HW4 Solutions

Problem 1.

1. $\operatorname{rank}(A) = \operatorname{rank}(QR) \le \operatorname{rank}(R)$

$\operatorname{rank}(R) = \operatorname{rank}(Q^{-1}QR) \le \operatorname{rank}(QR) = \operatorname{rank}(A)$

Therefore, $\operatorname{rank}(A) = \operatorname{rank}(R)$.

Columns of A are lin. indep

$$R = \begin{bmatrix} r_{11} & r_{12} & \\ & r_{22} & \\ & & \ddots \end{bmatrix}$$

iff

A is of full rank

iff

R is of full rank

iff

$$\det(R) = \prod_{i=1}^{n} r_{ii} \neq 0$$

iff

all diagonal entries of R are nonzero.

Note. There are alternative solutions not using determinants.

2. We prove that the nullspace of A^TA is zero (i.e. the null vector space).

Suppose $x \in \text{null}(A^TA)$. Therefore, $A^TAx = 0$.

Hence $x^TA^TAx = 0$, which implies

$$\|Ax\|^2 = \langle Ax, Ax \rangle = x^TA^TAx = 0$$

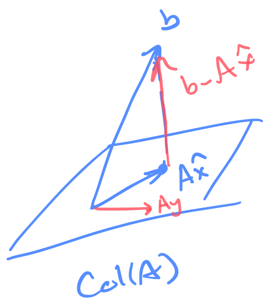
Hence, $Ax = 0$. Since A has lin. indep columns it follows that $x = 0$.

We conclude that $\text{null}(A^TA) = 0$.

Hence A^TA is a square matrix w/ $\text{null}(A^TA) = 0$. So, A^TA is invertible.

3. $$\hat{x} = (A^TA)^{-1} b$$

$$(A^TA)\,\hat{x} = (A^TA)(A^TA)^{-1} b = b$$



$$A^T(A\hat{x} - b) = 0$$

$$A^T(b - A\hat{x}) = 0$$

$y^TA^T(b - A\hat{x}) = 0$ for all vectors y.

$\langle Ay, b - A\hat{x} \rangle = 0$, for all vectors y.

$$\text{Col}(A) \perp b - A\hat{x}$$

4. $A^T A \hat{x} = A^T b$ and $A = \hat{Q}\hat{R}$

where $(\hat{Q}, \hat{R})$ are reduced factorization of A. In particular R is an invertible square matrix by part (i).

$A^T A \hat{x} = A^T b$ can be rewritten as

$$\hat{R}^T \hat{Q}^T \hat{Q} \hat{R} \hat{x} = \hat{R}^T \hat{Q}^T b$$

LHS is equal to $\hat{R}^T \hat{R} x$, since $\hat{Q}$ is orthogonal.

Therefore, $\hat{R}^T \hat{R} \hat{x} = \hat{R}^T \hat{Q}^T b$.

Since $\hat{R}^T$ is lower triangular with non-zero diagonal entries, we have

$$\hat{R} \hat{x} = \hat{Q}^T b.$$

Therefore

$$A\hat{x} = \hat{Q}\hat{R}\hat{x} = \underbrace{\hat{Q}\hat{Q}^T}\, b = P_A b$$

This is the projection matrix to the column space of A.

Note $\hat{Q}\hat{Q}^T$ is not necessarily the identity matrix as the following example shows.

$$\underset{3\times 2}{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}} \underset{2\times 3}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

5. Since $A^T b = A^T A \hat{x}$, we have

$$\|A\hat{x} - b\|^2 =$$

$$(A\hat{x} - b)^T (A\hat{x} - b) =$$

$$(\hat{x}^T A^T - b^T)(A\hat{x} - b) =$$

$$\hat{x}^T A^T A\hat{x} - \hat{x}^T A^T b - b^T A\hat{x} + b^T b =$$

$$\hat{x}^T A^T A\hat{x} - \hat{x}^T A^T A\hat{x} - (A^T b)^T \hat{x} + b^T b =$$

$$b^T b - (A^T b)^T \hat{x} =$$

$$\|b\|^2 - (A^T A\hat{x})^T \hat{x} =$$

$$\|b\|^2 - \hat{x}^T A^T A\hat{x} =$$

$$\|b\|^2 - \|A\hat{x}\|^2 =$$

$$\|b\|^2 - \|QQ^T b\|^2 =$$

$$\|b\|^2 - \|Q^T b\|^2$$

This is because Q is an orthonormal matrix and as such it preserves vector norms, i.e.

