EECS 16B Designing Information Devices and Systems II
Fall 2021 Note 16: SVD Properties

In this note, we'd like to explore and collect the fundamental properties of the SVD, so that from now on we'll be able to use it in a variety of contexts.

The following exposition derives heavily from Prof. Arcak's EECS16B reader.

Throughout this note, let's suppose A is an $m \times n$ matrix, with $\text{rank}(A) = r$. Note that $r \leq \min\{m, n\}$.

1 SVD Form

The *full SVD* (or just *SVD*) of A is the following decomposition of A:

$$A = U\Sigma V^\top = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix} \tag{1}$$

$$= \underbrace{\left[\begin{array}{ccc|ccc} | & & | & | & & | \\ \vec{u}_1 & \cdots & \vec{u}_r & \vec{u}_{r+1} & \cdots & \vec{u}_m \\ | & & | & | & & | \end{array}\right]}_{m\times m} \underbrace{\left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0_{r\times(n-r)} \\ & & \sigma_r & \\ \hline & 0_{(m-r)\times r} & & 0_{(m-r)\times(n-r)} \end{array}\right]}_{m\times n} \underbrace{\left[\begin{array}{c} - \ \vec{v}_1^\top \ - \\ \vdots \\ - \ \vec{v}_r^\top \ - \\ \hline -\vec{v}_{r+1}^\top- \\ \vdots \\ - \ \vec{v}_n^\top \ - \end{array}\right]}_{n\times n} \tag{2}$$

where U, V, and Σ are chosen such that

- U is an $m \times m$ matrix with orthonormal columns $\vec{u}_1, \ldots, \vec{u}_m$ that live in $\mathbb{R}^m$.
- V is an $n \times n$ matrix with orthonormal columns $\vec{v}_1, \ldots, \vec{v}_n$ that live in $\mathbb{R}^n$.
- Σ is an $m \times n$ matrix which has an $r \times r$ diagonal block Σ_r in the upper left, and 0 elsewhere.
- U_r is an $m \times r$ matrix with the first r orthonormal columns $\vec{u}_1, \ldots, \vec{u}_r$ of U.
- V_r is an $n \times r$ matrix with the first r orthonormal columns $\vec{v}_1, \ldots, \vec{v}_r$ of V.
- Σ_r is an $r \times r$ matrix with the largest r *singular values* $\sigma_1 \geq \cdots \geq \sigma_r > 0$ of A.[1]
- U_{m-r} is an $m \times (m-r)$ matrix with the last $m-r$ orthonormal columns $\vec{u}_{r+1}, \ldots, \vec{u}_m$ of U.
- V_{n-r} is an $n \times (n-r)$ matrix with the last $n-r$ orthonormal columns $\vec{v}_{r+1}, \ldots, \vec{v}_n$ of V.

[1] We also consider the singular values $\sigma_{r+1} = \cdots = \sigma_{\min\{m,n\}} = 0$, although they don't show up explicitly in the matrix. We can consider them as the singular values associated with $\vec{u}_{r+1}, \ldots, \vec{u}_{\min\{m,n\}}$ and $\vec{v}_{r+1}, \ldots, \vec{v}_{\min\{m,n\}}$, and thus $\sigma_{r+1}, \ldots, \sigma_{\min\{m,n\}}$ can be thought of as the $(r+1)^{\text{th}}, \ldots, \min\{m,n\}^{\text{th}}$ entries on the diagonal of Σ.

Our matrices U, V, Σ are carefully constructed to have particular linear-algebraic properties. Some of these are listed below.

- The columns of U are the orthonormal eigenvectors of $AA^\top$.
- The columns of V are the orthonormal eigenvectors of $A^\top A$.
- The diagonal entries in Σ are the *square roots* of the eigenvalues of $AA^\top$ or $A^\top A$.
- $\mathrm{Col}(U_r) = \mathrm{span}(\vec{u}_1, \ldots, \vec{u}_r) = \mathrm{Col}(A)$.
- $\mathrm{Col}(U_{m-r}) = \mathrm{span}(\vec{u}_{r+1}, \ldots, \vec{u}_m) \perp \mathrm{Col}(A)$[2].
- $\mathrm{Col}(V_r) = \mathrm{span}(\vec{v}_1, \ldots, \vec{v}_r) \perp \mathrm{Null}(A)$.
- $\mathrm{Col}(V_{n-r}) = \mathrm{span}(\vec{v}_{r+1}, \ldots, \vec{v}_n) = \mathrm{Null}(A)$.

