PSET 6. SOLUTIONS

Problem 1. Since $A = QR$ is upper-triangular and $R$ is upper-triangular we get that the result of orthogonalization $Q = AR^{-1}$ is upper-triangular as well (since the inverse of an upper-triangular matrix is upper-triangular and the product of two upper-triangular matrices is again upper-triangular). $Q$ is orthonormal, so $Q^TQ = I$. Rewriting this as $Q^T = Q^{-1}$ we get that $Q^T$ is also upper-triangular. It follows that $Q$ is diagonal:

$$Q = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n \end{bmatrix}$$

But we know $Q^TQ = I$, so $x_i^2 = 1$ for each $i$. It means that $Q$ is diagonal with the diagonal entries being $\pm 1$.

Problem 2. $A = QR$, $A$ is Hessenberg, $R^{-1}$ is upper-triangular since $R$ is and $Q = AR^{-1}$ is a product of a Hessenberg matrix and upper-triangular matrix which is again Hessenberg. So $Q$ is Hessenberg.

Problem 3. No, it does not. Say you added $x$ to the top-right corner. The lowest row has all elements being 0 except (possibly) the last (rightmost) one. If it is zero as well, then we have a row full of zeros and the determinant is 0 both before and after adding $x$. If it is not zero and is equal to some $a_{nn}$, then we can take the lowest row, multiply it by $-x \cdot a_{nn}^{-1}$ and add to the first row. This will undo the addition of $x$ and so the determinants are equal.

Problem 4. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn-1} & a_{nn} \end{bmatrix}$$

be our Hessenberg matrix and let’s consider also a matrix $A’$ where we replace the first row with $[0 \ 0 \ \ldots \ 0 \ x]$:

$$A’ = \begin{bmatrix} 0 & 0 & \cdots & x \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn-1} & a_{nn} \end{bmatrix}$$

In the matrix where we add $x$ to top right corner of $A$ is all rows except the top one stay the same as in $A$ and the top one is the sum of top rows of $A$ and $A’$. From the properties of the determinant it follows that the determinant of this matrix is $\det A + \det A’$. In other words the determinant changes exactly on $\det A’$, so we need to compute it.

For this we swap the first row with second one, second with third and so on, until the top row becomes the bottom row; each swap changes the sign of the determinant and together we made $n - 1$ swaps. So we get

$$\det(A’) = (-1)^{n-1} \cdot \det \begin{bmatrix} a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn-1} & a_{nn} \\ 0 & 0 & \cdots & 0 & x \end{bmatrix}$$

and the last determinant is easy to compute: it is just $a_{21}a_{32} \ldots a_{nn-1}x$. Answer: If we change the top right entry by $x$ the determinant will change on $x$ times the product of terms on the diagonal below the main one.
Problem 5. We would like to find a vector $w$ such that the sum $\sum_i \|w^T x_i + b - y_i\|^2$ is minimal. This can be reformulated using linear algebra: we need to look at

$$X \cdot \begin{bmatrix} b \\ w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Indeed the rows of the product are exactly of the form $w^T x_i + b$ and so we are trying to find the least-squares approximation $X \cdot z$ to $y$ (where $y$ is the vector of $y_i$'s).

Problem 6. This is not always invertible. Just take $A$ to be a row vector with length $n > 1$ (say $A = [1 \ 1]$), then the rows are independent (since there is only one) and $A^T A$ is of size $n \times n$ and is a product of a column vector on a row vector, so has rank 1 ($A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$: not invertible).

Problem 7. a) $Q = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, then $QQ^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \neq I$

b) Take $v_1$ to be any vector and $v_2 = 0$. Then $v_1$ is orthogonal to $v_2$: $v_1^T v_2 = 0$. But they are linearly independent.

c) Take $q_1 = \begin{bmatrix} \sqrt{3} \\ \sqrt{3} \end{bmatrix}$, $q_2 = \begin{bmatrix} 2/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}$, $q_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, then $q_1^T q_1 = (\frac{1}{\sqrt{3}})^2 + (\frac{1}{\sqrt{3}})^2 + (\frac{1}{\sqrt{3}})^2 = 1$, $q_1^T q_2 = \frac{1}{\sqrt{3}} \cdot \frac{2}{\sqrt{6}} - \frac{1}{\sqrt{3}} \cdot \frac{2}{\sqrt{6}} - \frac{1}{\sqrt{3}} \cdot \frac{2}{\sqrt{6}} = 0$, $q_1^T q_3 = \frac{1}{\sqrt{3}} \cdot 1 \cdot \frac{2}{\sqrt{6}} = 0$, $q_2^T q_2 = (\frac{2}{\sqrt{6}})^2 + (\frac{4}{\sqrt{6}})^2 + (\frac{4}{\sqrt{6}})^2 = \frac{6}{3} = 1$, $q_2^T q_3 = \frac{2}{\sqrt{6}} \cdot \frac{2}{\sqrt{6}} + \frac{2}{\sqrt{6}} \cdot \frac{2}{\sqrt{6}} = 1$.

