Minimum mean square error estimate (MMSE)

- Let us say $\underline{Z} = \begin{bmatrix} \underline{y} \\ \underline{x} \end{bmatrix}$; there is a joint pdf. $p_z(z)$.
- We observe $\underline{y}$ and want to get an estimate of $\underline{x}$ as $\hat{\underline{x}}$.
- We want to minimize $E\{\|\underline{x} - \hat{\underline{x}}\|^2\}$.
- The MMSE Estimate $\hat{x} = E\{\underline{x} | \underline{y}\}$.

$$= \int_{\underline{x}} \underline{x} \; p(\underline{x}|\underline{y}) \; d\underline{x}$$

Gaussian PDF

$$Z = \begin{bmatrix} \underline{y} \\ \underline{x} \end{bmatrix} \in R^N$$

$$p(z) = \frac{1}{(2\pi)^{N/2} |C_z|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(z-\mu_z)^T C_z^{-1} (z-\mu_z)\right] = \mathcal{N}(\mu_z, C_z)$$

Theorem 10.2 (Conditional PDF of Multivariate Gaussian):

If $\underline{x}$ and $\underline{y}$ are jointly Gaussian, where $\underline{x}$ is $k \times 1$ and $\underline{y}$ is $l \times 1$, with mean vector $[E(\underline{x})^T \; E(\underline{y})^T]^T$ and partitioned covariance matrix

$$C = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} = \begin{bmatrix} k \times k & k \times l \\ l \times k & l \times l \end{bmatrix}$$

Theorem is taken from Estimation Theory book of Steven Kay so that

$$p(\underline{x},\underline{y}) = \frac{1}{(2\pi)^{\frac{k+l}{2}} |C|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(\begin{bmatrix}\underline{x}-E(\underline{x})\\ \underline{y}-E(\underline{y})\end{bmatrix}\right)^T C^{-1}\left(\begin{bmatrix}\underline{x}-E(\underline{x})\\ \underline{y}-E(\underline{y})\end{bmatrix}\right)\right],$$

then the conditional PDF $p(\underline{y}|\underline{x})$ is a Gaussian and

$$E(\underline{y}|\underline{x}) = E(\underline{y}) + C_{yx} C_{xx}^{-1}(\underline{x}-E(\underline{x})) \text{ and}$$

$$C_{y|x} = C_{yy} - C_{yx} C_{xx}^{-1} C_{xy}.$$

10.6 Bayesian Linear Model

(Kay book)

$$\underline{x} = H\underline{\theta} + \underline{w}$$

where $\underline{x}$ is $N\times 1$ data vector, H is a known $N\times p$ matrix, $\underline{\theta}$ is a $p\times 1$ random vector with pdf $\mathcal{N}(\mu_\theta, C_\theta)$ and $\underline{w}$ is an $N\times 1$ noise vector with pdf $\mathcal{N}(\underline{0}, C_w)$ and independent of $\underline{\theta}$. This data model is called Baysian Linear Model.

- What we need? $\hat{\underline{\theta}} = E\{\underline{\theta}|\underline{x}\}$
- So, we must need to know the posterior pdf $p(\underline{\theta}|\underline{x})$.
- Let $Z = \begin{bmatrix}\underline{x}\\ \underline{\theta}\end{bmatrix}$

$$= \begin{bmatrix} H\underline{\theta}+\underline{w} \\ \underline{\theta}\end{bmatrix} = \begin{bmatrix} H_{N\times p} & I_{N\times N} \\ I_{p\times p} & 0_{p\times N}\end{bmatrix}\begin{bmatrix}\underline{\theta}\\ \underline{w}\end{bmatrix}$$

- $\underline{\theta}$ and $\underline{w}$ are independent statistically. So they are jointly Gaussian
- Observe $\underline{z}$ is a linear transformation of a jointly Gaussian random vector. So $\underline{z}$ is also Gaussian.
- So, Theorem 10.2 applies directly, and we need to determine the mean and covariance of posterior pdf. We identify $\underline{x}$ as $H\underline{\theta}+\underline{w}$ and $\underline{y}$ as $\underline{\theta}$ to obtain the means

$$E(\underline{x}) = E(H\underline{\theta}+\underline{w}) = H\,E(\underline{\theta}) + E(\underline{w}) = H E(\underline{\theta}) = H\mu_\theta$$

$$E(\underline{y}) = E(\underline{\theta}) = \mu_\theta$$

and covariances

$$
\begin{aligned}
C_{xx} &= E\left[(x - E(x))(x - E(x))^T\right] \\
&= E\left[(H\theta + w - H\mu_\theta)(H\theta + w - H\mu_\theta)^T\right] \\
&= E\left[(H(\theta - \mu_\theta) + w)(H(\theta-\mu_\theta)+w)^T\right] \\
&= E\Big[H(\theta-\mu_\theta)(\theta-\mu_\theta)^T H^T + H(\theta-\mu_\theta)w^T \\
&\qquad + w(\theta-\mu_\theta)^T H^T + w w^T\Big] \\
&\underset{\substack{\uparrow \\ \text{statistical} \\ \text{independence}}}{=} H\,E\left[(\theta-\mu_\theta)(\theta-\mu_\theta)^T\right]H^T + E\left[w w^T\right] \\
&= H\,C_\theta\,H^T + C_w .
\end{aligned}
$$

The cross-covariance matrix is

$$C_{yx} = E\left[(\underline{y} - E(\underline{y}))(\underline{x} - E(\underline{x}))^T\right]$$

$$= E\left[(\underline{\theta} - \mu_\theta)(H\underline{\theta} + \underline{w} - H\mu_\theta)^T\right]$$

$$= E\left[(\underline{\theta} - \mu_\theta)(H(\underline{\theta} - \mu_\theta) + \underline{w})^T\right]$$

$$= E\left[(\underline{\theta} - \mu_\theta)(\underline{\theta} - \mu_\theta)^T\right] H^T = C_\theta H^T.$$

statistical independence

Theorem 10.3 (Posterior pdf for the Bayesian Linear Model)

Same book of Kay

If the observed data $\underline{x}$ can be modeled as

$$\underline{x} = H\underline{\theta} + \underline{w}$$

where $\underline{x}$ is an $N \times 1$ data vector, H is a known $N \times p$ matrix, $\underline{\theta}$ is a $p \times 1$ random vector with pdf $\mathcal{N}(\mu_\theta, C_\theta)$, and $\underline{w}$ is an $N \times 1$ noise vector with pdf $\mathcal{N}(\underline{0}, C_w)$ and independent of $\underline{\theta}$, then the posterior pdf $p(\underline{\theta}|\underline{x})$ is Gaussian with mean

$$E(\underline{\theta}|\underline{x}) = \mu_\theta + C_\theta H^T (H C_\theta H^T + C_w)^{-1} (\underline{x} - H\mu_\theta)$$

and covariance

$$C_{\theta|x} = C_\theta - C_\theta H^T (H C_\theta H^T + C_w)^{-1} H C_\theta .$$

$$\boxed{\text{MMSE Estimate: } E(\underline{\theta}|\underline{x}) = \mu_\theta + C_\theta H^T (H C_\theta H^T + C_w)^{-1} (\underline{x} - H\mu_\theta)}$$