ROB 501 - Mathematics for Robotics
HW #8

Due 23:59 on Thursday, Nov. 07, 2024
To be submitted on Canvas

Remarks: Problem 2 is very important. Please spend extra time on it as it is very helpful for understanding the Kalman Filter. Problem 3 explains why we can often obtain recursion relations of the type: $\widehat{x}_{k+1}$ is a linear combination of $\widehat{x}_k$ and the "innovation" or new measured information $(y_{k+1} - \widehat{y}_{k+1|k})$. If you are pressed for time, skip Problem 3 and study the solutions. It is better to spend your time on Problem 2.

1. x has been bewitched to give the following data:

$$y = Ax + \epsilon$$

with

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 0 \\ 0 & 6 \end{bmatrix} \quad y = \begin{bmatrix} 1.5377 \\ 3.6948 \\ -7.7193 \\ 7.3621 \end{bmatrix} \quad \text{and} \quad E\{\epsilon\epsilon^\top\} = Q = \begin{bmatrix} 1.00 & 0.50 & 0.50 & 0.25 \\ 0.50 & 2.00 & 0.25 & 1.00 \\ 0.50 & 0.25 & 2.00 & 1.00 \\ 0.25 & 1.00 & 1.00 & 4.00 \end{bmatrix}$$

As in class, $E\{\epsilon\} = 0$.

(a) Find the Best Linear Unbiased Estimate (BLUE) for x, using only the first two values of y. Also compute the covariance of the estimate.

(b) Find the Best Linear Unbiased Estimate (BLUE) for x, using only the first three values of y. Also compute the covariance of the estimate.

(c) Find the Best Linear Unbiased Estimate (BLUE) for x, using all the values of y. Also compute the covariance of the estimate.

Note: For (a), you use the first 2 rows of y and A and the upper 2×2 part of Q. For (b), you use the first 3 rows of y and C, as well as the upper 3×3 part of Q. You see the pattern, I hope. Do all the calculations in MATLAB. You do not have to turn in your code.

2. Read the Handout `GaussianRandomVariablesAndVectors.pdf`, which you can find on CANVAS. We consider three jointly normal random variables (X, Y, Z), with

$$\text{mean} \quad \mu = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and covariance} \quad \Sigma = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 2 \end{bmatrix}$$

(a) Compute the conditional distribution of $\begin{bmatrix} X \\ Y \end{bmatrix} |_{\{Z=z\}}$., the conditional distribution of the vector $[X, Y]^\top$ given $Z = z$, which is the same as the joint distribution of the normal random variables $X|_{\{Z=z\}}$ and $Y|_{\{Z=z\}}$. To be extra clear, give the mean vector and covariance matrix for $[X, Y]^\top$ given $Z = z$.

(b) Compute the distribution of $X|_{\{Z=z\}}$ conditioned on $Y|_{\{Z=z\}} = y$.

(c) Compute the conditional distribution of $X|_{\left[\begin{array}{c} Y=y \\ Z=z \end{array}\right]}$, or more compactly, $X|_{Y=y,Z=z}$, the conditional distribution of X given the vector $[Y=y, Z=z]^\top$.

(d) Compare your answers for (b) and (c).

3. Let $(\mathcal{X}, \mathbb{R}, < \cdot, \cdot >)$ be a finite-dimensional inner product space. Our objective is to understand recursion relations similar to what we see in RLS and the Kalman filter. With that in mind, let $\{y_1, \cdots, y_N\}$ be a linearly independent set in $\mathcal{X}$. For $1 \leq k \leq N$ define

$$M_k := \text{span}\{y_1, \cdots, y_k\}$$

and for $x \in \mathcal{X}$ define $\widehat{x}_k := \arg\min_{m \in M_k} ||x - m||$, which is the orthogonal projection of x onto M_k.

(a) Suppose that for some $1 \leq k < N$, we have $y_{k+1} \perp M_k$ (interpretation: the new measurement y_{k+1} is orthogonal to the previous measurements). Show that there exists $\beta \in \mathbb{R}$ such that

$$\widehat{x}_{k+1} = \widehat{x}_k + \beta y_{k+1},$$

and give a formula for β.

(b) We no longer make any hypothesis about y_{k+1} being orthogonal to M_k. What we do now is, for each $1 \leq k < N$, define

$$\widehat{y}_{k+1|k} = \arg\min_{m \in M_k} ||y_{k+1} - m||,$$

which we interpret as the orthogonal projection of the new measurement y_{k+1} onto the subspace generated by the previous measurements. Show that there exists $\beta \in \mathbb{R}$ such that

$$\widehat{x}_{k+1} = \widehat{x}_k + \beta(y_{k+1} - \widehat{y}_{k+1|k}),$$

and give a formula for β.

