Rob 501 Fall 2014
Lecture 15
Typeset by: Connie Qiu
Proofread by: Bo Lin
Revised by Grizzle on 29 October 2015

Minimum Variance Estimator

$$y = Cx + \epsilon, y \in \mathbb{R}^m, x \in \mathbb{R}^n, \text{and } \epsilon \in \mathbb{R}^m.$$

Stochastic assumptions:

$E\{x\} = 0, E\{\epsilon\} = 0$ (means).

$E\{\epsilon\epsilon^\top\} = Q, E\{xx^\top\} = P, E\{\epsilon x^\top\} = 0$ (covariances).

Remark: $E\{\epsilon x^\top\} = 0$ implies that the states and noise are uncorrelated. Recall that uncorrelated does NOT imply independence, except for Gaussian random variables.

Assumptions: $Q \geq 0, P \geq 0, CPC^\top + Q > 0$. (will see why later)

Objective: minimize the variance

$$E\{\|\hat{x} - x\|^2\} = E\{\sum_{i=1}^{n}(\hat{x}_i - x_i)^2\} = \sum_{i=1}^{n} E\{(\hat{x}_i - x_i)^2\}.$$

We see that there are n separate optimization problems.

Remark: suppose $\hat{x} = Ky$. It is automatically unbiased, because

$$E\{\hat{x}\} = E\{Ky\} = E\{KCx + K\epsilon\} = KCE\{x\} + KE\{\epsilon\} = 0 = E\{x\}$$

Problem Formulation: We will pose this as a minimum norm problem in a vector space of random variables.

$$\mathcal{F} = \mathbb{R},$$

$$\mathcal{X} = span\{x_1, x_2, \ldots, x_n, \epsilon_1, \epsilon_2, \ldots, \epsilon_m\},$$

where

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_m \end{bmatrix}.$$

For $z_1, z_2 \in \mathcal{X}$, we <u>define their inner product</u> by:

$$< z_1, z_2 >= E\{z_1 z_2\}$$

$M = span\{y_1, y_2, \ldots, y_m\} \subset \mathcal{X}$ (measurements),

$y_i = C_i x + \epsilon_i = \sum\limits_{j=1}^{n} C_{ij} x_j + \epsilon_i, 1 \le i \le m$, ($i$-th row of y)

$\hat{x}_i = \arg\min\limits_{m \in M} \|x_i - m\| = d(x, M)$.

Fact: $\{y_1, y_2, \ldots, y_m\}$ is linearly independent if, and only if, $CPC^\top + Q$ is positive definite. This is proven below when we compute the Gram matrix. (Recall, $\{y_1, y_2, \ldots, y_m\}$ linearly independent if, and only if G is full rank, where $G_{ij} :=< y_i, y_j >$.)

Solution via the Normal Equations

By the normal equations,

$$\hat{x}_i = \hat{\alpha}_1 y_1 + \hat{\alpha}_2 y_2 + \cdots + \hat{\alpha}_m y_m$$

where $G^\top \hat{\alpha} = \beta$.

$$\begin{aligned}
G_{ij} =< y_i, y_j >= E\{y_i y_j\} &= E\{[C_i x + \epsilon_i][C_j x + \epsilon_j]\} \\
&= E\{[C_i x + \epsilon_i][C_j x + \epsilon_j]^\top\} \\
&= E\{[C_i x + \epsilon_i][x^\top C_j{}^\top + \epsilon_j]\} \\
&= E\{C_i x x^\top C_j^\top\} + E\{C_i x \epsilon_j\} + E\{\epsilon_i x^\top C_j^\top\} + E\{\epsilon_i \epsilon_j\} \\
&= C_i E\{x x^\top\} C_j^\top + E\{\epsilon_i \epsilon_j\} \\
&= C_i P C_j^\top + Q_{ij} \\
&= [CPC^\top + Q]_{ij}
\end{aligned}$$

where we have used the fact that x and ϵ are uncorrelated. We conclude that

$$G = CPC^\top + Q.$$

We now turn to computing β. Let's note that x_i, the i-th component of x is equal to $x^\top e_i$, where e_i is the standard basis vector in $\mathbb{R}^n$.

$$\begin{aligned}
\beta_j =< x_i, y_j > &= E\{x_i y_j\} \\
&= E\{x_i [C_j x + \epsilon_j]\} \\
&= E\{x_i C_j x\} + E\{x_i \epsilon_j\} \\
&= C_j E\{x x_i\} \\
&= C_j E\{x x^\top e_i\} \\
&= C_j E\{x x^\top\} e_i \\
&= C_j P e_i \\
&= C_j P_i
\end{aligned}$$

where $P = [P_1 | P_2 | \ldots | P_n]$.

