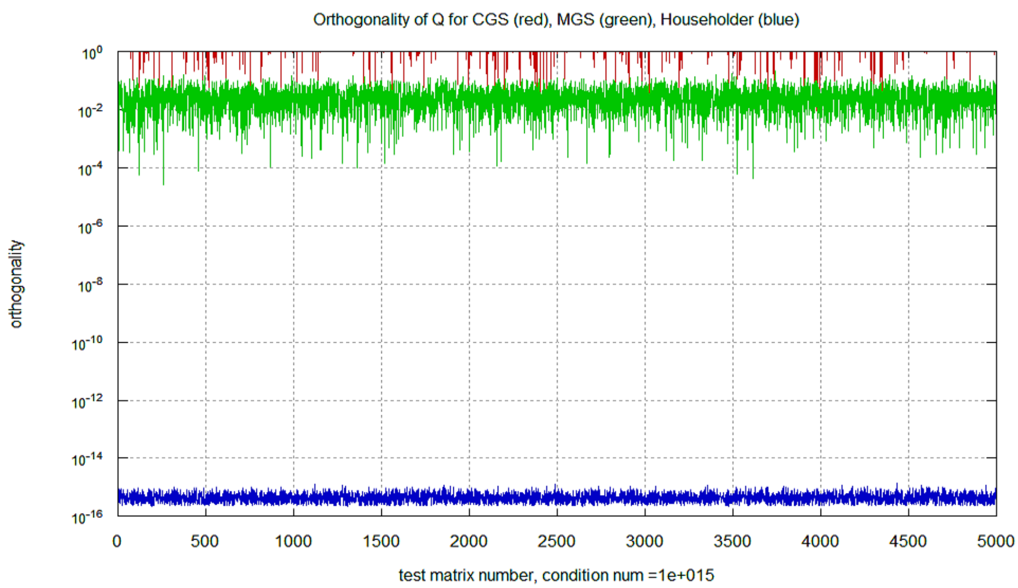
MA 522 Solution for Homework Assignment #8
(Based on the solutions of Isaac J. Lee)

---

1. *This question will illustrate the difference in numerical stability among three algorithms for computing the QR factorization of a matrix: Householder QR, CGS, and MGS. Obtain the Matlab program QRStability.m. This program generates random matrices with user-specified dimensions m and n and condition number cnd, computes their QR decomposition using the three algorithms, and measures the accuracy of the results. It does this with the residual, $||A - Q \cdot R||/||A||$, which should be around machine epsilon ϵ for a stable algorithm, and the orthogonality of Q, $||Q^T \cdot Q - I||$, which should also be around ϵ. Run this program for small matrix dimensions (such as $m = 6$ and $n = 4$), modest numbers of random matrices (samples = 20), and condition numbers ranging from $cnd = 1$ up to $cnd = 10^{15}$. Describe what you see. Which algorithms are more stable than others? See if you can describe how large $||Q^T \cdot Q - I||$ can be as a function of choice of algorithm, cnd, and ϵ.*

For parameters of $m = 6$, $n = 4$, $samples = 5000$, and $cnd = 10^{15}$, I obtained maximum residuals of 1.8222×10^{-16}, 1.7764×10^{-16}, and 1.0829×10^{-15} for CGS, MGS, and Householder, respectively. These residuals are well around $\epsilon \approx 2.2204 \times 10^{-16}$, indicating that the three algorithms are stable for ill-conditioned matrices. However, the maximum orthogonality norms were 2.0021×10^{0}, 2.1840×10^{-1}, and 1.3810×10^{-15}, respectively, suggesting that the Householder algorithm should be preferred over the other two in order to generate Q that is orthogonal when A is ill-conditioned.



The following table displays the mean orthogonality norms from $samples = 500$.

cnd	CGS	MGS	Householder
10^1	8.4195×10^{-16}	5.8226×10^{-16}	4.5702×10^{-16}
10^2	2.8653×10^{-14}	4.1783×10^{-15}	4.5801×10^{-16}
10^3	2.1063×10^{-12}	3.8649×10^{-14}	4.5510×10^{-16}
10^4	1.3925×10^{-10}	3.4739×10^{-13}	4.5514×10^{-16}
10^5	1.5974×10^{-8}	3.4043×10^{-12}	4.6579×10^{-16}
10^6	1.3581×10^{-6}	3.4731×10^{-11}	4.6286×10^{-16}
10^7	1.1418×10^{-4}	3.2090×10^{-10}	4.6095×10^{-16}
10^8	9.1207×10^{-3}	3.4948×10^{-9}	4.6477×10^{-16}
10^9	2.0275×10^{-1}	3.2533×10^{-8}	4.5732×10^{-16}
10^{10}	5.5719×10^{-1}	3.2155×10^{-7}	4.4746×10^{-16}
$\|\|Q^TQ - I\|\|$	$O\left(cnd^2 \cdot \epsilon\right)$	$O\left(cnd \cdot \epsilon\right)$	$O\left(\epsilon\right)$