Remark: The error term $(y_{k+1} - \widehat{y}_{k+1|k})$ is the "innovations" in the case of the Kalman filter and RLS. What is particularly nice in the case of RLS and the Kalman filter, where we have a model such as $y_k = C_k x_k + v_k$, we can compute $\widehat{y}_{k+1|k}$ directly from $\widehat{x}_k$, often with a formula such as $\widehat{y}_{k+1|k} = C_{k+1}\widehat{x}_k$. In other words, we do not have to solve an extra optimization problem in order to compute $\widehat{y}_{k+1|k}$; instead, we can bootstrap from the previous solution to the optimization problem.

4. You are given the data

$$y = Cx + \epsilon$$

and

$$C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 0 \\ 0 & 6 \end{bmatrix} \quad y = \begin{bmatrix} 1.5377 \\ 3.6948 \\ -7.7193 \\ 7.3621 \end{bmatrix} \quad E\{\epsilon\epsilon^\top\} = Q = \begin{bmatrix} 1.00 & 0.50 & 0.50 & 0.25 \\ 0.50 & 2.00 & 0.25 & 1.00 \\ 0.50 & 0.25 & 2.00 & 1.00 \\ 0.25 & 1.00 & 1.00 & 4.00 \end{bmatrix} \quad E\{xx^\top\} = P = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$$

As in class, $E\{x\} = 0$ and $E\{\epsilon\} = 0$.

(a) Find the Minimum Variance Estimate for x, using only the first value of y and the upper left entry of Q. Also compute the covariance of the estimate.

(b) Find the Minimum Variance Estimate for x, using only the first two values of y and the upper left 2×2 part of Q. Also compute the covariance of the estimate.

(c) Find the Minimum Variance Estimate for x, using only the first three values of y and the upper left 3×3 part of Q. Also compute the covariance of the estimate.

(d) Find the Minimum Variance Estimate for x, using all the values of y and Q. Also compute the covariance of the estimate.

5. This problem reuses some of the data in Problem 1, namely the FULL vector $y \in \mathbb{R}^4$ and the 4×2 matrix C.

   (a) Ignore all of the stochastic data, and do a standard least squares approximation of x, using the inner product $< x, y >= x^\top y$. Yes, the problem is then our usual over determined system of equations.

   (b) Find a BLUE of x assuming $Q = E\{\epsilon\epsilon^\top\} = I$.

   (c) Find the Minimum Variance Estimate for x, assuming $E\{\epsilon\epsilon^\top\} = Q = I$ and $P = E\{xx^\top\} = 100I$ (identity matrix times 100). Repeat for $P = 10^6 I$. (Conceptually, you are taking $P \to \infty I$.)

   (d) Compare all of your estimates[1].

6. This problem reuses the data from **Problem 4**, but this time, the **means are no longer zero**. The minimum variance estimator (MVE) becomes[2]

$$\widehat{x} = \bar{x} + PC^\top (CPC^\top + Q)^{-1}(y - \bar{y}) \text{ and } E\{(x - \hat{x})(x - \hat{x})^\top\} = P - PC^\top (CPC^\top + Q)^{-1}CP,$$

where $\bar{x} = E\{x\}$, $\bar{\epsilon} = E\{\epsilon\}$ and $\bar{y} = C\bar{x} + \bar{\epsilon}$. Assuming the data in Problem 1, plus

$$\bar{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ and } \bar{\epsilon} = 0,$$

determine the MVE for x using the full vector y. Turn in $\widehat{x}$; you do not have to provide the covariance.

[1] How would your comparisons have changed if we had assumed $Q \succ 0$ not equal to the identity matrix, as long as in (a) we used $< x, y >= x^\top Q^{-1} y$? Almost no change at all. BLUE equals weighted deterministic least squares with the weighting equal to the inverse of the covariance matrix of the noise. And MVE reduces to BLUE when the covariance of the unknown x becomes very large. In this case, reading the fine print gave you the answer to the question. :) !

[2] Notice the form of the estimate: best estimate of x given no measurements, plus a gain times the measurement minus its best estimate given the prior knowledge. Compare to RLS, compare to Kalman filter, compare to result in HW #7.

Hints

Hints: Prob. 1 Recall our formulas (using $C := A$ for convenience):

$$\widehat{K} = (C^\top Q^{-1} C)^{-1} C^\top Q^{-1} \quad \text{and} \quad E\{(\widehat{x} - x)(\widehat{x} - x)^\top\} = (C^\top Q^{-1} C)^{-1}$$

For (a), you use the first 2 rows of y and C and the upper 2×2 part of Q. For (b), you use the first 3 rows of y and C, as well as the upper 3×3 part of Q. You see the pattern, I hope. Do all the calculations in MATLAB. You do not have to turn in your code.

