Singular Value Decomposition and Applications

Kim Gwang Woo

October 23 2020

1 Introduction

In many mathematical fields, the Singular Value Decomposition (SVD) plays a very important role. Therefore, it is worth studying it. We will establish some preliminaries, prove the theorem and exercises, and then explore where we can apply it.

2 The Related Results

Theorem 2.1. Let T be a linear operator on a finite-dimensional complex (or real) inner product space V. Then T is normal (or self-adjoint) if and only if there exists an orthonormal basis for V consisting of eigenvectors of T.

Our goal starts from Theorem 2.1. But we use it without proving (basically, the proof is very tedious.)

Lemma 2.2. Let $T : V \to W$ be a linear transformation, where V and W are finite dimensional inner product spaces. If we denote the adjoint of T by T^*, then the following are valid.

(a) T^*T and TT^* are positive semidefinite.

(b) $\mathrm{rank}(T^*T) = \mathrm{rank}(TT^*) = \mathrm{rank}(T)$.

Proof. To verify that T^*T is positive semidefinite, we prove that it is self-adjoint and $\langle T^*T(x), x\rangle \geq 0$ for all $x \in V$. Since $(T^*T)^* = T^*T^{**} = T^*T$, the first assertion follows and the second one does from the fact that $\langle T^*T(x), x\rangle = \langle T(x), T(x)\rangle = \|T(x)\|^2 \geq 0$. The other case is similar.

We claim that $T(x) = 0$ if and only if $T^*T(x) = 0$. For if $T^*T(x) = 0$, then $\langle T^*T(x), x\rangle = \langle T(x), T(x)\rangle = \|T(x)\|^2 = 0$ and hence $T(x) = 0$. Because the converse is trivial, the claim is complete. This claim tells us that T and T^*T have the same null space and so do same rank. Using the fact $rank(T) = rank(T^*)$ and applying the preceding argument to T^* and TT^*, we get the remaining equality. ☐

3 Singular Value Theorem

Now we are going to state and prove the *Singular Value Theorem for Linear Transformations* and *Singular Value Decomposition Theorem for Matrices* as a corollary of that.

Theorem 3.1 (**Singular Value Theorem for Linear Transformations**)**.** Let V and W be finite-dimensional inner product spaces, and let $T : V \to W$ be a linear transformation of rank r. Then there exist orthonormal bases $\{v_1, v_2, ..., v_n\}$ for V and $\{w_1, w_2, ..., w_m\}$ for W and positive scalars $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r$ such that

$$T(v_i) = \begin{cases} \sigma_i w_i & (1 \leq i \leq r) \\ 0 & (i > r). \end{cases} \tag{1}$$

Proof. To establish the existance, note that $T^*T : V \to V$ is positive-semidefinite and has rank r (Lemma 2.2), so it follows from Theorem 2.1 that there exists an orthonormal basis $\{v_1, v_2, ..., v_n\}$ for V consisting of eigenvectors of T^*T with corresponding eigenvalues λ_i (here, all of λ_i's are nonnegative and $\lambda_i = 0$ for $i > r$.) Assume $\lambda_1 \geq ... \geq \lambda_r > 0$. For $1 \leq i \leq r$, define $\sigma_i = \sqrt{\lambda_i}$ and $w_i = \dfrac{T(v_i)}{\sigma_i}$. Then $\{w_1, w_2, ..., w_n\}$ is an orthonormal set in W, because

$$\langle w_k, w_j \rangle = \frac{1}{\sigma_k \sigma_j} \langle T(v_k), T(v_j) \rangle = \frac{1}{\sigma_k \sigma_j} \langle T^*T(v_k), v_j \rangle = \frac{\sigma_k^2}{\sigma_k \sigma_j} \langle v_k, v_j \rangle = \delta_{kj}. \quad (2)$$

Thus, we can extend $\{w_1, w_2, ..., w_n\}$ to an orthonormal basis $\{w_1, w_2, ..., w_m\}$. By our constructions, this basis is the required one, since $T^*T(v_i) = 0$ implies $T(v_i) = 0$. □

Theorem 3.2 (Uniqueness of σ_i's). Suppose the conclusion in Theorem 3.1 holds. Then for $1 \leq i \leq n$, v_i is an eigenvector of T^*T with corresponding eigenvalues σ_i^2 for $1 \leq i \leq r$ and 0 if $i > r$. Therefore the scalars $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r$ are uniquely determined by T.