We have already derived several of these properties using the construction in our previous note. For completeness, when we put together an algorithm for constructing the SVD, we will provide the proof that our algorithm supplies vectors with these properties.

2 Space Efficiency: Outer Product Form

This particular decomposition requires us to store m^2 numbers for U, mn numbers for Σ, and n^2 numbers for V. So we need to store $m^2 + mn + n^2$ numbers total. But a lot of these entries of Σ are 0s. This seems redundant. The fact that Σ is so structured means that there is probably a way to simplify this representation. Let's try to discover this way by trying to do the multiplication $A = U\Sigma V^\top$, and simplify if possible.

$$
\begin{aligned}
A &= U\Sigma V^\top && (3)\\
&= \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix} && (4)\\
&= \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \left(\begin{bmatrix} \Sigma_r \\ 0_{(m-r)\times r} \end{bmatrix} V_r^\top + \begin{bmatrix} 0_{r\times(n-r)} \\ 0_{(m-r)\times(n-r)} \end{bmatrix} V_{n-r}^\top \right) && (5)\\
&= \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} \Sigma_r \\ 0_{(m-r)\times r} \end{bmatrix} V_r^\top && (6)\\
&= \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} \Sigma_r V_r^\top \\ 0_{(m-r)\times r} V_r^\top \end{bmatrix} && (7)\\
&= \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} \Sigma_r V_r^\top \\ 0_{(m-r)\times n} \end{bmatrix} && (8)\\
&= U_r \Sigma_r V_r^\top + U_{m-r} 0_{(m-r)\times n} && (9)\\
&= U_r \Sigma_r V_r^\top. && (10)
\end{aligned}
$$

[2]This notation may be unfamiliar. By saying that a subspace is $\perp$ (orthogonal to) another subspace, we mean every vector in the first subspace is orthogonal to every vector in the second subspace, or vice versa.

This form of the SVD is the *compact* SVD, and while we are not going to work with it past this section (and it's out of scope for everything else in the class), it has its own useful properties (U_r and V_r have orthonormal columns[3], and Σ_r is square and invertible) and is less expensive to compute/store, especially when r is small.

Motivated by the fact that Σ_r is diagonal and still has a lot of 0s and therefore a lot of redundancy that we would like to optimize out, we attempt to complete the multiplication. This time we look at the columns of each matrix.

$$
\begin{aligned}
A &= U_r \Sigma_r V_r^\top && (11)\\
&= \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} \vec{v}_1^\top \\ \vdots \\ \vec{v}_r^\top \end{bmatrix} && (12)\\
&= \begin{bmatrix} \sigma_1 \vec{u}_1 & \cdots & \sigma_r \vec{u}_r \end{bmatrix} \begin{bmatrix} \vec{v}_1^\top \\ \vdots \\ \vec{v}_r^\top \end{bmatrix} && (13)\\
&= \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top. && (14)
\end{aligned}
$$

This is the *outer product form of the SVD*.

Let's look at how much storage this requires. For each $\vec{u}_i$, we require m numbers. For each $\vec{v}_i$, we require n numbers. And each σ_i is one number. We need r of each, so we need to store $r(m+n+1)$ numbers, a huge improvement from $m^2 + n^2 + mn$ numbers as before. This is especially true when r is small. In fact, by storing $\sigma_i \vec{u}_i$ and $\vec{v}_i$ separately as two sets of vectors, we could represent A using $r(m+n)$ entries, whereas before we needed mn entries! This is another huge improvement and benefit of the SVD.