Problem 8. We proceed step by step:

$$q_1 = \frac{a}{|a|} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$u_2 = b - (q_1^T b) q_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix}$$

$$u_2 = \frac{u_2}{|u_2|} = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

next step $u_3 = c - (q_2^T c) q_1 - (q_2^T c) q_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - 0 \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix} + \frac{2}{\sqrt{6}} \begin{bmatrix} 1 \\ \frac{1}{\sqrt{6}} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$

$$q_3 = \frac{u_3}{|u_3|} = \begin{bmatrix} \frac{2}{\sqrt{2}} \\ 0 \\ \sqrt{2} \\ 0 \end{bmatrix}$$

Since the span of $a$, $b$ and $c$ is equal to the span of $q_1$, $q_2$ and $q_3$ it is enough to find a vector perpendicular to $a$, $b$ and $c$. Since the sum of coordinates for any of these vectors is 0 we can take $d = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ (dot product with $d$ is exactly the sum of coordinates).

Problem 9. It is not true that $Q = U$ for example because the first column vector of $A$ (which up to scalar is the first column of $Q$) is not necessarily the first left-singular vector (which should be a vector
whose length increases the most). As a particular example let’s take $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, then $Q = I$ and

$$U = \begin{bmatrix} 0.7370 & 0.5910 & 0.3280 \\ 0.5910 & -0.3280 & -0.7370 \\ 0.3280 & -0.7370 & 0.5910 \end{bmatrix}.$$  

Considering the second part, since $Q$ and $U$ are orthogonal, the square $n \times n$-matrices $UU^T$ and $QQ^T$ are projections on the column space of $U$ and $Q$ correspondingly. So if column vectors of $A$ are linearly independent, all singular values are non-zero, the column spaces of $U$ and $Q$ coincide. It follows that $UU^T$ and $QQ^T$ are projections on the same subspace and so are equal.

**Problem 10.** The $i$-th column of matrix $M$ has coordinates $(-1)^{i-1}, (-1+h)^{i-1}, (-1+2h)^{i-1}, \ldots, (1-h)^{i-1}, 1^{i-1}$. When we take the dot product of $i$-th column with $j$-th one we get the sum $(-1)^{i-1} \cdot (-1)^{j-1} + (-1+h)^{i-1} \cdot (-1+h)^{j-1} + (-1+2h)^{i-1} \cdot (-1+2h)^{j-1} + \ldots + (1)^{i-1} \cdot (1)^{j-1}$ which is almost the integral $\int_{-1}^{1} x^{i-1}x^{j-1} dx$ except that to approximate the integral (by the sum of areas of rectangles with horizontal side $h$ sitting under the graph) we need to multiply the sum above by $h$. This means that the dot product of columns of the matrix $M$ is (almost) equal to the scalar products of polynomials of the forms $x^i$ which given by $(p(x), q(x)) = \frac{1}{h} \int_{-1}^{1} p(x)q(x) dx$. Now we orthogonalize $M$ obtaining $Q$, take $q_n$ and express it as a linear combination of $m_1, m_2, \ldots, m_n$: $q_n = a_1 m_1 + a_2 m_2 + \ldots + a_n m_n$. Returning to the comparison of the integral and scalar product, the polynomials $L_n(x) = a_1 + a_2 x + \ldots + a_n x^{n-1}$ will be (almost) orthogonal to each other and we will also have $\int_{-1}^{1} L_k(x) L_k(x) dx = h$. This means $L_k(x)$ are very close to be the Legendre polynomials except that the length $\int_{-1}^{1} L_k(x) L_k(x) dx$ is not 1, but $h$. This is easily improved by dividing all vectors by $\sqrt{h}$ and this is why it is done in the program code.  

**Problem 11.** It is enough to show that the null-space of $X^T X + \lambda I$ is 0. Let’s suppose there is a non-zero vector $v$ such that $(X^T X + \lambda I)v = 0$. Then $X^T X v = -\lambda v$. Let’s multiply both sides on $v^T$ on the left, then we will get $v^T X^T X v = -\lambda v^T v$. The left side is $(Xv)^T (Xv)$ and so is nonnegative (it can be 0 if $Xv = 0$), while $v^T v = ||v||^2$ is strictly positive and so (using $\lambda > 0$) $-\lambda v^T v$ is strictly negative and we get a contradiction. So $X^T X + \lambda I$ is invertible.