Optimal Linear Filtering

3 Types of Estimation Problems:

(i) Filtering – estimate $\underline{x}_t$ given $\{y(\tau), \tau < t\}$.

(ii) Prediction – estimate $\underline{x}_{t+\delta}$ given $\{y(\tau), \tau < t\}$.

(iii) Smoothing – estimate $\underline{x}_t$ given $\{y(\tau), \tau \in [t_0, T]\}$ and $t \in [t_0, T]$.

Solution approach:

- Define optimization problem.
- Solve to obtain minimizing estimator.

The optimal estimator must be

- Unbiased : $E\{\hat{\underline{x}}_t\} = E\{\underline{x}_t\}$
- Minimum variance : Its error variance must be less than the error variance of all other possible estimators.
- Consistent : $\hat{\underline{x}}_t \rightarrow \underline{x}_t$ as the number of measurements increases.

I. We begin with a simple case: Static linear observation.

$\underline{x}$ unknown r.v.; $\underline{x} \in \mathbb{R}^n$; $\underline{v} \in \mathbb{R}^\ell$ a r.v. with $E[\underline{v}] = \underline{0}$.

Measurement is linear, corrupted by additive noise:

$$\underline{z} = \underline{H}\underline{x} + \underline{v}$$

Let $\hat{\underline{x}}$ be the estimate of $\underline{x}$ given measurement $\underline{z}$.

Define

$$J = (\underline{z} - \underline{H}\hat{\underline{x}})^T(\underline{z} - \underline{H}\hat{\underline{x}})$$

= sum of squared errors between measurement $\underline{z}$ and predicted value of $\underline{z}$ given $\hat{\underline{x}}$.

Note that $J \geq 0$.

Solve for $\hat{\underline{x}} = \arg\min\limits_{\hat{\underline{x}}} J$

Expanding,

$$J = \underline{z}^T\underline{z} - 2\hat{\underline{x}}^T\underline{H}^T\underline{z} + \hat{\underline{x}}^T\underline{H}^T\underline{H}\hat{\underline{x}}$$

$$\frac{\partial J}{\partial \hat{\underline{x}}} = \underline{0} = -2\underline{H}^T\underline{z} + 2\underline{H}^T\underline{H}\hat{\underline{x}}$$

Solving for $\hat{\underline{x}}$: $\hat{\underline{x}} = (\underline{H}^T\underline{H})^{-1}\underline{H}^T\underline{z}$ if $(H^TH)^{-1}$ exists

A better way:

$$(\underline{U}, \underline{\Sigma}, \underline{V}) = \mathrm{svd}(\underline{H})$$

$(\underline{H} = \underline{U}\underline{\Sigma}\underline{V}^T;$
$\underline{\Sigma} = \mathrm{diag}(\sigma_i))$

$$\hat{\underline{x}} = \underline{V}\underline{\Sigma}^+\underline{U}^T\underline{z}$$

where $\underline{\Sigma}^+ = \mathrm{diag}(\sigma_i^+)$

$$\sigma_i^+ = \begin{cases} 1/\sigma_i & \text{for } \sigma_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

Weighted Least Squares :

$$J = (\underline{z} - \underline{H}\hat{\underline{x}})^T \underline{R}^{-1} (\underline{z} - \underline{H}\hat{\underline{x}}) \qquad ; \; \underline{R} > 0$$

$$= \underline{z}^T \underline{R}^{-1} \underline{z} - 2\hat{\underline{x}}^T \underline{H}^T \underline{R}^{-1} \underline{z} + \hat{\underline{x}}^T \underline{H}^T \underline{R}^{-1} \underline{H} \hat{\underline{x}}$$

$$\frac{\partial J}{\partial \hat{\underline{x}}} = 0 = -2\underline{H}^T \underline{R}^{-1} \underline{z} + 2\underline{H}^T \underline{R}^{-1} \underline{H} \hat{\underline{x}}$$

$$\hat{\underline{x}} = (\underline{H}^T \underline{R}^{-1} \underline{H})^{-1} \underline{H}^T \underline{R}^{-1} \underline{z} \qquad \text{if } (\underline{H}^T \underline{R}^{-1} \underline{H})^{-1} \text{ exists}$$

A better way: Let $\underline{S}$ be the Cholesky decomposition of R : $\underline{R} = \underline{S}\underline{S}^T$

Then

$$J = (\underline{z} - \underline{H}\hat{\underline{x}})^T \underline{S}^{-T} \underline{S}^{-1} (\underline{z} - \underline{H}\hat{\underline{x}})$$

$$= (\underline{S}^{-1}\underline{z} - \underline{S}^{-1}\underline{H}\hat{\underline{x}})^T (\underline{S}^{-1}\underline{z} - \underline{S}^{-1}\underline{H}\hat{\underline{x}})$$

Therefore, solve

(1) $\underline{S}\underline{y} = \underline{z}$ for $\underline{y}$

(2) $\underline{S}\underline{M} = \underline{H}$ for $\underline{M}$

(3) $(\underline{U}, \underline{\Sigma}, \underline{V}) = \text{svd}(\underline{M})$; $\underline{M} = \underline{U}\underline{\Sigma}\underline{V}^T$

(4) $\hat{\underline{x}} = \underline{V}\underline{\Sigma}^{+}\underline{U}^T\underline{y}$

These problems can be formulated & solved (at least) 3 ways:

(A) Least-Squares Estimate

(B) Probabilistic Estimate (Bayes Rule or Minimum Variance)

(C) Maximum-Likelihood Estimate

The results for the linear case with Gaussian r.v.'s (zero mean) are the same!

