MAS4115 Project
Singular Value Decomposition

Nicklas Rice

April 29, 2024

Much of our later course material was predicated on the existence of a certain matrix decomposition called singular value decomposition (SVD). In this paper, we will prove the existence of a singular value decomposition for real matrices. We begin by recalling the definition of *inner product*.

Definition. *Let V be a vector space over $\mathbb{R}$. An* ***inner product*** *on V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ such that for all $x, y, z \in V$ and all $c \in \mathbb{R}$ the following hold:*

1. $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$.
2. $\langle cx, y \rangle = c\langle x, y \rangle$.
3. $\langle x, y \rangle = \langle y, x \rangle$.
4. $\langle x, x \rangle > 0$ *if* $x \neq 0$.

An important, albeit trivial, consequence of the above definition is that if $a_1, a_2, ..., a_k \in \mathbb{R}$ and $y, v_1, v_2, ..., v_k \in V$, then

$$\left\langle \sum_{i=1}^{k} a_i v_i, y \right\rangle = \sum_{i=1}^{k} a_i \langle v_i, y \rangle.$$

Moreover, if $y, z \in V$ are such that $\langle x, y \rangle = \langle x, z \rangle$ for all $x \in V$, then it must be the case that $y = z$. To see why, note that, in particular, we have $\langle y - z, y \rangle = \langle y - z, z \rangle$. This implies that $\langle y - z, y - z \rangle = 0$, which is true only if $y = z$.

We refer to a vector space V over $\mathbb{R}$ endowed with a specific inner product as a **real inner product space**, or simply an **inner product space**.

Definition. *Let V be an inner product space. For $x \in V$, we define the* ***norm*** *or* ***length*** *of x by $\|x\| = \sqrt{\langle x, x \rangle}$.*

Definition. *Let V be an inner product space. Vectors x and y in V are* ***orthogonal*** *if $\langle x, y \rangle = 0$. A subset S of V is* ***orthogonal*** *if any two distinct vectors in S are orthogonal. A vector x in V is a* ***unit vector*** *if $\|x\| = 1$. Finally, a* *subset S of V is* ***orthonormal*** *if S is orthogonal and consists entirely of unit vectors.*

Observe that multiplying vectors by nonzero scalars does not affect their orthogonality and that if x is any nonzero vector, then $(1/\|x\|)x$ is a unit vector. The process of multiplying a nonzero vector by the reciprocal of its length is called **normalizing**. Hence, given an orthogonal set S of nonzero vectors, we can find an orthonormal set S' of nonzero vectors such that $\text{span}(S') = \text{span}(S)$ simply by normalizing each vector in S.

Theorem 1. *Let V be an inner product space, and let $S = \{v_1, v_2, ..., v_k\}$ be an orthogonal subset of V consisting of nonzero vectors. If $y \in span(S)$, then*

$$y = \sum_{i=1}^{k} \frac{\langle y, v_i \rangle}{\|v_i\|^2} v_i.$$

Proof. Since $y \in \text{span}(S)$, we may write $y = \sum_{i=1}^{k} a_i v_i$, where $a_1, a_2, ..., a_k \in \mathbb{R}$. Then, for $1 \leqslant j \leqslant k$, we have

$$\langle y, v_j \rangle = \left\langle \sum_{i=1}^{k} a_i v_i, v_j \right\rangle = \sum_{i=1}^{k} a_i \langle v_i, v_j \rangle = a_j \langle v_j, v_j \rangle = a_j \|v_j\|^2.$$

So $a_j = \frac{\langle y, v_j \rangle}{\|v_j\|^2}$ and the result follows. □

If, in the preceding Theorem, the set S is orthonormal, then the formula simplifies to

$$y = \sum_{i=1}^{k} \langle y, v_i \rangle v_i.$$

The following result can be obtained by following the proof of Theorem 1 with $y = 0$.