Proof. Since $T^*(w_i)$ is an element of V, it can be written as a linear combination of $\{v_1, w_2, ..., v_n\}$ and so let $T^*(w_i) = \sum \langle T^*(w_i), v_j \rangle v_j$. Then, for $1 \leq i \leq r$, we obtain

$$\begin{aligned} T^*T(v_i) &= \sigma_i T^*(w_i) \quad \text{(by assumption and linearity)} \\ &= \sigma_i \sum \langle T^*(w_i), v_j \rangle v_j \\ &= \sigma_i \sum \langle w_i, T(v_j) \rangle v_j \\ &= \sigma_i \sum \sigma_j \langle w_i, w_j \rangle v_j \\ &= \sigma_i^2 v_i. \end{aligned} \quad (3)$$

If $i > r$, then $T^*T(v_i) = T^*(0) = 0$, as desired. □

Definition 3.3. The unique scalars $\sigma_1, ..., \sigma_r$ in Theorem 3.1 and 3.2 are called the **singular values** of T. If r is less than both m and n, then the term singular value is extended to include $\sigma_{r+1} = ... = \sigma_k = 0$, where k is the minimum of m and n.

Definition 3.4. Let A be an $m \times n$ matrix. We define the singular values of A to be the singular values of the linear transformation L_A, where the map $L_A : F^n \to F^m$ is defined by $L_A(x) = Ax$.

Corollary 3.5 (Singular Value Decomposition Theorem for Matrices). Let A be an $m \times n$ matrix of rank r with the positive singular values $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r$, and let Σ be the $m \times n$ matrix defined by

$$\Sigma_{ij} = \begin{cases} \sigma_i & \text{if } i = j \leq r \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Then there exist an $m \times m$ unitary matrix U and an $n \times n$ unitary matrix V such that $A = U\Sigma V^*$

Proof. Let β' and γ' be the standard bases for F^n and F^m, respectively. Applying Singular Value Theorem to L_A, we get orthonormal bases β and γ for F^n and F^m, respectively. Therefore, using change of a basis, we know that

$$A = [L_A]_{\beta'}^{\gamma'} = [I]_{\gamma}^{\gamma'}[L_A]_{\beta}^{\gamma}[I]_{\beta'}^{\beta}, \tag{5}$$

where I is the identity transformation of F^n or F^m. If we let $U = [I]_{\gamma}^{\gamma'}$ and $V = [I]_{\beta}^{\beta'}$, then the equation (5) is $A = U\Sigma V^{-1}$. Since the column of U and V forms an orthonormal basis, they are unitary and it gives $A = U\Sigma V^*$ □

4 Examples and Exercises

In this section, we will calculate the SVD of a matrix and examine some related exercises.

Example 4.1. We now compute the SVD of $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix}$.

Note that $A^*A = \begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}$, so its eigenvalues are 6 and 0 with eigenvectors $\beta = \{\frac{1}{\sqrt{2}}(1,1), \frac{1}{\sqrt{2}}(1,-1)\}$, respectively. Next, we can extend $L_A(\beta)$ to an orthonormal basis $\gamma = \{\frac{1}{\sqrt{3}}(1,1,-1), \frac{1}{\sqrt{2}}(1,-1,0), \frac{1}{\sqrt{6}}(1,1,2)\}$. Therefore, we have the singular value decomposition of A, $A = U\Sigma V^*$, where

$$U = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sqrt{6} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}. \tag{6}$$

Remark. Of course, the above orthonormal basis is not unique. we can actually pick another orthonormal one and the decomposition is therefore not unique. □

Exercise 4.2. Let T be a normal linear operator on an n-dimensional inner product space with eigenvalues $\lambda_1, ..., \lambda_n$. Prove that the singular values of T are $|\lambda_1|, ..., |\lambda_n|$.

