An Exploration of Singular Values and the Singular Value Decomposition of Matrices

Julia Trimmer

March 12, 2021

North Carolina Wesleyan College

Dr. Brent Dozier

Introduction

Linear algebra is a branch of mathematics that relies greatly on the construction and manipulation of vectors and matrices in order to solve systems of linear equations. A matrix is defined as a rectangular array of real or complex numbers that are arranged into rows and columns, and the size of a matrix can be expressed as $m \times n$, where m is the number of rows and n is the number of columns. For mathematicians, it is always useful to find a way to simplify and decompose equations and concepts for easier computations and to see the inner workings. This provides a means to understand the important information the equation or concept can provide. This is no different for matrices in linear algebra. There are many ways matrices can be decomposed into its individual elements, with one of the most commonly used methods being diagonalization through the determination of the eigenvalues and eigenvectors of the matrix. However, there are constraints to this method that make it useful for only a limited number of matrices. Therefore, it is necessary to have another method of breaking down matrices. Singular values of a matrix and Singular Value Decomposition generalize the concept of diagonalization to all matrices, including non-diagonalizable matrices, and have many useful real-world applications. This paper will begin with a discussion of eigenvalues and diagonalization, followed by the limitations of eigenvalues and how singular values and the Singular Value Decomposition overcomes these limitations, along with examples and lastly the applications of the Singular Value Decomposition.

Eigenvalues

Before discussing singular values of a matrix, we must first discuss eigenvalues and their importance. Consider the following:

Example 1

Suppose we have a 2×2 matrix $A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$. This matrix can also be written as

$$A = \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} -2 & -1 \\ 1 & 1 \end{bmatrix},$$

which is known as its diagonalized form. The middle matrix is what is known as a diagonal matrix, which is a square matrix in which all entries that are not on the main diagonal are zero. We can express the form above in Example 1 in general terms as $A = PDP^{-1}$, where P^{-1} is the inverse of matrix P (Andrilli & Hecker, 2010). If we multiply this out, we will see that it is equivalent to the original matrix A given. If we were to square matrix A, we would have

$$A^2 = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} = \begin{bmatrix} -2 & -3 \\ 6 & 7 \end{bmatrix}.$$

This computation can also be accomplished by squaring the D matrix in the diagonalized form:

$$A^2 = \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} (-1)^2 & 0 \\ -0 & (-2)^2 \end{bmatrix} \begin{bmatrix} -2 & -1 \\ 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -0 & 4 \end{bmatrix} \begin{bmatrix} -2 & -1 \\ 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -2 & -3 \\ 6 & 7 \end{bmatrix}.$$

This is because $A^2 = (PDP^{-1})^2 = PDP^{-1}PDP^{-1} = PDDP^{-1} = PD^2P^{-1}$, and to square a diagonal matrix one just needs to square the diagonal entries. Similarly, A^3 can be computed by the same method, where the D matrix is cubed when A is in the diagonalized form. In general, for any natural number n, $A^n = PD^nP^{-1}$. So, the benefits of this diagonalized form is that once diagonalized, computations such as raising a matrix to a given power are greatly simplified, in that, the power is applied only to the entries, specifically those along the main diagonal, of *D* (Margalit &Rabinoff, 2019). Diagonalization of a matrix is dependent on determining the eigenvalues and eigenvectors of the given matrix.

Determination of Eigenvalues and Eigenvectors

The values along the main diagonal of the *D* matrix are known as the eigenvalues of matrix *A* and can be defined as follows. For an $n \times n$ matrix *A*, a real or complex number λ is an eigenvalue of *A* if and only if there is a nonzero *n*-vector *X* such that $AX = \lambda X$. The vector *X* is known as an eigenvector corresponding to the eigenvalue λ. In other words, the product of *AX* is equivalent to the scaling of *X* by λ. The columns of the *P* matrix mentioned previously are the eigenvectors of the matrix *A*. First, one must determine the eigenvalues of a matrix *A* by using the following theorem:

Theorem 1: If *A* is an $n \times n$ matrix and λ is a real number, then λ is an eigenvalue of *A* if and only if $|\lambda I_n - A| = 0$ (where $|\cdot|$ represents determinant, and I_n is the $n \times n$ identity matrix).

Therefore, the eigenvalues are the roots of the characteristic polynomial of *A*, which is defined as

$$p(x) = |xI_n - A|.$$

An eigenvector *X* corresponding to an eigenvalue λ is a solution to the homogeneous equation $(\lambda I_n - A)X = 0_n$, where 0_n is the zero *n*-vector (Andrilli & Hecker, 2010). Looking back at Example 1, we can find the eigenvalues for the matrix $A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$. The characteristic polynomial can be written as

$$p(x) = \left|\begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}\right| = \left|\begin{bmatrix} x & -1 \\ 2 & x+3 \end{bmatrix}\right|.$$

Taking the determinant, we get the equation $p(x) = x\,(x + 3) - (2)(-1) = x^2 + 3x + 2$. The roots of this equation are $x = -2$ and $x = -1$. Therefore, the eigenvalues of matrix A are $\lambda_1 = -1$ and $\lambda_2 = -2$. From here, the eigenvectors can be found by solving the homogeneous system $(\lambda I_2 - A)X = 0$, where $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, for each eigenvalue. So, for $\lambda_1 = -1$, the system would become $((-1)I_2 - A)X = 0$. Since

$$(-1)I_2 - A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 2 & 2 \end{bmatrix},$$

the augmented matrix is

$$[(-1)I_2 - A|0] = \left[\begin{array}{cc|c} -1 & -1 & 0 \\ 2 & 2 & 0 \end{array}\right],$$

and row reduces to $\left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$. This row reduced matrix represents the linear system

$$\begin{aligned} 1x_1 + 1x_2 &= 0 \\ 0x_1 + 0x_2 &= 0 \end{aligned}$$

We see that there are an infinite number of solutions, since x_2 is any real number. So $x_1 = -x_2$ and $x_2 = x_2$. Therefore, $X = x_2\begin{bmatrix} -1 \\ 1 \end{bmatrix}$, for any nonzero number x_2, and if we let $x_2 = 1$, then we generate one eigenvector for $\lambda_1 = -1$, which is $X_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Any nonzero scalar multiple of X_1 is also an eigenvector for the eigenvalue. This is true for all eigenvectors, since multiplying the equation $AX = \lambda X$ by a scalar c on both sides gives $cAX = c\lambda X$, which is the same as $A(cX) = \lambda(cX)$. Therefore, cX is also an eigenvector relative to λ. Through the same process, we can solve the homogeneous system for $\lambda_2 = -2$. The system would be

$$((-2)I_2 - A)X = 0 \text{ and } (-2)I_2 - A = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} = \begin{bmatrix} -2 & -1 \\ 2 & 1 \end{bmatrix}.$$

The augmented matrix is

$$[(-2)I_2 - A|0] = \left[\begin{array}{cc|c} -1 & -1 & 0 \\ 2 & 2 & 0 \end{array}\right],$$

which row reduces to $\left[\begin{array}{cc|c} 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{array}\right]$. Here, $x_2 = x_2$ and $x_1 = -\frac{1}{2}x_2$. If we let $x_2 = 2$, then the eigenvectors for $\lambda_2 = -2$ are the scalar multiples of $X_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$. Knowing the eigenvalues and eigenvectors for a given matrix allows us to easily construct its diagonalized form.

Diagonalization of a Matrix

The method for the diagonalization of a matrix A states that if A and P are $n \times n$ matrices such that each column of P is an eigenvector for A, and P is nonsingular, meaning it has an inverse (denoted here as P^{-1}) then $D = P^{-1}AP$ is a diagonal matrix. D and A are said to be similar matrices. The main diagonal entries of D are the eigenvalues of A such that the ith main diagonal entry d_{ii} of D is the eigenvalue corresponding to the eigenvector forming the ith column of P. The equation for D can be arranged to give the diagonalized form of matrix A given previously, which is $A = PDP^{-1}$, by left multiplying by P and right multiplying by P^{-1} on both sides we get

$$PDP^{-1} = P(P^{-1}AP)P^{-1} = A$$

since $PP^{-1} = I$. So $PDP^{-1} = A$ (Andrilli & Hecker, 2010). Returning to Example 1, we can now diagonalize matrix A given the eigenvalues and eigenvectors found. Matrix P will be composed of the eigenvectors of A beginning with the X_1 in first column and X_2 in the second. Matrix D will contain the eigenvalues along the main diagonal in the order corresponding to the eigenvectors in P. Finally, matrix P^{-1} is the inverse of P. Thus, the matrix components of the diagonalization are

$$P = \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix}, \quad D = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}, \quad P^{-1} = \begin{bmatrix} -2 & -1 \\ 1 & 1 \end{bmatrix},$$

and the complete diagonalization, as shown previously, is

$$A = \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} -2 & -1 \\ 1 & 1 \end{bmatrix}.$$

Example 2

For our second example, we have the matrix

$$A = \begin{bmatrix} 4 & -4 & 6 \\ -1 & 2 & -1 \\ -1 & 4 & -3 \end{bmatrix},$$

which can be diagonalized. The characteristic polynomial can be found by

$$p(x) = \left| \begin{bmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{bmatrix} - \begin{bmatrix} 4 & -4 & 6 \\ -1 & 2 & -1 \\ -1 & 4 & -3 \end{bmatrix} \right|.$$

Taking the determinant, the equation becomes

$$p(x) = x^3 - 3x^2 - 4x + 12 = (x + 2)(x - 2)(x - 3).$$

Therefore, $\lambda_1 = 3$, $\lambda_2 = 2$, and $\lambda_3 = -2$. So, $D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$. For $\lambda_1 = 3$, an eigenvector can be determined by solving the homogenous system $(3I_3 - A)X = 0$. So,

$$3I_3 - A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} - \begin{bmatrix} 4 & -4 & 6 \\ -1 & 2 & -1 \\ -1 & 4 & -3 \end{bmatrix} = \begin{bmatrix} -1 & 4 & -6 \\ 1 & 1 & 1 \\ 1 & -4 & 6 \end{bmatrix}.$$

The augmented matrix is $\left[\begin{array}{ccc|c} -1 & 4 & -6 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & -4 & 6 & 0 \end{array}\right]$, which row reduces to $\left[\begin{array}{ccc|c} 1 & 0 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$. Therefore, $x_1 = -2x_3$, $x_2 = x_3$, and $x_3 = x_3$. So, $X = x_3\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and if we let $x_3 = 1$, then $X_1 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$. For $\lambda_2 = 2$, the homogeneous system is $(2I_3 - A)X = 0$, and

$$2I_3 - A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} - \begin{bmatrix} 4 & -4 & 6 \\ -1 & 2 & -1 \\ -1 & 4 & -3 \end{bmatrix} = \begin{bmatrix} -2 & 4 & -6 \\ 1 & 0 & 1 \\ 1 & -4 & 5 \end{bmatrix}.$$

The augmented matrix row reduces to $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$, so $x_1 = -x_3$, $x_2 = x_3$, and $x_3 = x_3$.