Corollary. *Let V be an inner product space and let S be an orthogonal subset of V consisting of nonzero vectors. Then S is linearly independent.*

The next theorem tells us how to construct an orthogonal set from a linearly independent set in such a way that both sets generate the same subspace.

Theorem 2 (Gram-Schmidt process)**.** *Let V be an inner product space, and let $S = \{w_1, w_2, ..., w_n\}$ be a linearly independent subset of V. Define $S' = \{v_1, v_2, ..., v_n\}$, where $v_1 = w_1$ and*

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j \tag{1}$$

for all $2 \leqslant k \leqslant n$. Then S' is an orthogonal set of nonzero vectors such that $span(S') = span(S)$.

Proof. We proceed by induction on n. If $n = 1$, then $S' = S$ obviously satisfies the conclusion of the theorem. Assume now that $n \geqslant 2$ and that the set $S'_{n-1} = \{v_1, v_2, ..., v_{n-1}\}$ with the desired properties has been constructed by repeated use of (1). We show that the set $S' = \{v_1, v_2, ..., v_{n-1}, v_n\}$ also has the desired properties, where v_n is obtained from S'_{n-1} by (1). If $v_n = 0$, then (1) implies that $w_n \in \mathrm{span}(S'_{n-1}) = \mathrm{span}(\{w_1, w_2, ..., w_{n-1}\})$, which contradicts our assumption that S is linearly independent. For $1 \leqslant i \leqslant n-1$, it follows from (1) that

$$\langle v_n, v_i \rangle = \langle w_n, v_i \rangle - \sum_{j=1}^{n-1} \frac{\langle w_n, v_j \rangle}{\|v_j\|^2} \langle v_j, v_i \rangle = \langle w_n, v_i \rangle - \frac{\langle w_n, v_i \rangle}{\|v_i\|^2} \|v_i\|^2 = 0,$$

since $\langle v_i, v_j \rangle = 0$ if $i \neq j$ by the inductive hypothesis. Hence S' is an orthogonal set of nonzero vectors. To complete the proof, note that, by (1), we have $\mathrm{span}(S') \subseteq \mathrm{span}(S)$. But by the Corollary to Theorem 1, S' is linearly independent; so $\dim(\mathrm{span}(S')) = \dim(\mathrm{span}(S)) = n$. Therefore $\mathrm{span}(S) = \mathrm{span}(S')$. ☐

The following theorem is immediate from Theorems 1 and 2.

Theorem 3. *Let V be a nonzero finite-dimensional inner product space. Then V has an orthonormal basis β. Furthermore, if $\beta = \{v_1, v_2, ..., v_n\}$ and $x \in V$, then*

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i.$$

Corollary. *Let V and W be a finite-dimensional inner product spaces with orthonormal bases $\beta = \{u_1, u_2, ..., u_n\}$ and $\gamma = \{v_1, v_2, ..., v_m\}$ respectively. Let $T : V \to W$ be linear, and let $A = [T]_\beta^\gamma$. Then for any i and j, $A_{ij} = \langle T(u_j), v_i \rangle_W$.*

Note that in the above corollary, we do not need β to be orthonormal. In fact, we do not need V to have an inner product at all.

Theorem 4. *Let V be a finite-dimensional inner product space over $\mathbb{R}$, and let $g : V \to \mathbb{R}$ be a linear transformation. Then there exists a unique vector $y \in V$ such that $g(x) = \langle x, y \rangle$ for all $x \in V$.*

Proof. Let $\beta = \{v_1, v_2, ..., v_n\}$ be an orthonormal basis for V, and let $y = \sum_{i=1}^{n} g(v_i) v_i$. Define $h : V \to \mathbb{R}$ by $h(x) = \langle x, y \rangle$, which is clearly linear. Furthermore for $1 \leqslant j \leqslant n$ we have

$$h(v_j) = \langle v_j, y \rangle = \left\langle v_j, \sum_{i=1}^{n} g(v_i) v_i \right\rangle = \sum_{i=1}^{n} g(v_i) \langle v_j, v_i \rangle = g(v_j).$$

