18.06 Problem Set 5 -SOLUTIONS

Problem 1. (a) The two equations are equivalent to

$$\begin{aligned}0 =& \frac{\partial E}{\partial C}\\ =&2(C + Dx_1 - y_1) + 2(C + Dx_2 - y_2) + 2(C + Dx_3 - y_3)\\ =&6C + 2(x_1 + x_2 + x_3)D - 2(y_1 + y_2 + y_3)\end{aligned}$$

and

$$\begin{aligned}0 =& \frac{\partial E}{\partial D}\\ =&2x_1(C + Dx_1 - y_1) + 2x_2(C + Dx_2 - y_2) + 2x_3(C + Dx_3 - y_3)\\ =&2(x_1 + x_2 + x_3)C + 2(x_1^2 + x_2^2 + x_3^2)D - 2(x_1y_1 + x_2y_2 + x_3y_3)\end{aligned}$$

respectively. Both equations are linear to C and D, and the system of equation can be written as

$$\begin{bmatrix} 3 & x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 & x_1^2 + x_2^2 + x_3^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} y_1 + y_2 + y_3 \\ x_1y_2 + x_2y_2 + x_3y_3 \end{bmatrix}.$$

(b) Given A and $\mathbf{b}$, one can check that $A^TA = \begin{bmatrix} 3 & x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 & x_1^2 + x_2^2 + x_3^2 \end{bmatrix}$ and $A^T\mathbf{b} = \begin{bmatrix} y_1 + y_2 + y_3 \\ x_1y_2 + x_2y_2 + x_3y_3 \end{bmatrix}$ holds. Thus, the equation $A^TA\begin{bmatrix} C \\ D \end{bmatrix} = A^T\mathbf{b}$ on C, D is just the same as our system of equations obtained in (a). Since E is minimized when both partial derivatives are equal to zero, C, D minimizes E when

$$\begin{aligned}&\frac{\partial E}{\partial C} = \frac{\partial E}{\partial D} = 0\\ \iff& \begin{bmatrix} 3 & x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 & x_1^2 + x_2^2 + x_3^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} y_1 + y_2 + y_3 \\ x_1y_2 + x_2y_2 + x_3y_3 \end{bmatrix}\\ \iff& A^TA\begin{bmatrix} C \\ D \end{bmatrix} = A^T\mathbf{b}.\end{aligned}$$

(c) There are various ways to show that A^TA is invertible. You can check that the two columns of A are linearly independent (x_1, x_2, x_3 are all distinct so that $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ is not a scalar multiplication of $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$) and follow the proof that A^TA is invertible when the columns of A are independent, which is given in the page 211 of textbook. Alternatively, you can apply a row operation and check that two pivots are 3 and $x_1^2+x_2^2+x_3^2 - \frac{1}{3}(x_1+x_2+x_3)^2$, where the latter is equal to $\frac{1}{3}((x_1-x_2)^2+(x_2-x_3)^2+(x_3-x_1)^2)$. This is equal to zero if only $x_1 = x_2 = x_3$, which is not true by the assumption. So A^TA has two pivots, which implies that it is invertible.

Once you deduce that A^TA is invertible, the rest is simple. You can find that E is minimized when C, D satisfy $\begin{bmatrix} C \\ D \end{bmatrix} = (A^TA)^{-1}A^T\mathbf{b}$, so such C, D exists uniquely.

Remark. (Optional) It is not always true that a multivariable function F on variables C, D is minimized when $\frac{\partial F}{\partial C} = \frac{\partial F}{\partial D} = 0$. There is a famous test that whether F is minimized, maximized, etc. – the second partial derivative test, which is one of important topics covered in 18.02. In our case $E = (C+Dx_1-y_1)^2+(C+Dx_2-y_2)^2+(C+Dx_3-y_3)^2$, the Hessian matrix $H(C,D)$ is just equal to $2A^TA$ whose determinant is equal to $4((x_1-x_2)^2+(x_2-x_3)^2+(x_3-x_1)^2)$. As we have seen in (c), the determinant is non-zero, and is non-negative since it is a sum of squares. Hence, the determinant of $H(C,D)$ would be positive at our C, D which concludes that E is minimized. For more information on the second partial derivative test, see `http://mathworld.wolfram.com/SecondDerivativeTest.html`.

