18.06 Exam 2 Solutions

Johnson, Spring 2022

1. To fit the given points $(x_k, y_k, z_k) \in \{(1,2,7),(0,0,2),(-1,0,3),(1,1,4),(2,-1,5)\}$, we have

$$\begin{cases} \alpha x_1 + \beta y_1 + \gamma = z_1, \\ \alpha x_2 + \beta y_2 + \gamma = z_2, \\ \alpha x_3 + \beta y_3 + \gamma = z_3, \\ \alpha x_4 + \beta y_4 + \gamma = z_4, \\ \alpha x_5 + \beta y_5 + \gamma = z_5. \end{cases}$$

Writing the above as a matrix equation, we have

$$\begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \\ x_4 & y_4 & 1 \\ x_5 & y_5 & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \end{pmatrix}.$$

In other words, we have

$$Ax = b$$

where

$$\boxed{A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 1 \\ 2 & -1 & 1 \end{pmatrix}}, \quad \boxed{b = \begin{pmatrix} 7 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}}.$$

But of course, this is overdetermined (more equations than unknowns) and is unlikely to have an exact solution. Instead, the problem requests the least-square solution, corresponding to minimizing $\|b - Ax\|^2$, which yields the normal equations:

$$\boxed{A^T A\hat{x} = A^T b},$$

where $\hat{x} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ are the best-fit parameters. Writing this out explicitly by plugging in the numbers (which was *not* required) yields:

$$\boxed{\begin{pmatrix} 1 & 0 & -1 & 1 & 2 \\ 2 & 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 1 \\ 2 & -1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 & 1 & 2 \\ 2 & 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 7 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}}.$$

■

2. (a) As $b \in C(A) = C(Q)$, we can write b as

$$b = QQ^T b = q_1(q_1^T b) + q_2(q_2^T b) + q_3(q_3^T b) = \boxed{3\sqrt{2}q_1 - 4q_2 + 8q_3},$$

recalling that the coefficients of an orthonormal basis are obtained merely by dot products (i.e. projections qq^T).

(b) Since $N(A^T) = C(A)^\perp$, we can get the orthogonal projection of $y = \begin{pmatrix} 2 \\ -2 \\ 2 \\ -2 \end{pmatrix}$ onto $N(A^T)$ by simply subtracting the projection of y onto the q's. In other words, the orthogonal projection of y onto $N(A^T)$ is

$$\begin{aligned}(I - QQ^T)y &= y - q_1(q_1^T y) - q_2(q_2^T y) - q_3(q_3^T y) \\ &= \begin{pmatrix} 2 \\ -2 \\ 2 \\ -2 \end{pmatrix} - 0 \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} - 2 \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} - 2 \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} = \boxed{\begin{pmatrix} 0 \\ -2 \\ 2 \\ 0 \end{pmatrix}}.\end{aligned}$$

(c) The terms $\boxed{q_2^T a_1, q_3^T a_1, q_3^T a_2}$ must be 0.

In general, for $A = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix}$ with linearly independent columns, the QR factorization obtained using Gram–Schmidt is

$$A = QR,$$

where $Q = \begin{pmatrix} q_1 & q_2 & \dots & q_n \end{pmatrix}$ is a $m \times n$ matrix with orthonormal columns spanning $C(A)$ and $R = \begin{pmatrix} r_{11} & r_{21} & \dots & r_{n1} \\ 0 & r_{22} & \dots & r_{n2} \\ & & \vdots & \\ 0 & 0 & \dots & r_{nn} \end{pmatrix}$ is an $n \times n$ invertible upper-triangular matrix, with $r_{ij} = q_i^T a_j$ for all $i \geq j$.

Another way of seeing the same thing is to recall the Gram–Schmidt process. By construction, q_1 is parallel to a_1, so q_2 and q_3 must be $\perp a_1$. a_2 is in the span of q_1 and q_2, so we must also have $q_3 \perp a_2$. ■

3. For $f(x) = (b - Ax)^T M(b - Ax)$, recall from class that $d(y^T My) = dy^T My + y^T M dy = 2dy^T My$ (using $M = M^T$). For $y = b - Ax$, we have $dy = -Adx$. Combining these equations yields:

$$df = 2dy^T My = 2(-Adx)^T M(b - Ax) = dx^T \underbrace{[-2A^T MA(b - Ax)]}_{\nabla f},$$

since the gradient is defined by $df = \nabla f^T dx = dx^T \nabla f$. Alternatively, going through all of the steps explicitly using the product rule, we have

$$\begin{aligned}
df &= d((b - Ax)^T M(b - Ax)) \\
&= (d(b - Ax)^T)M(b - Ax) + (b - Ax)^T(dM)(b - Ax) + (b - Ax)^T M(d(b - Ax)) \\
&= -(Adx)^T M(b - Ax) + 0 - (b - Ax)^T MAdx \quad \text{(since } dA, db, dM \text{ all vanish)} \\
&= -(M(b - Ax))^T(Adx) - (b - Ax)^T MAdx \quad \text{(since } x^T y = y^T x \text{ for column vectors } x, y\text{)} \\
&= -((b - Ax)^T M^T A + (b - Ax)^T MA)dx \\
&= -2(b - Ax)^T M^T Adx \qquad \text{(since } M^T = M\text{)} \\
&= \underbrace{(-2A^T M(b - Ax))}_{\nabla f}{}^T dx.
\end{aligned}$$

