Problem Set 5 Solutions

1. (34 points) Consider the following linear system $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} -1 & 2 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -10 \\ 15 \\ 5 \end{bmatrix} \tag{1}$$

   (a) By performing the Gaussian Elimination, show that the system does not have a solution.

$$\begin{bmatrix} -1 & 2 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{pmatrix} \\ +R_1 \\ +2R_1 \end{pmatrix} \longrightarrow \begin{bmatrix} -1 & 2 \\ 0 & 3 \\ 0 & 5 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 2 \\ 0 & 3 \\ 0 & 5 \end{bmatrix} \begin{pmatrix} \\ \\ -5/3R_2 \end{pmatrix} \longrightarrow \begin{bmatrix} -1 & 2 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} = U$$

Then $L = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ -2 & \frac{5}{3} \end{bmatrix}$.

Then we would need to solve $L\mathbf{y} = \mathbf{b}$, or

$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \\ -2 & \frac{5}{3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -10 \\ 15 \\ 5 \end{bmatrix}$$

From this, we get $x_1 = -10$ and then $10 + x_2 = 15$, which gives us $x_2 = 5$.

Plugging in these values, the last equation gives us $20 + \frac{5}{3}(5) = 5$, which is not true. Therefore, there is no solution to this system.

   (b) Construct the normal system and find the least-squares solution $\mathbf{x}_{\text{lsq}}$.

$$A^\intercal A = \begin{bmatrix} -1 & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix}$$

Then $A^\intercal \mathbf{b} = \begin{bmatrix} -1 & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} -10 \\ 15 \\ 5 \end{bmatrix} = \begin{bmatrix} 35 \\ 0 \end{bmatrix}$

Therefore, the normal equations are $\begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 35 \\ 0 \end{bmatrix}$

By subtracting $\frac{1}{6}R_1$ from R_2, we get $L = \begin{bmatrix} 1 & 0 \\ \frac{1}{6} & 1 \end{bmatrix}$ and $U = \begin{bmatrix} 6 & 1 \\ 0 & \frac{35}{6} \end{bmatrix}$

Then $L\mathbf{y} = \mathbf{b}$ gives us $y_1 = 35$ and $\frac{35}{6} + y_2 = 0 \implies y_2 = -\frac{35}{6}$.

Then $U\mathbf{x} = \begin{bmatrix} 35 \\ -\frac{35}{6} \end{bmatrix}$ gives us $\frac{35}{6}x_2 = -\frac{35}{6} \implies x_2 = -1$ and $6x_1 - 1 = 35 \implies x_1 = 6$.

Therefore, the least squares solution is $\mathbf{x}_{lsq} = \begin{bmatrix} 6 \\ -1 \end{bmatrix}$.

(c) Compute the residual vector $\mathbf{r} = \mathbf{b} - A\mathbf{x}_{\text{lsq}}$.

$$\mathbf{r} = \begin{bmatrix} -10 \\ 15 \\ 5 \end{bmatrix} - \begin{bmatrix} -1 & 2 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ -1 \end{bmatrix} = \begin{bmatrix} -10 \\ 15 \\ 5 \end{bmatrix} - \begin{bmatrix} -8 \\ 5 \\ 11 \end{bmatrix} = \begin{bmatrix} -2 \\ 10 \\ -6 \end{bmatrix}$$

(d) Calculate the relative size of the residual,

$$\epsilon_{\text{rel}} = \frac{\|\mathbf{b} - A\mathbf{x}_{\text{lsq}}\|_2}{\|\mathbf{b}\|_2} \tag{2}$$

$$\epsilon_{\text{rel}} = \frac{\sqrt{4 + 100 + 35}}{\sqrt{100 + 225 + 25}} \approx 0.63$$

(e) Show that the residual is orthogonal to each column vector of $A = [\mathbf{a}_1, \mathbf{a}_2]$.

$$\mathbf{r} \cdot \mathbf{a}_1 = \begin{bmatrix} -2 \\ 10 \\ -6 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} = 2 + 10 - 12 = 0$$

$$\mathbf{r} \cdot \mathbf{a}_2 = \begin{bmatrix} -2 \\ 10 \\ -6 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = -4 + 10 - 6 = 0$$

Therefore, $\mathbf{r}$ is orthogonal to both column vectors of A.