$X = x_3\begin{bmatrix}-1\\1\\1\end{bmatrix}$, so $X_2 = \begin{bmatrix}-1\\1\\1\end{bmatrix}$ if we let $x_3 = 1$. Finally, for $\lambda_3 = -2$, the homogeneous system is

$(-2I_3 - A)X = 0$, and

$$-2I_3 - A = \begin{bmatrix}-2 & 0 & 0\\0 & -2 & 0\\0 & 0 & -2\end{bmatrix} - \begin{bmatrix}4 & -4 & 6\\-1 & 2 & -1\\-1 & 4 & -3\end{bmatrix} = \begin{bmatrix}-6 & 4 & -6\\1 & -4 & 1\\1 & -4 & 1\end{bmatrix}.$$

The augmented matrix row reduces to $\left[\begin{array}{ccc|c}1 & 0 & 1 & 0\\0 & 1 & 0 & 0\\0 & 0 & 0 & 0\end{array}\right]$, therefore $x_1 = -x_3$, $x_2 = 0$, and $x_3 = x_3$. So,

$X = x_3\begin{bmatrix}-1\\0\\1\end{bmatrix}$, and if we let $x_3 = 1$, then $X_3 = \begin{bmatrix}-1\\0\\1\end{bmatrix}$. So,

$$P = \begin{bmatrix}-2 & -1 & -1\\1 & 1 & 0\\1 & 1 & 1\end{bmatrix} \text{ and } P^{-1} = \begin{bmatrix}-1 & 0 & -1\\1 & 1 & 1\\0 & -1 & 1\end{bmatrix}.$$

Therefore, the diagonalization of A is

$$A = \begin{bmatrix}-2 & -1 & -1\\1 & 1 & 0\\1 & 1 & 1\end{bmatrix}\begin{bmatrix}3 & 0 & 0\\0 & 2 & 0\\0 & 0 & -2\end{bmatrix}\begin{bmatrix}-1 & 0 & -1\\1 & 1 & 1\\0 & -1 & 1\end{bmatrix}.$$

Limitations of Eigenvalues

It has been shown how the determination of the eigenvalues of a matrix is useful, particularly when it comes to the diagonalization of the given matrix. However, there are some limitations when it comes to the computation of eigenvalues and their use in the diagonalization of matrices. First, in order to determine the eigenvalues of a matrix, the given matrix must be square. By definition, a matrix with dimensions other than $n \times n$ cannot have eigenvalues or eigenvectors. Also, in general, a matrix with complex eigenvalues is not diagonalizable (Huang). Finally, if there are fewer fundamental eigenvectors than eigenvalues, the matrix is nondiagonalizable (Andrilli & Hecker, 2010). These limitations eliminate a great number of matrices from the advantages of eigenvalues and diagonalization. Therefore, a method is needed to break down and utilize matrices that do not fit the requirements for eigenvalues and diagonalization.

Singular Values

Another method for decomposing a matrix into its separate components like the diagonalization of a matrix is known as the Singular Value Decomposition, which has components related to eigenvalues, called singular values. In this section, this matrix decomposition will be described, and examples will be given to show its benefits in comparison to diagonalization. We will also discuss a method for constructing a matrix with specified singular values and size using the Singular Value Decomposition and some of its real-world applications.

A Brief History of the SVD

The SVD was discovered in the 1870s independently by Eugenio Beltrami (1835-1899) and Camille Jordan (1839-1921), both for the simplification of bilinear forms. James Joseph Sylvester (1814-1897) wrote two papers focusing on the SVD as a technique for diagonalizing a quadratic form, and later extending it to include bilinear forms as well. Erhard Schmidt (1876-1959) approached the SVD through integral equations. Schmidt generalized the SVD to infinite-dimensional spaces and showed how it could be used to find the best low-rank approximations to an operator. Eckart and Young extended the SVD to rectangular matrices. The name "singular value" appears to have come from their use in integral equations and is first used by Bateman and Picard. In 1937, Smithies uses the term in the same way it is used today (Stewart, 1993).

Definitions and Tools

In this short section we will discuss some matrix terminology that is needed, and a matrix theory that is sometimes needed when decomposing a matrix by means of the Singular Value Decomposition.

Orthogonal set of Vectors – A set of vectors $\{x_1, x_2, \ldots, x_k\}$ is an orthogonal set if and only if the dot product of any two distinct vectors in the set is zero, that is $x_i^T x_j = 0$.

Unit Vector – Any vector of length 1, is a unit vector, where length of a vector is the square root of $x^T x$.

Normalization – Dividing a vector by its length to obtain a unit vector in the same direction. The equation for normalizing a vector x is $u = \frac{1}{\sqrt{x^T x}} x$, where u is the unit vector in the direction of x.

Orthonormal Set of Vectors – A set of vectors $\{v_1, v_2, \ldots, v_k\}$ is an orthonormal set if and only if each vector v_i is a unit vector.

Orthogonal Matrix – A matrix (nonsingular and square) is orthogonal if and only if $A^T = A^{-1}$. In other words, the columns form an orthonormal set of vectors.

Gram-Schmidt Process

The Gram-Schmidt Process takes a set of m linearly independent vectors $x_1, x_2, \ldots, x_m \in \mathbb{R}^n$ and produces a set of m orthogonal vectors $v_1, v_2, \ldots, v_m \in \mathbb{R}^n$. Normalizing each of the vectors as they are produced gives an orthonormal set of vectors, The process is as follows:

Let $w_1 = x_2$ and then normalize $\Rightarrow v_1 = \frac{1}{\sqrt{w_1^T w_1}} w_1$.

Let $w_2 = x_2 - (x_2^T v_1) v_1$ and then normalize $\Rightarrow v_2 = \frac{1}{\sqrt{w_2^T w_2}} w_2$.

Let $w_3 = x_3 - (x_3^T v_1)v_1 - (x_3^T v_2)v_2$ and then normalize $\Rightarrow v_3 = \frac{1}{\sqrt{w_3^T w_3}} w_3$.

Let $w_4 = x_4 - (x_4^T v_1)v_1 - (x_4^T v_2)v_2 - (x_4^T v_3)v_3$ and then normalize $\Rightarrow$

$v_4 = \frac{1}{\sqrt{w_4^T w_4}} w_4$.

In general, $w_k = x_k - \sum_{i=1}^{k-1}(x_k^T v_i)v_i$, and then $v_k = \frac{1}{\sqrt{w_k^T w_k}} w_k$, for

$k = 2,3,\ldots,m$.

Then $\{v_1, v_2, \ldots, v_m\} \subseteq \mathbb{R}^n$ is an orthogonal set (Andrilli & Hecker, 2010).

See Result 1 in the Appendix for further explanation of the Gram-Schmidt process. Now that we have these tools defined, we are prepared to describe the Singular Value Decomposition as well as the procedure to derive it.

Singular Value Decomposition

A technical description of singular values and the Singular Value Decomposition will now be given. Let A be any $m \times n$ matrix. Singular values are defined as the square roots of the eigenvalues shared by the matrices A^TA and AA^T. Singular values are usually denoted by σ_1, σ_2, …, σ_n and listed in decreasing order (Bretscher, 2009). It is important to note that, by definition, AA^T and A^TA are symmetric matrices. This means they are equal to their transpose, since

$$(AA^T)^T = (A^T)^TA^T = AA^T \text{ and } (A^TA)^T = A^T(A^T)^T = A^TA.$$

The eigenvalues of a real symmetric matrix are always real and non-negative and the Spectral Theorem states that all symmetric matrices are diagonalizable. So, an $n \times n$ matrix with real entries has n linearly independent eigenvectors and for all distinct eigenvalues (non-repeated), the corresponding eigenvectors will be orthogonal. This shows that the singular values of any matrix are always real (Jerison).

The Singular Value Decomposition, or SVD, is a generalization of the diagonalization of a matrix. Any real matrix can be factored using the SVD. The SVD decomposes a matrix into the product of three matrices. Let A be an $m\times n$ matrix, and let $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k > \sigma_{k+1} = \ldots = \sigma_n = 0$ be the singular values of A. Let V be the $n\times n$ orthogonal matrix whose columns are composed of right singular vectors of A, $\{v_1,\ldots v_n\}$, where v_i corresponds to σ_i. The vectors $\{v_1,\ldots v_n\}$ are also an orthonormal set of eigenvectors for the symmetric matrix A^TA, where v_i corresponds to λ_i. Let U be the $m\times m$ orthogonal matrix whose columns are composed of left singular vectors of A, $\{u_1,\ldots u_m\}$, where u_i corresponds to σ_i. Here, $u_i = \frac{1}{\sigma_i}Av_i$ for $1 \leq i \leq k$. Finally, Σ is the $m\times n$ matrix whose main diagonal entries are the singular values in decreasing order and all other entries are zero. Then the complete SVD of matrix A is $A = U\Sigma V^T$ (Andrilli & Hecker, 2010). See Appendix Result 3 for further derivation of the SVD.

An example, Example 3, will now be shown to visualize the components of the SVD.

<u>Example 3</u>

Given the 2×3 matrix

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix},$$

the SVD, $A = U\Sigma V^T$, is

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

Here, we can see the three matrices U, Σ, and V are

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & -\frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & \frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & \frac{1}{3} \end{bmatrix}.$$

Therefore, by the definition of SVD, the entries along the main diagonal of the Σ matrix are the singular values ($\sigma_1 = 5$ and $\sigma_2 = 3$) of A. The columns of matrix U (u_1 and u_2) are left singular vectors of A and the columns of V (v_1, v_2, and v_3) are right singular vectors of A. We will return to this example and the derivation of its SVD following the procedure for determining the SVD for any matrix.