3 An Algorithm for Computing the SVD

So how do we actually calculate the SVD? We will work with the $n \times n$ symmetric matrix $A^\top A$ or $m \times m$ symmetric matrix $AA^\top$[4]. The one that is smaller is preferable to work with, since it requires less computation and storage. We will give a justification for the case that we're working with $A^\top A$, but provide algorithms for both cases. The justification for the case of $AA^\top$ is left to discussion section.

By the Spectral Theorem for real symmetric matrices (discussed in Note 14), both of these matrices have all real eigenvalues, r of which are positive and the remaining are 0. Thus we can do either of the following procedures:

Method 1 To Find The Full SVD:

[3]Note that U_r is $m \times r$ and V_r is $n \times r$. They're generally not square matrices, so we can't say $U_r^{-1} = U_r^\top$, because U_r^{-1} and V_r^{-1} don't exist. But because they have orthonormal columns, we can say that $U_r^\top U_r = V_r^\top V_r = I_r$.

[4]To check that they're symmetric, take the transpose of each matrix, and observe that the matrix equals its transpose. For example, we can check that $A^\top A$ is symmetric:

$$(A^\top A)^\top = (A)^\top (A^\top)^\top = A^\top A.$$

You can try the case for $AA^\top$ yourself.

(a) Find eigenvalues $\lambda_1, \ldots, \lambda_n$ of $A^\top A$ and order them such that $\lambda_1 \geq \cdots \geq \lambda_r > 0$ and $\lambda_{r+1} = \cdots = \lambda_n = 0$.

(b) Find orthonormal eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$ such that

$$A^\top A\vec{v}_i = \lambda_i \vec{v}_i \qquad \text{for } i = 1, \ldots, n. \tag{15}$$

(c) Define $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, \ldots, \min\{m, n\}$.

(d) Find orthonormal vectors $\vec{u}_1, \ldots, \vec{u}_m$, obtaining $\vec{u}_1, \ldots, \vec{u}_r$ by the equation

$$\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i} \qquad \text{for } i = 1, \ldots, r \tag{16}$$

and finding $\vec{u}_{r+1}, \ldots, \vec{u}_m$ by Gram-Schmidt.

Method 2 To Find The Full SVD:

(a) Find eigenvalues $\lambda_1, \ldots, \lambda_m$ of $AA^\top$ and order them such that $\lambda_1 \geq \cdots \geq \lambda_r > 0$ and $\lambda_{r+1} = \cdots = \lambda_m = 0$.

(b) Find orthonormal eigenvectors $\vec{u}_1, \ldots, \vec{u}_m$ such that

$$AA^\top \vec{u}_i = \lambda_i \vec{u}_i \qquad \text{for } i = 1, \ldots, m. \tag{17}$$

(c) Define $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, \ldots, \min\{m, n\}$.

(d) Find orthonormal vectors $\vec{v}_1, \ldots, \vec{v}_n$, obtaining $\vec{v}_1, \ldots, \vec{v}_r$ by the equation

$$\vec{v}_i = \frac{A^\top \vec{u}_i}{\sigma_i} \qquad \text{for } i = 1, \ldots, r \tag{18}$$

and finding $\vec{v}_{r+1}, \ldots, \vec{v}_n$ by Gram-Schmidt.

If we wanted to form the compact SVD, no problem! We could just skip finding the "extra" vectors $\vec{u}_{r+1}, \ldots, \vec{u}_m$ and $\vec{v}_{r+1}, \ldots, \vec{v}_n$, and it wouldn't affect the computation of $\vec{u}_1, \ldots, \vec{u}_r$ and $\vec{v}_1, \ldots, \vec{v}_r$ at all.

4 Proving That the Algorithm Correctly Calculates the Full SVD

We have some things to show about this construction (specifically Method 1). First, we need to show that if U, Σ, and V are defined as in eq. (1) and eq. (2) using the vectors provided by our algorithm, then A really does equal $U\Sigma V^\top$. This is necessary to prove because otherwise the decomposition is wrong from the start. After we show $A = U\Sigma V^\top$, we would like to prove the list of properties we listed on the first couple pages.