Proof. Because T is normal (that is, $TT^* = T^*T$), $T(x) = \lambda x$ implies $T^*(x) = \bar{\lambda}x$ and thus the eigenvalues of T^*T are $|\lambda_1|^2, ..., |\lambda_n|^2$. It follows that the singular values of T are $|\lambda_1|, ..., |\lambda_n|$. □

Exercise 4.3. Let A be a normal matrix with an orthonormal basis of eigenvectors $\beta = \{v_1, ..., v_n\}$ and corresponding eigenvalues $\lambda_1, ..., \lambda_n$. Let V be the $n \times n$ matrix whose columns are the vectors in β. Prove that for each i there is a scalar θ_i of absolute value 1 such that if U is the $n \times n$ matrix with $\theta_i v_i$ as column i and Σ is the diagonal matrix such that $\Sigma_{ii} = |\lambda_i|$ for each i, then $U\Sigma V^*$ is a singular value decomposition of A.

Proof. Let $\theta_i = \frac{\lambda_i}{|\lambda_i|}$ if $\lambda_i \neq 0$ and $\theta_i = 1$ otherwise. Then it is immediate that $Av_i = \lambda_i v_i = |\lambda_i| \theta_i v_i = |\lambda_i| Ue_i = U\Sigma_i$ for all i. This shows that $AV = U\Sigma$ and so $A = U\Sigma V^*$. Since $\|\theta_i v_i\| = |\theta_i| \|v_i\| = 1$, the column of U forms an orthonormal basis, which menas that it is a unitary matrix. Therefore, $A = U\Sigma V^*$ is the SVD of A □

Exercise 4.4. Prove that if A is a positive semidefinite matrix, then the singular values of A are the same as the eigenvalues of A.

Proof. Because A is positive semidefinite, it is self-adjoint(hence, normal) and all of its eigenvalues are non-negative. From this and Exercise 4.2, the result follows directly. □

Exercise 4.5. Prove that if A is a positive definite matrix and $A = U\Sigma V^*$ is a singular value decomposition of A, then $U = V$.

Proof. Because A is self-adjoint (that is, $A = A^*$), we have

$$A^2 = A^*A = V\Sigma^*U^*U\Sigma V^* = V\Sigma^2V^* = (V\Sigma V^*)^2. \tag{7}$$

Since A is positive definite, so is Σ and it follows that $\langle V\Sigma V^*x, x\rangle = \langle \Sigma V^*x, V^*x\rangle$ > 0 for all x. This proves that $V\Sigma V^*$ is positive definite. Let $W = V\Sigma V^*$. We assert that $A = W$. If so, it gives that

$$U\Sigma V^* = A = W = V\Sigma V^* \Rightarrow U\Sigma V^* = V\Sigma V^*. \tag{8}$$

Note that positive-definiteness of A implies that all of singular values are positive, hence $\det\Sigma \neq 0$ and so the matrix is invertible. Multiplying both right sides of (8) by $V\Sigma^{-1}$, we get $U = V$, as required.

Now, to prove the assertion, suppose $A^2v = \lambda v$ with $\lambda > 0$. Then $(A^2 - \lambda I)v = (A + \sqrt{\lambda}I)(A - \sqrt{\lambda}I)v = 0$, and if $\det(A + \sqrt{\lambda}I) = 0$, $-\sqrt{\lambda}$ is an eigenvalue of A, contradicting positive-definiteness of A. Thus $A + \sqrt{\lambda}I$ is invertible, and we then arrive at $(A - \sqrt{\lambda}I)v = 0$. This shows that v is an eigenvector of A corresponding to an eigenvalue $\sqrt{\lambda}$. Therefore, we can conclude that A, A^2, W, and W^2 have the same eigenvectors and in particular A and W have the same eigenvalues. By diagonalizability of them (Theorem 2.1), $A = W$ □

