Practice Problems

1. Consider the singular value decomposition $A = U\Sigma V^T$ where

$$U = \begin{pmatrix} u_1 & u_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & 0 \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} = \begin{pmatrix} 100 & 0 \\ 0 & 1 \end{pmatrix} \qquad V = \begin{pmatrix} v_1 & v_2 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}.$$

a) What orthonormal bases does this give for $C(A)$ and $C(A^T)$?

b) Write A as $\sigma_1 R_1 + \sigma_2 R_2$ where R_1 and R_2 are rank 1 matrices.

c) What is a good rank 1 approximation for A?

d) If you apply A to a unit circle in $\mathbb{R}^2$, what is the output? (A vague answer is fine.)

e) Why not choose $V = I$, which is another orthonormal basis for $C(A^T)$? What does A do to the columns of I?

Solution. a): Given an SVD $A = U\Sigma V^T$, we know that $C(A^T) = C(V)$ and $C(A) = C(U)$. (As a sanity check, $C(A)$ is a subspace of $\mathbb{R}^3$ and $C(A^T)$ is a subspace of $\mathbb{R}^2$, so this is reasonable.) So the orthonormal basis for $C(A)$ is $\{u_1, u_2\}$, and the orthonormal basis for $C(A^T)$ is $\{v_1, v_2\}$.

b): We can read R_1 and R_2 straight off of the formula $A = U\Sigma V^T$.

$$A = (u_1\ u_2)\Sigma\begin{pmatrix} v_1^T \\ v_2^T \end{pmatrix} = u_1\sigma_1 v_1^T + u_2\sigma_2 v_2^T$$

so

$$R_1 = u_1 v_1^T = \frac{1}{5\sqrt{3}}\begin{pmatrix} 3 & 4 \\ 3 & 4 \\ -3 & -4 \end{pmatrix} \qquad \text{and} \qquad R_2 = \frac{1}{5\sqrt{2}}\begin{pmatrix} 4 & -3 \\ 0 & 0 \\ 4 & -3 \end{pmatrix}.$$

c): Since $\sigma_1 > \sigma_2$, the better rank 1 approximation of A is $\sigma_1 R_1$.

d): The output is an ellipse in $\mathbb{R}^3$.

e): The point of choosing V as in the problem is that A takes the columns of V to another set of orthogonal vectors (more precisely, $Av_i = \sigma_i u_i$, so the σ_i's tell us how long the output vectors are). If we chose $V = I$, then the output vectors Ae_1, Ae_2 are just the columns of the matrix A (which you can compute by multiplying out $U\Sigma V^T$). Because σ_1 is so much larger than σ_2, we can see that the first and second columns of A are quite close together, so are far from orthogonal.

2. Suppose A is square and upper triangular (with nonzero diagonal entries). If you perform Gram-Schmidt on the columns of A, what can you say about the square matrix Q whose columns are the Gram-Schmidt vectors?

Solution. Performing Gram-Schmidt is the same as factoring A into $A = QR$ where Q has orthonormal columns and R is a square upper-triangular matrix. If A is already upper-triangular, then Q will be very close to the identity matrix; it will be a diagonal matrix with ± 1's on the diagonal. If this is confusing, try running Gram-Schmidt on a small square upper-triangular matrix, like

$$\begin{pmatrix} -2 & 3 \\ 0 & 4 \end{pmatrix}$$

3. a) Give a 4×3 matrix A with 3 different, nonzero columns such that blindly applying Gram-Schmidt to the columns of A will lead you to divide by zero.

b) What property of A causes Gram-Schmidt to fail?

c) To find an orthonormal basis for $C(A)$, you should instead apply Gram-Schmidt to what matrix B?

Solution. For a): Say the columns of A are $a_1, \ldots, a_n$. In the Gram-Schmidt algorithm, you take a_i and subtract the projection of a_i onto the span of $a_1, \ldots, a_{i-1}$ to get some vector v_i. Then you divide by the norm of v_i in order to get a vector of norm 1. We want to construct A so that at some point in this process, we divide by zero. In the algorithm, we only ever divide by the norm of the vector v_i. So we'd like to choose A so that some v_i is the zero vector; that is, some column a_i is in the span of the previous columns. One example is

$$A = \begin{pmatrix} 1 & 2 & 5 \\ 1 & 2 & 5 \\ 1 & 1 & 3 \end{pmatrix}.$$

For this A, $a_3 = a_1 + 2a_2$. There are many other examples.

For b): The reasoning in the above paragraph tells us that the reason Gram-Schmidt failed (i.e. the reason we divided by zero) is that A does not have full column rank.

For c): We just need to throw out enough columns that we're left with a basis for the original column space. Applying Gram-Schmidt to the resulting matrix B won't run into any problems. For the example above,

$$B = \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 1 \end{pmatrix}$$

would work (so would the matrix consisting of any two columns of A).