01 November 2018

Review / Background

- Deterministic Weighted Least Squares

$$y = Cx + e \quad , \quad x \in \mathbb{R}^n, y \in \mathbb{R}^m \qquad (e \in \mathbb{R}^m)$$

Columns of C linearly indep ($\Leftrightarrow$ rank $C = n$), $\underset{m \times m}{S} > 0$

x = unknown, y = measurement, e "explains" why $\nexists\, x \in \mathbb{R}^n$ s.t. $y - Cx = 0$ (overdetermined equations)

$$\hat{x} := \underset{x \in \mathbb{R}^n}{\arg\min} \| \underbrace{y - Cx}_{e} \|_S^2 = (C^T S C)^{-1} C^T S y$$

$$\|e\|_S^2 := \langle e, e \rangle_S := e^T S e$$

BLUE (Best Linear Unbiased Estimator)

$$y = Cx + \varepsilon \quad , \quad y \in \mathbb{R}^m, \ x \in \mathbb{R}^n, \ \varepsilon \in \mathbb{R}^m$$

$E\{\varepsilon\} = 0$, $\operatorname{cov}\{\varepsilon\} = E\{\varepsilon \varepsilon^T\} = Q > 0$

rank $C = n$ (as in Weighted Least Squares)

y = measurement, x = unknown (determinitic) constant and ε = noise model to explain $\nexists$ $x \in \mathbb{R}^n$ such that $y - Cx = 0$.

$\hat{x} = Ky$ (Linear), $E\{\hat{x}\} = x$ (unbiased)

$$\Leftrightarrow KC = I_{n \times n}$$

Seek optimal gain $\hat{K}$ to minimize the variance of $\hat{x}$

$$\hat{K} = \underset{KC = I_{n\times n}}{\arg\min} \; Var(\tilde{x})$$

$$= \underset{KC = I_{n\times n}}{\arg\min} \; E\{(\hat{x}-x)(\hat{x}-x)^T\}$$

$$\vdots$$

$$= \underset{KC = I_{n\times n}}{\arg\min} \; tr(KQK^T)$$

$\therefore \hat{x} = \hat{K}y$

$$\hat{K} = (C^TQ^{-1}C)^{-1}C^TQ^{-1}$$

$$\boxed{\therefore \hat{x} = (C^TQ^{-1}C)^{-1}C^TQ^{-1}y}$$

Weighted Least Squares = BLUE when $S = Q^{-1}$ = Information Matrix of the noise term.

Let $X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$, $\mu = E\{X\} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$

$$\begin{aligned} var(X) &= E\{(X-\mu)^T \cancel{(X-\mu)^T}\} \\ &= E\left\{\sum_{i=1}^{n} (X_i-\mu_i)^2\right\} \\ &= \sum_{i=1}^{n} E\{(X_i-\mu_i)^2\} \\ &= \sum_{i=1}^{n} var(X_i) \end{aligned}$$

$\therefore$ Variance of a Vector is the Sum of the Variances of its Components.

Minimum Variance Estimator (MVE)

$y = Cx + \varepsilon \qquad y \in \mathbb{R}^m,\ x \in \mathbb{R}^n,\ \varepsilon \in \mathbb{R}^m$

Stochastic Assumptions

Means: $E\{x\} = 0$, $E\{\varepsilon\} = 0$

Covariances

$$E\{\varepsilon\varepsilon^T\} = Q \quad , \quad E\{xx^T\} = P \quad , \quad E\{\varepsilon x^T\} = 0$$

One says that ε and x are uncorrelated.

$Q \geq 0$, $P \geq 0$, $CPC^T + Q > 0$ (later we will linear independence of some vectors !!!)

Objective: Find estimate of x, based on y, the minimizes the variance.

Call the estimate $\hat{x}$

$$\text{minimize} \quad E\{(x-\hat{x})^T(x-\hat{x})\}$$

Remark $\hat{x} = Ky$ is automatically unbiased ; $E\{\hat{x}\} = KE\{y\}$

$$= K\,E\{Cx+\varepsilon\}$$

$= KCE\{x\} + K\,E\{\varepsilon\}$

$= 0 = E\{x\}.$

Reminder: $\operatorname{span}\{1, t, t^2, \dots, t^n\}.$

Random variables are functions!

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad x_i : \Omega \longrightarrow \mathbb{R}$$

For normal = Gaussian random variables,

$\Omega = \mathbb{R}$

Ω = Probability Space.

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

$\mathcal{X} = \operatorname{span}\{x_1, x_2, \dots, x_n, \varepsilon_1, \dots, \varepsilon_m\}$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{bmatrix}, \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Define the inner product so that the norm squared of a vector in $\mathcal{X}$ is equal to its variance.