SVD Procedure

The procedure for computing the SVD of a matrix can now be outlined. Without loss of generality, we will assume $m \leq n$ (since SVD for $A = U\Sigma V^T$ automatically produces SVD for $A^T = V\Sigma U^T$).

Step 1: The matrix AA^T, which is $m \times m$, is formed and its eigenvalues are found (all eigenvalues of a symmetric matrix are real and positive). The square roots of these eigenvalues are the singular values, denoted by σ_i, which can be used to form the Σ matrix.

Step 2: Form the $n \times n$ matrix A^TA and compute all n fundamental eigenvectors for this matrix (a symmetric matrix such as this will always have a full set of linearly independent eigenvectors).

Step 3: From these n fundamental eigenvectors, we form an orthonormal set of vectors, which are the columns of V, denoted by v_i (eigenvectors corresponding to distinct eigenvalues are orthogonal, but one may need to use the Gram-Schmidt procedure on the eigenvectors corresponding to repeated eigenvalues, and all eigenvectors must be normalized).

Step 4: For all $\sigma_i > 0$, compute column u_i of U by the formula $u_i = \frac{1}{\sigma_i} A v_i$.

Step 5: For all $\sigma_i = 0$, find the fundamental eigenvectors of the AA^T matrix, corresponding to the zero eigenvalues (one may need to use the Gram-Schmidt process in the case of repeated eigenvalues and, as always, the vectors must be normalized). This will give the rest of the u_i for matrix U.

This completes the procedure to obtain $A = U\Sigma V^T$.

It is important to note that the columns of U, composed of left singular vectors of A, must correspond to the right singular vectors of A in the columns of V. Thus, we use the formula $u_i = \frac{1}{\sigma_i} A v_i$ to determine left singular vectors. This formula can be derived from the formula for the SVD, $A = U\Sigma V^T$.

Let u_i be the columns of U, v_i be the columns of V, and Σ_i be the columns Σ. Since $A = U\Sigma V^T$, we can right multiply by V on both sides such that

$$AV = U\Sigma V^T V$$

And since $V^T V$ is the identity,

$$AV = U\Sigma$$

The ith column of AV is Av_i and the ith column of $U\Sigma$ is $U\Sigma_i$ therefore,

$$Av_i = U\Sigma_i.$$

As mentioned before, the columns of Σ consist of zeros everywhere except along the main diagonal, which contains the σ_i entries. So $\Sigma_i = \begin{bmatrix} 0 \\ \vdots \\ \sigma_i \\ \vdots \\ 0 \end{bmatrix}$, where σ_i is in the ith position.

So, we have

$$Av_i = U\begin{bmatrix} 0 \\ \vdots \\ \sigma_i \\ \vdots \\ 0 \end{bmatrix}$$

and we can factor out the σ_i coefficient such that

$$Av_i = \sigma_i U\begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \text{ or,}$$

$$Av_i = \sigma_i u_i.$$

This final equation can be rearranged such that we get the formula for the columns of U:

$$u_i = \frac{1}{\sigma_i} Av_i.$$

Example for Determining the SVD of a Matrix

We will now return to Example 3 and show the process of decomposing it according to the SVD method. Following the procedure, we will begin by determining the eigenvalues of matrix $AA^T = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$. To determine the eigenvalues, we can find the characteristic polynomial, $p(x) = |xI_2 - AA^T|$, or

$$p(x) = \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} - \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix} = \left|\begin{bmatrix} x-17 & 8 \\ 8 & x-17 \end{bmatrix}\right|.$$

Taking the determinant, we get the equation $p(x) = (x - 17)(x - 17) - 64 = x^2 - 34x + 225$, the roots of which are the eigenvalues of AA^T. These eigenvalues are $\lambda_1 = 25$ and $\lambda_2 = 9$. The singular values of matrix A are the square roots of these eigenvalues, thus $\sigma_1 = \sqrt{25} = 5$ and

$\sigma_2 = \sqrt{9} = 3$. From this, we can form the Σ matrix, which will be a rectangular matrix of the same size as A, 2×3, and singular values along the main diagonal, with all other entries being zero. Therefore, for this example,

$$\Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}.$$

We can now determine the columns of V by normalizing the fundamental eigenvectors of matrix A^TA.

$$A^TA = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} = \begin{bmatrix} 13 & 12 & 2 \\ 12 & 13 & -2 \\ 2 & -2 & 8 \end{bmatrix}$$

Beginning with $\lambda_1 = 25$, we need to solve the homogeneous system $(25I_3 - A^TA)X = 0$, where $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. The augmented matrix is

$$[25I_3 - A^TA|0] = \left[\begin{array}{ccc|c} 12 & -12 & -2 & 0 \\ -12 & 12 & 2 & 0 \\ -2 & 2 & 17 & 0 \end{array}\right],$$

and row reduces to $\left[\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$, which represents the linear system

$$\begin{array}{l} 1x_1 - 1x_2 + 0x_3 = 0 \\ 0x_1 + 0x_2 + 1x_3 = 0. \\ 0x_1 + 0x_2 + 0x_3 = 0 \end{array}$$

We see that there are an infinite number of solutions, since x_2 is any real number. So, $x_1 = x_2$, $x_2 = x_2$, and $x_3 = 0$, therefore $X = x_2\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. If we let $x_2 = 1$, we generate X_1, the eigenvector for $\lambda_1 = 25$. This vector must now be normalized in order to make it a unit vector. Normalizing vector X_1 produces the vector $v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$, making this a right singular vector corresponding to the singular value $\sigma_1 = 5$. The same process can be completed for $\lambda_2 = 9$ and $\lambda_3 = 0$, resulting in the right singular vectors

$$v_2 = \begin{bmatrix} \frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \end{bmatrix} \text{ and } v_3 = \begin{bmatrix} -\frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}.$$

Therefore, we know have matrix V and, in turn, matrix V^T:

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & -\frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & \frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & \frac{1}{3} \end{bmatrix} \text{ and } V^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

We can now construct the U matrix in relation to the V matrix by the formula $u_i = \frac{1}{\sigma_i} A v_i$. We can find the first column of U, u_1, using $\sigma_1 = 5$ and $v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$:

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{5} \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix},$$

and $u_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$. Similarly, we can use $\sigma_2 = 3$ and $v_2 = \begin{bmatrix} \frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{18}} \\ \frac{4}{\sqrt{18}} \end{bmatrix}$ to find that $u_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$, and

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Putting all the components together, we get the Singular Value Decomposition of A as

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

More Examples

Example 4

We will use matrix $A = \begin{bmatrix} 0 & 1 & 1 \\ \sqrt{2} & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$ to display the SVD. We can begin by determining the matrix AA^T and determining the eigenvalues of this matrix:

$$AA^T = \begin{bmatrix} 0 & 1 & 1 \\ \sqrt{2} & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & \sqrt{2} & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 2 \end{bmatrix}.$$

To find the eigenvalues, we can find the characteristic polynomial, $p(x) = |xI_3 - AA^T|$, or

$$p(x) = \left| \begin{bmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{bmatrix} - \begin{bmatrix} 2 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 2 \end{bmatrix} \right| = \left| \begin{bmatrix} x-2 & -2 & -2 \\ -2 & x-6 & -2 \\ -2 & -2 & x-2 \end{bmatrix} \right|.$$

Finding the determinant, we get $p(x) = x^3 - 10x^2 + 16x$. Factoring, we find that the eigenvalues of AA^T are $\lambda_1 = 8$, $\lambda_2 = 2$, and $\lambda_3 = 0$. Therefore, the singular values of A are $\sigma_1 = 2\sqrt{2}$, $\sigma_2 = \sqrt{2}$, and $\sigma_3 = 0$ and

$$\Sigma = \begin{bmatrix} 2\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Right singular vectors now need to be determined using matrix A^TA.

$$A^TA = \begin{bmatrix} 0 & \sqrt{2} & 0 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ \sqrt{2} & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2\sqrt{2} & 0 \\ 2\sqrt{2} & 6 & 2 \\ 0 & 2 & 2 \end{bmatrix}$$

The augmented matrix when solving the homogeneous system for $\lambda_1 = 8$ is

$$[8I_3 - A^TA|0] = \left[\begin{array}{ccc|c} 6 & -2\sqrt{2} & 0 & 0 \\ -2\sqrt{2} & 2 & -2 & 0 \\ 0 & -2 & 6 & 0 \end{array}\right],$$

which row reduces to $\left[\begin{array}{ccc|c} 1 & 0 & -\sqrt{2} & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$. Thus, an eigenvector for A^TA corresponding to $\lambda_1 = 8$ is

$X_1 = \begin{bmatrix} \sqrt{2} \\ 3 \\ 1 \end{bmatrix}$ and a right singular vector corresponding to $\sigma_1 = 2\sqrt{2}$ is $v_1 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{3}{\sqrt{12}} \\ \frac{1}{\sqrt{12}} \end{bmatrix}$. We also find that

$v_2 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ 0 \\ \frac{2}{\sqrt{6}} \end{bmatrix}$ and $v_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$. Therefore, matrix $V = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{12}} & 0 & -\frac{1}{2} \\ \frac{1}{\sqrt{12}} & \frac{2}{\sqrt{6}} & \frac{1}{2} \end{bmatrix}$ and

$$V^T = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{3}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\ -\frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Left singular vectors can now be determined using the equation $u_i = \frac{1}{\sigma_i} A v_i$. We can begin by

determining the u_1 matrix using $\sigma_1 = 2\sqrt{2}$ and $v_1 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{3}{\sqrt{12}} \\ \frac{1}{\sqrt{12}} \end{bmatrix}$:

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{2\sqrt{2}} \begin{bmatrix} 0 & 1 & 1 \\ \sqrt{2} & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{3}{\sqrt{12}} \\ \frac{1}{\sqrt{12}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}.$$

Similarly, we can use this equation along with σ_2 and v_2 to find that $u_2 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$. For $\sigma_3 = 0$, we must find a fundamental eigenvector for AA^T corresponding to the eigenvalue $\lambda_3 = 0$. Solving for the homogeneous system $(0I_3 - A^TA)X = 0$, we get the augmented matrix

$$[0I_3 - AA^T\,|0] = \left[\begin{array}{ccc|c} -2 & -2 & -2 & 0 \\ -2 & -6 & -2 & 0 \\ -2 & -2 & -2 & 0 \end{array}\right].$$

