least-squares

goal: optimal estimation through linear observations

ref: Stengel Ch 4.1

---

previously: use observations $y = Cx$ to estimate $\hat{x}$

- solve linear equations
- applying observer

now: use noisy observations $z = Cx + \eta$ to estimate $\hat{x}$

where η is an external disturbance / measurement noise

- we want to choose $\hat{x}$ to minimize $\underbrace{\|x - \hat{x}\|}$, i.e. estimation error

  but: we don't know x, so can't evaluate

- instead, we'll minimize $\|z - C\hat{x}\|$, i.e. observation error

$\longrightarrow$ solve NLP $\min\limits_{\hat{x}} \|z - C\hat{x}\|_2^2 = (z - C\hat{x})^T (z - C\hat{x})$

→ solve NLP $\min_{\hat{x}} \|z - C\hat{x}\|_2^2 = (z - C\hat{x})^T(z - C\hat{x})$

– let $J(\hat{x}) = \|z - C\hat{x}\|_2^2$, then minima are <u>stationary</u>:

$$D_{\hat{x}} J = C^T C \hat{x} - C^T z$$

$$= 0 \underset{\uparrow}{\iff} \hat{x} = (C^T C)^{-1} C^T z \leftarrow \text{"least squares" estimate}$$

assume $\operatorname{rank} C^T C = \dim \hat{x}$ – necessary that # rows $C \geq$ # cols C

– to confirm $\hat{x}$ is minimum: $D_{\hat{x}}^2 J(\hat{x}) = C^T C > 0$

→ confirm that $C^T C > 0$, i.e. all eigenvalues are positive

– let $v \neq 0$, $\lambda \in \mathbb{R}$ be eigenvalue/eigenvector pair for $C^T C$

so $C^T C v = \lambda v$

so $0 \leq \|Cv\|^2 = v^T C^T C v = \lambda \|v\|^2$; since $\|v\| \neq 0$, $\lambda > 0$

→ compute $\hat{x}$ for $z = Cx + \eta$, $x \in \mathbb{R}$, $z \in \mathbb{R}^k$, $C = \mathbb{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$
(this should be familiar)

– $\hat{x} = \underbrace{(C^T C)^{-1}}_{=k} \underbrace{C^T z}_{= \sum_{i=1}^{k} z_i} = \frac{1}{k} \sum_{\ell=1}^{k} z_\ell$, i.e. sample mean

e.g. x is the length of something, z is measurement of that length

◦ now consider weighted least-squares objective:

$$J(\hat{x}) = (z - C\hat{x})^T S^{-1} (z - C\hat{x}), \quad S = S^T > 0$$

→ verify that $S^{-1} = (S^{-1})^T > 0$

→ solve NLP $\min_{\hat{x}} J(\hat{x})$

– $\hat{x} = (C^T S^{-1} C)^{-1} C^T S^{-1} z$ ← agrees with previous answer when $S = I$

– could have changed coordinates to $\varsigma = \text{"}\sqrt{S}\text{"} \cdot z$ and applied

– $\varsigma = M z \Rightarrow z = M^{-1} \varsigma$

so $J(\hat{x}) = (M^{-1}\varsigma - C\hat{x})^T (M^{-1}\varsigma - C\hat{x})$

$$= (\varsigma - MC\hat{x})^T \underbrace{(M^{-1})^T M^{-1}}_{S^{-1}, \text{ i.e. } M^T M = S, \text{ i.e. } \text{"}M = \sqrt{S}\text{"}} (\varsigma - M\hat{x})$$

<u>fact:</u> if $S = Cov[\eta]$, $E[\eta] = 0$, then

$\hat{x}$ has the minimum (co-)variance out of all unbiased estimators

– note that $E[\hat{x}] = (C^T S^{-1} C)^{-1} C^T S^{-1} E[z]$,

$$E[z] = E[Cx] + \cancel{E[\eta]} \to 0$$

so $E[\hat{x}] = (C^T S^{-1} C)^{-1} (C^T S^{-1} C) E[x] = E[x]$

- suppose we estimate $\underset{\in \mathbb{R}^d}{\hat{x}_1}$ using batch of measurements $\underset{\in \mathbb{R}^{k_1}}{z_1} = C_1 x + \eta_1$

and subsequently obtain new measurements $z_2 = C_2 x + \eta_2$

- then we want to minimize $J(\hat{x}) = \| z - C\hat{x} \|^2_{S^{-1}}$

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

where $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$, $C = \cancel{\begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix}}$ $S = \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}$, $\hat{x} = \cancel{\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix}}$

$C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$ $\hat{x} = \hat{x}_2$

* we want to solve for $\hat{x}_2$ in terms of $\hat{x}_1$ that minimizes J

→ solve $D_{\hat{x}_2} J = 0$ for $\hat{x}_2$ in terms of $\cancel{\hat{x}_1}$ z_1, z_2

- $\hat{x}_2 = (C_1^T S_1^{-1} C_1 + C_2^T S_2^{-1} C_2)^{-1} (C_1^T S_1^{-1} z_1 + C_2^T S_2^{-1} z_2)$

- $J(\hat{x}) = (z - C\hat{x})^T S^{-1} (z - C\hat{x}) = \left\| \begin{bmatrix} z_1 - C_1 \hat{x}_2 \\ z_2 - C_2 \hat{x}_2 \end{bmatrix} \right\|^2_{S^{-1}}$

$\Rightarrow D_{\hat{x}_2} J = -(z_1 - C_1 \hat{x}_2)^T S_1^{-1} C_1 - (z_2 - C_2 \hat{x}_2)^T S_2^{-1} C_2$

$= 0 \iff C_1^T S_1^{-1} C_1 \hat{x}_2 + C_2^T S_2^{-1} C_2 \hat{x}_2$

$$= C_1^T S_1^{-1} z_1 + C_2^T S_2^{-1} z_2$$

$$\Longleftrightarrow \quad \hat{x}_2 = \left(C_1^T S_1^{-1} C_1 + C_2^T S_2^{-1} C_2\right)^{-1} \left(C_1^T S_1^{-1} z_1 + C_2^T S_2^{-1} z_2\right)$$

→ apply matrix inversion Lemma with $P_1^{-1} = C_1^T S_1^{-1} C_1$

$$(A + U H V)^{-1} = A^{-1} - A^{-1} U \left(H^{-1} + V A^{-1} U\right)^{-1} V A^{-1}$$

$$- \left(\underbrace{C_1^T S_1^{-1} C_1}_{P_1^{-1}} + C_2^T S_2^{-1} C_2\right)^{-1} = P_1 - P_1 C_2^T \left(C_2 P_1 C_2^T + S_2\right)^{-1} C_2 P_1$$

→ substitute into formula for $\hat{x}_2$ w/ $\hat{x}_1 = \left(C_1^T S_1^{-1} C_1\right)^{-1} C_1^T S_1^{-1} z_1$

$$- \quad \hat{x}_2 = \hat{x}_1 + K_2\left(z_2 - C_2 \hat{x}_1\right), \quad K_2 = P_1 C_2^T \left(C_2 P_1 C_2^T + S_2\right)^{-1}$$