1a) Let \( A = [\tilde{a}_1 \ldots \tilde{a}_n] \) with \( A_j = [A_{1j} \ A_{2j} \vdots \ A_{mj}] \).

Then \( (A^T A)_{jj} = \tilde{q}_j^T \tilde{q}_j \)

\[
\begin{bmatrix}
A_{1j} \\
A_{2j} \\
\vdots \\
A_{mj}
\end{bmatrix}
\begin{bmatrix}
A_{1j} \\
A_{2j} \\
\vdots \\
A_{mj}
\end{bmatrix}
= A_{1j}^2 + \cdots + A_{mj}^2 = \sum_{i=1}^{m} A_{ij}^2
\]

Then \( \text{trace} \ (A^T A) = \sum_{j=1}^{n} A_{jj} = \sum_{j=1}^{n} \sum_{i=1}^{m} A_{ij}^2 = \|A\|^2_F \)

The other case is similar.

(b) \( \|A\|_2^2 = \text{trace} \ (A^T U(U^T A)) = \text{trace} \ A U U^T A^T \)

\( = \text{trace} \ A A^T = \|A\|_2^2 \) using part (a) and \( U \) orthogonal.

The other case is similar.

(c) If \( A = U \Sigma V^T \) full SVD with \( U \) and \( V \) orthogonal, using (b)

\( \|A\|_F^2 = \|U \Sigma V^T\|_F^2 = \|\Sigma\|_F^2 = \sigma_1^2 + \cdots + \sigma_r^2 \)

Since these are the only nonzero elements, \( r \leq n \).
\(2a\) \quad A \hat{x} = A(\hat{x} + \epsilon \hat{y}) - A \hat{x} = A \hat{x} + A \epsilon \hat{y} - A \hat{x} = \epsilon \hat{y} \) and \(2 - \hat{x} = \hat{x} + \epsilon \hat{y} - \hat{x} = \epsilon \hat{y} \)

Thus \( \frac{\|A \hat{x} - A \hat{y}\|}{\|A \hat{x}\|} = \frac{\|\epsilon \hat{y}\|}{\|A \hat{x}\|} = \frac{\|\epsilon \hat{y}\|}{\|A \hat{x}\|} = \frac{\|\hat{y}\|}{\|A \hat{x}\|} \)

\(= \frac{\|A \hat{y}\|}{\|A \hat{x}\|} \) since \(1 = \|\hat{x}\| = \|\hat{y}\| \)

2g: For random \(\hat{y}\) and \(\hat{x}\), one expects that \(\|A \hat{y}\| \approx \|A \hat{x}\|\) so \(X(\hat{x}, \hat{y})\) typically around 1.

But \(\hat{v}_i\) is the direction which gets stretched the most by \(A\), \(A \hat{v}_i = \hat{v}_i \hat{u}_i\) so

\(\|A \hat{v}_i\| = \|\hat{v}_i \hat{u}_i\| = \|\hat{v}_i\|\)

\(\|A \hat{v}_i\| = \|\hat{v}_i \hat{u}_i\| = \|\hat{v}_i\|\)

The direction in which \(A\) shrinks the most, \(A \hat{v}_S = \hat{v}_5 \hat{v}_5\) and so \(\|A \hat{v}_S\| = \|\hat{v}_5 \hat{v}_5\| = \|\hat{v}_5\|\)

And so \(X(\hat{x}, \hat{y}) = \frac{\|\hat{v}_5\|}{\|\hat{v}_5\|}\), the condition number.

So the worst case is realized by \(\hat{y} = \hat{v}_i\) and \(\hat{x} = \hat{v}_5\).