Since g and h both agree on β, we have that $g = h$. To show that y is unique, suppose that $g(x) = \langle x, y' \rangle$ for all x. Then $\langle x, y \rangle = \langle x, y' \rangle$ for all x, so $y = y'$ by the discussion following the definition of inner product. ☐

Theorem 5. *Let V and W be finite-dimensional inner product spaces, and let $T : V \to W$ be linear. Then there exists a unique function $T^* : W \to V$, called the* ***adjoint*** *of T, such that $\langle T(x), y\rangle_W = \langle x, T^*(y)\rangle_V$ for all $x \in V$ and $y \in W$. Furthermore, T^* is linear.*

Proof. Let $y \in W$. Define $g : V \to \mathbb{R}$ by $g(x) = \langle T(x), y\rangle_W$ for all $x \in V$. It is easy to verify that g is linear. We now apply Theorem 4 to obtain a unique vector $x' \in V$ such that $g(x) = \langle x, x'\rangle_V$; that is, $\langle T(x), y\rangle_W = \langle x, x'\rangle_V$ for all $x \in V$. Now if we define $T^* : W \to V$ by $T^*(y) = x'$, we have $\langle T(x), y\rangle_W = \langle x, T^*(y)\rangle_V$. To show that T^* is linear, let $y_1, y_2 \in W$ and $c \in \mathbb{R}$. Then for any $x \in V$, we have

$$\begin{aligned}
\langle x, T^*(cy_1 + y_2)\rangle_V &= \langle T(x), cy_1 + y_2\rangle_W \\
&= c\langle T(x), y_1\rangle_W + \langle T(x), y_2\rangle_W \\
&= c\langle x, T^*(y_1)\rangle_V + \langle x, T^*(y_2)\rangle_V \\
&= \langle x, cT^*(y_1) + T^*(y_2)\rangle_V.
\end{aligned}$$

Since x was arbitrary, we again have by the discussion following the definition of inner product that $T^*(cy_1 + y_2) = cT^*(y_1) + T^*(y_2)$, so T^* is linear. Finally, we need to show that T^* is unique. Suppose that $U : W \to V$ is linear and satisfies $\langle T(x), y\rangle_W = \langle x, U(y)\rangle_V$ for all $x \in V$. Then $\langle x, T^*(y)\rangle = \langle x, U(y)\rangle$ for all $x, y \in V$, so $T^* = U$. □

The following properties of the adjoint are easy to prove and will be used later.

Theorem 6. *Let U, V, and W be inner product spaces, and let $T : U \to V$ and $U : V \to W$ be linear. Then the following hold:*

1. *$(UT)^* = T^*U^*$.*
2. *$T^{**} = T$.*

Note that, because the inner product is symmetric, the adjoint T^* of a linear transformation $T : V \to W$ also satisfies $\langle y, T(x)\rangle_W = \langle T^*(y), x\rangle_V$ for all $x \in V$ and all $y \in W$. The next result relates the matrix representation of the adjoint of a linear transformation to the matrix representation of the transformation itself. However, we first need a definition.

Definition. *Let $A \in M_{m\times n}(\mathbb{R})$. We define the* ***transpose*** *of A to be the $n \times m$ matrix A^T such that $(A^T)_{ij} = A_{ji}$ for all i, j.*

In other words, we obtain A^T from A by interchanging the rows and columns. Now we are ready for the result.

