ROB 501 Exam-II Solutions

16 December 2019 (Prof. Gregg)

Problem 1:

(a) True. For a symmetric positive definite matrix A, the SVD is equivalent to diagonalization via eigen-decomposition.

(b) False. The QR decomposition only exists with invertible R if the columns of A are linearly independent.

(c) False. As a counterexample, let $E = \sigma_r u_r v_r^T$. Then $\|E\| = \sigma_r$, and $\text{rank}(A - E) = r - 1$.

(d) True. The statement "$m = n$ and $r = p$" tells us that the matrix is square and all its singular values are greater than zero, so matrix A is full rank. Since A is square and invertible, σ_r, the smallest singular value, is the distance from A to the nearest singular matrix.

Problem 2:

(a) False. That consecutive terms of the sequence become arbitrarily close together is not sufficient for the sequence to be Cauchy. We must have $|\sqrt{n} - \sqrt{m}| \to 0$ for all $n, m \geq N < \infty$. A counter example is for any $m > (\sqrt{(n)} + \epsilon)^2$, $|\sqrt{n} - \sqrt{m}| > \epsilon$.

(b) True. Since S is open and $x^* \in S$, $\exists\ \epsilon > 0$ such that $B_\epsilon(x^*) \subset S$. Because $x_k \to x^*$, there exists $N < \infty$ such that $n \geq N$ implies $x_n \in B_\epsilon(x^*) \subset S$, so $d(x_n, \sim S) \geq \epsilon$.

(c) False. $\mathring{\mathbb{Q}} = \{x \in \mathbb{R} \,|\, \mathrm{d}(x, \sim \mathbb{Q}) > 0\} = \{x \in \mathbb{R} \,|\, \mathrm{d}(x, \mathbb{I}) > 0\} = \{x \in \mathbb{Q} \,|\, \mathrm{d}(x, \mathbb{I}) > 0\} \bigcup \{x \in \mathbb{I} \,|\, \mathrm{d}(x, \mathbb{I}) > 0\} = \emptyset$.

(d) True. The normed space $(\mathbb{R}, |\cdot|)$ is finite-dimensional. By the Bolzano-Weierstrass theorem, if S is a closed and bounded subset of a finite-dimensional normed space, every sequence (x_n) in S has a convergent subsequence with limit $x_0 \in S$. This also means that S is compact.
For completeness, we are interested in convergent sequences (x_n), where $x_n \to x_0$ and we want to check if $x_0 \in S$. If $x_n \to x_0$, then for a subsequence (x_{n_i}), $x_{n_i} \to x_0$. By Bolzano-Weierstrass and the fact that the set is closed and bounded, $\exists\ x_{n_i} \to x_0$, such that $x_0 \in S$. Therefore, S is complete. In fact, every compact normed space is also complete.

Problem 3:

(a) True. This is the contraction mapping theorem. For complete set $S \subset \mathcal{X}$, $T : S \to S$ such that $\forall x, y \in S$, $\|T(x) - T(y)\| \leq c\|x - y\|$, $0 \leq c < 1$ is a contraction mapping. According to the contraction mapping theorem, if we define sequence $x_n = T(x_{n-1})$, then (x_n) is Cauchy and converges to unique fixed point $x^* = T(x^*)$ for any initial point $x_0 \in S$.

(b) False. The first part says function $f(x)$ is discontinuous at x_0. The next part says there is no sequence $x_n \to x_0$ such that $f(x_n) \to f(x_0)$. The true statement is that for *at least one* such sequence $x_n \to x_0$, $f(x_n) \nrightarrow f(x_0)$. As a counter example, the following function is right-continuous at $x_0 = 1$ for $c \neq 0$:

$$f(x) = \begin{cases} c, & x \geq 1 \\ 0, & x < 1 \end{cases}$$

Sequence $x_n = 1 + \frac{1}{n}$ converges to $x_0 = 1$ from the right, and $f(x_n) \to f(x_0) = c$. However, sequence $x_n = 1 - \frac{1}{n}$ converges to $x_0 = 1$ from the left, but in this case, $f(x_n) \to 0 \neq f(x_0)$.

(c) True. From the Balzano-Weierstrass theorem, S is a compact set since it is a closed and bounded subset of a finite-dimensional normed space $(\mathcal{X}, \mathbb{R}, \|\cdot\|)$. Also, $f(x) = \|x\|$ is a continuous, real-valued function over S. Thus, the infimum of $f(x) = \|x\|$ can be achieved on S, i.e., there exists $x^* \in S$ such that $\|x^*\| = \inf_{y \in S} \|y\|$, based on the Weierstrass theorem.

