Rob 501 - Mathematics for Robotics
Recitation #8

Nils Smit-Anseeuw (Courtesy: Abhishek Venkataraman, Wubing Qin)

Nov 13, 2018

1 Application of Projection Theorem

1. **Theorem:** In an inner product space $(\mathbb{R}^n, \mathbb{R}, < \cdot, \cdot >)$ with $< x, y >= x^\top Qy$ where $Q \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix, and $\|x\|_Q = \sqrt{< x, x >}$. Given a matrix $A \in \mathbb{R}^{m\times n}$, $m \geq n$, $\text{rank}\{A\} = n$, and vector $b \in \mathbb{R}^m$, then

   $$\hat{x} := \arg\min_x \|Ax - b\|_Q^2$$

   exists, is unique and given by $\hat{x} = (A^\top QA)^{-1}A^\top Qb$.

   Note: In this case, $Ax = b$ is an over-determined or over-constrained equation, thus there is no such x satisfying $Ax = b$, but we can find a solution $\hat{x}$ such that the error $A\hat{x} - b$ has the minimum norm.

2. **Theorem:** In an inner product space $(\mathbb{R}^n, \mathbb{R}, < \cdot, \cdot >)$ with $< x, y >= x^\top Qy$ where $Q \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix, and $\|x\|_Q = \sqrt{< x, x >}$. Given a matrix $A \in \mathbb{R}^{m\times n}$, $m \leq n$, $\text{rank}\{A\} = m$, and vector $b \in \mathbb{R}^m$, then

   $$\hat{x} := \arg\min_{Ax=b} \|x\|_Q^2$$

   exists, is unique and given by $\hat{x} = Q^{-1}A^\top(AQ^{-1}A^\top)^{-1}b$.

   Note: In this case, $Ax = b$ is an under-determined or under-constrained equation, thus there are infinitely many solutions x satisfying $Ax = b$, and among all the solutions we can find a solution $\hat{x}$ such that $\hat{x}$ has the minimum norm.

3. **Least squares:** Model:

   - $y_i = C_i x + \epsilon_i$, $x \in \mathbb{R}^n$, $y_i \in \mathbb{R}^{m_i}$, $\epsilon_i \in \mathbb{R}^{m_i}$, $C_i \in \mathbb{R}^{m_i\times n}$
   - i is time index, y_i is measurement, x is a deterministic but unknown vector that we would like to estimate, ϵ_i is the unknown measurement noise.

- Compute an estimate of x at time k, using all the available measurements y_i, $i = 1, 2, \ldots, k$ such that

$$\hat{x}_k = \arg\min_{x\in\mathbb{R}^n} \left(\sum_{i=1}^{n} (y_i - C_i x)^\top S_i (y_i - C_i x) \right)$$

  where S_i is the weighting factor and $S_i > 0$ is symmetric positive definite .

Solution: Let $m_1 + m_2 + \ldots + m_k = m$ and suppose $m \geq n$. If we define

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} \in \mathbb{R}^m,\ C = \begin{bmatrix} C_1 \\ \vdots \\ C_k \end{bmatrix} \in \mathbb{R}^{m\times n},\ \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_k \end{bmatrix} \in \mathbb{R}^m,\ R = \begin{bmatrix} S_1 & & \\ & \ddots & \\ & & S_k \end{bmatrix} \in \mathbb{R}^{m\times m},$$

Then the measurements become $y = Cx + \epsilon$.
In the inner product space $(\mathbb{R}^n, \mathbb{R}, < \cdot, \cdot >)$, we define $< x, y >= x^\top R y$ and $\|x\|_R = \sqrt{< x, x >}$ where $Q \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix. The the minimization problem can be re-interpreted as

$$\hat{x}_k = \arg\min_{x\in\mathbb{R}^n} \|Cx - y\|_R^2$$

The least squares solution gives us $\hat{x}_k = (C^\top RC)^{-1}C^\top R\,y$, i.e., $\hat{x}_k = K\,y$ where $K = (C^\top RC)^{-1}C^\top R$. You can compare this result with those given by BLUE and MVE later.