Problem 2. (a) Since $q_1, q_2, \cdots, q_n$ are orthonormal, we have $q_i \cdot q_i = 1$ for all i and $q_i \cdot q_j = 0$ for all $i \neq j$. Thus, we have

$$\begin{aligned} ||c_1q_1 + c_2q_2 + \cdots + c_nq_n||^2 =& (c_1q_1 + c_2q_2 + \cdots + c_nq_n) \cdot (c_1q_1 + c_2q_2 + \cdots + c_nq_n) \\ =& \sum_{i=1}^{n} c_i^2 (q_i \cdot q_i) + \sum_{i<j} 2c_ic_j(q_i \cdot q_j) = \sum_{i=1}^{n} c_i^2. \end{aligned}$$

(b) Suppose that there are some real numbers $c_1, c_2, \ldots, c_n$ such that $c_1q_1 + c_2q_2 + \cdots + c_nq_n = 0$. We have $0 = ||c_1q_1 + \cdots + c_nq_n||^2 = \sum_{i=1}^{n} c_i^2$, which forces all c_i to be zeroes. Hence, $c_1q_1 + c_2q_2 + \cdots + c_nq_n = 0$ implies $c_1 = c_2 = \cdots = c_n = 0$, so that $q_1, q_2, \ldots, q_n$ are independent by definition.

Problem 3. (a) Let the three column vectors be $\mathbf{a} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$. We follow the Gram-Schmidt process on those. We obtain $\mathbf{A} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$, $\mathbf{B} = \mathbf{b} - \frac{\mathbf{A}^T\mathbf{b}}{\mathbf{A}^T\mathbf{A}}\mathbf{A} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \frac{6}{9}\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 2/3 \\ -1/3 \end{bmatrix}$, and $\mathbf{C} = \mathbf{c} - \frac{\mathbf{A}^T\mathbf{c}}{\mathbf{A}^T\mathbf{A}}\mathbf{A} - \frac{\mathbf{B}^T\mathbf{c}}{\mathbf{B}^T\mathbf{B}}\mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} - \frac{6}{9}\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} - \frac{-1}{1}\begin{bmatrix} 2/3 \\ 2/3 \\ -1/3 \end{bmatrix} = \begin{bmatrix} 4/3 \\ -2/3 \\ 4/3 \end{bmatrix}$. After rescaling, we have orthogonal matrix $Q = \begin{bmatrix} -1/3 & 2/3 & 2/3 \\ 2/3 & 2/3 & -1/3 \\ 2/3 & -1/3 & 2/3 \end{bmatrix}$, and $R = Q^TA = \begin{bmatrix} 3 & 2 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \end{bmatrix}$.

(b) Similarly, we obtain $\mathbf{A} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} -1 \\ -2/5 \\ 4/5 \end{bmatrix}$, and $\mathbf{C} = \begin{bmatrix} 4/3 \\ -2/3 \\ 4/3 \end{bmatrix}$. After rescaling, we obtain $Q = \begin{bmatrix} 0 & -\frac{5}{3\sqrt{5}} & \frac{2}{3} \\ \frac{2}{\sqrt{5}} & -\frac{2}{3\sqrt{5}} & -\frac{1}{3} \\ \frac{1}{\sqrt{5}} & \frac{4}{3\sqrt{5}} & \frac{2}{3} \end{bmatrix}$ and $R = Q^TA = \begin{bmatrix} \sqrt{5} & \frac{6}{\sqrt{5}} & \frac{3}{\sqrt{5}} \\ 0 & \frac{3}{\sqrt{5}} & \frac{4}{\sqrt{5}} \\ 0 & 0 & 2 \end{bmatrix}$.

Remark. You can see that if you switch the first two columns, the result of Gram-Schmidt process is changed. However, you can see that the third columns of both Q are the same. This is not by accident. Recall that $\mathbf{C}$ is the error vector for the projection of $\mathbf{c}$ onto the subspace spanned by $\mathbf{a}$ and $\mathbf{b}$. Even if you exchange $\mathbf{a}$ and $\mathbf{b}$, the subspace stays unchanged. Thus, the projection of $\mathbf{c}$ (unchanged third column) onto the subspace is also unchanged, and so is the error vector $\mathbf{C}$.

Problem 4. We have

$$\begin{aligned}
a_k =& \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx dx = \frac{1}{\pi}\int_0^{\pi}\cos kx dx \\
=& \frac{1}{\pi}\left[\frac{1}{k}\sin kx\right]_0^{\pi} = \frac{1}{k\pi}(\sin k\pi - \sin 0) \\
=& 0, \\
b_k =& \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx dx = \frac{1}{\pi}\int_0^{\pi}\sin kx dx \\
=& \frac{1}{\pi}\left[-\frac{1}{k}\cos kx\right]_0^{\pi} = \frac{1}{k\pi}(-\cos k\pi + \cos 0) \\
=& \frac{1}{k\pi}(1-(-1)^k), \\
a_0 =& \frac{1}{2\pi}\int_0^{2\pi} f(x)dx = \frac{1}{2}
\end{aligned}$$

for $k \geq 1$. You can also write b_k as 0 if k is even and $\frac{2}{k\pi}$ if k is odd.