Theorem 10.2: Conditional PDF of Multivariate Gaussian

Let $\mathbf{X}$ ($k\times 1$) and $\mathbf{Y}$ ($l\times 1$) be random vectors distributed jointly Gaussian with mean vector $[E\{\mathbf{X}\}^T \;\; E\{\mathbf{Y}\}^T]^T$ and covariance matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{C_{XX}} & \mathbf{C_{XY}} \\ \mathbf{C_{YX}} & \mathbf{C_{YY}} \end{bmatrix} = \begin{bmatrix} (k\times k) & (k\times l) \\ (l\times k) & (l\times l) \end{bmatrix}$$

Then $p(\mathbf{y}|\mathbf{x})$ is also Gaussian with mean vector and covariance matrix given by:

$$E\{\mathbf{Y} \mid \mathbf{X} = \mathbf{x}_o\} = E\{\mathbf{Y}\} + \mathbf{C_{YX}}\mathbf{C_{XX}^{-1}}\left(\mathbf{x}_o - E\{\mathbf{X}\}\right)$$

$$\mathbf{C}_{\mathbf{Y|X=x}_o} = \mathbf{C_{YY}} - \mathbf{C_{YX}}\mathbf{C_{XX}^{-1}}\mathbf{C_{XY}}$$

$$E\{Y \mid X = x_o\} = E\{Y\} + \frac{\sigma_{XY}}{\sigma_X^2}\left(x_o - E\{X\}\right)$$

$$\operatorname{var}\{Y \mid X = x_o\} = \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}$$

Compare to Bivariate Results

For the Gaussian case… the cond. covariance does not depend on the conditioning x-value!!!

10.6 Bayesian Linear Model

Now we have all the machinery we need to find the MMSE for the "Bayesian Linear Model"

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

- $\mathbf{x}$: $N \times 1$
- $\mathbf{H}$: $N \times p$, known
- $\boldsymbol{\theta}$: $p \times 1$, $\sim N(\boldsymbol{\mu}_\theta, \mathbf{C}_\theta)$
- $\mathbf{w}$: $N \times 1$, $\sim N(\mathbf{0}, \mathbf{C}_w)$

Clearly, $\mathbf{x}$ is Gaussian and θ is Gaussian…

But are they <u>jointly</u> Gaussian???

If yes… then we can use Theorem 10.2 to get the MMSE for $\boldsymbol{\theta}$!!!

Answer = Yes!!

Bayesian Linear Model is Jointly Gaussian

θ and $\mathbf{w}$ are each Gaussian and are independent

Thus their joint PDF is a product of Gaussians…

…which has the form of a jointly Gaussian PDF

Can now use: a linear transform of jointly Gaussian is jointly Gaussian

$$\begin{bmatrix} \mathbf{x} \\ \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{H} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{w} \end{bmatrix}$$

Jointly Gaussian

Thus, Thm. 10.2 applies! Posterior PDF is…

- Joint Gaussian
- Completely described by its mean and variance

Conditional PDF for Bayesian Linear Model

To apply Theorem 10.2, notationally let $\mathbf{X} = \mathbf{x}$ and $\mathbf{Y} = \theta$.

First we need
$$E\{\mathbf{X}\} = \mathbf{H}\, E\{\theta\} + E\{\mathbf{w}\} = \mathbf{H}\mu_\theta$$

$$E\{\mathbf{Y}\} = E\{\theta\} = \mu_\theta$$

And also $\mathbf{C}_{\mathbf{YY}} = \mathbf{C}_\theta$

$$\begin{aligned}\mathbf{C}_{\mathbf{XX}} &= E\left\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})^T\right\} \\ &= E\left\{[\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) + \mathbf{w}][\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) + \mathbf{w}]^T\right\} \\ &= \mathbf{H}\underbrace{E\left\{(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)^T\right\}}_{\mathbf{C}_\theta}\mathbf{H}^T + E\left\{\mathbf{w}\mathbf{w}^T\right\}\end{aligned}$$

Cross Terms are Zero because θ and $\mathbf{w}$ *are independent*

$$\mathbf{C}_{\mathbf{XX}} = \mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + E\left\{\mathbf{w}\mathbf{w}^T\right\}$$

Similarly… $\mathbf{C}_{\mathbf{YX}} = \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}} = E\left\{(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T\right\}$

Use $E\{\boldsymbol{\theta}\mathbf{w}\} = \mathbf{0}$, $E\{\boldsymbol{\mu}_{\boldsymbol{\theta}}\mathbf{w}\} = \mathbf{0}$

$$= E\left\{(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})(\mathbf{H}\boldsymbol{\theta} + \mathbf{w} - \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{\theta}})^T\right\}$$

$$= E\left\{(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^T \mathbf{H}^T\right\}$$

$\Rightarrow$ $\mathbf{C}_{\boldsymbol{\theta}\mathbf{x}} = \mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T$

Then Theorem 10.2 gives the conditional PDF's mean and cov
(and we know the conditional mean is the MMSE estimate)

Posterior Mean:

$$\hat{\boldsymbol{\theta}}_{MMSE} = E\{\boldsymbol{\theta} \mid \mathbf{x}\} = \boldsymbol{\mu}_{\boldsymbol{\theta}} + \mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_{\mathbf{w}}\right)^{-1}(\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{\theta}})$$

- *Cross Correlation $\mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}$* — $\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T$
- *Relative Quality* — $\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_{\mathbf{w}}\right)^{-1}$
- *a priori estimate* — $\boldsymbol{\mu}_{\boldsymbol{\theta}}$
- *Update Transformation* / *Maps unpredictable part* — $\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_{\mathbf{w}}\right)^{-1}$
- *Data Prediction Error* — $(\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{\theta}})$

Bayesian MMSE Estimator

Posterior Covariance:

$$\mathbf{C}_{\boldsymbol{\theta}|\mathbf{x}} = \mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_{\mathbf{w}}\right)^{-1}\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}$$

- *a priori covariance* — $\mathbf{C}_{\boldsymbol{\theta}}$
- *Reduction Due to Data* — $\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_{\mathbf{w}}\right)^{-1}\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}$