Math 662 Homework 4

Siyang Jing

3.1

Claim: For MGS, after iteration $i=k$, where $1 \leqslant k \leqslant n$, $\{q_1, \ldots, q_k\}$ are orthonormal.

We prove by induction on k.

Base Case: $k=1$.

$q_1 = a_1 / r_{11} = a_1 / \|a_1\|_2$. Therefore q_1 is normal.

Inductive Step: Assume true for $k < m$. Need to show $\{q_1, \ldots, q_m\}$ are orthonormal after iteration $i=m$.

Claim 2: (For MGS, inside iteration $i=m$) After iteration $j=l$, where $l < m$, q_m is orthogonal to $q_1, \ldots, q_l$.

We show this Claim by induction on l. Base case is trivially true.

Inductive Step: Assume true for $l < p$.

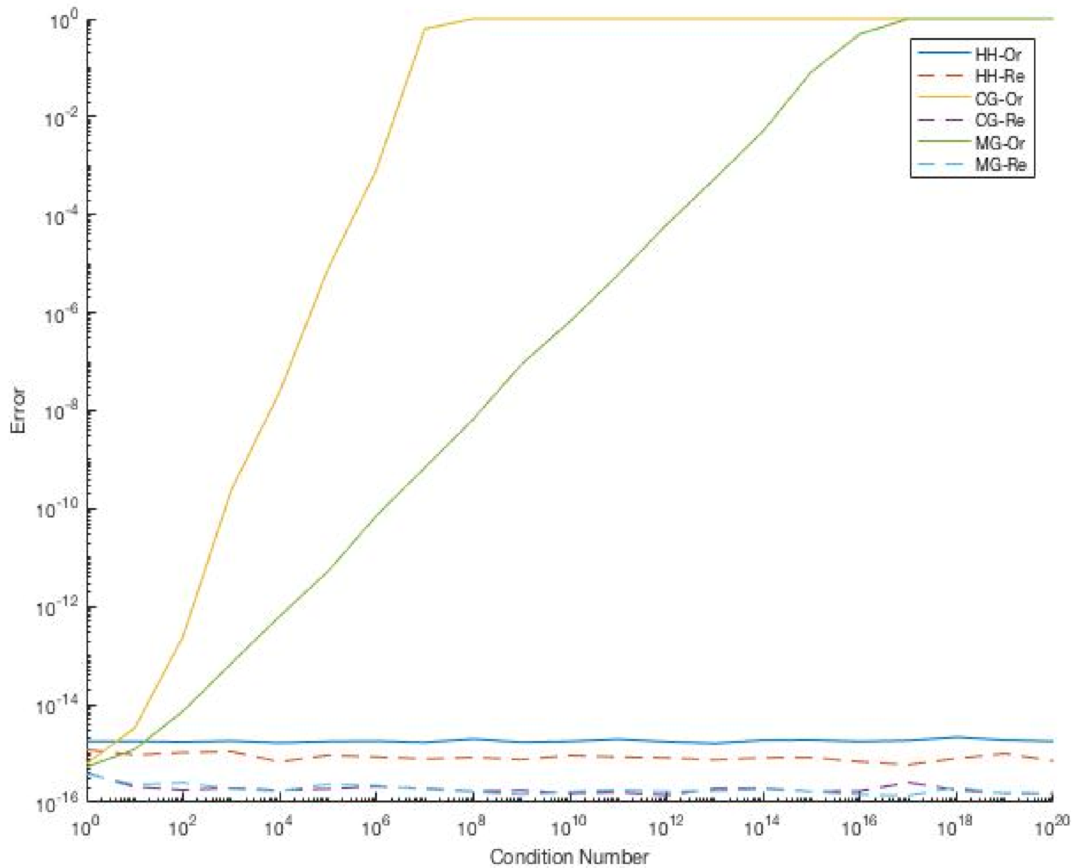
In $j=p$ iteration, $q_p^T(q_m - r_{pm} q_p) = r_{pm}(1 - q_p^T q_p)$. Since $p < m$, and $\{q_1, \ldots, q_{m-1}\}$ are orthonormal, this inner product is just zero, and therefore q_m is orthogonal to q_p.

For each $l < p$, $q_l^T(q_m - r_{pm} q_p) = q_l^T q_m - r_{pm} q_l^T q_p = 0$, by inductive assumption and that $\{q_1, \ldots, q_{m-1}\}$ are orthogonal.

Then, $q_m = q_m / r_{mm} = q_m / \|q_m\|_2$, is normalized. And since $q_1, \ldots, q_{m-1}$ are not changed, $\{q_1, \ldots, q_m\}$ are orthonormal.

In the i loop, in the j loop, for each r_{ji}, for MGS, $r_{ji} = q_j^T q_i = q_j^T(a_i - r_{j-1,i}q_{j-1} - r_{j-2,i}q_{j-2} - \ldots - r_{1i}q_1)$. Since $j < i$, and by Claim, we have $\{q_1, \ldots, q_j\}$ orthogonal, therefore $r_{ji} = q_j^T q_i = q_j^T a_i$. Therefore, MGS and CGS are equivalent.

3.2



Approximately, CGS: $1E-15 \times e^{c^2}$, MGS: $1E-15 \times e^{c}$, Householder: constant

3.3

1. $\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}\begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix} \rightarrow \begin{cases} r + Ax = b \\ A^T r = 0 \end{cases} \rightarrow A^T r + A^T A x = A^T b \xrightarrow{A \text{ has full rank}} x = (A^T A)^{-1} A^T b$

   This is simply the normal equation, and it minimizes $\|Ax - b\|_2$.