Therefore, when $\nabla f = 0$, we have

$$-2A^T M(b - Ax) = 0 \iff \boxed{A^T MAx = A^T Mb}.$$

■

4. (a) If $A = \begin{pmatrix} a_1 & a_2 \end{pmatrix}$, the projection matrix onto $C(A)$ is given by $\frac{a_1 a_1^T}{a_1^T a_1} + \frac{a_2 a_2^T}{a_2^T a_2}$ only when $\boxed{a_1, a_2 \text{ are orthogonal}}$ ($\neq$ orthonormal).

In general, we have $P = A(A^TA)^{-1}A^T = \begin{pmatrix} a_1 & a_2 \end{pmatrix} \begin{pmatrix} a_1^T a_1 & a_1^T a_2 \\ a_2^T a_1 & a_2^T a_2 \end{pmatrix}^{-1} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, which would have terms involving both a_1 and a_2 if they are not orthogonal.

(b) If S and T are orthogonal subspaces of a vector space V, then

(i) their intersection (vectors in both S and T) is the set $\boxed{\{\vec{0}\}}$.
Note that if $x \in S \cap T$ then $x^T x = 0 \Rightarrow x = 0$.

(ii) (dimension of S) + (dimension of T) must be $\boxed{\leq}$ (dimension of V).

(The sum = dimension V only when S and T are *orthogonal complements*, not merely orthogonal.) For example, $S = \text{span}\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right\}$ and $T = \text{span}\left\{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right\}$ are two orthogonal subspaces of $V = \mathbb{R}^3$, and we have (dimension of S) + (dimension of T) $= 1 + 1 = 2 \leq 3$.

(c) For the vector space $\mathbb{R}^3$, give projection matrices onto:

(i) any 0-dimensional subspace: $\boxed{P = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}$, i.e. the 3×3 zero matrix.

(Note that the only 0-dimensional subspace is $\{\vec{0}\}$.)

(ii) any 1-dimensional subspace: $\boxed{P = \frac{aa^T}{a^Ta}}$ for $S = \text{span}\{a\}$ with some $a \neq \vec{0}$.

A specific example is $\boxed{P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}$ for $S = \text{span}\left\{\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right\}$.

(iii) any 3-dimensional subspace: $\boxed{P = I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}$, i.e. the 3×3 identity matrix. Note that the only subpsace of $\mathbb{R}^3$ with dimension 3 is $\mathbb{R}^3$ itself.

(d) We must have $Q^TQ = I$ for orthonormal columns, but $\boxed{QQ^T \neq I}$ is possible whenever Q is not square (not unitary), in which case QQ^T is the projection matrix onto a lower-dimensional subspace $C(Q)$ of the whole space. In particular, you just need any "tall" Q matrix: orthonormal columns, but fewer columns than rows, such as the Q matrix of problem 2.

The simplest example is a Q matrix with only a *single* orthonormal column, in which QQ^T is projection onto a 1d subspace, such as:

$$\boxed{Q = \begin{pmatrix} 1 \\ 0 \end{pmatrix}}, \quad QQ^T = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq I.$$

(e) A is a 7×5 matrix of rank 4.

(i) Give the size and rank of the following projection matrices:

i. $P_1 =$ projection onto $C(A)$: $\boxed{\text{size} = 7 \times 7,\ \text{rank } = 4}$

ii. $P_2 =$ projection onto $C(A^T)$: $\boxed{\text{size} = 5 \times 5,\ \text{rank } = 4}$

iii. $P_3 =$ projection onto $N(A)$: $\boxed{\text{size} = 5 \times 5,\ \text{rank } = 5 - 4 = 1}$

iv. $P_4 =$ projection onto $N(A^T)$: $\boxed{\text{size} = 7 \times 7,\ \text{rank } = 7 - 4 = 3}$

(ii) Give a sum or product of two of these P matrices that must $= 0$ (a zero matrix): Note that $\boxed{P_1P_4 = 0}$ as $C(A)$ and $N(A^T)$ are orthogonal complements. Similarly, we have $\boxed{P_4P_1 = 0}$, $\boxed{P_2P_3 = 0}$, $\boxed{P_3P_2 = 0}$.

(iii) Give a sum or product of two of these P matrices that must $= I$ (an identity matrix): As $C(A)$ and $N(A^T)$ are orthogonal complements, we have $P_4 = I - P_1$. Therefore, $\boxed{P_1 + P_4 = I}$. Similarly, $\boxed{P_2 + P_3 = I}$. ■