Putting all this together, we have

$$\begin{aligned}
G^\top \hat{\alpha} &= \beta \\
&\Updownarrow \\
[CPC^\top + Q]\hat{\alpha} &= CP_i \\
&\Updownarrow \\
\hat{\alpha} &= [CPC^\top + Q]^{-1} CP_i
\end{aligned}$$

$\hat{x}_i = \hat{\alpha}_1 y_1 + \hat{\alpha}_2 y_2 + \cdots + \hat{\alpha}_m y_m = \hat{\alpha}^\top y =$ (row vector $\times$ column vector.)

$$\hat{\alpha} = \begin{bmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_m \end{bmatrix}.$$

We now seek to identify the gain matrix K so that

$$\hat{x} = Ky \Leftrightarrow \hat{x}_i = K_i y, \text{where} \;\; K = \begin{bmatrix} K_1 \\ \hline K_2 \\ \hline \vdots \\ \hline K_n \end{bmatrix};$$

that is, K_i is the i-th row of K.

$$\begin{aligned}
K_i^\top = \hat{\alpha} &= [CPC^\top + Q]^{-1} CP_i \\
[K_1^\top | \dots | K_n^\top] &= [CPC^\top + Q]^{-1} CP \\
K &= PC^\top [CPC^\top + Q]^{-1}
\end{aligned}$$

$$\boxed{\hat{x} = Ky = PC^\top [CPC^\top + Q]^{-1} y}$$

Remarks:

1. Exercise: $E\{(\hat{x}-x)(\hat{x}-x)^\top\} = P - PC^\top[CPC^\top + Q]^{-1}CP$
2. The term $PC^\top[CPC^\top + Q]^{-1}CP$ represents the "value" of the measurements. It is the reduction in the variance of x given the measurement y.
3. If $Q > 0$ and $P > 0$, then from the Matrix Inversion Lemma

$$\boxed{\hat{x} = Ky = [C^\top Q^{-1}C + P^{-1}]^{-1}C^\top Q^{-1}y.}$$

   This form of the equation is useful for comparing BLUE vs MVE
4. <u>BLUE vs MVE</u>

   - **BLUE:** $\hat{x} = [C^\top Q^{-1}C]^{-1}C^\top Q^{-1}y$
   - **MVE:** $\hat{x} = [C^\top Q^{-1}C + P^{-1}]^{-1}C^\top Q^{-1}y$
   - Hence, BLUE = MVE when $P^{-1} = 0$.
   - $P^{-1} = 0$ roughly means $P = \infty I$, that is infinite covariance in x, which in turn means *no idea* about how x is distributed!
   - For BLUE to exist, we need $\dim(y) \geq \dim(x)$
   - For MVE to exist, we can have $\dim(y) < \dim(x)$ as long as

$$(CPC^\top + Q) > 0$$

Solution to Exercise

We seek $E\{(\hat{x}-x)(\hat{x}-x)^\top\}$ To get started, let's note that

$$\hat{x}-x = Ky - x = KCx + K\epsilon - x = (KC-I)x + K\epsilon$$

and thus

$$(\hat{x}-x)(\hat{x}-x)^\top = (KC-I)xx^\top(KC-I)^\top + K\epsilon\epsilon^\top K^\top - 2(KC-I)x\epsilon^\top K^\top$$

Taking expectations, and recalling that x and ϵ are uncorrelated, we have

$$\begin{aligned} E\{(\hat{x}-x)(\hat{x}-x)^\top\} &= (KC-I)P(KC-I)^\top + KQK^\top \\ &= KCPC^\top K^\top + P - 2PC^\top K^\top + KQK^\top \\ &= P + K[CPC^\top + Q]K^\top - 2PC^\top K^\top \end{aligned}$$

substituting with $K = PC^\top[CPC^\top + Q]^{-1}$ and simplifying yields the result.

Solution to MIL

We will show that if $Q > 0$ and $P > 0$, then

$$PC^\top[CPC^\top + Q]^{-1} = [C^\top Q^{-1}C + P^{-1}]^{-1}C^\top Q^{-1}$$

MIL: Suppose that A, B, C and D are compatible[1] matrices. If A, C, and $(C^{-1} + DA^{-1}B)$ are each square and invertible, then $A + BCD$ is invertible and

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

We apply the MIL to $[C^\top Q^{-1}C + P^{-1}]^{-1}$, where we identify $A = P^{-1}, B = C^\top, C = Q^{-1}, D = C$. This yields

$$[C^\top Q^{-1}C + P^{-1}]^{-1} = P - PC^\top[Q + CPC^\top]^{-1}CP$$

Hence

$$\begin{aligned}
[C^\top Q^{-1}C + P^{-1}]^{-1}C^\top Q^{-1} &= PC^\top Q^{-1} - PC^\top[Q + CPC^\top]^{-1}CPC^\top Q^{-1} \\
&= PC^\top \left[I - [Q + CPC^\top]^{-1}CPC^\top\right] Q^{-1} \\
&= PC^\top\left[[Q+CPC^\top]^{-1}[Q+CPC^\top] - [Q+CPC^\top]^{-1}CPC^\top\right]Q^{-1} \\
&= PC^\top[Q + CPC^\top]^{-1} \left[[Q + CPC^\top] - CPC^\top\right] Q^{-1} \\
&= PC^\top[Q + CPC^\top]^{-1} \left[Q + CPC^\top - CPC^\top\right] Q^{-1} \\
&= PC^\top[Q + CPC^\top]^{-1} [Q]\, Q^{-1} \\
&= PC^\top[Q + CPC^\top]^{-1}
\end{aligned}$$

[1] The sizes are such the matrix products and sum in $A + BCD$ make sense.