(f) Compute $\|A\mathbf{x}_{\text{lsq}}\|_2^2$ and $\|\mathbf{b}\|_2^2$ and show

$$\|\mathbf{b} - A\mathbf{x}_{\text{lsq}}\|_2^2 + \|A\mathbf{x}_{\text{lsq}}\|_2^2 = \|\mathbf{b}\|_2^2. \tag{3}$$

$||A\mathbf{x}_{lsq}||_2^2 = 64 + 25 + 121 = 210$

$||\mathbf{b}||_2^2 = 100 + 225 + 25 = 350$

Then $||\mathbf{r}||_2^2 + ||A\mathbf{x}_{lsq}||_2^2 = 140 + 210 = 350 = ||\mathbf{b}||_2^2$

(g) With a picture, explain the geometric meaning of the relation (3). Hint: The relation has the same form as the Pythagorean theorem.

Equation (3) corresponds to the fact that the residual is perpendicular to the range of A. Since $A\mathbf{x}_{lsq}$ is in the range of A (and is the projection of b onto the Range of A), the residual is perpendicular to it, recovering the usual Pythagorean Theorem since the three vectors form a right triangle.

2. (13 points) Consider the following vectors:

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}. \tag{4}$$

(a) Show that $\{\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3\}$ forms an orthogonal basis for $\mathbb{R}^3$.

First, we show that they are orthogonal:

$\mathbf{v}_1 \cdot \mathbf{v}_2 = 2 - 4 + 2 = 0$, so $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal.

$\mathbf{v}_1 \cdot \mathbf{v}_3 = -4 + 2 + 2 = 0$, so $\mathbf{v}_1$ and $\mathbf{v}_3$ are orthogonal.

$\mathbf{v}_2 \cdot \mathbf{v}_3 = -2 - 2 + 4 = 0$, so $\mathbf{v}_2$ and $\mathbf{v}_3$ are orthogonal.

Next, we show that they are linearly independent. We can do this by looking at a matrix $A = [\mathbf{v}_1 \mathbf{v}_2 \mathbf{v}_3]$ and then by either doing row operations and checking that we get an upper triangular matrix with no zero rows or we can find the determinant of A and check that it is nonzero. Either of these will confirm that the rank of A is 3 and since the vectors are in $\mathbb{R}^3$, they would then span all of $\mathbb{R}^3$.

$$\begin{vmatrix} 2 & 1 & -2 \\ 2 & -2 & 1 \\ 1 & 2 & 2 \end{vmatrix} = 2\begin{vmatrix} -2 & 1 \\ 2 & 2 \end{vmatrix} - \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} - 2\begin{vmatrix} 2 & -2 \\ 1 & 2 \end{vmatrix} = 2(-4-2) - (4-1) - 2(4+2) =$$
$$-12 - 3 - 12 = -27 \neq 0$$

Therefore, the vectors span $\mathbb{R}^3$, and they are orthogonal, so they form an orthogonal basis for $\mathbb{R}^3$.

(b) By normalizing $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, construct an orthonormal basis $\{\mathbf{q}_1,\ \mathbf{q}_2,\ \mathbf{q}_3\}$ for $\mathbb{R}^3$.

$||\mathbf{v}_1||_2 = ||\mathbf{v}_2||_2 = ||\mathbf{v}_3||_2 = \sqrt{4+4+1} = 3$

Therefore, $\mathbf{q}_i = \dfrac{\mathbf{v}_i}{||\mathbf{v}_i||_2}$

Then, $\mathbf{q}_1 = \begin{bmatrix} 2/3 \\ 2/3 \\ 1/3 \end{bmatrix}, \quad \mathbf{q}_2 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{q}_3 = \begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix}.$

3. (22 points) Consider the following basis for $\mathbb{R}^2$:

$$\mathbf{a}_1 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \tag{5}$$

(a) Use the Gram Schmidt orthogonalization process to find the set $\{\mathbf{q}_1, \mathbf{q}_2\}$ that forms an orthonormal basis for $\mathbb{R}^2$:

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{||\mathbf{a}_1||_2} = \frac{\mathbf{a}_1}{5} = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix}$$