This row reduces to $\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$, so an eigenvector corresponding to $\lambda_3 = 0$ is $X_1 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, and a left singular vector corresponding to $\sigma_3 = 0$ is $u_3 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}$. So, the complete SVD of matrix A is

$$A = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{3}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\ -\frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Example 5

The next example of the SVD is for matrix

$$A = \begin{bmatrix} 18 & -12 & -18 & -9 \\ -24 & -9 & 24 & 12 \end{bmatrix}.$$

To begin, we can find AA^T in order to determine its eigenvalues:

$$AA^T = \begin{bmatrix} 18 & -12 & -18 & -9 \\ -24 & -9 & 24 & 12 \end{bmatrix} \begin{bmatrix} 18 & -24 \\ -12 & -9 \\ -18 & 24 \\ -9 & 12 \end{bmatrix} = \begin{bmatrix} 873 & -864 \\ -864 & 1{,}377 \end{bmatrix}.$$

The characteristic polynomial is $p(x) = |xI_2 - AA^T|$, or

$$p(x) = \left| \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} - \begin{bmatrix} 873 & -864 \\ -864 & 1{,}377 \end{bmatrix} \right| = \left| \begin{bmatrix} x - 873 & 864 \\ 864 & x - 1{,}377 \end{bmatrix} \right|.$$

Taking the determinant, we get $p(x) = (x\text{-}873)(x\text{-}1{,}377) - 864^2 = x^2 - 2{,}250x + 455{,}625$. Factoring, we find the eigenvalues of AA^T to be $\lambda_1 = 2025$ and $\lambda_2 = 225$, making the singular values of A $\sigma_1 = 45$ and $\sigma_2 = 15$. The right singular vectors, and thus the V matrix, can be constructed by first finding the eigenvectors of A^TA. First, we can find A^TA as

$$A^TA = \begin{bmatrix} 18 & -24 \\ -12 & -9 \\ -18 & 24 \\ -9 & 12 \end{bmatrix} \begin{bmatrix} 18 & -12 & -18 & -9 \\ -24 & -9 & 24 & 12 \end{bmatrix} = \begin{bmatrix} 900 & 0 & -900 & -450 \\ 0 & 225 & 0 & 0 \\ -900 & 0 & 900 & 450 \\ -450 & 0 & 450 & 225 \end{bmatrix}.$$

Substituting $\lambda_1 = 2025$ into the homogeneous system $(2025I_4 - A^TA)X = 0$, where $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$, gives the augmented matrix

$$[2025I_4 - A^TA|0] = \left[\begin{array}{cccc|c} 1{,}125 & 0 & 900 & 450 & 0 \\ 0 & 1{,}800 & 0 & 0 & 0 \\ 900 & 0 & 1{,}125 & -450 & 0 \\ 450 & 0 & -450 & 1{,}800 & 0 \end{array}\right],$$

which row reduces to $\left[\begin{array}{cccc|c} 1 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$. So, $x_1 = -2x_4$, $x_2 = 0$, $x_3 = 2x_4$ and $x_4 = x_4$. Thus,

$X = x_4\begin{bmatrix} -2 \\ 0 \\ 2 \\ 1 \end{bmatrix}$, and if we let $x_4 = 1$, then an eigenvector corresponding to $\lambda_1 = 2025$ for A^TA is

$X_1 = \begin{bmatrix} -2 \\ 0 \\ 2 \\ 1 \end{bmatrix}$. Normalizing this vector, we get $v_1 = \begin{bmatrix} -\frac{2}{3} \\ 0 \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$, which is a right singular vector for A. We can repeat this process using $\lambda_2 = 225$, which produces the right singular vector $v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$.

Because $m = 2$ and $n = 4$, the difference in dimension produces two more zero eigenvalues for A^TA than AA^T has. If we substitute $\lambda_3 = \lambda_4 = 0$ into the homogeneous system, the augmented matrix is

$$[0I_4 - A^TA|0] = \left[\begin{array}{cccc|c} -900 & 0 & 900 & 450 & 0 \\ 0 & -225 & 0 & 0 & 0 \\ 900 & 0 & -900 & -450 & 0 \\ 450 & 0 & -450 & -225 & 0 \end{array}\right].$$

This matrix row reduces to $\left[\begin{array}{cccc|c} 1 & 0 & -1 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$, therefore $x_1 = x_3 + \frac{1}{2}x_4$, $x_2 = 0$, $x_3 = x_3$, and $x_4 = x_4$. So, $X = x_3\begin{bmatrix}1\\0\\1\\0\end{bmatrix} + x_4\begin{bmatrix}\frac{1}{2}\\0\\0\\1\end{bmatrix}$, and if we let $x_3 = 1$ and $x_4 = 2$, we get the eigenvectors corresponding to the eigenvectors $\lambda_3 = \lambda_4 = 0$ as $X_3 = \begin{bmatrix}1\\0\\1\\0\end{bmatrix}$ and $X_4 = \begin{bmatrix}1\\0\\0\\2\end{bmatrix}$. However, these eigenvectors are not orthogonal to each other, and we must therefore use the Gram-Schmidt process to create the orthogonal basis. The first three vectors are orthogonal to one another, and if we normalize X_3 in the same way we normalized the first two, we get $v_3 = \begin{bmatrix}\frac{1}{\sqrt{2}}\\0\\\frac{1}{\sqrt{2}}\\0\end{bmatrix}$. If we substitute these orthonormal vectors into the Gram-Schmidt formula

$$w_4 = x_4 - (x_4^T v_1)v_1 - (x_4^T v_2)v_2 - (x_4^T v_3)v_3,$$

we get

$$w_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \end{bmatrix} - 0 - 0 - \frac{1}{\sqrt{2}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \\ 2 \end{bmatrix}.$$

We can scale this matrix by 2, such that $w_4 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 4 \end{bmatrix}$. Finally, w_4 can be normalized such that the unit vector is $v_4 = \begin{bmatrix} \frac{1}{3\sqrt{2}} \\ 0 \\ -\frac{1}{3\sqrt{2}} \\ \frac{4}{3\sqrt{2}} \end{bmatrix}$. This vector is orthogonal to the other three and is therefore the final right singular vector. Therefore, $V^T = \begin{bmatrix} -\frac{2}{3} & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 1 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{3\sqrt{2}} & 0 & -\frac{1}{3\sqrt{2}} & \frac{4}{3\sqrt{2}} \end{bmatrix}$. Left singular vectors (columns of *U*) can be determined using the formula $u_i = \frac{1}{\sigma_i} A v_i$. The first column, u_1, can be found using σ_1 = 45 and $v_1 = \begin{bmatrix} -\frac{2}{3} \\ 0 \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$, so

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{5} \begin{bmatrix} 18 & -12 & -18 & -9 \\ -24 & -9 & 24 & 12 \end{bmatrix} \begin{bmatrix} -\frac{2}{3} \\ 0 \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix},$$

and $u_1 = \begin{bmatrix} -\frac{3}{5} \\ \frac{4}{5} \end{bmatrix}$. Using σ_1 = 15 and $v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, we find that $u_2 = \begin{bmatrix} -\frac{4}{5} \\ -\frac{3}{5} \end{bmatrix}$.

Therefore,

$$U = \begin{bmatrix} -\frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix}.$$

The final SVD of matrix A is

$$A = \begin{bmatrix} -\frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix} \begin{bmatrix} 45 & 0 & 0 & 0 \\ 0 & 15 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{2}{3} & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 1 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{3\sqrt{2}} & 0 & -\frac{1}{3\sqrt{2}} & \frac{4}{3\sqrt{2}} \end{bmatrix}.$$

Example 6

For this example, we will use matrix $A = \begin{bmatrix} 96 & 80 & 90 \\ 120 & 100 & -69 \\ -24 & 156 & 137 \\ 24 & -68 & -27 \\ 24 & -68 & -27 \end{bmatrix}$. This time, we start by using matrix

$$A^TA = \begin{bmatrix} 96 & 120 & -24 & 24 & 24 \\ 80 & 100 & 156 & -68 & -68 \\ 90 & -69 & 137 & -27 & -27 \end{bmatrix} \begin{bmatrix} 96 & 80 & 90 \\ 120 & 100 & -69 \\ -24 & 156 & 137 \\ 24 & -68 & -27 \\ 24 & -68 & -27 \end{bmatrix} = \begin{bmatrix} 25{,}344 & 12{,}672 & -4{,}224 \\ 12{,}672 & 49{,}984 & 25{,}344 \\ -4{,}224 & 25{,}344 & 33{,}088 \end{bmatrix}$$

to find the singular values. So, $p(x) = |xI_3 - A^TA|$, or

$$p(x) = \left| \begin{bmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{bmatrix} - \begin{bmatrix} 25{,}344 & 12{,}672 & -4{,}224 \\ 12{,}672 & 49{,}984 & 25{,}344 \\ -4{,}224 & 25{,}344 & 33{,}088 \end{bmatrix} \right| =$$

$$\left| \begin{bmatrix} x - 25{,}344 & -12{,}672 & 4{,}224 \\ -12{,}672 & x - 49{,}984 & -25{,}344 \\ 4{,}224 & -25{,}344 & x - 33{,}088 \end{bmatrix} \right|,$$

and taking the determinant yields $p(x) = x^3 - 108{,}416x^2 + 2{,}938{,}507{,}264x - 16{,}718{,}547{,}124{,}224$.