Theorem 7. *Let V and W be finite-dimensional inner product spaces with orthonormal bases β and γ respectively. If $T : V \to W$ is linear, then*

$$[T^*]_\gamma^\beta = \left([T]_\beta^\gamma\right)^T.$$

Proof. Let $A = [T]_\beta^\gamma, B = [T^*]_\gamma^\beta, \beta = \{u_1, u_2, ..., u_n\}$, and $\gamma = \{v_1, v_2, ..., v_n\}$. Then from the Corollary to Theorem 3, we have

$$B_{ij} = \langle T^*(v_j), u_i \rangle_V = \langle T(u_i), v_j \rangle_W = A_{ji}.$$

Hence $B = A^T$. □

Corollary. *Let $A \in M_{m\times n}(\mathbb{R})$. Then $L_{A^T} = (L_A)^*$.*

Proof. If β and γ are the standard ordered bases for F^n and F^m respectively, then $[L_A]_\beta^\gamma = A$. Hence $[(L_A)^*]_\gamma^\beta = ([L_A]_\beta^\gamma)^T = A^T = [L_{A^T}]_\gamma^\beta$, and so $L_{A^T} = (L_A)^*$. □

We now become interested in finding a necessary and sufficient condition for an inner product space V to have an orthonormal basis of eigenvectors of a given linear transformation T. We begin to answer this question with a result of Schur which we will not prove for purposes of brevity and scope.

Theorem 8 (Schur)**.** *Let T be a linear operator on a finite-dimensional inner product space V. Suppose that the characteristic polynomial of T splits. Then there exists an orthonormal basis β for V such that the matrix $[T]_\beta$ is upper triangular.*

The criterion we are looking for, it turns out, is given in the following definition.

Definition. *Let T be a linear operator on an inner product space V. We say that T is **self-adjoint** if $T = T^*$. A matrix $A \in M_{n\times n}(\mathbb{R})$ is **self-adjoint** if L_A is self-adjoint.*

Note that, by the Corollary to Theorem 7, the condition that $A^T{=}A$ is equivalent to A being self-adjoint. Matrices that satisfy this condition are called **symmetric**.

It follows from Theorem 7 that, if β is an orthonormal basis, then T is self-adjoint if and only if $[T]_\beta$ is symmetric. We will assume that the characteristic polynomial of any self-adjoint operator on a (real) inner product space splits. Although this is not hard to prove, it does involve working with complex numbers and so we will not do it. Now we present the result we have been building toward.

Theorem 9. *Let T be a linear operator on a finite-dimensional inner product space V. Then T is self-adjoint if and only if there exists an orthonormal basis β for V consisting of eigenvectors of T.*

Proof. Suppose that T is self-adjoint. By the assumption we made above, we may apply Schur's theorem to obtain an orthonormal basis β for V such that the matrix $A = [T]_\beta$ is upper triangular. But

$$A^T = ([T]_\beta)^T = [T^*]_\beta = [T]_\beta = A.$$

So A and A^T are both upper triangular, and therefore A is a diagonal matrix. Thus β must consist of eigenvectors of T.

For the converse, let β be an orthonormal basis consisting of eigenvectors of T. Then $[T]_\beta$ is diagonal and hence symmetric. It follows that T is self-adjoint. □

We are almost at a point where we can prove the existence of an SVD, but we need one more definition first.

Definition. *A linear operator T on a finite-dimensional inner product space is called **positive definite** [**positive semidefinite**] if T is self-adjoint and $\langle T(x), x\rangle > 0$ [$\langle T(x), x\rangle \geqslant 0$] for all $x \neq 0$.*

The following result provides an alternative definition.

Theorem 10. *Let T be a self-adjoint linear operator on a finite-dimensional inner product space V. Then T is positive definite [semidefinite] if and only if all of its eigenvalues are positive [nonnegative].*

Proof. Let $\beta = \{v_1, v_2, ..., v_n\}$ be a basis of orthonormal eigenvectors corresponding to eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$. Let $x \neq 0$ be arbitrary and write

$$x = \sum_{i=1}^{n} a_i v_i,$$

where $a_1, a_2, ..., a_n \in \mathbb{R}$. We now compute

$$T(x) = \sum_{i=1}^{n} a_i T(v_i) = \sum_{i=1}^{n} a_i \lambda_i v_i$$

so that

$$\langle T(x), x\rangle = \left\langle \sum_{i=1}^{n} a_i \lambda_i v_i, \sum_{i=1}^{n} a_i v_i \right\rangle = \sum_{i=1}^{n} a_i^2 \lambda_i \langle v_i, v_i\rangle,$$

which is positive [nonnegative] if all the eigenvalues of T are positive [nonnegative].