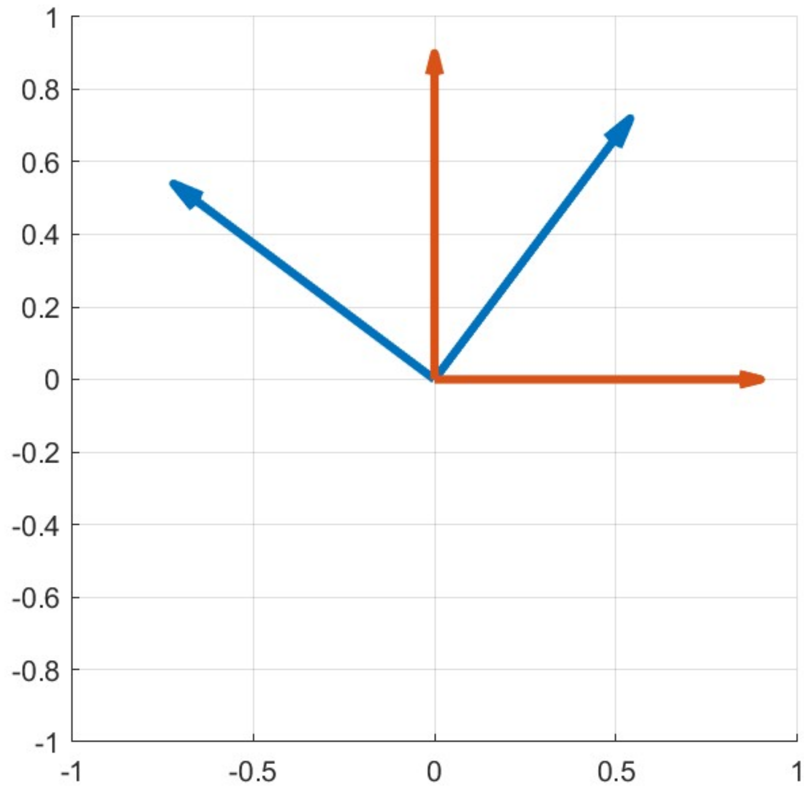
$$\mathbf{v}_2 = \mathbf{a}_2 - (\mathbf{q}_1^\intercal \mathbf{a}_2)\mathbf{q}_1 = \mathbf{a}_2 + \frac{2}{5}\mathbf{q}_1 = \begin{bmatrix} -2 + \frac{6}{25} \\ 1 + \frac{8}{25} \end{bmatrix} = \begin{bmatrix} -44/25 \\ 33/25 \end{bmatrix}$$

Then $\mathbf{q}_2 = \frac{\mathbf{v}_2}{||\mathbf{v}_2||_2} = \frac{\mathbf{a}_1}{\sqrt{44^2+33^2/25}} = \frac{\mathbf{a}_1}{11/5} = \begin{bmatrix} -4/5 \\ 3/5 \end{bmatrix}$

Therefore, the orthonormal basis is $\mathbf{q}_1 = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix}, \quad \mathbf{q}_2 = \begin{bmatrix} -4/5 \\ 3/5 \end{bmatrix}$

(b) By plotting $\mathbf{q}_1$ and $\mathbf{q}_2$ in the $\mathbb{R}^2$ plane, explain how geometrically that $\{\mathbf{q}_1, \ \mathbf{q}_2\}$ forms an orthonormal basis for $\mathbb{R}^2$.

As seen in the figure below, the vectors (shown in blue) are perpendicular. Just as any vector in $\mathbb{R}^2$ can be written with x and y components, by simply rotating the usual axis vectors (shown in red), we could also write any vector with components according to these two basis vectors instead.



(c) For $\mathbf{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, compute its coordinates with respect to the basis $\{\mathbf{v}_1, \ \mathbf{v}_2\}$. In other words, compute $\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$ satisfying $\mathbf{w} = \alpha_1 \mathbf{q}_1 + \alpha_2 \mathbf{q}_2$.

We then want to solve the equations: $1 = \frac{3}{5}\alpha_1 - \frac{4}{5}\alpha_2, \qquad 2 = \frac{4}{5}\alpha_1 + \frac{3}{5}\alpha_2$. Solving this however you like, you should get $\alpha_1 = \frac{11}{5}$ and $\alpha_2 = \frac{2}{5}$.