Therefore, after factoring, we get eigenvalues $\lambda_1 = 69{,}696$, $\lambda_2 = 30{,}976$, and $\lambda_3 = 7{,}744$. So, the singular values of matrix A are $\sigma_1 = 264$, $\sigma_2 = 176$, and $\sigma_3 = 88$. Next, right singular vectors of A and (the columns of matrix V) will be found. Substituting $\lambda_1 = 69{,}696$ into the homogeneous system $(69{,}696I_3 - A^TA)X = 0$, where $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, we get the augmented matrix

$$[69{,}696I_3 - A^TA\,|0] = \left[\begin{array}{ccc|c} 44{,}352 & -12{,}672 & 4{,}224 & 0 \\ -12{,}672 & 19{,}712 & -25{,}344 & 0 \\ 4224 & -25{,}344 & 36{,}608 & 0 \end{array}\right],$$

which row reduces to $\left[\begin{array}{ccc|c} 1 & 0 & -\frac{1}{3} & 0 \\ 0 & 1 & -\frac{3}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$. So, $x_1 = \frac{1}{3}x_3$, $x_2 = \frac{3}{2}x_3$, and $x_3 = x_3$. Therefore, $X = x_3\begin{bmatrix} \frac{1}{3} \\ \frac{3}{2} \\ 1 \end{bmatrix}$,

and if we let $x_3 = 6$, then $X_1 = \begin{bmatrix} 2 \\ 9 \\ 6 \end{bmatrix}$. Normalized, we get a right singular vector $v_1 = \begin{bmatrix} \frac{2}{11} \\ \frac{9}{11} \\ \frac{6}{11} \end{bmatrix}$. Repeating this process for the eigenvalues $\lambda_2 = 30{,}976$ and $\lambda_3 = 7{,}744$, we get

$$V = \begin{bmatrix} \frac{2}{11} & -\frac{9}{11} & \frac{6}{11} \\ \frac{9}{11} & -\frac{2}{11} & -\frac{6}{11} \\ \frac{6}{11} & \frac{6}{11} & \frac{7}{11} \end{bmatrix} \text{ and } V^T = \begin{bmatrix} \frac{2}{11} & \frac{9}{11} & \frac{6}{11} \\ -\frac{9}{11} & -\frac{2}{11} & \frac{6}{11} \\ \frac{6}{11} & -\frac{6}{11} & \frac{7}{11} \end{bmatrix}.$$

Next, we can use the formula $u_i = \frac{1}{\sigma_i}Av_i$ to find the columns of U. For u_1, $\sigma_1 = 264$ and $v_1 = \begin{bmatrix} \frac{2}{11} \\ \frac{9}{11} \\ \frac{6}{11} \end{bmatrix}$, so the formula becomes

$$u_1 = \frac{1}{264}\begin{bmatrix} 96 & 80 & 90 \\ 120 & 100 & -69 \\ -24 & 156 & 137 \\ 24 & -68 & -27 \\ 24 & -68 & -27 \end{bmatrix}\begin{bmatrix} \frac{2}{11} \\ \frac{9}{11} \\ \frac{6}{11} \end{bmatrix},$$

So $u_1 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \\ \frac{3}{4} \\ -\frac{1}{4} \\ -\frac{1}{4} \end{bmatrix}$. This equation can also be used to find u_2 and u_3, which are $u_2 = \begin{bmatrix} -\frac{2}{8} \\ -\frac{7}{8} \\ \frac{3}{8} \\ -\frac{1}{8} \\ -\frac{1}{8} \end{bmatrix}$ and

$u_3 = \begin{bmatrix} \frac{6}{8} \\ -\frac{3}{8} \\ -\frac{1}{8} \\ \frac{3}{8} \\ \frac{3}{8} \end{bmatrix}$. Since there are two more rows in this U matrix than there are in the V matrix, there are two more singular values associated with it. Both of the additional singular values are equal to zero, so we can say $\sigma_4 = \sigma_5 = 0$. Left singular vectors associated with these additional singular values must be found by computing the fundamental eigenvectors of AA^T associated with the zero eigenvalues. Substituting $\lambda_4 = \lambda_5 = 0$ into the homogeneous system produces the augmented matrix

$$[0I_3 - AA^T|0] = \left[\begin{array}{ccccc|c} -23{,}716 & -13{,}310 & -22{,}506 & 5{,}566 & 5{,}566 & 0 \\ -13{,}310 & -29{,}161 & -3{,}267 & 2{,}057 & 2{,}057 & 0 \\ -22{,}506 & -3{,}267 & -43{,}681 & 14{,}883 & 14{,}883 & 0 \\ 5{,}566 & 2{,}057 & 14{,}883 & -5{,}929 & -5{,}929 & 0 \\ 5{,}566 & 2{,}057 & 14{,}883 & -5{,}929 & -5{,}929 & 0 \end{array}\right].$$

This matrix row reduces to $\left[\begin{array}{ccccc|c} 1 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 1 & 0 & -\frac{1}{6} & -\frac{1}{6} & 0 \\ 0 & 0 & 1 & -\frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$, which represents the linear system

$$1x_1 + 0x_2 + 0x_3 + \frac{1}{3}x_4 + \frac{1}{3}x_5 = 0$$
$$0x_1 + 1x_2 + 0x_3 - \frac{1}{6}x_4 - \frac{1}{6}x_5 = 0$$
$$0x_1 + 0x_2 + 1x_3 - \frac{1}{2}x_4 - \frac{1}{2}x_5 = 0$$
$$0x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 = 0$$
$$0x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 = 0$$

There are two free variables, x_4 and x_5. Therefore, $x_1 = -\frac{1}{3}x_4 - \frac{1}{3}x_5$, $x_2 = \frac{1}{6}x_4 + \frac{1}{6}x_5$, $x_3 = \frac{1}{2}x_4 + \frac{1}{2}x_5$, $x_4 = x_4$, and $x_5 = x_5$. Letting $x_4 = 6$ and $x_5 = 0$, gives a fundamental eigenvector $X_4 = \begin{bmatrix} -2 \\ 1 \\ 3 \\ 6 \\ 0 \end{bmatrix}$. Normalizing gives unit vector $u_4 = \begin{bmatrix} -\frac{2}{\sqrt{50}} \\ \frac{1}{\sqrt{50}} \\ \frac{3}{\sqrt{50}} \\ \frac{6}{\sqrt{50}} \\ 0 \end{bmatrix}$. For the final fundamental eigenvector, we can let $x_4 = 0$ and $x_5 = 6$, which produces the vector $X_5 = \begin{bmatrix} -2 \\ 1 \\ 3 \\ 0 \\ 6 \end{bmatrix}$. All vectors $u_1, u_2, u_3,$ and u_4 are orthogonal to one another. However, u_5 is not orthogonal to u_4. Therefore, we must use the Gram-Schmidt process to produce an orthogonal vector u_5 to form the orthonormal basis for matrix *U*. Using the formula, we get

$$w_5 = x_5 - (x_5^T u_1)u_1 - (x_5^T u_2)u_2 - (x_5^T u_3)u_3 - (x_5^T u_4)u_4, \text{ or}$$

$$w_5 = \begin{bmatrix} -2 \\ 1 \\ 3 \\ 0 \\ 6 \end{bmatrix} - 0 - 0 - 0 - \frac{14}{\sqrt{50}} \begin{bmatrix} -\frac{2}{\sqrt{50}} \\ \frac{1}{\sqrt{50}} \\ \frac{3}{\sqrt{50}} \\ \frac{6}{\sqrt{50}} \\ 0 \end{bmatrix} = \frac{1}{25} \begin{bmatrix} -36 \\ 18 \\ 54 \\ -42 \\ 150 \end{bmatrix}.$$

Normalizing this vector produces a final left singular vector, $u_5 = \begin{bmatrix} -\frac{6}{\sqrt{800}} \\ \frac{3}{\sqrt{800}} \\ \frac{9}{\sqrt{800}} \\ -\frac{7}{\sqrt{800}} \\ \frac{25}{\sqrt{800}} \end{bmatrix}$. Therefore,

$$U = \begin{bmatrix} \frac{1}{2} & -\frac{2}{8} & \frac{6}{8} & -\frac{2}{\sqrt{50}} & -\frac{6}{\sqrt{800}} \\ \frac{1}{4} & -\frac{7}{8} & -\frac{3}{8} & \frac{1}{\sqrt{50}} & \frac{3}{\sqrt{800}} \\ \frac{3}{4} & \frac{3}{8} & -\frac{1}{8} & \frac{3}{\sqrt{50}} & \frac{9}{\sqrt{800}} \\ -\frac{1}{4} & -\frac{1}{8} & \frac{3}{8} & \frac{6}{\sqrt{50}} & -\frac{7}{\sqrt{800}} \\ -\frac{1}{4} & -\frac{1}{8} & \frac{3}{8} & 0 & \frac{25}{\sqrt{800}} \end{bmatrix},$$

and the final SVD of matrix A is

$$A = \begin{bmatrix} \frac{1}{2} & -\frac{2}{8} & \frac{6}{8} & -\frac{2}{\sqrt{50}} & -\frac{6}{\sqrt{800}} \\ \frac{1}{4} & -\frac{7}{8} & -\frac{3}{8} & \frac{1}{\sqrt{50}} & \frac{3}{\sqrt{800}} \\ \frac{3}{4} & \frac{3}{8} & -\frac{1}{8} & \frac{3}{\sqrt{50}} & \frac{9}{\sqrt{800}} \\ -\frac{1}{4} & -\frac{1}{8} & \frac{3}{8} & \frac{6}{\sqrt{50}} & -\frac{7}{\sqrt{800}} \\ -\frac{1}{4} & -\frac{1}{8} & \frac{3}{8} & 0 & \frac{25}{\sqrt{800}} \end{bmatrix} \begin{bmatrix} 264 & 0 & 0 \\ 0 & 176 & 0 \\ 0 & 0 & 88 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{11} & \frac{9}{11} & \frac{6}{11} \\ -\frac{9}{11} & -\frac{2}{11} & \frac{6}{11} \\ \frac{6}{11} & -\frac{6}{11} & \frac{7}{11} \end{bmatrix}.$$

The previous two examples have been engineered to give answers that are easier to handle using a method that will be discussed later. However, an example will now be shown where a random matrix was chosen, resulting in less guarantee for relatively easy computations. A potential problem that arises when finding the SVD for any matrix is round-off error when solving the homogeneous system. If the exact eigenvalues are not used, such as when rounded, it can result in the trivial solution, since any slight change in the values causes the system to become nonsingular. An example will now be shown to display the possible issues that may arise with round-off error when computing the SVD for a matrix in which random integers were chosen as the data entries.