(d) False. By the Weierstrass theorem, only a function that is continuous everywhere in C is guaranteed to achieve its extreme values.

Problem 4:

(a) True. The MVE solution is equivalent to the solution to BLUE when P is infinitely large. However, the BLUE estimate only exists if C has linearly independent columns (meaning the estimate might not be unbiased).

(b) False. We also require that $C^\top PC + Q > 0$ in order to invert it in the solution.

(c) True. The minimum variance estimate minimizes the variance of the error vector or the trace of the covariance matrix of the error vector. The mean-squared error is $\mathcal{E}\{\|\hat{x} - x\|^2\} = \mathcal{E}\{< \hat{x} - x, \hat{x} - x > \} = \mathcal{E}\{(\hat{x} - x)^\top (\hat{x} - x)\} = \mathrm{tr}(\mathcal{E}\{(\hat{x} - x)(\hat{x} - x)^\top\})$. Hence, all of these are equivalent.

(d) False. The columns of C must be independent, not the rows.

Problem 5:

(a) True. All covariance matrices are positive semi-definite.

(b) False. For the case of jointly *normally* distributed random vectors, they are uncorrelated if and only if they are independent.

(c) False. We can define a new random variable $Y = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ by marginalizing out X_2, with $\mu_Y = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ and $\Sigma_Y = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$. We can use the conditional covariance formula:

$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba}$$

To find $\Sigma_{3|1}$,

$$\Sigma_{3|1} = 4 - 1(2)^{-1}1 = 4 - 0.5 = 3.5$$

(d) True. This is using the result from HW7, Q5.

Problem 6:

$\hat{x}_{5\|5} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$	$K_6 = \begin{bmatrix} 1/2 \\ 1/8 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.125 \end{bmatrix}$	$\mathcal{E}\left\{ x_6 \| Y_5 \;\middle\|\; y_6 \| Y_5 \right\} = \begin{bmatrix} 1 \\ 5/8 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.625 \end{bmatrix}$

(a) Here we just reverse the Prediction Step:

$$\hat{x}_{5|5} = A^{-1}\hat{x}_{6|5} = \frac{4}{3}\begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1 \end{bmatrix} \hat{x}_{6|5} = \begin{bmatrix} 2/3 & -2/3 \\ 2/3 & 4/3 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

→ 2 points for the formula (with correct indices), 1 point for correct answer.

(b) We first need to perform the Prediction Step on the covariance matrix:

$$\begin{aligned} P_{6|5} =& AP_{5|5}A^T + GRG^T \\ =& \begin{bmatrix} 1 & 0.5 \\ -0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & -0.5 \\ 0.5 & 0.5 \end{bmatrix} + \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} 3 \begin{bmatrix} 0.5 & 0 \end{bmatrix} \\ =& \begin{bmatrix} 3.25 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 0.75 & 0 \\ 0 & 0 \end{bmatrix} \\ =& \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \end{aligned}$$

Then use the formula for the Kalman gain during the Measurement Update Step:

$$\begin{aligned} K_6 =& P_{6|5}C^T(CP_{6|5}C^T + Q)^{-1} \\ =& \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} (CP_{6|5}C^T + Q)^{-1} \\ =& \begin{bmatrix} 4 \\ 1 \end{bmatrix} (4+4)^{-1} \\ =& \begin{bmatrix} 1/2 \\ 1/8 \end{bmatrix} \end{aligned}$$

→ 3 points for each formula (with correct indices), 1 point for each correct answer.

(c) First note that $\mathcal{E}\left\{ x_6|Y_5 \;\middle|\; y_6|Y_5 \right\} = \mathcal{E}\left\{ x_6|Y_6 \right\} = \hat{x}_{6|6}$ (remember this was the main trick behind the Kalman filter). Hence, we need to perform a Measurement Update on the state:

$$\begin{aligned} \hat{x}_{6|6} =& \hat{x}_{6|5} + K_6(y_6 - C\hat{x}_{6|5}) \\ =& \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} + \begin{bmatrix} 1/2 \\ 1/8 \end{bmatrix} (1.5 - \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}) \\ =& \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} + \begin{bmatrix} 1/2 \\ 1/8 \end{bmatrix} (1.5 - 1/2) \\ =& \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} + \begin{bmatrix} 1/2 \\ 1/8 \end{bmatrix} \\ =& \begin{bmatrix} 1 \\ 5/8 \end{bmatrix} \end{aligned}$$

→ 3 points for writing the formula (with correct indices), 1 point for the correct answer.