2

	BLUE	MVE
Model	$y = Cx + \epsilon$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $\epsilon \in \mathbb{R}^m$, $C \in \mathbb{R}^{m \times n}$, $m > n$, $\mathrm{rank}(C) = n$	
Assumption	x is deterministic but unknown	$\mathbb{E}[x] = 0$, $\mathbb{E}[xx^\top] = \mathrm{cov}(x,x) = P \geq 0 \in \mathbb{R}^{n \times n}$
	$\mathbb{E}[\epsilon] = 0$, $\mathbb{E}[\epsilon\epsilon^\top] = \mathrm{cov}(\epsilon,\epsilon) = Q \geq 0 \in \mathbb{R}^{m \times m}$	
	$\mathbb{E}[\epsilon x^\top] = \mathbb{E}[\epsilon]x^\top = 0$	$\mathbb{E}[\epsilon x^\top] = 0$, $Q + CPC^\top > 0$
Objective	Given a linear estimator $\hat{x} = Ky$, $K \in \mathbb{R}^{n \times m}$, that is unbiased ($\mathbb{E}[x - \hat{x}] = \mathbb{E}[(I - KC)x] - K\mathbb{E}[\epsilon] = (I - KC)\mathbb{E}[x] = 0$) Find the best $\hat{K} = \arg\min_K \mathbb{E}[\|x - \hat{x}\|_2^2] = \arg\min_K \mathbb{E}[(x - \hat{x})^\top (x - \hat{x})] = \arg\min_K \mathbb{E}[\mathrm{trace}\big((x - \hat{x})^\top (x - \hat{x})\big)]$ $= \arg\min_K \mathbb{E}[\mathrm{trace}\big((x - \hat{x})(x - \hat{x})^\top\big)] = \arg\min_K \mathrm{trace}\big(\mathbb{E}[(x - \hat{x})(x - \hat{x})^\top]\big)$	
Unbiased Estimator	$(I - KC)\mathbb{E}[x] = 0 \implies KC = I$	$(I - KC)\mathbb{E}[x] = 0$ is automatically true (since $\mathbb{E}[x] = 0$)
Covariance matrix of estimation error	$\mathrm{cov}(x - \hat{x}, x - \hat{x}) = \mathbb{E}[(x - \hat{x})(x - \hat{x})^\top] = \mathbb{E}[(x - Ky)(x - Ky)^\top] = \mathbb{E}[(x - KCx - K\epsilon)(x - KCx - K\epsilon)^\top]$ $= \mathbb{E}\left[(I - KC)xx^\top (I - KC)^\top - (I - KC)x\epsilon^\top K^\top - K\epsilon x^\top (I - KC)^\top + K\epsilon\epsilon^\top K^\top\right] \qquad (*)$	
	$(*) = (I - KC)xx^\top (I - KC)^\top + KQK^\top$ $\mathbb{E}[\|x - \hat{x}\|_2^2] = \|(I - KC)x\|_2^2 + \mathrm{trace}(KQK^\top)$ $= \mathrm{trace}(KQK^\top)$ (since $KC = I$)	$(*) = (I - KC)P(I - KC)^\top + KQK^\top$ $\mathbb{E}[\|x - \hat{x}\|_2^2] = \mathrm{trace}\big((I - KC)P(I - KC)^\top + KQK^\top\big)$ $= \mathrm{trace}\left(\begin{bmatrix} I - KC & K \end{bmatrix} \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} (I - KC)^\top \\ K^\top \end{bmatrix}\right)$ $= \mathrm{trace}\left(\left(\begin{bmatrix} C^\top \\ I \end{bmatrix} K^\top - \begin{bmatrix} I \\ 0 \end{bmatrix} \right)^\top \underbrace{\begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}}_{:=R} \left(\underbrace{\begin{bmatrix} C^\top \\ I \end{bmatrix}}_{:=A} K^\top - \underbrace{\begin{bmatrix} I \\ 0 \end{bmatrix}}_{:=B} \right) \right)$
Re-interpretation	$\min_K \mathbb{E}[\|x - \hat{x}\|_2^2] = \min_K \mathrm{trace}(KQK^\top) = \min_K \left(\sum_i^n K_i Q K_i^\top \right)$ $= \sum_i^n \min_{K_i} (K_i Q K_i^\top)$ (K_i is the i-th row) $\hat{K}_i = \arg\min_{K_i} (K_i Q K_i^\top)$ s.t. $C^\top K_i^\top = e_i$	$\min_K \mathbb{E}[\|x - \hat{x}\|_2^2] = \min_K \mathrm{trace}\Big(\big(AK^\top - B\big)^\top R \big(AK^\top - B\big)\Big)$ $= \sum_i^n \min_{K_i} \Big(\big(AK_i^\top - Be_i\big)^\top R \big(AK_i^\top - Be_i\big)\Big)$ $\hat{K}_i = \arg\min_{K_i} \Big(\big(AK_i^\top - Be_i\big)^\top R \big(AK_i^\top - Be_i\big)\Big)$