We begin by proving that $A = U\Sigma V^\top$. We start with the fact that V has orthogonal columns and is square, so $V^\top = V^{-1}$. Thus $V^\top V = VV^\top = I_n$, so

$$A = AVV^\top = A\begin{bmatrix} V_r & V_{n-r} \end{bmatrix} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix} \tag{19}$$

$$= \begin{bmatrix} AV_r & AV_{n-r} \end{bmatrix} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix}. \tag{20}$$

Notice that in the algorithm, we defined $\vec{u}_i$ and $\vec{v}_i$ in a very particular way. That is, for $i \leq r$, we defined $\vec{u}_i$ and $\vec{v}_i$ such that $A\vec{v}_i = \sigma_i \vec{u}_i$. Therefore

$$AV_r = A \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_r \end{bmatrix} = \begin{bmatrix} A\vec{v}_1 & \cdots & A\vec{v}_r \end{bmatrix} = \begin{bmatrix} \sigma_1 \vec{u}_1 & \cdots & \sigma_r \vec{u}_r \end{bmatrix} = U_r \Sigma_r. \tag{21}$$

We sort of did this computation in reverse in eq. (13), so we are allowed to make this simplification.

Note also that for $i > r$, we know $\vec{v}_i$ is an eigenvector of $A^\top A$ with eigenvalue 0. Thus

$$A^\top A \vec{v}_i = 0\vec{v}_i = \vec{0}. \tag{22}$$

Left-multiplying by $\vec{v}_i^\top$, we get

$$\vec{v}_i^\top A^\top A \vec{v}_i = \vec{v}_i^\top \vec{0} = 0. \tag{23}$$

But this leftmost term can be simplified further! It is indeed the squared norm of $A\vec{v}_i$:

$$0 = \vec{v}_i^\top A^\top A \vec{v}_i = (A\vec{v}_i)^\top (A\vec{v}_i) = \|A\vec{v}_i\|^2. \tag{24}$$

Since the squared norm of $A\vec{v}_i$ is 0, the norm of $A\vec{v}_i$ is 0 (by taking square roots):

$$\|A\vec{v}_i\|^2 = 0 \implies \|A\vec{v}_i\| = 0. \tag{25}$$

But the norm of $A\vec{v}_i$ is the length of the vector $A\vec{v}_i$. And the only vector with length 0 is the zero vector $\vec{0}$, so we must have $A\vec{v}_i = \vec{0}$. Therefore

$$AV_{n-r} = A \begin{bmatrix} \vec{v}_{r+1} & \cdots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} A\vec{v}_{r+1} & \cdots & A\vec{v}_n \end{bmatrix} = \begin{bmatrix} \vec{0} & \cdots & \vec{0} \end{bmatrix} = 0_{m\times(n-r)}. \tag{26}$$

Returning to our original computation in eq. (20), we use the results of the previous two calculations to get

$$\begin{aligned} A &= \begin{bmatrix} AV_r & AV_{n-r} \end{bmatrix} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix} && (27) \\ &= \begin{bmatrix} U_r \Sigma_r & 0_{m\times(n-r)} \end{bmatrix} \begin{bmatrix} V_r^\top \\ V_{n-r}^\top \end{bmatrix} && (28) \\ &= U_r \Sigma_r V_r^\top + 0_{m\times(n-r)} V_{n-r}^\top && (29) \\ &= U_r \Sigma_r V_r^\top. && (30) \end{aligned}$$

We have shown that A is equal to its *compact* SVD. To ensure that A is equal to its *full* SVD, we just need to show that the compact SVD is exactly equal to the full SVD. But, we already did this, in particular in eq. (3) through eq. (10)! There, we've shown that the algorithm given by Method 1 produces the correct SVD.

5 Proving Properties of the SVD

The only thing that's left to prove now is the laundry list of properties on the first and second pages, for the construction of the SVD defined by Method 1.

We will do the proofs in a different order than the facts were presented, but this is only to avoid getting any cyclic proofs.

- V is an $n \times n$ matrix with orthonormal columns $\vec{v}_1, \ldots, \vec{v}_n$ that live in $\mathbb{R}^n$.

