

$$V(X+Y) = V(X) + V(Y)$$

The Choice of $C = W^T W$

$\vec{e} = (e_1, \cdots, e_m)^T$: errors

$E(e_i) = 0$: unbiased

covariance $= E(e_i e_j) = \iint (e_i)(e_j) \cdot \begin{pmatrix} \text{joint probability} \\ \text{of } e_i \text{ and } e_j \end{pmatrix}$

(A) In the independent case,

$$E(e_i e_j) = E(e_i) E(e_j) = 0 \cdot 0 = 0$$

$\Rightarrow$ The right weights are $w_i = 1/\sigma_i$.

Least squares minimizes

$$\|W(\vec{b} - A\vec{x})\|^2 = \|W\vec{e}\|^2 = \frac{e_1^2}{\sigma_1^2} + \cdots + \frac{e_n^2}{\sigma_n^2}$$

W and $C = W^T W$ are diagonal matrices

$\hookrightarrow$ with diagonal $1/\sigma_i^2$

(B) In the general (dependent) case,

$V = E(\vec{e}\,\vec{e}^T)$: covariance matrix

$$\begin{cases} V_{ii} = E(e_i^2) : \text{diagonal entries} \\ V_{ij} = E(e_i e_j) : \text{off-diagonal entries.} \end{cases}$$

$\vec{b}$: measurements

$\vec{e} = \vec{b} - A\vec{x}$: errors

$E[\vec{e}] = E[\vec{b} - A\vec{x}] = \vec{0}$: unbiased

$\Rightarrow \hat{\vec{x}} = L\vec{b}$: estimation of the true but unknown parameters $\vec{x}$ from the measurments $\vec{b}$

① linear (L is a matrix)

② unbiased ($E[\vec{x} - \hat{\vec{x}}] = \vec{0}$)

$$\vec{0} = E[\vec{x} - \hat{\vec{x}}] = E[\vec{x} - L\vec{b}]$$
$$= E[\vec{x} - L(A\vec{x} + \vec{e})]$$
$$= E[\vec{x} - LA\vec{x}] - L\underbrace{E[\vec{e}]}_{=\vec{0}}$$

$\therefore L$ is unbiased if $\underline{LA = I}$

(L : a left inverse of A)

o. The best linear and unbiased estimate (BLUE) is the one with $C = V^{-1}$. The optimal estimate $\hat{\vec{x}}$ and the optimal matrix L_0 are

$$\hat{\vec{x}} = L_0\vec{b} = (A^T V^{-1} A)^{-1} A^T V^{-1} \vec{b}$$

o. This choice minimizes the expected error in the estimate, measured by the covariance matrix

$$P = E[(\vec{x} - \hat{\vec{x}})(\vec{x} - \hat{\vec{x}})^T].$$

o. The covariance matrix P for the error in $\hat{\vec{x}}$ is

$$P = (A^T V^{-1} A)^{-1}$$

$P^{-1} = (A^T V^{-1} A)$: the information matrix

$= (A^T C A)$ measuring the information content

<proof>

To minimize

$$
\begin{aligned}
P &= E[(\vec{x}-\hat{\vec{x}})(\vec{x}-\hat{\vec{x}})^T] \\
&= E[(\vec{x}-L\vec{b})(\vec{x}-L\vec{b})^T] \\
&= E[(\vec{x}-LA\vec{x}-L\vec{e})(\vec{x}-LA\vec{x}-L\vec{e})^T] \\
&= E[(L\vec{e})(L\vec{e})^T] \quad (\because \vec{x}=LA\vec{x}) \\
&= LE[\vec{e}\,\vec{e}^T]L^T \quad (\because L: \text{linear}) \\
&= LVL^T \\
&= [L_0+(L-L_0)]V[L_0+(L-L_0)]^T \\
&= L_0VL_0^T + \underbrace{(L-L_0)VL_0^T + L_0V(L-L_0)^T}_{=0} \text{ ①} \\
&\quad + (L-L_0)V(L-L_0)^T \\
&= L_0VL_0^T + (L-L_0)V(L-L_0)^T \geqq L_0VL_0^T
\end{aligned}
$$

(equality holds if $L=L_0$)

$$
\begin{aligned}
\ulcorner \text{①} \quad (L-L_0)VL_0^T &= (L-L_0)VV^{-1}A(A^TV^{-1}A)^{-1} \\
&= (L-L_0)A(A^TV^{-1}A)^{-1} \\
&= (I-I)(A^TV^{-1}A)^{-1} = 0 \lrcorner \\
&(\because LA = L_0A = I)
\end{aligned}
$$

$$
\begin{aligned}
P &= L_0VL_0^T \\
&= (A^TV^{-1}A)^{-1}A^T\overbrace{V^{-1}V}^{=I}V^{-1}A(A^TV^{-1}A)^{-1} \\
&= (A^TV^{-1}A)^{-1}\underbrace{A^TV^{-1}A(A^TV^{-1}A)^{-1}}_{=I} \\
&= (A^TV^{-1}A)^{-1}
\end{aligned}
$$

↳ P gives the expected errors in $\hat{\vec{x}}$ just as V gave the expected errors in $\vec{b}$

Remark

Let $V^{-1} = C = W^T W$,

$$\hat{\vec{e}} = W\vec{e} = W(\vec{b} - A\vec{x}) \Rightarrow E[\hat{\vec{e}}] = 0$$

$$E[\hat{\vec{e}} \cdot \hat{\vec{e}}^T] = E[(W\vec{e})(W\vec{e})^T]$$

$$= W E[\vec{e} \cdot \vec{e}^T] W^T$$

$$= W V W^T = I$$

A change of variables reduces the problem to a unit covariance problem

Example:

m different measurments,

each equally reliable $\Rightarrow$ $\hat{x}$ will be the average

$E[(x - x_i)^2] = \sigma^2$

$$P^{-1} = A^T V^{-1} A = [1 \cdots 1] \begin{bmatrix} 1/\sigma^2 & & \\ & \ddots & \\ & & 1/\sigma^2 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{m}{\sigma^2}$$

$V = \sigma^2 \Rightarrow P = \sigma^2/m$