Hints: Prob. 2 Print out the handout on Jointly Gaussian Random Vectors, and read it carefully. Note **Fact 1:** Conditional Distributions of Gaussian Random Vectors

(a) Identify $X_1 = \begin{bmatrix} X \\ Y \end{bmatrix}$ and $X_2 = Z$. Based on this, identify and write down Σ_{11}, Σ_{12}, Σ_{21}, and Σ_{22}, and then μ_1 and μ_2, and finally, note that $x_2 = z$. Now apply the formulas for $\mu_{1|2}$ and $\Sigma_{1|2}$. These are the mean and covariance of the jointly normally distributed random variables $X_{|Z=z}$ and $Y_{|Z=z}$.

(b) From (a), we know $X_{|Z=z}$ and $Y_{|Z=z}$ are jointly distributed normal random variables, and we know their mean and covariance. Rename the mean μ and the covariance Σ (i.e., $\mu_{1|2} \to \mu$ and $\Sigma_{1|2} \to \Sigma$). Using Fact 1, identify $X_1 = X_{|Z=z}$ and $X_2 = Y_{|Z=z}$, and then identify and write down Σ_{11}, Σ_{12}, Σ_{21}, and Σ_{22}, and then μ_1 and μ_2, and finally, note that $x_2 = y$. Now apply the formulas for $\mu_{1|2}$ and $\Sigma_{1|2}$. These are the mean and covariance of a normally distributed random variable. Which one? If you do not know, work part (c) and then return here. If you do know, still work part (c).

(c) Go back to the very beginning with our three jointly normal random variables, and this time identify $X_1 = X$ and $X_2 = \begin{bmatrix} Y \\ Z \end{bmatrix}$. Based on this, identify and write down Σ_{11}, Σ_{12}, Σ_{21}, and Σ_{22}, and then μ_1 and μ_2, and finally, note that $x_2 = \begin{bmatrix} y \\ z \end{bmatrix}$. Now apply the formulas for $\mu_{1|2}$ and $\Sigma_{1|2}$. These are the mean and covariance of the normally distributed random variable $X_{|Y=y,Z=z}$.

(d) Read again the handout on Jointly Gaussian Random Vectors. Go back to the beginning and repeat as necessary: **FACT 4:** *If we have jointly distributed normal random vectors, when we condition one block of vectors on another, we always obtain either a jointly distributed normal random vector or, if only a scalar quantity is left, a normally distributed random variable.* This is an amazingly useful property of Gaussian (i.e., normal) random variables.

Hints: Prob. 3 There are several ways to approach part (a):

- Method 1: Let G_k be the Gram matrix for M_k and G_{k+1} be the Gram matrix for M_{k+1}. Then, using $y_{k+1} \perp M_k$, you deduce that

$$G_{k+1} = \begin{bmatrix} G_k & 0_{k\times 1} \\ 0_{1\times k} & < y_{k+1}, y_{k+1} > \end{bmatrix}$$

  Use this block diagonal structure to relate the solution of the normal equations for M_{k+1} to the solution of the normal equations for M_k.

- Method 2: The solution of the optimization problem depends on the subspaces M_k and not on the bases you use for them. Apply Gram Schmidt to produce an ortho**normal** basis for M_k. Relate it to an orthornormal basis for M_{k+1}. Then use what you know about the orthogonal projection of x onto sets with orthonormal bases.

For part (b), we know from the Projection Theorem that $y_{k+1} - \widehat{y}_{k+1|k}$ is orthogonal to M_k. If you need a second hint, note that

$$M_{k+1} = M_k \oplus \text{span}\{y_{k+1}\} = M_k \oplus \text{span}\{y_{k+1} - v\},$$

for any $v \in M_k$. Hence, because $\widehat{y}_{k+1|k} \in M_k$ by definition, we have

$$M_{k+1} = M_k \oplus \text{span}\{y_{k+1} - \widehat{y}_{k+1|k}\} \quad \text{and} \quad M_k \perp (y_{k+1} - \widehat{y}_{k+1|k}).$$

You can now apply your result from (a).

Hints: Prob. 4 Recall our formulas

$$\widehat{K} = PC^\top(CPC^\top + Q)^{-1} = (C^\top Q^{-1}C + P^{-1})^{-1}C^\top Q^{-1}$$

and

$$E\{(\hat{x} - x)(\hat{x} - x)^\top\} = P - PC^\top(CPC^\top + Q)^{-1}CP$$

Hints: Prob. 6 The problem is as simple as it looks. Its purpose is to make you aware of the MVE when the means are non-zero. If you compare this formula to the measurement update step of the Kalman filter, you will see that they are basically the same. This is because computing a conditional expectation with Gaussian random vectors is really an orthogonal projection. You can learn more about this in EECS 564, Estimation, Filtering, and Detection.