$z_1, z_2 \in \mathcal{X}$, $\langle z_1, z_2 \rangle := E\{z_1 z_2\}$.

Note: $z \in \mathcal{X} \Rightarrow E\{z\} = 0$ because

$z = \gamma_1 x_1 + \cdots + \gamma_n x_n + \gamma_{n+1} \varepsilon_1 + \cdots + \gamma_{n+m} \varepsilon_m$

$\|z\|^2 = \langle z, z \rangle = E\{z^2\} = var(z)$ because z has zero mean.

$$\mathcal{M} = \underbrace{span\{y_1, \ldots, y_m\}}_{\text{subspace}} \subset \mathcal{X}$$

$$\hat{x}_i = \underset{m \in \mathcal{M}}{\arg\min} \|x_i - m\|^2$$

To solve this we write down the normal equations!!!

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

$$= \hat{\alpha}^T y \qquad \hat{\alpha} = \begin{bmatrix} \hat{\alpha}_1 \\ \vdots \\ \vdots \\ \hat{\alpha}_m \end{bmatrix}$$

$$G^T \hat{\alpha} = \beta$$

$$G^T = G$$

$$[G]_{ij} = E\{y_i y_j\} = \langle y_i, y_j \rangle$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = CX + \varepsilon = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix} X + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

$$C = \begin{bmatrix} \underline{c_1} \\ \underline{c_2} \\ \vdots \\ \overline{c_m} \end{bmatrix}$$

$$G_{ij} = \langle y_i, y_j \rangle = E\{y_i y_j\}$$

$$= E\{(c_i X + \varepsilon_i)(c_j X + \varepsilon_j)\}$$

$$= E\{(c_i X)(c_j X) + \cancel{c_i X \varepsilon_j}^{\,0} + \cancel{\varepsilon_i c_j X}^{\,0} + \varepsilon_i \varepsilon_j\}$$

$$= E\{(c_i X)(c_j X)^T + \varepsilon_i \varepsilon_j\}$$

$$= E\{c_i X X^T c_j^T + \varepsilon_i \varepsilon_j\}$$

$$= C_i \, E\{xx^T\} \, C_j^T + Q_{ij}$$

$$= C_i P C_j^T + Q_{ij}$$

$$= \left[CPC^T + Q \right]_{ij}$$

$G = CPC^T + Q > 0$ $\{y_1, \ldots, y_m\}$ are linearly independent !!!

$$\beta_j = \langle x_i, y_j \rangle$$

$$= E\{ x_i (C_j x + \varepsilon_j) \}$$

$$= E\{ x_i \underbrace{C_j}_{} x \} + \cancelto{0}{E\{ x_i \varepsilon_j \}}$$

$$= E\{ x_i x^T C_j^T \}$$

$$= P_i C_j^T \qquad P = \left[P_1 | P_2 | \cdots | P_n \right]$$

~~T~~ I did not want to write this way ☹

$$\beta_j = \langle x_i, y_j \rangle$$
$$= E\{x_i (C_j x + \varepsilon_j)\}$$
$$= E\{x_i C_j x + x_i \varepsilon_j\}$$
$$= C_j E\{x_i x\} + \cancel{E\{x_i \varepsilon_j\}} \to 0$$
$$= C_j P_i \qquad P = [P_1 | \cdots | P_n]$$

$$G \hat{\alpha} = \beta$$
$$\hat{\alpha} = G^{-1} \beta$$
$$\hat{\alpha} = [C P C^T + Q]^{-1} C P_i$$

A few more steps gives

$$\boxed{\begin{aligned} \hat{x} &= K y \\ K &= P C^T [C P C^T + Q]^{-1} \end{aligned}}$$

A priori uncertainty in X

Posteriori uncertaint

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

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

$$Z = \begin{bmatrix} x \\ y \end{bmatrix}, \quad E\{Z\} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{aligned} \mathrm{cov}(Z) &= E\{ZZ^T\} \\ &= E\left\{ \begin{bmatrix} x \\ y \end{bmatrix} \begin{bmatrix} x^T & y^T \end{bmatrix} \right\} \\ &= E\left\{ \begin{bmatrix} xx^T & xy^T \\ yx^T & yy^T \end{bmatrix} \right\} \\ &= \text{(exercise)} \\ &= \begin{bmatrix} P & PC^T \\ CP & CPC^T + \Phi \end{bmatrix} \end{aligned}$$

Schur complement of P in $\begin{bmatrix} P & PC^T \\ CP & CPC^T+\Phi \end{bmatrix}$

$$P - PC^T\left[CPC^T + \Phi\right]^{-1}CP$$

$\boxed{\text{Done with MUE.}}$