  - We know that $A^\top A$ is an $n \times n$ real symmetric matrix. Thus by the spectral theorem for real symmetric matrices, $A^\top A$ has n real eigenvalues $\lambda_1, \dots, \lambda_n$ and n orthonormal eigenvectors $\vec{v}_1, \dots, \vec{v}_n$. Our construction sets those as the columns of V in eq. (15), so the columns of V are orthonormal, as we desired.

- U is an $m \times m$ matrix with orthonormal columns $\vec{u}_1, \dots, \vec{u}_m$ that live in $\mathbb{R}^m$.

  - We know from the construction of the algorithm that U is an $m \times m$ matrix and can be written as
$$U = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} = \left[\begin{array}{ccc|ccc} \vec{u}_1 & \cdots & \vec{u}_r & \vec{u}_{r+1} & \cdots & \vec{u}_m \end{array}\right]. \tag{31}$$
To show that U has orthonormal columns, we must show that $U^\top U = I_m$. To do this, we can do the computation:
$$U^\top U = \begin{bmatrix} U_r^\top \\ U_{m-r}^\top \end{bmatrix} \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} = \begin{bmatrix} U_r^\top U_r & U_r^\top U_{m-r} \\ U_{m-r}^\top U_r & U_{m-r}^\top U_{m-r} \end{bmatrix}. \tag{32}$$
Since U_{m-r} was created by Gram-Schmidt to have $m-r$ orthonormal columns $\vec{u}_{r+1}, \dots, \vec{u}_m$, each of which is orthogonal with the columns $\vec{u}_1, \dots, \vec{u}_r$ of U_r, we know that
$$U_r^\top U_{m-r} = \begin{bmatrix} \vec{u}_1^\top \\ \vdots \\ \vec{u}_r^\top \end{bmatrix} \begin{bmatrix} \vec{u}_{r+1} & \cdots & \vec{u}_m \end{bmatrix} = \begin{bmatrix} \vec{u}_1^\top \vec{u}_{r+1} & \cdots & \vec{u}_1^\top \vec{u}_m \\ \vdots & \ddots & \vdots \\ \vec{u}_r^\top \vec{u}_{r+1} & \cdots & \vec{u}_r^\top \vec{u}_m \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} = 0_{r \times (m-r)}. \tag{33}$$
By adapting this calculation, we can also show that
$$U_{m-r}^\top U_r = 0_{(m-r) \times r} \qquad \text{and} \qquad U_{m-r}^\top U_{m-r} = I_{m-r}.$$
The last quantity to compute is $U_r^\top U_r$, and the calculation is slightly different. We can again go column-by-column:
$$U_r^\top U_r = \begin{bmatrix} \vec{u}_1^\top \\ \vdots \\ \vec{u}_r^\top \end{bmatrix} \begin{bmatrix} \vec{u}_1 & \cdots & \vec{u}_r \end{bmatrix} = \begin{bmatrix} \vec{u}_1^\top \vec{u}_1 & \cdots & \vec{u}_r^\top \vec{u}_1 \\ \vdots & \ddots & \vdots \\ \vec{u}_1^\top \vec{u}_r & \cdots & \vec{u}_r^\top \vec{u}_r \end{bmatrix}. \tag{34}$$
This time, we haven't computed any vectors by Gram-Schmidt, so we can't say that everything is immediately zero. Instead, what we can do is use our construction for $\vec{u}_i$, introduced in eq. (16), i.e., $\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i}$ for $i \leq r$. Then we can take the inner product of any two (not necessarily different) $\vec{u}_i$ to get
$$\vec{u}_i^\top \vec{u}_j = \left(\frac{A\vec{v}_i}{\sigma_i}\right)^\top \left(\frac{A\vec{v}_j}{\sigma_j}\right) \tag{35}$$
$$= \frac{\vec{v}_i^\top A^\top A \vec{v}_j}{\sigma_i \sigma_j}. \tag{36}$$
At this point we recall our definition of $\vec{v}_j$ as an eigenvector of $A^\top A$ with eigenvalue λ_j, so we