Problem 7: The given data is that X_1, X_2, X_3 are ZERO MEAN, jointly normal, random variables with covariance matrix

$$\Sigma = \begin{bmatrix} 6 & 2 & 4 \\ 2 & 5 & 1 \\ 4 & 1 & 3 \end{bmatrix}.$$

(a) We express $Y = \begin{bmatrix} X_1 - X_2 \\ X_1 + X_2 \end{bmatrix} = A \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$, with $A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$. From the handout on Gaussian Random Vectors, we have that

$$\Sigma_Y = \mathrm{cov}(Y) = \mathcal{E}\{YY^\top\} = A\Sigma A^\top.$$

Doing the calculations yields

$$\Sigma_Y = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 6 & 2 & 4 \\ 2 & 5 & 1 \\ 4 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 4 & -3 & 3 \\ 8 & 7 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 7 & 1 \\ 1 & 15 \end{bmatrix}$$

A more difficult solution is to write the covariance as

$$\begin{bmatrix} \mathcal{E}\{(X_1 - X_2)^2\} & \mathcal{E}\{(X_1 - X_2)(X_1 + X_2)\} \\ \mathcal{E}\{(X_1 + X_2)(X_1 - X_2)\} & \mathcal{E}\{(X_1 + X_2)^2\} \end{bmatrix}$$

and then expand it out

$$\begin{bmatrix} \mathcal{E}\{(X_1)^2 - 2X_1X_2 + (X_2)^2\} & \mathcal{E}\{(X_1)^2 - (X_2)^2\} \\ \mathcal{E}\{(X_1)^2 - (X_2)^2\} & \mathcal{E}\{(X_1)^2 + 2X_1X_2 + (X_2)^2\} \end{bmatrix}$$

and then use linearity of the expectation to read the individual terms from Σ. It is easy to make mistakes here.

Notes:

- If you expressed $Y = AX$ correctly, and then said $\Sigma_Y = A\Sigma A^\top$, and then made a numerical mistake, you earned 7 of the 8 points.
- If you expressed $Y = AX$ correctly, and then put the transpose on the wrong side, as in $\Sigma_Y = A^\top \Sigma A$, or messed up the transpose itself, you earned 6 of the 8 points.
- If you went the hard way, and got it right, then fine. If you messed it up, you were penalized for not showing knowledge of the first solution given, which is one of the three key facts about Gaussian Random Vectors.
- Several of you proposed the solution

  $$Y = \begin{bmatrix} X_1 - X_2 \\ X_1 + X_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} X_1 + \begin{bmatrix} -1 \\ 1 \end{bmatrix} X_2,$$

  and then said

  $$\Sigma_Y = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Sigma_{11} \begin{bmatrix} -1 & 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix} \Sigma_{22} \begin{bmatrix} -1 & 1 \end{bmatrix} = \begin{bmatrix} 11 & 1 \\ 1 & 11 \end{bmatrix}.$$

  You can see that this removes all of the cross-correlation terms, such as $\mathcal{E}\{X_1X_2\}$. Nevertheless, you did show some familiarity with the "$\Sigma_Y = A\Sigma A^\top$" concept, and hence were awarded half credit (4 points).

(b) This problem really ties together many topics of the course. From the Projection Theorem, we have that $X^* = \alpha_1(X_1 + X_2) + \alpha_2 X_3$, where

$$\begin{bmatrix} < X_1 + X_2, X_1 + X_2 > & < X_1 + X_2, X_3 > \\ < X_3, X_1 + X_2 > & < X_3, X_3 > \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} < X_1 - X_2, X_1 + X_2 > \\ < X_1 - X_2, X_3 > \end{bmatrix},$$