5 Applications

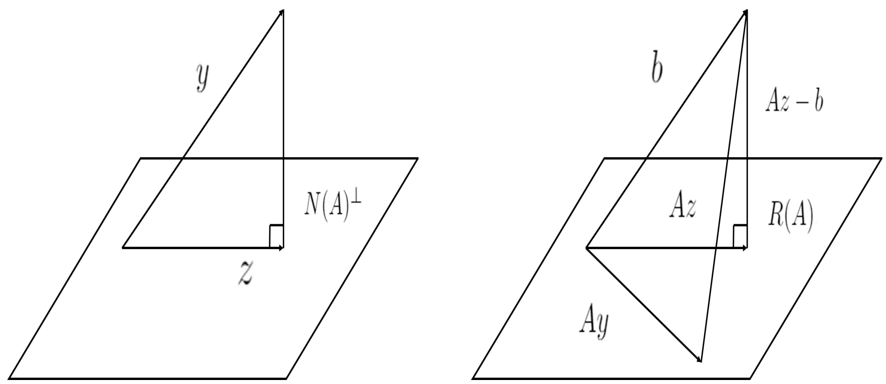
SVD is applied in many fields. For example, polar decomposition, image processing, the effective rank, principle component analysis (PCA), etc. But it's hard to deal with these topics here, hence we only focus on *the pseudoinverse*. Its concept is as follows: If we define the inverse matrix of *any* $m \times n$ matrix A, denoted by $A^\dagger$, the linear system $Ax = b$ might have the unique solution $x = A^\dagger b$. Even if the system $Ax = b$ is inconsistent, then $A^\dagger b$ still exists. How is the vector $A^\dagger b$ related to the system of linear equations $Ax = b$? We will define pseudoinverse and investigate some interesting results. Also, we will discuss how SVD and the psuedoinverse are related.

Definition 5.1. Let V and W be finite-dimensional inner product spaces over the same field, and let $T : V \to W$ be a linear transformation. Let $L : N(T)^\perp \to R(T)$ be the linear transformation defined by $L(x) = T(x)$ for all $x \in N(T)^\perp$. The pseudoinverse of T, denoted by $T^\dagger$, is defined as the unique linear transformation from W to V such that

$$T^\dagger(y) = \begin{cases} L^{-1}(y) & \text{for} \quad y \in R(T) \\ 0 & \text{for} \quad y \in R(T)^\perp. \end{cases} \tag{9}$$

Remark. The pseudoinverse of a linear transformation T on a finite-dimensional inner product space exists even if T is not invertible. Furthermore, if T is invertible, then $T^\dagger = T^{-1}$ because $N(T)^\perp = V$, and L (as just defined) coincides with T.

Definition 5.2. Let A be an $m \times n$ matrix. Then there exists a unique $n \times m$ matrix B such that $L_A^\dagger : F^m \to F^n$ is equal to the left-multiplication transformation L_B. We call B the pseudoinverse of A and denote it by $B = A^\dagger$.



(a) Figure 1 (b) Figure 2

Theorem 5.3. Consider the system of linear equations $Ax = b$, where A is an $m \times n$ matrix and $b \in F^m$. If $z = A^{\dagger}b$, then z has the following properties.

(a) If $Ax = b$ is consistent, then z is the unique solution to the system having minimum norm.

(b) If $Ax = b$ is inconsistent, then z is the unique best approximation to a solution having minimum norm. Furthermore, if $Az = Ay$, then $\|z\| \leq \|y\|$ with equality if and only if $z = y$.

Proof. For (a), suppose $Ax = b$ is consistent and let $z = A^{\dagger}b$. Then $Az = AA^{\dagger}b = b$, which shows z is a solution to the system. To prove that it is the unique solution having minimum norm, suppose y is another solution to the system, that is, $Ay = b$. Multiplying both left sides of that by $A^{\dagger}$, it gives $A^{\dagger}Ay = A^{\dagger}b = z$ and so z is the orthogonal projection of y on $N(A)^{\perp}$, because $A^{\dagger}A$ is the orthogonal projection of F^n on $N(A)^{\perp}$. Therefore, $\|z\| \leq \|y\|$ with equality if and only if $z = y$ (See Figure 1).
For (b), suppose $Ax = b$ is inconsistent. Note that $AA^{\dagger}$ is the orthogonal projection of F^m on $R(A)$, hence it follows from $AA^{\dagger}b = Az$ that Az is the vector in $R(A)$ nearest b (See Figure 2). Finally, we contend that the norm of z is minimal, that is, $\|z\| \leq \|y\|$ whenever $Az = Ay$. Let $Az = Ay = c$ and consider the linear system $Ax = c$. By part (a), $A^{\dagger}c = A^{\dagger}Az = A^{\dagger}AA^{\dagger}b = A^{\dagger}b = z$ is the unique solution to the system $Ax = c$ and $\|z\| \leq \|y\|$. $\square$

According to Theorem 5.3, any linear system $Ax = b$ can be solved in a reasonable manner (even if the system is inconsistent, the theorem provides the best approximation.) But how can we calculate the pseudoinverse $A^\dagger$? As mentioned earlier, this is related to the SVD. In fact, $A^\dagger$ can be calculated very easily by using the SVD. As a result, if $A = U\Sigma V^*$, then $A^\dagger = V\Sigma^\dagger U^*$ is known. The proof can be obtained by slightly modifying the arguments of Section 3. But we does not prove it.