2-1. *Let A be m-by-n, $m \geq n$, and have full rank. Show that* $\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \vec{r} \\ \vec{x} \end{bmatrix} = \begin{bmatrix} \vec{b} \\ \vec{0} \end{bmatrix}$ *has a solution where $\vec{x}$ minimizes $||A\vec{x} - \vec{b}||_2$.*

We have,

$$\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \vec{r} \\ \vec{x} \end{bmatrix} = \begin{bmatrix} \vec{r} + A\vec{x} \\ A^T\vec{r} \end{bmatrix} = \begin{bmatrix} \vec{b} \\ \vec{0} \end{bmatrix}$$

$$\Rightarrow \begin{cases} \vec{r} = \vec{b} - A\vec{x} \\ A^T\vec{r} = \vec{0} \end{cases}$$

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

$$\Rightarrow A^T A\vec{x} = A^T\vec{b}.$$

Since A has full rank of n, $A^T A$ is invertible (see Homework #7), so we get $\vec{x} = (A^T A)^{-1} A^T \vec{b}$ which minimizes $||A\vec{x} - \vec{b}||_2$.

2-2. *Give an explicit expression for the inverse of the coefficient matrix, as a block 2-by-2 matrix. Where have we previously seen the (2, 1) block entry?*

Let $B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$ be the inverse of the coefficient matrix. Then, we have

$$\begin{bmatrix} I_m & A \\ A^T & 0 \end{bmatrix} \cdot \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} I_m & 0 \\ 0 & I_n \end{bmatrix}$$

$$\Rightarrow \begin{cases} B_{11} + AB_{21} = I_m & \text{Multiply both sides by } A^T. \\ B_{12} + AB_{22} = 0 & \text{Multiply both sides by } A^T. \\ A^T B_{11} = 0 \\ A^T B_{12} = I_n \end{cases}$$

$$\Rightarrow \begin{cases} \underbrace{A_T B_{11}}_{=0} + A_T A B_{21} = A_T \\ \underbrace{A_T B_{12}}_{=I_n} + A_T A B_{22} = 0 \end{cases}$$

$$\Rightarrow \begin{cases} B_{21} = (A^T A)^{-1} A^T \\ B_{22} = -(A^T A)^{-1} \end{cases}$$

$$\Rightarrow \begin{cases} B_{11} = I_m - A(A^T A)^{-1} A^T \\ B_{12} = A(A^T A)^{-1} \end{cases} .$$

Note, B_{21} is the pseudoinverse of A.

3. *Let A be m-by-n, with SVD $A = U\Sigma V^T$. Compute the SVDs of the following matrices in terms of U, Σ, and V.*

1. $(A^TA)^{-1}$

$$\begin{aligned}
(A^TA)^{-1} &= (V\Sigma^T U^T U\Sigma V^T)^{-1} && U^TU = I_n;\ \Sigma^T = \Sigma \\
&= (V\Sigma^2 V^T)^{-1} \\
&= (V^T)^{-1}(\Sigma^2)^{-1}V^{-1} && V^T = V^{-1} \\
&= V(\Sigma^2)^{-1}V^T.
\end{aligned}$$

2. $(A^TA)^{-1}A^T$

$$\begin{aligned}
(A^TA)^{-1}A^T &= V(\Sigma^2)^{-1}V^T \cdot (V\Sigma^T U^T) && V^TV = I_n;\ \Sigma^T = \Sigma \\
&= V(\Sigma^2)^{-1} \cdot \Sigma U^T \\
&= V\Sigma^{-1}U^T.
\end{aligned}$$

3. $A(A^TA)^{-1}$

$$A(A^TA)^{-1} = (U\Sigma V^T) \cdot V(\Sigma^2)^{-1}V^T = U\Sigma^{-1}V^T.$$

4. $A(A^TA)^{-1}A^T$

$$A(A^TA)^{-1}A^T = U\Sigma^{-1}V^T \cdot (V\Sigma^T U^T) = \begin{bmatrix} U & \tilde{U} \end{bmatrix} \begin{bmatrix} I_n & \\ & 0 \end{bmatrix} \begin{bmatrix} U & \tilde{U} \end{bmatrix}^T = UU^T.$$