4. (10 points) Let $\{\mathbf{q}_1,\ \mathbf{q}_2\}$ be an orthonormal basis for $\mathbb{R}^2$. Consider $\mathbf{w} = \alpha_1\mathbf{q}_1 + \alpha_2\mathbf{q}_2$.

   (a) Express $\mathbf{w}$ in terms of the matrix $Q = [\mathbf{q}_1\ \mathbf{q}_2]$ and vector $\alpha = \begin{bmatrix}\alpha_1 \\ \alpha_2\end{bmatrix}$.

   $w_1 = \alpha_1 Q_{11} + \alpha_2 Q_{12}$ and $w_2 = \alpha_1 Q_{21} + \alpha_2 Q_{22}$

   In other words, $w_i = \sum_{j=1}^{2} Q_{ij}\alpha_j$. We can recognize this is a matrix vector product.

   Therefore, $\mathbf{w} = Q\alpha$.

   (b) By using the result of (a) and the fact that $Q^T Q = QQ^T = I$, express α in terms of Q^T and $\mathbf{w}$.

   $\mathbf{w} = Q\alpha \implies Q^\intercal \mathbf{w} = \boldsymbol{Q}^\intercal Q\alpha = \alpha$

   Therefore, $\alpha = Q^\intercal \mathbf{w}$.

5. (10 points) A square matrix Q is orthogonal if $Q^\intercal = Q^{-1}$. Show that the condition number (in the 2-norm) of an orthogonal matrix is 1. Note: This helps to explain why, unlike in using the normal equations, the conditioning of the linear problem is not worsened by using QR Factorization.

   $\kappa(Q) = ||Q||_2||Q^{-1}||_2 = ||Q||_2||Q^\intercal||_2$.

   Then, $||Q||_2 = \sqrt{\max \lambda_i}$, where λ_i are the eigenvalues of $Q^\intercal Q$. Since Q is orthogonal, $Q^\intercal Q = I$, which has eigenvalues of 1. Therefore, $||Q||_2 = 1$.

   The same logic can be used to find that $||Q^\intercal||_2 = 1$. Therefore, $\kappa(Q) = 1$.

6. (10 points) Explain what may happen during the course of using the Gram-Schmidt process to solve the least squares problem if the matrix A is rank deficient (i.e. if the columns are not linearly independent). What kind of fix could you put in your code to remedy this?

   If there is a column that can be written as a linear combination of the other columns, then the Gram-Schmidt orthogonalization process will produce a column vector $\mathbf{v}_i$ that is all zeros (since after subtracting off its projections onto the other basis vectors, there will be nothing left). Furthermore, when the algorithm attempts to normalize it to find $\mathbf{q}_i$, there will be division by 0. In terms of $\widehat{Q}$ and $\widehat{R}$, this will result in a zero/infinite column in $\widehat{Q}$ and a zero diagonal element in $\widehat{R}$, further causing a failure when using backward substitution.

Since the Range of A is given by the range of its column vectors, which is unchanged by the inclusion or exclusion of an "extra" vector that is not linearly independent, we can fix this issue in our code by reducing the size of $\widehat{Q}$ and $\widehat{R}$ to exclude those that result in a zero vector.

However, there may be a bit of an issue with "losing information." For example, if it is a polynomial fit, it may mean that one of the x values has been used twice. Therefore, if we just delete one of the rows or columns we would lose some information and/or change the size of the polynomial we are finding. Therefore, in a real problem, you would probably want to investigate *why* the columns are not linearly independent and then make decisions about what to do, such as taking the average of the y values if there were two with the same x or throw away one of the points.

7. (12 points) In class, we claimed that $P = \widehat{Q}\widehat{Q}^\intercal$ is an orthogonal projector, where $\widehat{Q}$ is defined as the matrix whose columns, $\{\mathbf{q}_1, ...\mathbf{q}_n\}$ ($\mathbf{q}_i \in \mathbb{R}^m$) give an orthonormal basis for the range of matrix A. Prove this fact about P.

In order to prove this, we must prove that (a) $P^2 = P$ and that (b)$P^\intercal = P$.