The converse was proven in class and will not be repeated here. □

Theorem 11. *Let $T : V \to W$ be a linear transformation, where V and W are finite-dimensional inner product spaces. Then T^*T is positive semidefinite with rank(T^*T) = rank(T).*

Proof. By Theorem 6, we have

$$(T^*T)^* = T^*T^{**} = T^*T,$$

so T^*T is self-adjoint. Moreover, for any $x \in V$, we have

$$\langle (T^*T)(x), x\rangle_V = \langle T^*(T(x)), x\rangle_V = \langle T(x), T(x)\rangle_W \geqslant 0.$$

It follows that T^*T is positive semidefinite. To show that $\text{rank}(T^*T) = \text{rank}(T)$, it suffices to show that $N(T^*T) = N(T)$. This can be done easily using the previous computation. $\square$

We conclude with two theorems that achieve our goal of proving the existence of SVD for real matrices.

Theorem 12 (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 $\{u_1, u_2, ..., u_m\}$ for W and positive scalars $\sigma_1 \geqslant \sigma_2 \geqslant ... \geqslant \sigma_r$ such that*

$$T(v_i) = \begin{cases} \sigma_i u_i & \text{if } 1 \leqslant i \leqslant r \\ 0 & \text{if } i > r. \end{cases}$$

Conversely, suppose that the preceding conditions are satisfied. Then for $1 \leqslant i \leqslant n$, v_i is an eigenvector of T^*T with corresponding eigenvalue σ_i^2 if $1 \leqslant i \leqslant r$ and 0 if $i > r$. Therefore the scalars $\sigma_1, \sigma_2, ..., \sigma_r$ are uniquely determined by T.

Proof. We first establish the existence of the bases and scalars. By Theorem 11, T^*T is a positive semidefinite linear operator of rank r on V. Hence there is an orthonormal basis $\{v_1, v_2, ..., v_n\}$ for V consisting of eigenvectors of T^*T with corresponding eigenvalues λ_i, where $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_r > 0$, and $\lambda_i = 0$ for $i > r$. For $1 \leqslant i \leqslant r$, define $\sigma_i = \sqrt{\lambda_i}$ and $u_i = \frac{1}{\sigma_i} T(v_i)$. We show that $S = \{u_1, u_2, ..., u_r\}$ is an orthonormal subset of W. Suppose $1 \leqslant i, j \leqslant r$. Then

$$\begin{aligned} \langle u_i, u_j \rangle_W &= \left\langle \frac{1}{\sigma_i} T(v_i), \frac{1}{\sigma_j} T(v_j) \right\rangle_W \\ &= \frac{1}{\sigma_i \sigma_j} \langle T^*T(v_i), v_j \rangle_V \\ &= \frac{1}{\sigma_i \sigma_j} \langle \lambda_i v_i, v_j \rangle_V \\ &= \frac{\sigma_i^2}{\sigma_i \sigma_j} \langle v_i, v_j \rangle_V \\ &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Hence, S is orthonormal. We can extend S to an orthonormal basis for W in the following way. By the replacement theorem, S can be extended to an ordered basis γ for W. Now apply the Gram-Schmidt process (Theorem 2) to γ to get a new orthogonal set $\gamma' = \{u_1, u_2, ..., u_r, w_{r+1}, ..., w_m\}$ that spans W (it can be shown by induction that the first r vectors are unaffected by the Gram-Schmidt process since they were already orthogonal). By normalizing the last $m - r$ vectors $w_{r+1}, ..., w_m$ of γ' we obtain an orthonormal basis $\{u_1, u_2, ..., u_m\}$ for

W. By definition $T(v_i) = \sigma_i u_i$ if $1 \leqslant i \leqslant r$. If $i > r$, then $T^*T(v_i) = 0$ and so $T(v_i) = 0$ since $N(T^*T) = N(T)$ as we noted in the proof of Theorem 11.