Gaussian r.v.'s : measurement z is Gaussian

$$p(\underline{z}|\underline{x}) = \frac{1}{(2\pi)^{\ell/2}|R|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{z}-\underline{H}\underline{x})^T \underline{R}^{-1}(\underline{z}-\underline{H}\underline{x})\right\}$$

Max-Like estimation: maximize $p(z|x)$ over x

$\Leftrightarrow$ minimizing

$$-\frac{1}{2}(\underline{z}-\underline{H}\hat{\underline{x}})^T \underline{R}^{-1}(\underline{z}-\underline{H}\underline{x}) \quad \text{over } \underline{x}$$

to obtain $\hat{\underline{x}}$.

Bayes' Rule : $p(\underline{x}|z) = \dfrac{p(z|x)\,p(\underline{x})}{p(\underline{z})} = \dfrac{p(z|x)\,p(\underline{x})}{\int_{\underline{x}} p(z|\underline{x})\,p(\underline{x})\,d\underline{x}}$

The $\hat{\underline{x}}$ that maximizes $p(\underline{x}|\underline{z})$ is the mean of $\underline{x}|\underline{z}$

$\Leftrightarrow$ multiply through $p(\underline{z}|\underline{x})p(\underline{x})$ and simplify to obtain single $\exp\{\ \}$ term, and minimize its argument to find $\hat{\underline{x}}$.

Minimum Variance : Define

$$J = E_{x|z}\left\{(\hat{\underline{x}}-\underline{x})^T \underline{S}(\hat{\underline{x}}-\underline{x})\right\} = E[x|z]$$

↑ independent of $\underline{S}$!

The General Approach:

(1) Define a <u>Loss</u> function $L(\hat{\underline{x}}-\underline{x})$

(2) Choose $\hat{\underline{x}}$ to minimize $E_{x|z}[L(\hat{\underline{x}}-\underline{x})]$.

Let

$$\underline{x} \sim N(\underline{0}, \underline{P}_0), \quad \underline{v} \sim N(\underline{0}, \underline{R}), \quad \underline{z} = \underline{H}\underline{x} + \underline{v}$$

then

$$p(\underline{z}|\underline{x}) = p_v(\underline{z} - \underline{H}\underline{x}) = \frac{1}{(2\pi)^{l/2}|\underline{R}|^{1/2}} \exp\left\{-\frac{1}{2}(\underline{z} - \underline{H}\underline{x})^T \underline{R}^{-1}(\underline{z} - \underline{H}\underline{x})\right\}$$

$$p(\underline{x}) = \frac{1}{(2\pi)^{n/2}|\underline{P}_0|^{1/2}} \exp\left\{-\frac{1}{2}\underline{x}^T \underline{P}_0^{-1}\underline{x}\right\}$$

$$p(\underline{z}|\underline{x})p(\underline{x}) = \frac{1}{(2\pi)^{\frac{l+n}{2}}|\underline{R}|^{1/2}|\underline{P}_0|^{1/2}} \exp\left\{-\frac{1}{2}\left[(\underline{z} - \underline{H}\underline{x})^T \underline{R}^{-1}(\underline{z} - \underline{H}\underline{x}) + \underline{x}^T \underline{P}_0^{-1}\underline{x}\right]\right\}$$

Minimize

$$J = (\underline{z} - \underline{H}\underline{x})^T \underline{R}^{-1}(\underline{z} - \underline{H}\underline{x}) + \underline{x}^T P_0^{-1} x \quad \text{over } \underline{x}$$

to obtain $\hat{x}$:

$$\frac{\partial J}{\partial \underline{x}} = -H^T 2\,\underline{R}^{-1}(\underline{z} - \underline{H}\underline{x}) + 2P_0^{-1}\underline{x} = 0$$

$$= -2H^T R^{-1} z + [H^T R^{-1} H + P_0^{-1}]\underline{x}$$

$$\boxed{\hat{\underline{x}} = [H^T R^{-1} H + P_0^{-1}]^{-1} H^T R^{-1} z}$$

Another line of reasoning:

Suppose $\underline{x}$ & $\underline{z}$ are correlated zero-mean Gaussian r.v.