Example 7

Given the 2×3 matrix

$$A = \begin{bmatrix} 2 & -1 & 5 \\ 6 & -4 & -4 \end{bmatrix},$$

we will now determine the SVD. To begin, we determine that

$$AA^T = \begin{bmatrix} 2 & -1 & 5 \\ 6 & -4 & -4 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ -1 & -4 \\ 5 & -4 \end{bmatrix} = \begin{bmatrix} 30 & -4 \\ -4 & 68 \end{bmatrix}.$$

The characteristic polynomial is $p(x) = |\, xI_2 - AA^T|$ or

$$p(x) = \left| \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} - \begin{bmatrix} 30 & -4 \\ -4 & 68 \end{bmatrix} \right| = \left| \begin{bmatrix} x - 30 & 4 \\ 4 & x - 68 \end{bmatrix} \right|.$$

Taking the determinant gives $p(x) = x^2 - 98x + 2024$, the roots of which are $x = 49 + \sqrt{377}$ and $x = 49 - \sqrt{377}$. Therefore, the eigenvalues for AA^T are $\lambda_1 \approx 68.42$ and $\lambda_2 \approx 29.58$, and the singular values for A are $\sigma_1 \approx 8.27$ and $\sigma_2 \approx 5.44$. Next, we want to find right singular vectors by finding the eigenvectors for

$$A^TA = \begin{bmatrix} 2 & 6 \\ -1 & -4 \\ 5 & -4 \end{bmatrix} \begin{bmatrix} 2 & -1 & 5 \\ 6 & -4 & -4 \end{bmatrix} = \begin{bmatrix} 40 & -26 & -14 \\ -26 & 17 & 11 \\ -14 & 11 & 41 \end{bmatrix}.$$

Here, it is important to be as exact as possible when solving the homogeneous system. Solving it by hand can lead to round off error, which can cause the matrix to be nonsingular and, in turn, the only solution to be the trivial solution. Using a TI-84 graphing calculator, we are able to plug in the exact (or at least as exact as the calculator's precision will allow) eigenvalues ($\lambda_1 = 49 + \sqrt{377}$ and $\lambda_2 = 49 - \sqrt{377}$) for the matrix into the homogeneous system. The eigenvalues for this matrix also include $\lambda_3 = 0$. By substituting the exact values for λ_1 and λ_2, as well as λ_3, into the homogeneous system individually, we find right singular vectors

$$v_1 \approx \begin{bmatrix} -0.696 \\ 0.468 \\ 0.543 \end{bmatrix}, v_2 \approx \begin{bmatrix} 0.480 \\ -0.259 \\ 0.838 \end{bmatrix}, \text{and } v_3 \approx \begin{bmatrix} -0.533 \\ -0.845 \\ 0.044 \end{bmatrix}.$$

Therefore,

$$V^T = \begin{bmatrix} -0.696 & 0.468 & 0.543 \\ 0.480 & -0.259 & 0.838 \\ -0.533 & -0.845 & 0.044 \end{bmatrix}.$$

We can now construct the *U* matrix using the formula $u_i = \frac{1}{\sigma_i} A v_i$. For σ_1 = 8.27, we get

$$u_1 = \frac{1}{8.27}\begin{bmatrix} 2 & -1 & 5 \\ 6 & -4 & -4 \end{bmatrix}\begin{bmatrix} -0.696 \\ 0.468 \\ 0.543 \end{bmatrix} = \begin{bmatrix} 0.103 \\ -0.994 \end{bmatrix}$$

and for $\sigma_2 \approx 5.44$, we get

$$u_2 = \frac{1}{5.44}\begin{bmatrix} 2 & -1 & 5 \\ 6 & -4 & -4 \end{bmatrix}\begin{bmatrix} 0.480 \\ -0.259 \\ 0.838 \end{bmatrix} = \begin{bmatrix} 0.994 \\ 0.103 \end{bmatrix}.$$

Thus, the *U* matrix is

$$U = \begin{bmatrix} 0.103 & 0.994 \\ -0.994 & 0.103 \end{bmatrix}$$

and the complete SVD is

$$U\Sigma V^T = \begin{bmatrix} 0.103 & 0.994 \\ -0.994 & 0.103 \end{bmatrix}\begin{bmatrix} 8.27 & 0 & 0 \\ 0 & 5.44 & 0 \end{bmatrix}\begin{bmatrix} -0.696 & 0.468 & 0.543 \\ 0.480 & -0.259 & 0.838 \\ -0.533 & -0.845 & 0.044 \end{bmatrix}.$$

If we compute the product of this matrix, which we will call A', we get

$$A' = \begin{bmatrix} 2.00 & -100 & 4.99 \\ 5.99 & -3.99 & -3.99 \end{bmatrix}.$$

Recall that the original matrix *A* is

$$A = \begin{bmatrix} 2 & -1 & 5 \\ 6 & -4 & -4 \end{bmatrix}$$

Therefore, in this case, we were able to construct a matrix using the SVD that is very close to the original matrix. However, round-off still has an effect on the results of the SVD and could potentially produce issues in matrices containing a larger variety of data types.

Looking Back at Diagonalization

The diagonalized form from before with eigenvalues was for A itself and $A = PDP^{-1}$. With singular values, two diagonalizations can be performed, one for A^TA and another for AA^T. By definition, $A = U\Sigma V^T$, so

$$A^TA = (U\Sigma V^T)^T U\Sigma V^T$$

$$A^TA = (V^T)^T \Sigma^T U^T U\Sigma V^T$$

U^TU is the identity, since U is orthogonal and $U^T = U^{-1}$, so

$$A^TA = V\Sigma^T I\Sigma V^T$$

$$A^TA = V\Sigma^T\Sigma V^T$$

Here, $\Sigma^T\Sigma$ is a diagonal matrix of size $n\times n$, which we can call D_n,where the entries are the singular values squared. Therefore, A^TA can be written as $A^TA = VD_nV^T$. We can similarly find the diagonalized form of matrix AA^T:

$$AA^T = U\Sigma V^T(U\Sigma V^T)^T$$

$$AA^T = U\Sigma V^T(V^T)^T\Sigma^T U^T$$

$$AA^T = U\Sigma V^T V\Sigma^T U^T$$

V^TV is the identity, therefore

$$AA^T = U\Sigma I\Sigma^T U^T$$

$$AA^T = U\Sigma\Sigma^T U^T$$

$\Sigma\Sigma^T$ is also a diagonal matrix with diagonal entries equal to the singular values squared. This diagonal matrix is of the size $m\times m$, so we will call it D_m, and $AA^T = UD_mU^T$.

Constructing a Matrix with Specified Singular Values and Size

Just as it is important to know how to decompose a matrix using the Singular Value Decomposition, it is also important to know how to construct a matrix of any size with any desired singular values. Given the singular values and the size of the matrix we wish to construct, we can choose any vector of size $m \times 1$ to generate matrix U and any vector of size $n \times 1$ to generate matrix V. To do this, we can use what is known as the Householder Transformation, H. This transformation is defined by the formula $H = I - \frac{2}{X^T X} XX^T$. A matrix generated using the Householder Transformation is symmetric and orthogonal. Matrices U and V in the SVD are required to be orthogonal, therefore this is a useful tool for generating these matrices when given the singular values to construct matrix A, rather than computing the decomposition A as previously discussed (see Appendix Result 2 for orthogonality of the Householder Transformation). This process is best shown through an example, which we will call Example 8.

Example 8

To begin, if we would like to have a matrix of size 3×2 with singular values, $\sigma_1 = 6$ and $\sigma_2 = 2$, we can first construct the Σ matrix, as it will be the same size as A with main diagonal entries equal to the singular values, therefore

$$\Sigma = \begin{bmatrix} 6 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}.$$

Next, we will generate the U matrix, which we know must be 3×3. So, we can take any vector of size 3×1 in order to generate the matrix. If we choose the vector $X = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, then

$$X^T X = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 6, \qquad XX^T = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix},$$

which produces the matrix *U*

$$U = I_3 - \frac{2}{6}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} = I_3 - \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{4}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ -\frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \end{bmatrix}.$$

This same process can be used to find the 2×2 *V* matrix. If the chosen 2×1 vector is $X = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$,

then $X^T X = \begin{bmatrix} 3 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = 13$ and $XX^T = \begin{bmatrix} 3 \\ -2 \end{bmatrix}\begin{bmatrix} 3 & -2 \end{bmatrix} = \begin{bmatrix} 9 & -6 \\ -6 & 4 \end{bmatrix}$. Therefore, the equation

for the *V* matrix is

$$V = I_2 - \frac{2}{13}\begin{bmatrix} 9 & -6 \\ -6 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{18}{13} & -\frac{12}{13} \\ -\frac{12}{13} & \frac{8}{13} \end{bmatrix} = \begin{bmatrix} -\frac{5}{13} & \frac{12}{13} \\ \frac{12}{13} & \frac{5}{13} \end{bmatrix}.$$

In this case, $V^T = V$. So, the completed decomposed matrix is

$$A = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & -\frac{1}{3} \\ -\frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \end{bmatrix}\begin{bmatrix} 6 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} -\frac{5}{13} & \frac{12}{13} \\ \frac{12}{13} & \frac{5}{13} \end{bmatrix}.$$

Multiplied out, we get that

$$A = \begin{bmatrix} -\frac{36}{13} & \frac{124}{39} \\ \frac{12}{13} & -\frac{154}{39} \\ -\frac{6}{13} & -\frac{92}{39} \end{bmatrix}.$$

As shown, it is possible to construct any matrix of desired size and with desired singular values using the Householder matrix to build the SVD in reverse.

