Homework 2

Julian Lehrer

Question 1. Suppose A is symmetric, so $A = A^T$. Let $A = Q^{-1}AQ$ be the Schur deomposition of A. Then $QA = QQ^{-1}UQ = IUQ = UQ \implies QAQ^{-1} = UQQ^{-1} = U$. We now show that U is diagonal. First, note that $U^T = (QAQ^{-1})^T = Q^TA^TQ^{-T} = Q^TAQ^{-T}$ is symmetric. Since U is symmetric, we haev that $u_{ij}, i > j = uiji < j$. But $a_{ij}, i < j = 0$ so $uij = 0, i \neq j$. Therefore U is diagonal. Finally, we have that $D = \lambda_1, ..., \lambda_m$ since by definition, A is diagonalizable.

Question 2. Consider the pertubation of the given system with $c\epsilon > 1$, so we have that

$$\begin{pmatrix} 1 & 1 \\ c\epsilon & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ c \end{pmatrix}$$

Then using Gaussian elimination with partial pivoting we have that

$$\begin{aligned} \begin{pmatrix} 1 & 1 \\ c\epsilon & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 2 \\ c \end{pmatrix} \quad r_1 \leftrightarrow r_2 \\ = \begin{pmatrix} c\epsilon & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} c \\ 2 \end{pmatrix} \quad r_1 - 1/c\epsilon r_1 \\ = \begin{pmatrix} 1 & 1/\epsilon \\ 0 & 1 - 1/\epsilon \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 1/\epsilon \\ 2 - 1/\epsilon \end{pmatrix} \end{aligned}$$

Now consider the case if $\epsilon \leq \epsilon_{\text{mach}}$. We'll have that this system is, in terms of floating point arithmatic, equivalent to

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

So $x = 0, y = 2$. Since $1/\epsilon = 0$. To alleviate this numerical issue, we can use implicit pivoting, where each row of the matrix is scaled by the largest absolute value in its entries. Therefore, we would have the system

$$\begin{pmatrix} 1/2 & 1/2 \\ \epsilon & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Which we previously showed was numerically stable with partial pivoting (see pg. 37 of chapter 2 notes).

Question 3. Let A be symmetric and positive-definite, so $x^*Ax > 0$. First, note that since A is symmetrics, $a_{ii} = a_{ii}^*$, so a_{ii} is real since the only way a complex number can be equal to it's conjugate transpose is if the imaginary part is zero. Then let $x = e_i$, the ith basis vector. So $e_i^* A e_i = e_i^T A e_i = a_{ii} > 0$, therefore the diagonals are both real and strictly postive.

Question 4.

a. We have that

$$\begin{aligned}
&\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A22 \end{pmatrix} \\
&= \begin{pmatrix} IA_{11} + 0A_{12} & IA_{12} + 0A_{22} \\ -A_{21}A_{11}^{-1}A_{11} + A_{21} & -A_{12}A_{21}A^{-1}A_{11} + A_{22} \end{pmatrix} \\
&= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} - A_{21}(A_{11}^{-1}A_{11}) & A_{22} - A_{12}A_{21}A^{-1}A_{11} \end{pmatrix} \\
&= \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{12}A_{21}A^{-1}A_{11} \end{pmatrix}
\end{aligned}$$

b. We want that the first block matrix of the second row A_{21} is eliminated. Therefore, we must multiple the first row by A_{11}^{-1} so the first block entry is $A_{11}^{-1}A_{11} = I$. Then we also multiply by A_{21}, so upon subtraction of row 1 from row 2 that $A_{21} = 0$. Therefore, we subtract by $[A_{21}A_{11}^{-1}A_{11} \quad A_{21}A_{11}^{-1}A_{12}]$ and obtain

$$\begin{pmatrix} A_{11} & C \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{pmatrix}$$

as desired.

Question 5.

a. Let $A = A_1 + iA_2$, $b = b_1 + ib_2$ and $x = x_1 + ix_2$. Then consider the system $(A_1 + iA_2)(x_1 + ix_2) = b_1 + ib_2$. Let $x = (x_1, ..., x_n)$ be our solution vector where $x_i = \alpha_i + i\beta_i$. Then consider $x' = (x_1, ..., x_n, x_{n+1}, ..., x_{2n})$, where the elements $x_{n+1}, ..., x_{2n}$ correspond to the $\beta_i, i = 1, ..., n$. And similarly for b'. Then we have that the system

$$[A_1 A_2]x' = b'$$

Is a system of a $2x \times 2n$ matrix with vectors x and b of length $2n$.

b. Suppose our Gaussian elimination for an $n \times n$ system takes n steps. Note that since complex numbers $x_1 = \alpha_1 + i\beta_1, x_2 = \alpha_2 + i\beta_2$, then x_1x_2 actually has a total of 4 multiplication operations, not one. The exact computational cost for the real valued system is $16n^3/3 = O(n^3)$ operations, where as the big-O computational cost of the complex system will be $O(n^4)$. Therefore, we have that for large n, it is better to use the real-valued equivalent of the system.