To establish uniqueness, suppose that $\{v_1, v_2, ..., v_n\}, \{u_1, u_2, ..., u_m\}$, and $\sigma_1, \sigma_2, ..., \sigma_r > 0$ satisfy the properties stated in the first part of the theorem. Then for $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$,

$$\langle T^*(u_i), v_j \rangle_V = \langle u_i, T(v_j) \rangle_W = \begin{cases} \sigma_i & \text{if } i = j \leqslant r \\ 0 & \text{otherwise.} \end{cases}$$

Hence for any $1 \leqslant i \leqslant m$, we have by Theorem 3 that

$$T^*(u_i) = \sum_{j=1}^{n} \langle T^*(u_i), v_j \rangle v_j = \begin{cases} \sigma_i v_i & \text{if } i = j \leqslant r \\ 0 & \text{otherwise} \end{cases}$$

So for $i \leqslant r$,

$$T^*T(v_i) = T^*(\sigma_i u_i) = \sigma_i T^*(u_i) = \sigma_i^2 v_i$$

and $T^*T(v_i) = T^*(0) = 0$ for $i > r$. Therefore each v_i is an eigenvector of T^*T with corresponding eigenvalue σ_i^2 if $i \leqslant r$ and 0 if $i > r$. ☐

Definition. *The unique scalars $\sigma_1, \sigma_2, ..., \sigma_r$ in Theorem 12 are called the **singular values** of T. For a matrix $A \in M_{m \times n}(\mathbb{R})$, we define the **singular values** of A to be the singular values of the linear transformation L_A.*

Definition. *A matrix $A \in M_{m \times n}(\mathbb{R})$ is **orthogonal** if $A^T A = AA^T = I$.*

Theorem 13 (Singular Value Decomposition Theorem for Matrices)**.** *Let $A \in M_{m \times n}(\mathbb{R})$ be a matrix of rank r with the positive singular values $\sigma_1 \geqslant \sigma_2 \geqslant ... \geqslant \sigma_r$, and let Σ be the $m \times n$ matrix defined by*

$$\Sigma_{ij} = \begin{cases} \sigma_i & \textit{if } i = j \leqslant r \\ 0 & \textit{otherwise.} \end{cases}$$

Then there exists an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

$$A = U \Sigma V^T.$$

Proof. Let $T = L_A : \mathbb{R}^n \to \mathbb{R}^m$. By Theorem 12, there exist orthonormal bases $\beta = \{v_1, v_2, ..., v_n\}$ for $\mathbb{R}^n$ and $\gamma = \{u_1, u_2, ..., u_m\}$ for $\mathbb{R}^m$ such that $T(v_i) = \sigma_i u_i$ for $1 \leqslant i \leqslant r$ and $T(v_i) = 0$ for $i > r$. Let U be the $m \times m$ matrix whose jth column is u_j for all j, and let V be the $n \times n$ matrix whose jth column is v_j for all j. Note that both U and V are orthogonal matrices.

The jth column of AV is $Av_j = \sigma_j u_j$, while the jth column of Σ is $\sigma_j e_j$, where e_j is the jth vector of the standard basis for $\mathbb{R}^m$. Hence the jth column of $U\Sigma$ is given by

$$U(\sigma_j e_j) = \sigma_j U(e_j) = \sigma_j u_j.$$

It follows that $AV = U\Sigma$ since their corresponding columns are equal. Therefore $A = AVV^T = U\Sigma V^T$. ☐

Definition. *The factorization of Theorem 13 is called a* ***singular value decomposition*** *for A.*

References

[1] Lawrence E. Spence Stephen H. Friedburg Arnold J. Insel. *Linear Algebra.* 4th ed. Pearson Education, 2003.