can write

$$\vec{u}_i^\top \vec{u}_j = \frac{\vec{v}_i^\top A^\top A \vec{v}_j}{\sigma_i \sigma_j} \tag{37}$$

$$= \frac{\vec{v}_i^\top \lambda_j \vec{v}_j}{\sigma_i \sigma_j} \tag{38}$$

$$= \frac{\lambda_j}{\sigma_i \sigma_j} \vec{v}_i^\top \vec{v}_j. \tag{39}$$

Here we know $\vec{v}_i$ and $\vec{v}_j$ are columns of V. But we have just proven that the columns of V are orthonormal! So we already know

$$\vec{v}_i^\top \vec{v}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}. \tag{40}$$

Putting it all together,

$$\vec{u}_i^\top \vec{u}_j = \frac{\lambda_j}{\sigma_i \sigma_j} \vec{v}_i^\top \vec{v}_j = \frac{\lambda_j}{\sigma_i \sigma_j} \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} = \frac{\lambda_i}{\sigma_i^2} \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{41}$$

where in the last equality we use step (c) of Method 1 to show that $\sigma_i^2 = \lambda_i$. Thus

$$U_r^\top U_r = \begin{bmatrix} \vec{u}_1^\top \vec{u}_1 & \cdots & \vec{u}_r^\top \vec{u}_1 \\ \vdots & \ddots & \vdots \\ \vec{u}_1^\top \vec{u}_r & \cdots & \vec{u}_r^\top \vec{u}_r \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} = I_r. \tag{42}$$

Given everything we have found out, we can start again from eq. (32) and get

$$U^\top U = \begin{bmatrix} U_r^\top U_r & U_r^\top U_{m-r} \\ U_{m-r}^\top U_r & U_{m-r}^\top U_{m-r} \end{bmatrix} \tag{43}$$

$$= \begin{bmatrix} I_r & 0_{r \times (m-r)} \\ 0_{(m-r) \times r} & I_{m-r} \end{bmatrix} \tag{44}$$

$$= I_m. \tag{45}$$

Thus the columns of U are orthonormal, as we wanted to show.

- Σ is an $m \times n$ matrix which has an $r \times r$ diagonal block Σ_r in the upper left, and 0 elsewhere. Σ_r is an $r \times r$ matrix with the largest r *singular values* $\sigma_1 \geq \cdots \geq \sigma_r > 0$ of A.
  - Since we get the singular values and lay it out in the prescribed form during our construction, we just need to show that we actually get exactly r positive singular values. We know from the spectral theorem for real symmetric matrices that $A^\top A$ has n real eigenvalues, and r nonzero eigenvalues.

    We now claim that these nonzero eigenvalues are all positive. In fact, let $\vec{v}$ be an eigenvector of $A^\top A$ with eigenvalue λ. Then

$$A^\top A \vec{v} = \lambda \vec{v}. \tag{46}$$

Left-multiplying by $\vec{v}^\top$,we get

$$\vec{v}^\top A^\top A\vec{v} = \lambda\vec{v}^\top\vec{v}. \tag{47}$$

But the left-hand side can be simplified:

$$\vec{v}^\top A^\top A\vec{v} = (A\vec{v})^\top(A\vec{v}) = \|A\vec{v}\|^2. \tag{48}$$

So can the right-hand side:

$$\lambda\vec{v}^\top\vec{v} = \lambda\|\vec{v}\|^2. \tag{49}$$

Thus

$$\|A\vec{v}\|^2 = \lambda\|\vec{v}\|^2. \tag{50}$$

We know that both $\|A\vec{v}\|^2$ and $\|\vec{v}\|^2$ are squared terms, so they are both ≥ 0. Thus $\lambda \geq 0$ as well. This is true for any arbitrary eigenvalue of $A^\top A$, so every eigenvalue of $A^\top A$ is non-negative. Thus if an eigenvalue is nonzero then it must be positive. Finally, in step (c) of Method 1, we set $\sigma_i = \sqrt{\lambda_i}$ for all eigenvalues λ_i, so there are r positive σ_i as well.

