8.2 Singular value decomposition and Principal component analysis

Here's the truth about SVD/PCA:

- Singular value decomposition, SVD, is a type of matrix manipulation, termed "matrix factorization", to help identify important aspects of the information contained within the matrix, typically with data about multiple variables and how they are related to each other.
- Principal component analysis, PCA, is a subset of SVD, which is more often referred to in the context of data science. Other mathematics and engineering courses may explore the comprehension of this technique. PCA is relevant in these contexts, where it is *always computed by a computer, NOT by hand, like here.*
- **SVD is NOT a major component of MATH 2400. It is rarely tested, highly unrelated to the other topics in the course, and often skipped by the professor during a semester.**
- SVD, and thus PCA, can be really overwhelming and confusing, even for me, to compute by hand. Use this example and information as a reference, but **do not** stress yourself out over this topic. There's other, more relevant topics to study. Refer to previously texted questions to see how professors have historically asked questions about SVD, if at all.

Okay, my rant is done. Let's take a *brief* dive into SVD.

Singular value decomposition is similar to diagonalization, which we first learned about in 6.1 and explored more with the concept of eigenvalues, eigenvectors, and $D = P^{-1}AP$ in 6.2. Here, our equation of interest is:

$$\mathbf{A} = \mathbf{Q\Sigma P^{-1}} \text{ or } \mathbf{A} = \mathbf{Q\Sigma P^{T}}$$

(P is an orthogonal matrix, and recall from 8.1 that for any orthogonal matrix, P, $P^{-1} = P^T$)

Thinking about it in terms of diagonalization:

Matrix	**Diagonalization**	**SVD**
original $m \times n$ matrix	A MUST be square $n \times n$	A can be non-square $m \times n$
contains scalars on diagonal	D eigenvalues on diagonal $n \times n$	Σ contains singular values on diagonal $m \times n$
contains eigenvectors	P eigenvectors of A $n \times n$	P eigenvectors of A^TA $n \times n$
vectors in $NullSp(A^T)$		Q orthogonal $m \times m$

Okay, now that we've familiarized ourselves with these matrices (a bit), let's dive into how to solve for them.

> SOLVING FOR SVD FOR A, $m \times n$, WHEN $\mathbf{n < m}$
> If $m > n$, follow these steps and take the transpose of your original matrix, A^T, and use that in these steps.
> IMPORTANT (if $m > n$): you must then transpose all matrices at the end to get the "correct" Q, Σ, P.
> If doing this, I recommend renaming this matrix such as $B = A^T$, finding the SVD of B, and then transposing, to avoid confusion.

Step 1: Calculate $A^T A$

Since $n < m$, $A^T A$ will be a $n \times n$ matrix, smaller than the $m \times m$ AA^T.

Step 2: Calculate singular values of $A^T A$

$A^T A$ and AA^T will have the same singular values, so in step 1 we choose the smaller matrix, $n \times n$, to make our lives easier.
Calculating singular values:

1. Calculate eigenvalues of $A^T A$ as previously described
2. Take the square root of each eigenvalue. These are the singular values, σ. This means singular values are always nonnegative.

Step 3: Form Σ

Σ is the same size as A with the singular values along the diagonal. # rows may exceed # σ's. Σ will be $m \times n$.

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Step 4: Find P

1. P is comprised of the **right singular vectors** of A, which are just the eigenvectors of $A^T A$. Find them as previously described. These belong to $NullSp(A^T A)$ by definition of an eigenvector.
2. If there exists σ with algebraic multiplicity > 1, ensure that its eigenvectors are orthogonal. (Recall that while there may be repeated eigenvalues, it will still be diagonal, i.e. no geometric multiplicity > 1 because $A^T A$ is orthogonal)
3. Normalize each eigenvector

4. Place these eigenvectors in the corresponding columns to their singular values in Σ. P will be $n \times n$.

$$P = \begin{bmatrix} \\ v_1 & v_2 & \dots & v_n \\ \\ \end{bmatrix}$$

Step 5: Generate first n columns of Q

Now to make our final matrix, Q, we must find our **left singular vectors** of A, which consists of vectors that belong to $NullSp(A^T)$.

1. We find the first n columns by computing $q_1 = Ap_1 ... q_n = Ap_n$.

2. Place these vectors in their corresponding columns in Q as they are in P, and make sure to normalize each vector

$$Q = \begin{bmatrix} \\ \\ q_1 & q_2 & \dots & q_n & \dots \\ \\ \\ \end{bmatrix}$$

Step 6: Generate last $m - n$ columns of Q

The remainder of Q is comprised of orthogonal vectors in $NullSp(A^T)$ that are also orthogonal to the first n vectors

1. Find A^T and thus the kernel of A^T

2. Make sure that these vectors are orthogonal to each other. If they are not (dot product $\neq 0$), use the **Gram-Schmidt process** to transform the $m - n$ vectors into a mutually orthogonal set

3. Make sure these vectors are orthogonal to the first n vectors in Q using the same logic in 2

   - NOTE that the $m - n$ column vectors will automatically be orthogonal to the first n column vectors if $Rank(A) = n$

$$Q = \begin{bmatrix} \\ \\ q_1 & \dots & q_n & \dots & q_m \\ \\ \\ \end{bmatrix}$$

The complete singular value decomposition of A is now complete as follows:

$$A = Q\Sigma P^T = \begin{bmatrix} \\ q_1 & \cdots & q_n & \cdots & q_m \\ \\ \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \\ v_1 & v_2 & \cdots & v_n \\ \\ \end{bmatrix}^T$$

Recall that from the first step, if you had to transpose the original matrix, A, to make a satisfactory $n < m$ matrix, you must transpose all individual matrices to make the correct Q, Σ, P.
That's the gist of SVD in the context of MATH 2400. Some further notes:

- Principal component analysis is an application of SVD, where you do the same steps, but the matrices have statistical meaning
- PCA is used to interpret data sets with higher dimensions and finds the fewest linear combinations of data columns to account for variation in observations.
- Here's more information about PCA (not relevant)
  - Given A, $m \times n$, where $n >> m$:
    - Mean: $m = \frac{1}{n}\Sigma x_j$
    - Centered data: $B_{m \times n} = [\hat{x_1}| \cdots |\hat{x_n}$ where $\hat{x_j} = x_j - m$
    - Covariance matrix: $S = \frac{1}{b-1}BB^T$
    - Variance: product of the diagonal of S
    - Total variance: $Tr(S)$
  - **Principal components** are the left singular vectors of B
  - PCA is used in a lot of data science, including in yield curves, image processing, genomic data, etc.