(a) $P^2 = \widehat{Q}\widehat{Q}^\intercal\widehat{Q}\widehat{Q}^\intercal$

Let's investigate the two interior matrices:

$\widehat{Q}^\intercal\widehat{Q} = \begin{bmatrix} \mathbf{q}_1^\intercal \\ \mathbf{q}_2^\intercal \\ \vdots \\ \mathbf{q}_n^\intercal \end{bmatrix} \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & ... & \mathbf{q}_n \end{bmatrix}$, where $\mathbf{q}_i^\intercal$ are row vectors and $\mathbf{q}_i$ are column vectors.

$$= \begin{bmatrix} \mathbf{q}_1^\intercal\mathbf{q}_1 & \mathbf{q}_1^\intercal\mathbf{q}_2 & ...\mathbf{q}_1^\intercal\mathbf{q}_n \\ \mathbf{q}_2^\intercal\mathbf{q}_1 & \mathbf{q}_2^\intercal\mathbf{q}_2 & ...\mathbf{q}_2^\intercal\mathbf{q}_n \\ \vdots & \ddots & \vdots \\ \mathbf{q}_n^\intercal\mathbf{q}_1 & \mathbf{q}_n^\intercal\mathbf{q}_2 & ...\mathbf{q}_n^\intercal\mathbf{q}_n \end{bmatrix}.$$

Then, since $\{\mathbf{q}_1, ...\mathbf{q}_n\}$ form an orthonormal basis, we have that $\mathbf{q}_i^\intercal\mathbf{q}_j$ is 1 if $i = j$ and 0 if $i \neq j$. Therefore, we get

$\widehat{Q}^\intercal\widehat{Q} = I.$

Then $P^2 = \widehat{Q}\widehat{Q}^\intercal\widehat{Q}\widehat{Q}^\intercal = \widehat{Q}I\widehat{Q}^\intercal = \widehat{Q}\widehat{Q}^\intercal = P.$

(b) $P^\intercal = (\widehat{Q}\widehat{Q}^\intercal)^\intercal = (Q^\intercal)^\intercal(\widehat{Q})^\intercal = \widehat{Q}\widehat{Q}^\intercal = P$

8. (12 points) In class, we defined a reflection defined by $F = I - 2P_v$, where P_v is the orthogonal projection onto vector $\mathbf{v}$. Prove that F is an orthogonal matrix.

For F to be an orthogonal matrix, we must prove that $F^\intercal F = I$ and $FF^\intercal = I$.

$F^\intercal F = (I - 2P_v)^\intercal(I - 2P_v)$

$= (I - 2P_v^\intercal)(I - 2P_v)$

$= (I - 2P_v)(I - 2P_v)$ (where we use that since P is an orthogonal projection, $P^\intercal = P$)

$= I - 4P_v + 4P_v^2$

$= I - 4P_v + 4P_v$ (where we use that since P is an orthogonal projection, $P^2 = P$)

$= I$

Similarly, $FF^\intercal = (I - 2P_v)(I - 2P_v)^\intercal$

$= (I - 2P_v)(I - 2P_v^\intercal)$

$= (I - 2P_v)(I - 2P_v)$

$= I - 4P_v + 4P_v^2$

$= I - 4P_v + 4P_v$

$= I$

Therefore, since $F^\intercal F = FF^\intercal = I$, $F^{-1} = F^\intercal$, F is an orthogonal matrix.

9. (a) (25 points) Write functions to perform the reduced QR factorization using the classical Gram-Schmidt algorithm and the modified Gram-Schmidt algorithm. Your code should take as input $m \times n$ matrix A with linearly independent column vectors (Note: this means you do NOT need to include your fix from Number (6)) and return matrices $\widehat{Q}$ and $\widehat{R}$. Also write functions to use these factorizations to solve the least squares problem. These functions should take as input matrices A and vector $\mathbf{b}$ and return the least squares solution to the system $A\mathbf{x} = \mathbf{b}$. Validate your codes using a random matrix of

size 50×10, and explain how you validated them. Note: Matlab's backslash performs least squares if $m > n$ for input matrix A.

(20 Bonus points) for making a version of the Classical Gram-Schmidt factorization function that includes your fix from Number (6), validating it, and explaining how you validated it. (Note: If you are checking if something is equal to 0, since there is numerical error, you may instead want to check if it is below some tolerance.)