Define
$$\underline{\Gamma}_{zz} = E[\underline{z}\underline{z}^T] \quad (\text{covariance of } \underline{z})$$
$$\underline{\Gamma}_{xz} = E[\underline{x}\underline{z}^T] \quad (\text{cross-covariance of } \underline{x} \text{ \& } \underline{z})$$

Let
$$\underline{\eta} = \underline{x} - \underline{\Gamma}_{xz}\underline{\Gamma}_{zz}^{-1}\underline{z}$$

Then
$$E[\underline{z}\underline{\eta}^T] = E[\underline{z}\underline{x}^T - \underline{z}\underline{z}^T\underline{\Gamma}_{zz}^{-1}\underline{\Gamma}_{zx}]$$
$$= \underline{\Gamma}_{zx} - \underline{\Gamma}_{zz}\underline{\Gamma}_{zz}^{-1}\underline{\Gamma}_{zx} = \underline{\Gamma}_{zx} - \underline{\Gamma}_{zx} = \underline{0}$$

$\Rightarrow$ $\underline{\eta}$ and $\underline{z}$ are independent (because they are Gaussian).

Three observations

(1) It is possible to remove the information in x about z (correlation) by using an affine operation.

(2) $\underline{\Gamma}_{xz}\underline{\Gamma}_{zz}^{-1}\underline{z}$ is the optimal estimate of $\underline{x}$ given the information in $\underline{z}$.

(3) $\underline{\eta}$ is the information that is unique to $\underline{x}$ (independent of $\underline{z}$). Think of $\underline{\eta}$ and $\underline{z}$ as orthogonal components of information whose linear combination forms $\underline{x}$.

~~If we know in addition that~~ Given

$$\underline{z} = \underline{H}\underline{x} + \underline{v}$$

where $\underline{v}$ is a zero-mean Gaussian r.v. independent of $\underline{x}$, then

$$\underline{\Gamma}_{zz} = E\left[(\underline{H}\underline{x} + \underline{v})(\underline{H}\underline{x} + \underline{v})^T\right] = H\underline{\Gamma}_{xx}H^T + \underline{\Gamma}_{vv}$$

$$\underline{\Gamma}_{xz} = E\left[\underline{x}(\underline{H}\underline{x} + \underline{v})^T\right] = \underline{\Gamma}_{xx}\underline{H}^T$$

and

$$\underline{\eta} = \underline{x} - \underline{\Gamma}_{xx}\underline{H}^T\left(\underline{H}\,\underline{\Gamma}_{xx}\underline{H}^T + \underline{\Gamma}_{vv}\right)^{-1}\underline{z} = \underline{x} - \underline{\Gamma}_{xz}\underline{\Gamma}_{zz}^{-1}\underline{z}$$

We have a second form of the optimal estimate of $\underline{x}$ given z:

$$\boxed{\hat{\underline{x}} = \underline{\Gamma}_{xx}\underline{H}^T\left[\underline{H}\,\underline{\Gamma}_{xx}\underline{H}^T + \underline{\Gamma}_{vv}\right]^{-1}\underline{z}}$$

Using the previous notation,

$$\boxed{\hat{\underline{x}} = \underline{P}_0\underline{H}^T\left[\underline{H}\underline{P}_0\underline{H}^T + \underline{R}\right]^{-1}\underline{z}}$$

Thus, it is evident that the two matrices

$$\underline{P}_0\underline{H}^T\left[\underline{H}\underline{P}_0\underline{H}^T + \underline{R}\right]^{-1} \quad \text{and} \quad \left[\underline{H}^T\underline{R}^{-1}\underline{H} + \underline{P}_0^{-1}\right]^{-1}\underline{H}^T\underline{R}^{-1}$$

should be equal, and in fact they are.

To show this, multiply $\left[\underline{H}^T\underline{R}^{-1}\underline{H} + \underline{P}_0^{-1}\right]^{-1}\underline{H}^T\underline{R}^{-1}$ on the right by $\left[\underline{H}\,\underline{P}_0\underline{H}^T + \underline{R}\right]$:

$$\left[\underline{H}^T\underline{R}^{-1}\underline{H} + \underline{P}_0^{-1}\right]^{-1}\underline{H}^T\underline{R}^{-1}\left[\underline{H}\,\underline{P}_0\underline{H}^T + \underline{R}\right]$$

$$= \left[\underline{H}^T\underline{R}^{-1}\underline{H} + \underline{P}_0^{-1}\right]^{-1}\left[\underline{H}^T\underline{R}^{-1}\underline{H}\,\underline{P}_0\underline{H}^T + \underline{P}_0^{-1}\underline{P}_0\underline{H}^T\right]$$

$$= \left[\underline{H}^T\underline{R}^{-1}\underline{H} + \underline{P}_0^{-1}\right]^{-1}\left[\underline{H}^T\underline{R}^{-1}\underline{H} + \underline{P}_0^{-1}\right]\left[\underline{P}_0\underline{H}^T\right] = \underline{P}_0\underline{H}^T$$

Multiplying on the right by $\left[\underline{H}\,\underline{P}_0\underline{H}^T + \underline{R}\right]^{-1}$ gives:

$$\boxed{\left[\underline{H}^T\underline{R}^{-1}\underline{H} + \underline{P}_0^{-1}\right]^{-1}\underline{H}^T\underline{R}^{-1} = \underline{P}_0\underline{H}^T\left[\underline{H}\,\underline{P}_0\underline{H}^T + \underline{R}\right]^{-1}}$$