2. By part 6 of theorem 3.3, the condition number is just the ratio of the largest to the smallest singular value, i.e. σ_1/σ_n.

3. $\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} I & A^{-1}B \\ C & D \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} I & A^{-1}B \\ 0 & D-CA^{-1}B \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ -CA^{-1} & 1 \end{bmatrix} \rightarrow$

   $\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{bmatrix} \rightarrow$

   $\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1}+A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{bmatrix}$

   For $\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}$, the inverse is $\begin{bmatrix} I-A(A^TA)^{-1}A^T & A(A^TA)^{-1} \\ (A^TA)^{-1}A^T & -(A^TA)^{-1} \end{bmatrix}$

   The $(2,1)$ entry is just the solution operator for the least square normal equation.

4. First obtain $A=QR$

   $x^{(1)}=R^{-1}Q^Tb$, $r^{(1)}=b-Ax^{(1)}$.

   Repeat

   i. Compute the residual vectors (possibly in double precision):

      $s^{(i)}=b-r^{(i)}-Ax^{(1)}$, $t^{(i)}=-A^Tr^{(i)}$

   ii. Solve for the corrections using QR factorization of A

      $$\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix}\begin{bmatrix} dr^{(i)} \\ dx^{(i)} \end{bmatrix} = \begin{bmatrix} s^{(i)} \\ t^{(i)} \end{bmatrix}$$

      $$\begin{bmatrix} dr^{(i)} \\ dx^{(i)} \end{bmatrix} = \begin{bmatrix} I-A(A^TA)^{-1}A^T & A(A^TA)^{-1} \\ (A^TA)^{-1}A^T & -(A^TA)^{-1} \end{bmatrix}\begin{bmatrix} s^{(i)} \\ t^{(i)} \end{bmatrix} = \begin{bmatrix} 0 & QR^{-T} \\ R^{-1}Q^T & -R^{-1}R^{-T} \end{bmatrix}\begin{bmatrix} s^{(i)} \\ t^{(i)} \end{bmatrix}$$

   iii. Update:

      $r^{(i+1)}=r^{(i)}+dr^{(i)}$

      $x^{(i+1)}=x^{(i)}+dx^{(i)}$

   iv. $i \leftarrow i+1$

   Until accurate enough

3.4

$\frac{\partial}{\partial x}\|Ax-b\|_C = \frac{\partial}{\partial x}(Ax-b)^TC(Ax-b) = A^T(C^T+C)(Ax-b)$

$\frac{\partial}{\partial x}\|A\hat{x}-b\|_C = 0 \rightarrow A^TC(A\hat{x}-b)=0 \rightarrow \hat{x}=(A^TCA)^{-1}A^TCb$

We notice that since C is SPD, we can apply Cholesky factorization to C such that $C=LL^T$.

Let $\tilde{A}=L^TA, \tilde{b}=L^Tb$, then the original problem becomes a least square problem of minimizing $\|\tilde{A}x-\tilde{b}\|_2$.

Then everything is the same as the last question, we just substitute $\tilde{A}$ for A and $\tilde{b}$ for b.

3.9

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

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

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

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

3.11

Notice that $\hat{X} = \underset{X}{\operatorname{argmin}} \|AX - I\|_F = \underset{X}{\operatorname{argmin}} \|AX - I\|_F^2$, since Frobenius norm is nonnegative.

Since this is quadratic in X, $\frac{\partial}{\partial X}\|AX - I\|_F^2$ equals zeros when evaluated at $X = \hat{X}$.

$\frac{\partial}{\partial X}\|AX - I\|_F^2 = \frac{\partial AX - I}{\partial X}\frac{\partial}{\partial AX - I}\|AX - I\|_F^2 = 2A^T(AX - I)$

$2A^T(A\hat{X} - I) = 0 \rightarrow \hat{X} = (A^TA)^{-1}A^T = V\Sigma^{-1}U^T$, is the Moore-Penrose pseudo inverse.

$\|A\hat{X} - I\|_F = \|U\Sigma V^TV\Sigma^{-1}U^T - I\|_F = 0$

3.14

Let A be a $m \times n$ matrix. WLOG, we assume $m \geqslant n$.

$$\begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} v_i \\ \pm u_i \end{bmatrix} = \begin{bmatrix} \pm A^Tu_i \\ Av_i \end{bmatrix} = \begin{bmatrix} \pm V\Sigma U^Tu_i \\ U\Sigma V^Tv_i \end{bmatrix} \quad (*)$$

Notice that since U and V are orthonormal matrices, $U^Tu_i = e_i$ and $V^Tv_i = e_i$, where e_i is an n dimensional vector whose entries are all 0 except the ith entry, which is equal to 1.

$$(*) = \begin{bmatrix} \pm V\Sigma e_i \\ U\Sigma e_i \end{bmatrix} = \begin{bmatrix} \pm\sigma_i Ve_i \\ \sigma_i Ue_i \end{bmatrix} = \pm\sigma_i\begin{bmatrix} v_i \\ \pm u_i \end{bmatrix}$$

3.15

We still have $A^TA\hat{x} = A^Tb$, but now A^TA is not invertible.

We first show existence of solution. A is full rank $\Rightarrow \exists y, s.t. Ay = b \Rightarrow \exists y, s.t. A^TAy = A^Tb$.

Since A is full rank and $m < n$, $\operatorname{rank}(A^TA) = \operatorname{rank}(A) = m$. Therefore $\dim(\ker(A^TA)) = n - m$. The solution space of $A^TAy = A^Tb$ has the same dimension as kernel, i.e. $n - m$.