	BLUE	MVE
solution	For vector space $(\mathbb{R}^m, \mathbb{R})$, define $< x, y >= \text{trace}(x^\top Q y)$ $\|\|x\|\|_Q = \sqrt{< x, x >}$	For vector space $(\mathbb{R}^{m+n}, \mathbb{R})$, define $< x, y >= \text{trace}(x^\top R y)$ $\|\|x\|\|_R = \sqrt{< x, x >}$
	$\hat{K}_i^\top = \arg \min\limits_{C^\top K_i^\top = e_i} \|K_i^\top\|_Q^2$ $\hat{K}_i^\top = Q^{-1} C (C^\top Q^{-1} C)^{-1} e_i$	$\hat{K}_i^\top = \arg \min\limits_{K_i^\top} \|A K_i^\top - B e_i\|_R^2$ $\hat{K}_i^\top = (A^\top R A)^{-1} A^\top R B e_i$
	solution is $\hat{K}^\top = Q^{-1} C (C^\top Q^{-1} C)^{-1}$ or $\hat{K} = (C^\top Q^{-1} C)^{-1} C^\top Q^{-1}$	solution is $\hat{K}^\top = (A^\top R A)^{-1} A^\top R b = (CPC^\top + Q)^{-1} CP$ or $\hat{K} = PC^\top (CPC^\top + Q)^{-1}$ $(**)$ $= (P^{-1} + C^\top Q^{-1} C)^{-1} C^\top Q^{-1}$ if P^{-1} exists
optimal Covariance matrix of estimation error	$(*) = KQK^\top = (C^\top Q^{-1} C)^{-1}$	$(*) = (I - KC) P (I - KC)^\top + KQK^\top$ $= P - KCP - PC^\top K^\top + KCPC^\top K^\top + KQK^\top$ $= P - PC^\top (CPC^\top + Q)^{-1} CP = (P^{-1} + C^\top Q^{-1} C)^{-1}$

Recall that from recitation 6, we have matrix identity $P(I + QP)^{-1} = (I + PQ)^{-1} P$.

$$(**) \quad \hat{K} = PC^\top (CPC^\top + Q)^{-1} = PC^\top (QQ^{-1} CPC^\top + Q)^{-1} = PC^\top \big(Q(Q^{-1} CPC^\top + I)\big)^{-1} = \underbrace{PC^\top}_{}(I + \underbrace{Q^{-1} C}_{} \underbrace{PC^\top}_{})^{-1} Q^{-1}$$

$$= (I + \underbrace{PC^\top}_{} \underbrace{Q^{-1} C}_{})^{-1} \underbrace{PC^\top}_{} Q^{-1} = \Big(P^{-1}(I + PC^\top Q^{-1} C)\Big)^{-1} C^\top Q^{-1} = (P^{-1} + C^\top Q^{-1} C)^{-1} C^\top Q^{-1}$$

The brief comparison among least squares, BLUE and MVE is listed below.

	Least Squares	BLUE	MVE
Model	$y = Cx + \epsilon$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $\epsilon \in \mathbb{R}^m$, $C \in \mathbb{R}^{m \times n}$, $m > n$, $\text{rank}(C) = n$		
	No info about x and ϵ	$\epsilon \sim N(0, Q)$, but no info about x	$\epsilon \sim N(0, Q)$ and $x \sim N(0, P)$
Linear estimator	$\hat{x} = Ky$		
The best linear estimator	$\hat{K} = (C^\top R C)^{-1} C^\top R$	$\hat{K} = (C^\top Q^{-1} C)^{-1} C^\top Q^{-1}$	$\hat{K} = (C^\top Q^{-1} C + P^{-1})^{-1} C^\top Q^{-1}$ $= PC^\top (CPC^\top + Q)^{-1}$
The optimal covariance	N/A	$(C^\top Q^{-1} C)^{-1}$	$(C^\top Q^{-1} C + P^{-1})^{-1}$ $= P - PC^\top (CPC^\top + Q)^{-1} CP$