$$\|Qy\| = \|y\| \text{ for all vectors } y.$$

6. (Reduced $\hat{Q}\hat{R}$)

$$\begin{array}{cc} a_1 & a_2 \end{array}$$

$$\begin{bmatrix} \boxed{\begin{matrix} 2 \\ -1 \\ 0 \end{matrix}} & \boxed{\begin{matrix} 0 \\ 1 \\ 2 \end{matrix}} \end{bmatrix} = \begin{bmatrix} | & | \\ q_1 & q_2 \\ | & | \end{bmatrix}_{3\times 2} \begin{bmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{bmatrix}_{2\times 2}$$

$$r_{11} q_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \|q_1\| = 1$$

Therefore, $r_{11} = \dfrac{\left\| \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \right\|}{\|q_1\|} = \dfrac{\sqrt{5}}{1} = \sqrt{5}$

$$q_1 = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} \\ 0 \end{bmatrix}$$

$$q_2 = \frac{a_2 - r_{12} q_1}{r_{22}}$$

$$r_{12} = \langle q_1, a_2 \rangle = \frac{-1}{\sqrt{5}}$$

$$q_2 \parallel a_2 - r_{12} q_1 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} - \left(\frac{-1}{\sqrt{5}}\right) \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 + \frac{2}{5} \\ 1 - \frac{1}{5} \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} \\ 4/5 \\ 2 \end{bmatrix}$$

Since $\|q_2\| = 1$ we have

$$q_2 = \frac{\begin{bmatrix} \frac{2}{5} \\ 4/5 \\ 2 \end{bmatrix}}{\left\| \begin{bmatrix} \frac{2}{5} \\ 4/5 \\ 2 \end{bmatrix} \right\|} = \frac{\begin{bmatrix} \frac{2}{5} \\ 4/5 \\ 2 \end{bmatrix}}{\sqrt{\frac{4+16+100}{25}}} = \frac{\sqrt{5}\begin{bmatrix} \frac{2}{5} \\ 4/5 \\ 2 \end{bmatrix}}{\sqrt{120}}$$

$$= \frac{5\begin{bmatrix} 2/5 \\ 4/5 \\ 2 \end{bmatrix}}{2\sqrt{30}} = \frac{5\begin{bmatrix} 1/5 \\ 2/5 \\ 1 \end{bmatrix}}{\sqrt{30}} = \begin{bmatrix} 1/\sqrt{30} \\ 2/\sqrt{30} \\ 5/\sqrt{30} \end{bmatrix}$$

and, $r_{22} = \|a_2 - r_{12}q_1\| = \frac{2}{5}\sqrt{30}$

So,

$$\hat{Q} = \begin{bmatrix} 2/\sqrt{5} & \frac{1}{\sqrt{30}} \\ -1/\sqrt{5} & \frac{2}{\sqrt{30}} \\ 0 & 5/\sqrt{30} \end{bmatrix} \qquad \hat{R} = \begin{bmatrix} \sqrt{5} & \frac{-1}{\sqrt{5}} \\ 0 & \frac{2}{5}\sqrt{30} \end{bmatrix}$$

7. $A\hat{x} = \hat{Q}\hat{Q}^T b$ is closest to the vector b in the sense that $\|Ax-b\|$ is minimized among all vectors x.