Applications of Singular Values

Application 1: Image Compression

An application of the SVD is in image compression. Any image on a computer is comprised of small red, blue, and green pixels arranged in a grid-like pattern. Due to this pattern, an image can be represented as data in a matrix. Each pixel is a value in the matrix and has a range of possible values depending on the desired color. For example, to make the color gray, a pixel can range in value from 0 to 255, depending on the shade. To store a grayscale image with dimensions 100×100 pixels, it would require storage of 10,000 different values. This number increases with the size of the image and for colored images. Computing the SVD of the matrix for an image and calculating a level of precision can help with saving storage. In the SVD, the data is arranged so that the most important data is stored close at the top. The most important data about the rows of matrix A is stored in the first column of matrix U, while the most important information about the columns of matrix A is stored in the first row of V^T. For matrix Σ, the most important data is located along the main diagonal, while the rest of the matrix is composed of zeros. This means that a close approximation of the original matrix A can be computed using the first components of the matrices comprising the SVD. The level of precision can also be called the mode of precision and refers to the number of columns of matrix U and rows of matrix V^T that are chosen to approximate the original matrix. For example, if we have a matrix with dimensions $m \times n$ and want a precision of 1 mode, we can multiply the first column of U by the first row of V^T as follows: $U'\Sigma'V^{T'} = [(m\times1)(1\times1)(1\times n)] = (m\times n)$. Therefore, the resulting image is the same size, but less data has been used. If we were to compress the 100×100 grayscale image from before with a precision of 10 modes, the SVD matrices would have the dimensions $U' = (100\times10)$, $\Sigma' = (10\times10)$, and $V^{T'} = (10\times100)$. So, there are now only 2,100 pixel data points to store rather than the original 10,000 (Mathews, 2014). Figure 1 shows an example of an image reconstructed using the SVD and various modes of precision. The more modes of precision used, the better quality, however because the most important data is stored near the top of the SVD matrices, relatively few modes are needed to maintain image quality. Therefore, the SVD can help in storage management of images by decreasing the amount of data required to store the image.

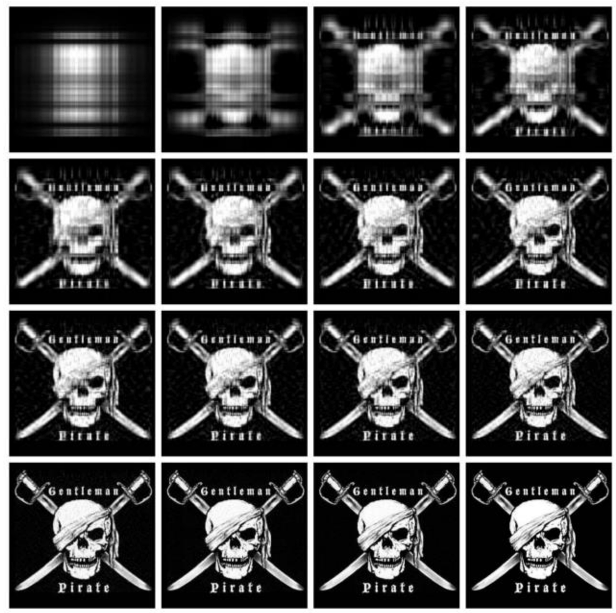

Figure 1. Image of size 250×236. Row 1 contains images constructed using modes 1, 2, 4 and 6 from left to right. Row 2 uses modes 8, 10, 12, and 14. Row 3 uses modes 16, 18, 20, and 25. Finally, row 4 uses modes 50, 75, and 100, and the final image is the original image (Mathews, 2014).

<u>Application 2: Filtering Noisy Signals</u>

Similarly, the SVD can be used for filtering noisy data and retain the relevant information. Say a matrix A of size $m \times n$ is representative of digital data from a video or audio sample and it is contaminated with noise. The SVD breaks the data in A down into r mutually orthogonal components as follows:

$$A = U\begin{bmatrix} D_{r\times r} & 0 \\ 0 & 0 \end{bmatrix}V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T = \sum_{i=1}^{r} \sigma_i Z_i$$

Where $Z_i = u_i v_i^T$ and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$. The noise contamination in A is random and thus the noise is distributed approximately uniformly across each Z_i matrix. Therefore, it is expected that each $\sigma_i Z_i$ term contains approximately the same noise level. If we denote the signal-to-noise ratio in $\sigma_i Z_i$ as SNR($\sigma_i Z_i$), then, in general

$$\mathrm{SNR}(\sigma_1 Z_1) \geq \mathrm{SNR}(\sigma_2 Z_2) \geq \ldots \geq \mathrm{SNR}(\sigma_r Z_r).$$

If a group of the singular values, for instance $\sigma_{k+1}, \ldots \sigma_r$, are small in relation to the (total noise)/r, then the SNR($\sigma_{k+1} Z_{k+1}$), …, SNR($\sigma_r Z_r$) are small. The removal of these terms ($\sigma_{k+1} Z_{k+1}$ through $\sigma_r Z_r$) would result in a small loss in total signal, but disproportionately large loss of total noise in A. Therefore, some noise can be filtered out without a significant loss of information regarding the signal in A (Meyer, 2000).

Application 3: Linear Least Squares Regression

The SVD can also be applied to solving linear least squares problems. Linear least squares regression is the procedure for finding the best-fitting curve for a given set of 2-variable data.

Example 9

Say we have the data plotted in Figure 2. The given x values are $x = \{1, 2, 3\}$ and the given y values are $y = \{11, 19, 31\}$. Using the linear least squares equation, $y = a + bx$, this means we have the linear system of equations

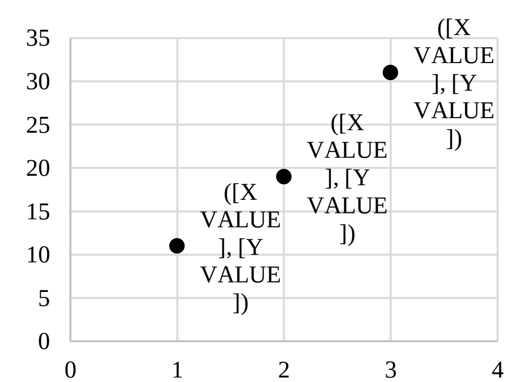

Figure 2. Data points for Example 9.

$$a + b(1) = 11$$
$$a + b(2) = 19$$
$$a + b(3) = 31$$

We can convert this linear system of equations into the form:

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 11 \\ 19 \\ 31 \end{bmatrix},$$

Where $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$, $X = \begin{bmatrix} a \\ b \end{bmatrix}$, and $B = \begin{bmatrix} 11 \\ 19 \\ 31 \end{bmatrix}$. The question becomes, can we solve the linear system $AX = B$ for vector X ? Since there are more equations than unknowns (a and b), it is unlikely that the data is exactly linear, and therefore unlikely the equation can be solved. However, it is possible to find an X such that the square of the error between AX and B is minimized.

> Linear Least Squares and SVD: Let A be an $m \times 2$ matrix, X be a 2×1 vector of unknowns, and B an $m \times 1$ vector of constants. We want to minimize $\|AX - B\|^2$ over all possible vectors X.

Suppose $A = U\Sigma V^T$ is the SVD for A. In this case, U is $m \times m$, Σ is $m \times 2$, and V is 2×2.

(Note: for an orthogonal matrix M and vector Y, $\|Y\|^2 = \|MY\|^2$ since

$\|MY\|^2 = (MY)^T MY = Y^T M^T MY = Y^T Y = \|Y\|^2$)

Then,

$$\|AX - B\|^2 = \|U^T(AX - B)\|^2$$

$$= \|U^T AX - U^T B\|^2$$

$$= \|U^T U \Sigma V^T X - U^T B\|^2$$

$$= \|\Sigma V^T X - U^T B\|^2$$

Note, $V^T X = \begin{bmatrix} v_1^T X \\ v_2^T X \end{bmatrix}$, where v_i is the ith column of V and $\Sigma V^T X$ is the $m \times 1$ vector with $\sigma_1 v_1^T X$ and $\sigma_2 v_2^T X$ as the first two entries and 0 for entries 3, 4, …, m. $U^T B$ is the $m \times 1$ vector with entries $u_i^T B$, where u_i is the ith column of u. Therefore,

$$\|AX - B\|^2 = \|\Sigma V^T X - U^T B\|^2$$

$$= \left\| \begin{bmatrix} \sigma_1 v_1^T X \\ \sigma_2 v_2^T X \\ 0 \\ \vdots \\ 0 \end{bmatrix} - \begin{bmatrix} u_1^T B \\ u_2^T B \\ \vdots \\ u_m^T B \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} \sigma_1 v_1^T X - u_1^T B \\ \sigma_2 v_2^T X - u_2^T B \\ -u_3^T B \\ \vdots \\ -u_m^T B \end{bmatrix} \right\|^2$$

$$= (\sigma_1 v_1^T X - u_1^T B)^2 + (\sigma_2 v_2^T X - u_2^T B)^2 + \sum_{i=3}^{m} (u_i^T B)^2$$

$$= \sum_{i=1}^{2} (\sigma_i v_i^T X - u_i^T B)^2 + \sum_{i=3}^{m} (u_i^T B)^2$$

Therefore, to minimize, it is necessary that $\sigma_i v_i^T X - u_i^T B = 0$ for *i* = 1, 2, or $v_i^T X = \frac{u_i^T B}{\sigma_i}$ for *i* = 1, 2 (Assume $\sigma_1 \geq \sigma_2 > 0$). If we define the *m* × 1 vector *Z* as $z_i = \frac{u_i^T B}{\sigma_i}$ for *i* = 1, 2, and 0 for all other entries, then all solutions, X, to this minimization problem must satisfy $z = V^T X$. Then,

$$Vz = VV^T X = X$$

is the solution, where *X* is the 2 × 1 vector given by the sum $X = \frac{u_1^T B}{\sigma_1} v_1 + \frac{u_2^T B}{\sigma_2} v_2$. Returning to Example 9, the approximate SVD of *A*, rounded to five decimal places, is

$$U\Sigma V^T = \begin{bmatrix} 0.32311 & -0.85379 & 0.40825 \\ 0.54751 & -0.18323 & -0.81650 \\ 0.77191 & 0.48732 & 0.40825 \end{bmatrix} \begin{bmatrix} 4.07914 & 0 \\ 0 & 0.60049 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0.40266 & 0.91535 \\ -0.91535 & 0.40266 \end{bmatrix}$$

We can use this information to determine the values of *a* and *b* in vector *X*. The singular values of matrix *A* are $\sigma_1 = 4.07914$ and $\sigma_2 = 0.60049$. We can transpose the first two columns of the *U* matrix, separately, to obtain the vectors u_1^T and u_2^T, and v_1 and v_2 are the columns of matrix *V*, which is $V = \begin{bmatrix} 0.40266 & -0.91535 \\ 0.91535 & 0.40266 \end{bmatrix}$. Therefore, using the formula $X = \frac{u_1^T B}{\sigma_1} v_1 + \frac{u_2^T B}{\sigma_2} v_2$, we get the vector

$$X = \begin{bmatrix} 0.33465 \\ 9.99948 \end{bmatrix}$$

This is close to the actual values of a and b for the line of best fit shown in Figure 3, which are $a = \frac{1}{3}$ and $b = 10$. This is an intensive method for solving least linear squares problems but can be useful for larger amounts of data.

