270A HW4 graded questions

4. (a) If \( L \) is nonsingular and lower triangular, then \( L \) has nonzero diagonals, \( L_{12} = 0 \). \( L_{11} \) and \( L_{22} \) are both nonsingular and lower triangular. From \( Lx = b \) we have \( L_{11}y = v \). If \( v = 0 \), then \( y = 0 \). Hence \( L_{21}y + L_{22}z = L_{22}z = w \). The first row of this equation is \((L_{k+1,k+1})z_1 = w_1\). Since \( L_{k+1,k+1} \neq 0 \), if \( w_1 \) is nonzero, then \( z_1 \) is also nonzero.

(b) Let \( A \) be SPD with Chelesky factorization \( A = R^T R \). Partition the equation as suggested and we have

\[
\begin{bmatrix}
    A_{11} & c \\
    c^T & A_{33}
\end{bmatrix}
= \begin{bmatrix} L_{11} & 0 & 0 \\
    s^T & b & 0 \\
    L_{31} & t & L_{33}
\end{bmatrix}
\begin{bmatrix}
    L_{11}^T & s & L_{13}^T \\
    0 & b & t^T \\
    0 & 0 & L_{33}^T
\end{bmatrix},
\]

where \( L_{11} \) is lower triangular. We have \( L_{11}s = c \). Notice that \( l_{k+1}(A) \) is the index of the first nonzero entry of \( c \), and \( l_{k+1}(R) \) is the index of the first nonzero entry of \( s \). From (a), we know that if the first \( m \) entries of \( c \) is zero, the \( m + 1 \)-th entry of \( c \) is nonzero, then the same pattern is true for \( s \). Therefore \( l_{k+1}(R) = l_{k+1}(A) \).

7. We first prove that matrix 2-norm is invariant under orthogonal transformation. Let \( Q \) be orthogonal matrix. Then

\[
\|QA\|_2 = \sup_{\|x\|_2 = 1} \|QAx\|_2 = \sup_{\|x\|_2 = 1} (x^T A^T Q^T Q A x)^{-1/2} = \sup_{\|x\|_2 = 1} (x^T A^T A x)^{-1/2} = \|A\|_2.
\]

In addition, we also have \( \|A\|_2 = \|A^T\|_2 = \sigma_1(A) \), \( \|AQ\|_2 = \|(Q A^T)^T\|_2 = \|QA^T\|_2 = \|A\|_2 \).

Our goal is to find \( B \in \mathbb{R}^{n \times n} \), \( \text{rank}(B) = k < n \), such that \( \|A - B\|_2 \) is minimized. We first find a lower bound of \( \|A - B\|_2 \), then we construct \( B^* \) to attain this lower bound.

From the hint, we know that \( \|A - B\|_2 = \sigma_1(A - B) \geq \sigma_{k+1}(A) \). Without loss of generality, suppose \( m \geq n \), otherwise we can consider \( A^T \). The SVD of \( A \) is given by \( A = U \Sigma V^T \), \( \Sigma = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times n}, \Sigma_{11} = \text{diag}\{\sigma_1, \ldots, \sigma_n\} \in \mathbb{R}^{n \times n} \).

Let \( B^* = U \Lambda V^T, \Lambda = \begin{bmatrix} \Lambda_{11} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times n}, \Lambda_{11} = \text{diag}\{\sigma_1, \ldots, \sigma_k, 0, \ldots, 0\} \in \mathbb{R}^{n \times n} \).

Then

\[
\|A - B^*\|_2 = \|U(\Sigma - \Lambda)V^T\|_2 = \|\Sigma - \Lambda\|_2 = \sigma_{k+1}(A).
\]