PSET 6. SOLUTIONS

Problem 1. Since $\mathbf{A} = \mathbf{QR}$ is upper-triangular and $\mathbf{R}$ is upper-triangular we get that the result of orthogonalization $\mathbf{Q} = \mathbf{AR}^{-1}$ 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). $\mathbf{Q}$ is orthonormal, so $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$. Rewriting this as $\mathbf{Q}^T = \mathbf{Q}^{-1}$ we get that $\mathbf{Q}^T$ is also upper-triangular. It follows that $\mathbf{Q}$ is diagonal:

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

But we know $\mathbf{Q}^T\mathbf{Q} = I$, so $x_i^2 = 1$ for each i. It means that $\mathbf{Q}$ is diagonal with the diagonal entries being ± 1.

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

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

be our Hessenberg matrix and let's consider also a matrix $\mathbf{A}'$ where we replace the first row with $\begin{bmatrix} 0 & 0 & \dots & 0 & x \end{bmatrix}$:

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

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

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

and the last determinant is easy to compute: it is just $a_{21}a_{32}\dots a_{nn-1}x$. **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 $\mathbf{w}$ such that the sum $\sum_i \| \mathbf{w}^T \mathbf{x}_i + b - y_i \|^2$ is minimal. This can be reformulated using linear algebra: we need to look at

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

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

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

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

b) Take $\mathbf{v}_1$ to be any vector and $\mathbf{v}_2 = 0$. Then $\mathbf{v}_1$ is orthogonal to $\mathbf{v}_2$: $\mathbf{v}_1^T \mathbf{v}_2 = 0$. But they are linearly independent

c) Take $\mathbf{q}_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$, $\mathbf{q}_2 = \begin{bmatrix} \frac{2}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \\ \frac{-1}{\sqrt{6}} \end{bmatrix}$, $\mathbf{q}_3 = \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$, then $\mathbf{q}_1^T \mathbf{q}_1 = (\frac{1}{\sqrt{3}})^2 + (\frac{1}{\sqrt{3}})^2 + (\frac{1}{\sqrt{3}})^2 = 1$, $\mathbf{q}_1^T \mathbf{q}_2 = \frac{1}{\sqrt{3}} \frac{2}{\sqrt{6}} - \frac{1}{\sqrt{3}} \frac{1}{\sqrt{6}} - \frac{1}{\sqrt{3}} \frac{1}{\sqrt{6}} = 0$, $\mathbf{q}_1^T \mathbf{q}_3 = \frac{1}{\sqrt{3}} \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{3}} \frac{1}{\sqrt{2}} = 0$, $\mathbf{q}_2^T \mathbf{q}_2 = (\frac{2}{\sqrt{6}})^2 + (\frac{-1}{\sqrt{6}})^2 + (\frac{-1}{\sqrt{6}})^2 = \frac{6}{6} = 1$, $\mathbf{q}_2^T \mathbf{q}_3 = \frac{-1}{\sqrt{6}} \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{6}} \frac{1}{\sqrt{2}}$, $\mathbf{q}_3^T \mathbf{q}_3 = (\frac{1}{\sqrt{2}})^2 + (\frac{1}{\sqrt{2}})^2 = 1$

Problem 8. We proceed step by step:

$$\mathbf{q}_1 = \frac{\mathbf{a}}{||\mathbf{a}||} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix}, \text{ then } \mathbf{u}_2 = \mathbf{b} - (\mathbf{q}_1^T \mathbf{b}) \mathbf{q}_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} + \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -1 \\ 0 \end{bmatrix}$$

$$\mathbf{q}_2 = \frac{\mathbf{u}_2}{||\mathbf{u}_2||} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} \\ 0 \end{bmatrix}, \text{ next step } \mathbf{u}_3 = \mathbf{c} - (\mathbf{q}_1^T \mathbf{c}) \mathbf{q}_1 - (\mathbf{q}_2^T \mathbf{c}) \mathbf{q}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} - 0 \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix} + \frac{2}{\sqrt{6}} \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \\ -1 \end{bmatrix}$$

$$\mathbf{q}_3 = \frac{\mathbf{u}_3}{||\mathbf{u}_3||} = \begin{bmatrix} \frac{1}{2\sqrt{3}} \\ \frac{1}{2\sqrt{3}} \\ \frac{1}{2\sqrt{3}} \\ \frac{-3}{2\sqrt{3}} \end{bmatrix}$$

Since the span of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is equal to the span of $\mathbf{q}_1$, $\mathbf{q}_2$ and $\mathbf{q}_3$ it is enough to find a vector perpendicular to $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$. Since the sum of coordinates for any of these vectors is 0 we can take $\mathbf{d} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ (dot product with $\mathbf{d}$ is exactly the sum of coordinates).

Problem 9. It is not true that $\mathbf{Q} = \mathbf{U}$ for example because the first column vector of $\mathbf{A}$ (which up to scalar is the first column of $\mathbf{Q}$) 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 $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, then $\mathbf{Q} = \mathbf{I}$ and $\mathbf{U} = \begin{bmatrix} 0.7370 & 0.5910 & 0.3280 \\ 0.5910 & -0.3280 & -0.7370 \\ 0.3280 & -0.7370 & 0.5910 \end{bmatrix}$.

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

Problem 10. The i-th column of matrix M has coordinates $(-1)^{i-1}, (-1+h)^{i-1}, , (-1+2h)^{i-1}, \dots, (1-h)^{i-1}, 1^{i-1}$. 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 orthogonolize $\mathbf{M}$ obtaining $\mathbf{Q}$, take $\mathbf{q}_n$ and express it as a linear combination of $\mathbf{m}_1, \dots, \mathbf{m}_n$: $\mathbf{q}_n = a_1\mathbf{m}_1 + a_2\mathbf{m}_2 + \ldots + a_n\mathbf{m}_n$. Returning to the comparison of the integral and scalar product, the polynomials $L_n(x) = a_1 + a_2x + \ldots + a_nx^{n-1}$ will be (almost) orthogonal to each other and we will also have $\int_{-1}^{1} L_k(x)L_k(x)dx \approx h$. 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 $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}$ is 0. Let's suppose there is a non-zero vector $\mathbf{v}$ such that $(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})\mathbf{v} = 0$. Then $\mathbf{X}^T\mathbf{X}\mathbf{v} = -\lambda\mathbf{v}$. Let's multiply both sides on $\mathbf{v}^T$ on the left, then we will get $\mathbf{v}^T\mathbf{X}^T\mathbf{X}\mathbf{v} = -\lambda\mathbf{v}^T\mathbf{v}$. The left side is $(\mathbf{X}\mathbf{v})^T(\mathbf{X}\mathbf{v})$ and so is nonnegative (it can be 0 if $\mathbf{X}\mathbf{v} = 0$), while $\mathbf{v}^T\mathbf{v} = ||v||^2$ is strictly positive and so (using $\lambda > 0$) $-\lambda\mathbf{v}^T\mathbf{v}$ is strictly negative and we get a contradiction. So $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}$ is invertible.