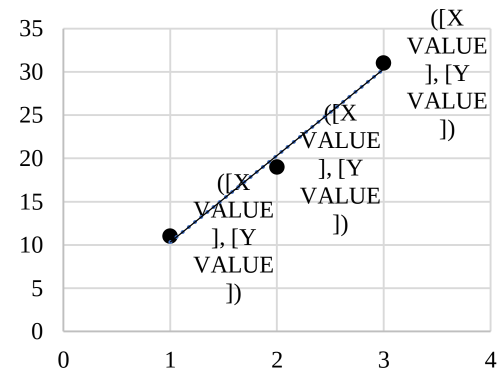

Figure 3. Line of best fit for Example 9 data.

Conclusion

Singular values of a matrix and the Singular Value Decomposition are great tools that generalize the ideas of eigenvalues and diagonalization of a matrix to any matrix of any size. By decomposing a matrix into the product three separate matrices, it is easier to examine the information a matrix holds and manipulate the data as needed. It is also useful to be able to construct a matrix with known singular values using the Householder Transformation. Finally, there are many real-world applications for the SVD, one of which is image compression. By storing a smaller portion of the image data from the decomposed matrix, rather than the entire image, storage space can be saved while much of the image's quality is retained. Another application is in noise filtering, where noisy data is filtered, while relevant data is retained. The final application discussed was the use of the SVD to approximate the linear least squares regression for a given set of data. The SVD is a useful process for decomposing matrices of any size containing any kind of data, allowing one to study, manipulate, and understand a matrix and its components in more depth, in contrast to diagonalization.

Upon entering this project, the main focus was on studying and understanding a higher-level mathematics concept that was not presented in any of the courses offered in the undergraduate course of study at North Carolina Wesleyan College, and relating it back to information that was presented. However, through researching this topic, it became apparent that there is very little approachable literature on the topic of singular values and the Singular Value Decomposition. This matrix decomposition can also be quite tedious to do by hand. There are software packages available for this purpose, however none were utilized during the course of this project. It was necessary throughout the course of researching and writing about this topic to create our own procedures for various methods discussed, including the SVD itself and constructing a matrix using the Householder Transformation, much of which came about through trial and error. The major takeaway from this project is that there are often multiple ways to approach a topic in mathematics and it may be necessary condense the information provided within the limited literature to look to for guidance in order to obtain an approachable method for understanding. However, this guarantees a deeper understanding of the topic at hand and of the building blocks used along the way.

References

Andrilli, S., & Hecker, D. (2010). *Elementary Linear Algebra* (4th ed.). Academic Press.

Bretscher, O. (2009). *Linear Algebra with Applications* (4th ed.). Pearson Education, Inc.

Gundersen, G. (2018). Singular Value Decomposition as simply as possible [Blog Post]. Retrieved from http://gregorygundersen.com/blog/2018/12/10/svd/

Huang, C. C. *Lecture 13: Complex Eigenvalues and Factorization*. Personal Collection of C. C. Huang, Wright State University, Dayton OH.

Jerison, D. *Math 2940: Symmetric Matrices have Real Eigenvalues*. Personal Collection of D. Jerison, Cornell University, Ithaca, NY.

Margalit, D., & Rabinoff, J. (2019). *Interactive Linear Algebra.* Georgia Institute of Technology.

Mathews, B. "Image Compression using Singular Value Decomposition (SVD)", 12 December 2014, The University of Utah.

Meyer, C. (2000). *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics.

Stewart, G. (1993). On the early history of the singular value decomposition. *SIAM Review,* 15(3), 5-13.

Appendix

Result 1: Reasoning for Gram-Schmidt Procedure.

To begin the Gram-Schmidt process, we start with the linearly independent set of vectors $x_1, x_2, \dots, x_m \in \mathbb{R}^n$ and normalize vector x_1, forming vector v_1. Note that vector x_2 can be decomposed into $x_2 = \alpha v_1 + w_2$, where αv_1 is a vector lying in the direction of v_1 and w_2 is a vector that is orthogonal to v_1. So, this equation can be solved such that $w_2 = x_2 - \alpha v_1$. The vector αv_1 is the projection of x_2 onto v_1, so it can also be written as $\alpha v_1 = \left(\frac{{x_2}^T v_1}{\|v_1\|^2}\right) v_1 = ({x_2}^T v_1) v_1$. Therefore, we now have the equation $w_2 = x_2 - ({x_2}^T v_1) v_1$, and normalizing w_2 gives the vector v_2. We can continue by decomposing vector x_k into a linear combination of the existing orthonormal vectors $v_1, v_2, \dots, v_{k-1}$ and the vector w_k, which is orthogonal each vector v_i. So, we get

$$x_k = \sum_{i=1}^{k-1} \mathrm{proj}_{v_i} x_k + w_k.$$

The formula for projection is $\mathrm{proj}_a b = \left(\frac{b^T a}{a^T a}\right) a$, so we have

$$\mathrm{proj}_{v_i} x_k = \left(\frac{{x_k}^T v_i}{{v_i}^T v_i}\right) v_i = ({x_k}^T v_i) v_i$$

and, in turn,

$$x_k = \sum_{i=1}^{k-1} ({x_k}^T v_i) v_i + w_k.$$

Solving for w_k yields

$$w_k = x_k - \sum_{i=1}^{k-1} ({x_k}^T v_i) v_i$$

and normalizing gives the vector v_k. Since the original set of vectors $\{x_1, x_2, \dots, x_m\}$ is linearly independent, the process is guaranteed to produce a new set of linearly independent vectors $\{v_1, v_2, \dots, v_m\}$, since linear combinations of linearly independent vectors are uniquely determined by the coefficients. Note also that w_k is orthogonal to each vector v_j for each vector $j = 1, 2, \dots, k-1$ since

$$w_k{}^T v_j = \left(x_k - \sum_{i=1}^{k-1} (x_k{}^T v_i) v_i \right)^T v_j$$

$$= \left(x_k{}^T - \sum_{i=1}^{k-1} (x_k{}^T v_i) v_i{}^T \right) v_j$$

$$= x_k{}^T v_j - \sum_{i=1}^{k-1} (x_k{}^T v_i) v_i{}^T v_j$$

$$= x_k{}^T v_j - \left(x_k{}^T v_j\right) v_j{}^T v_j = x_k{}^T v_j - x_k{}^T v_j = 0.$$

(Recall that the vectors v_i are orthonormal, so, by definition, $v_i{}^T v_i = 1$ and $v_i{}^T v_j = 0$, for $i \neq j$)

Therefore, the Gram-Schmidt method produces an orthonormal set of vectors $v_1, v_2, \dots, v_m \in \mathbb{R}^n$.

Result 2: Householder Matrix Is Symmetric and Orthogonal.

Let $x^T x = 1$ and form the Householder matrix $U = I - 2xx^T$.

1) $(xx^T)^T = (x^T)^T x^T = xx^T$. Therefore, xx^T is symmetric. $-2xx^T$ is also symmetric since a scalar multiple of a symmetric matrix is symmetric. Finally, $I - 2xx^T = I + (-2xx^T)$ is the sum of symmetric matrices, which is also symmetric. Therefore, U is symmetric.

2) $UU^T = U^2$ since U is symmetric. So,

$$UU^T = (I - 2xx^T)(I - 2xx^T)$$

$$UU^T = I - 2xx^T - 2xx^T + 4xx^T xx^T$$

$$UU^T = I - 4xx^T + 4x(x^T x)x^T$$

$$UU^T = I - 4xx^T + 4xx^T$$

$$UU^T = I.$$

Therefore, U is an orthogonal matrix.

Result 3: Reasoning for SVD.

To begin, a pair of orthogonal vectors v_1 and v_2 in two-dimensional space forms a basis for that space. If we apply a matrix transformation A to these vectors, we get vectors Av_1 and Av_2. These vectors can also be written as unit vectors u_1 and u_2 times their respective magnitudes, which we will call σ_1 and σ_2. Thus, these two transformed vectors can now be written as $Av_1 = u_1\sigma_1$ and $Av_2 = u_2\sigma_2$. Any vector x in the space can be expressed using the basis vectors v_1 and v_2 as follows:

$$x = (x^T v_1)v_1 + (x^T v_2)v_2,$$

where $x^T v_1$ represents the dot product between the two vectors and is a scalar value. If we left-multiply both sides by A, we get

$$Ax = (x^T v_1)Av_1 + (x^T v_2)Av_2.$$

Since $Av_i = u_i\sigma_I$ as shown before, we can rewrite the equation as

$$Ax = (x^T v_1)\, u_1\sigma_1 + (x^T v_2)\, u_2\sigma_2.$$

Since the dot product is commutative, and since this is a scalar value, it can be moved to the end of the each expression in which it appears in the equation as follows:

$$Ax = u_1\sigma_1 v_1^{\mathrm{T}} x + u_2\sigma_2\, v_2^{\mathrm{T}} x = (u_1\sigma_1 v_1^{\mathrm{T}} + u_2\sigma_2 v_2^{\mathrm{T}})x.$$

Since x is arbitrary (see Appendix Result 4), we have

$$A = u_1\sigma_1 v_1^{\mathrm{T}} + u_2\sigma_2 v_2^{\mathrm{T}}.$$

Rewritten into matrix form, we get SVD form for all 2×2 matrices:

$$A = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix}$$

This can be generalized for any matrix *A* with dimensions $m \times n$ to give the SVD equation $A = U\Sigma V^T$ (Gundersen, 2018).

Result 4: Matrix Claim.

Let *A* and *B* be $m \times n$ matrices. If $Ax = Bx$ for all $n \times 1$ vectors *x*, then $A = B$.

Proof: Suppose $Ax = Bx$ for all $n \times 1$ vectors *x*, then $Ae_i = Be_i$, where e_i is the *i*th standard basis vector for $\mathbb{R}^n$. Ae_i is the *i*th column of *A* and Be_i is the *i*th column of *B*. Therefore, each column of *A* is equal to the corresponding column of *B*. Therefore $A = B$. ■