the famous normal equations. Computing the indicated inner products, we have

$$\begin{aligned} < X_1 + X_2, X_1 + X_2 > &= \mathcal{E}\{(X_1 + X_2)(X_1 + X_2)\} = \mathcal{E}\{(X_1)^2\} + 2\mathcal{E}\{X_1 X_2\} + \mathcal{E}\{(X_2)^2\} = \Sigma_{11} + 2\Sigma_{12} + \Sigma_{22} = 15 \\ < X_1 + X_2, X_3 > &= \mathcal{E}\{(X_1 + X_2)(X_3)\} = \mathcal{E}\{X_1 X_3\} + \mathcal{E}\{X_2 X_3\} = \Sigma_{13} + \Sigma_{23} = 5 \\ < X_3, X_1 + X_2 > &= < X_1 + X_2, X_3 > = 5 \\ < X_3, X_3 > &= \mathcal{E}\{(X_3)^2\} = \Sigma_{33} = 3 \end{aligned}$$

$$\begin{aligned} < X_1 - X_2, X_1 + X_2 > &= \mathcal{E}\{(X_1 - X_2)(X_1 + X_2)\} = \mathcal{E}\{(X_1)^2\} - \mathcal{E}\{(X_2)^2\} = \Sigma_{11} - \Sigma_{22} = 1 \\ < X_1 - X_2, X_3 > &= \mathcal{E}\{(X_1 - X_2)(X_3)\} = \mathcal{E}\{X_1 X_3\} - \mathcal{E}\{X_2 X_3\} = \Sigma_{13} - \Sigma_{23} = 3. \end{aligned}$$

Hence, we need to solve

$$\begin{bmatrix} 15 & 5 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

to obtain $X^* = -0.6(X_1 + X_2) + 2X_3$.

Remark: The solution above is the one we expected most of you to give. There is a more clever solution for the Gram matrix! Let's write $Y = \begin{bmatrix} X_1 + X_2 \\ X_3 \end{bmatrix} = A \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$, with $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Then the Gram matrix is the covariance of Y. Just as in part (a), we can use

$$\Sigma_Y = \text{cov}(Y) = \mathcal{E}\{YY^\top\} = A\Sigma A^\top.$$

Doing the calculations yields the same answer in a simpler manner. What about the "β' terms....is there a short cut there too? Well, yes, if you introduce instead $Z = \begin{bmatrix} X_1 + X_2 \\ X_3 \\ X_1 - X_2 \end{bmatrix} = A \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$, with $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$, then computing $\Sigma_Z = \text{cov}(Z) = \mathcal{E}\{ZZ^\top\} = A\Sigma A^\top$ will give all of the terms! We'll let you work it out. Impressively, a few of you used this method!

Notes:

- Writing down the the Normal Equations, namely $X^* = \alpha_1(X_1 + X_2) + \alpha_2 X_3$, where $G\alpha = \beta$, and giving the expressions for the terms that show up in G and β, earned you 6 of the 12 points.
- It is important to express the answer as $X^* = \alpha_1(X_1 + X_2) + \alpha_2 X_3$ and NOT as $X^* = [\alpha_1, \alpha_2]^\top$, as several of you did. Why? Because the answer is a random variable and not a vector in $\mathbb{R}^2$.
- Correctly computing the inner products in the Gram matrix, G, earned 4 points.
- Correctly computing β earned 1 point.
- Getting the α's correct was the final point!

Problem 8:

(a) FALSE, because A does not have linearly independent columns, which is required for R to be invertible in the factorization $A = QR$. To elaborate, we start with the least squares solution $A^T A\hat{x} = A^T b$ where $A^T A = R^T Q^T QR = R^T R$ due to orthogonality of Q, and $A^T b = R^T Q^T b$. Then we have $R^T R\hat{x} = R^T Q^T b$, but we cannot invert R^T to obtain $R\hat{x} = Q^T b$.

(b) TRUE, because $x_0 = x_M + x_{M^\perp}$ for $x_M \in M$ and $x_{M^\perp} \in M^\perp$. In particular,

$$x_M = \arg\min_{y \in M} ||x_0 - y|| \text{ and } x_{M^\perp} = \arg\min_{z \in M^\perp} ||x_0 - z||.$$

The normal equations provide the unique solution $x_M = \alpha_1 y^1 + \ldots + \alpha_k y^k$, and hence

$$x_M^\perp = x_0 - (\alpha_1 y^1 + \ldots + \alpha_k y^k).$$

(c) FALSE. Consider 3/4 of a circle (e.g., a partially eaten apple pie) in 2D real vector space. This is non-convex. Rotate that pie 90 degrees, and it is still non-convex. The intersection of the pie with the rotated pie is half a pie, which is convex. Yum!

✻ ✻ ✻ On behalf of the instructional team for ROB 501, we sincerely hope that you enjoyed the class and have happy holidays! ✻ ✻ ✻