See ClassicalGS.m, ModifiedGS.m, LeastSquares_Classical GS.m, and LeastSquares_ModifiedGS.m, and LeastSquares_ClassicalGS2.m (for bonus part) for the relevant functions. See PS5N9A.m for the test scripts. For the QR Factorizations, I validated my code by finding the component of maximum absolute value of $QR - A$. I consistently got around ϵ_{mach}. For the least squares solves, I found the relative difference between my solution and the solution from using Matlab's backslash. I got values around 10^{-15}. For the Bonus part, I created a random matrix A and then replaced two of the columns with a linear combination of other columns in order to ensure that the matrix did not have all linearly independent columns. Then, after doing the QR factorization, for which I included an output that gives the dependent vector(s), I deleted the corresponding columns in A and then again checked the maximum difference between A and QR, and I got $4.4(10)^{-16}$. I also then compared the solution to the solution I got in Matlab using the corrected matrix and the command backslash. I got a relative difference of $1.8(10)^{-15}$.

(b) (12 points) Write a function to solve the least squares problem using the Householder method. For this you may want to simply write one function, taking as input the $m \times n$ matrix A and vector $\mathbf{b}$ and return the least squares solution to the system $A\mathbf{x} = \mathbf{b}$. Validate your code and explain how you validated it.

See file LeastSquares_Householder.m for the function and PS5N9B.m for the testing script. I validated it by finding the relative difference between my solution and the solution from using Matlab's backslash. I got values around ϵ_{mach}.

(c) (25 points) Find an 11th degree polynomial that approximates the function $\cos(4x)$ evaluated at 50 equally spaced points on the interval $[0, 1]$ by solving the discrete least squares problem. You will be finding the 12 coefficients of $p(x) = c_0 + c_1 x + ... + c_{11}x^{11}$. The Vandermonde matrix that arises in the discrete problem is 50×12. Set up the linear leqst squares problem for the coefficients, and solve it using the methods below:

i. Form the normal equations and solve them.
ii. Use a QR factorization generated by classical Gram-Schmidt.
iii. Use a QR factorization generated by modified Gram-Schmidt.
iv. Use a QR factorization generated by Householder reflectors.
v. Use Matlab's least squares solver by `A\b`. This is also based on QR, but it includes pivoting for added stability.

For each method, give the 2-norm of the residual. Additionally, estimate what the relative error of the coefficients will be using condition numbers of the appropriate matrices. (You can skip this step for the Matlab function.) Lastly, estimate the relative error of the coefficients using the coefficients provided below, and in the file coefficients.mat. Give your results in a table, with one row for each method.

See PS5N9C.m for script. Below is the table.

Method	$\|\|\mathbf{r}\|\|_2$	Est. Rel. Error using κ
Normal Eqns	8.842466657274009e-08	3.129869347179043
Classical G-S	1.378698175913915e-08	2.601866637737549e-08
Modified G-S	1.378698175913915e-08	2.601866637737549e-08
Householder	7.999154279301983e-09	2.601866638308609e-08
Matlab \	7.999154060299545e-09	

(d) (15 points) Comment on the performance of each method. Things to keep in mind/discuss: In class we discussed that the classical Gram-Schmidt method is unstable. Additionally, the Householder method is more stable than the modified Gram-Schmidt method. Lastly, using the condition numbers only allows you to *predict/estimate* the relative error of your coefficients. (Is there one error prediction that you should not believe? Which one and why?)

The largest residual comes from the using the normal equations, and this matches the fact that because of the poor conditioning of this method, the relative error is expected to be the worst, so the residual is also affected negatively. The other methods have fairly comparable residual sizes, with the more stable method of Householder (and Matlab) resulting in the smallest residuals. The residuals all being about the same size makes sense, as in all problems we are setting up an equation that will minimize this quantity.

The real question then becomes how accurate are the solutions (ie the polynomial coefficients) we obtain from this minimization via the different methods. We see that the error from using the normal equations is expected to be large since the matrix we invert is poorly conditioned. Next, since we are theoretically inverting the same matrix in both forms of Gram-Schmidt, the expected relative error from the condition number of that matrix ($\widehat{R}$) is the same in both cases. However, in general, we cannot *believe* the error prediction for the classical method. We briefly mentioned in class that the "rule of thumb" only applies for a special case of stable algorithms, and this algorithm is unstable. It appears that for this particular problem, it gives the same coefficients as the other methods, but in general, we cannot expect it to get a low error for an arbitrary problem.