✓ See VMLS pp.227 Example

$$\min \|Ax-b\|^2 = \|A\hat{x}-b\|^2$$

$$= \|b\|^2 - \|Q^Tb\|^2 \quad \text{(from part 5)}$$

$$Q^Tb = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ -\frac{4}{\sqrt{30}} \end{bmatrix}$$

$$\|Q^Tb\|^2 = \left(\frac{2}{\sqrt{5}}\right)^2 + \left(-\frac{4}{\sqrt{30}}\right)^2 = \frac{4}{3}$$

So $\min_x \|Ax-b\|^2 = (1^2+(-1)^2) - \frac{4}{3} = \frac{2}{3}$

So $\min_x \|Ax-b\| = \sqrt{\frac{2}{3}}$

Problem 2 (Ridge Regression)

1. The loss function to be minimized is

$$\| Ax - y \|^2 + \lambda \|x\|^2$$ where x, y are n-dim vectors and $\lambda : \mathbb{R}$.

$$\| Ax - y \|^2 + \lambda \|x\|^2 =$$

$$\| y - Ax \|^2 + \lambda \|x\|^2 =$$

$$\langle y - Ax, y - Ax \rangle + \lambda \langle x, x \rangle =$$

$$(y - Ax)^T (y - Ax) + \lambda x^T x =$$

$$y^T y - 2 x^T A y + x^T (A^T A + \lambda I) x \qquad (*)$$

We are looking for $x = (x_1, \dots, x_n)^T$ which minimizes the value $(*)$. To this we set the derivative in the direction of vector x to zero.

Claim: $$\frac{d}{dx} \left(y^T y - 2 x^T A^T y + x^T (A^T A + \lambda I) x \right) =$$

$$-2 A^T y + 2 (A^T A + \lambda I) x$$

It is clear that $$\frac{d(-2 x^T A^T y)}{dx} = -2 A^T y$$

by examining partial derivatives

To see why this is true we first observe that for any matrix M

(Lemma)
$$\frac{d\, x^T M x}{dx} = (M+M^T)\, x$$

proof:
$$\frac{d\, x^T M x}{dx} = \begin{pmatrix} \frac{\partial x^T M x}{\partial x_1} \\ \vdots \\ \frac{\partial x^T M x}{\partial x_n} \end{pmatrix}$$

The i-th row of M is the i-th column of M^T.

Therefore the i-th row of M is denoted by M_i^T.

Note that

$$x^T M x = (x_1\ x_2\ \cdots\ x_n) \begin{pmatrix} \langle M_1^T, x\rangle \\ \vdots \\ \langle M_n^T, x\rangle \end{pmatrix}$$

$$= x_1 \langle M_1^T, x\rangle + \cdots + x_n \langle M_n^T, x\rangle$$

Therefore by the product rule of derivatives

$$\frac{\partial x^T M x}{\partial x_i} = x_1 \frac{\partial \langle M_1^T, x\rangle}{\partial x_i} + \cdots + x_n \frac{\partial \langle M_n^T, x\rangle}{\partial x_i}$$

$$+ \langle M_i^T, x\rangle$$

$$= x_1 m_{1i} + x_2 m_{2i} + \cdots + x_n m_{ni}$$

$$+ x_1 m_{i1} + x_2 m_{i2} + \cdots + x_n m_{in}$$

which is the i-th component of $(M+M^T)x$.

Now we apply this lemma to the symmetric matrix

$A^T A + \lambda I$.

$$\frac{d}{dx}\left(x^T (A^T A + \lambda I) x \right) =$$

$$2(A^T A + \lambda I)\, x$$

Setting

$$\frac{d}{dx}\left(y^Ty - 2x^TA^Ty + x^T(A^TA+\lambda I)x\right) =$$

$$-2A^Ty + 2(A^TA+\lambda I)x = 0$$

implies

$$A^Ty = (A^TA+\lambda I)x.$$

Therefore, $\boxed{x = (A^TA+\lambda I)^{-1}A^Ty}$

2. When A has linearly indep. columns

A^TA is invertible and therefore,

$$A^+ = \lim_{\lambda\to 0}(A^TA+\lambda I)^{-1}A^T = (A^TA)^{-1}A^T$$

$$\lim_{\lambda\to 0} x^{(\lambda)}_{ridge} = \lim_{\lambda\to 0}(A^TA+\lambda I)^{-1}A^Tb$$

$$= (A^TA)^{-1}A^Tb = \hat{x}$$