Example 5.4. Let $A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \end{pmatrix}$. We now find its pseudoinverse $A^\dagger$.

Note that

$$A = U\Sigma V^* = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{6} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}^* . \tag{10}$$

Hence

$$\begin{aligned} A^\dagger = V\Sigma^\dagger U^* &= \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{-1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \\ &= \frac{1}{6} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix}. \end{aligned} \tag{11}$$

□

Example 5.5. Consider the linear system

$$x_1 + x_2 - x_3 = 1 \tag{12}$$

$$x_1 + x_2 - x_3 = 1. \tag{13}$$

The system has infinitely many solutions. Let A be the coefficient matrix and let $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. From the preceding example, we know $A^\dagger = \frac{1}{6} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix}$, and thus

$$z = A^\dagger b = \frac{1}{6}\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}. \tag{14}$$

is the solution of minimal norm. □

Example 5.6. Consider the linear system

$$x_1 + x_2 - x_3 = 1 \tag{15}$$

$$x_1 + x_2 - x_3 = 2. \tag{16}$$

The system clearly has no solutions. Let A be the coefficient matrix and let $b = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Therefore,

$$z = A^\dagger b == \frac{1}{2}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}. \tag{17}$$

is the best approximation to a solution having minimum norm, although it is not a solution to the system. □

We finally refer to algebraic property of the pseudoinverse $A^\dagger$. Fortunately, most of the properties are valid, such as A^{-1}. For example, $\left(A^\dagger\right)^\dagger = A$, $\left(A^t\right)^\dagger = \left(A^\dagger\right)^t$, and $\left(\bar{A}\right)^\dagger = \overline{(A^\dagger)}$. But $(AB)^\dagger = B^\dagger A^\dagger$ is not valid unless either of A or B is a good matrix.

Exercise 5.7. Exhibit matrices A and B such that AB is defined, but $(AB)^\dagger \neq B^\dagger A^\dagger$.

Proof. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, and $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then $AB = B$ and we show that $(AB)^\dagger x = B^\dagger x \neq B^\dagger A^\dagger x$ by definition. Note that

$$N(A)^\perp = \{(x,y) : x = y\} \quad \text{and} \quad N(A) = \{(x,y) : x + y = 0\}$$
$$R(A) = \{(x,0) : x \in R\} \quad \text{and} \quad R(A)^\perp = \{(0,y) : y \in R\}$$
$$N(B)^\perp = \{(x,0) : x \in R\} \quad \text{and} \quad N(B) = \{(0,y) : y \in R\}$$
$$R(B) = \{(x,0) : x \in R\} \quad \text{and} \quad R(B)^\perp = \{(0,y) : y \in R\}.$$

Therefore, $B^\dagger x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $B^\dagger A^\dagger x = B^\dagger \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}$, as desired. □

Exercise 5.8. Let A be an $m \times n$ matrix. Prove the following results.

(a) For any $m \times m$ unitary matrix G, $(GA)^\dagger = A^\dagger G^*$.

(b) For any $n \times n$ unitary matrix H, $(AH)^\dagger = H^* A^\dagger$.

Proof. Because G preserves orthogonality (that it, $\langle x, y\rangle = \langle Gx, Gy\rangle$), $R(A) \oplus R(A)^\perp$ is mapped by G into $G(R(A)) \oplus G(R(A))^\perp$ and it follows from this fact that the result (a) is obvious by definition of the pseudoinverse. The result (b) is analogous. □

References

[1] *Linear algebra*, Spence, Lawrence E., Arnold J. Insel, and Stephen H. Friedberg, Prentice Hall, 2002.

[2] *Linear algebra and its applications*, Strang, Gilbert, 2006.