- U_r, V_r, U_{m-r}, V_{n-r} are sub-matrices of U and V with orthonormal columns.
  - This sub-matrix breakdown is just how we construct U and V. The orthonormality comes from the fact that the columns of U and V are orthonormal, so any subset of them will also be orthonormal. This includes the subsets of columns that become the columns of U_r, etc.
- $\text{Col}(V_{n-r}) = \text{Null}(A)$.
  - We first want to show that $\text{Col}(V_{n-r}) \subseteq \text{Null}(A)$, then $\text{Null}(A) \subseteq \text{Col}(V_{n-r})$. This implies that $\text{Col}(V_{n-r}) = \text{Null}(A)$.

    First, take any $\vec{v} \in \text{Col}(V_{n-r})$. We want to show that $\vec{v} \in \text{Null}(A)$. Let $\vec{w}$ be such that $\vec{v} = V_{n-r}\vec{w}$; we know $\vec{w}$ exists by the definition of $\text{Col}(V_{n-r})$. We want to show that $\vec{v} \in \text{Null}(A)$, so the natural thing to do is to multiply by A:

$$\begin{aligned} A\vec{v} &= AV_{n-r}\vec{w} & (51)\\ &= A\begin{bmatrix}\vec{v}_{r+1} & \cdots & \vec{v}_n\end{bmatrix}\vec{w} & (52)\\ &= \begin{bmatrix}A\vec{v}_{r+1} & \cdots & A\vec{v}_n\end{bmatrix}\vec{w}. & (53)\end{aligned}$$

    At this point we would like to stop and consider what each of these columns are. We know that since $\vec{v}_{r+1}, \ldots, \vec{v}_n$ are eigenvectors of $A^\top A$ with eigenvalue 0,

$$A^\top A\vec{v}_{r+1} = \cdots = A^\top A\vec{v}_n = \vec{0}. \tag{54}$$

    Therefore $\vec{v}_{r+1}, \ldots, \vec{v}_n \in \text{Null}\left(A^\top A\right)$. But from EECS16A Note 23, we know that for any matrix A, we have

$$\text{Null}(A) = \text{Null}\left(A^\top A\right) \tag{55}$$

    Thus $\vec{v}_{r+1}, \ldots, \vec{v}_n \in \text{Null}(A)$, and so

$$\begin{aligned} A\vec{v} &= \begin{bmatrix}A\vec{v}_{r+1} & \cdots & A\vec{v}_n\end{bmatrix}\vec{w} & (56)\\ &= \begin{bmatrix}\vec{0} & \cdots & \vec{0}\end{bmatrix}\vec{w} & (57)\end{aligned}$$

$$= \vec{0}. \tag{58}$$

Thus $v \in \text{Null}(A)$. Since v is an arbitrary vector in $\text{Col}(V_{n-r})$,

$$\text{Col}(V_{n-r}) \subseteq \text{Null}(A). \tag{59}$$

Now we want to try the other direction. Take any $\vec{v} \in \text{Null}(A)$. We want to show that $\vec{v} \in \text{Col}(V_{n-r})$. Since $\vec{v} \in \text{Null}(A)$,

$$A\vec{v} = \vec{0} \tag{60}$$

Left-multiplying by $A^\top$,

$$A^\top A\vec{v} = A^\top \vec{0} = \vec{0} = 0\vec{v}. \tag{61}$$

Thus $\vec{v}$ is an eigenvector of $A^\top A$ with eigenvalue 0, so it's contained in the span of the eigenvectors of $A^\top A$ associated with the eigenvalue 0. But a basis for these eigenvectors is exactly $\vec{v}_{r+1}, \ldots, \vec{v}_n$, and so this span is $\text{Col}(V_{n-r})$. Thus $\vec{v} \in \text{Col}(V_{n-r})$. Since $\vec{v}$ is an arbitrary vector in $\text{Null}(A)$,

$$\text{Null}(A) \subseteq \text{Col}(V_{n-r}). \tag{62}$$

Thus by eq. (59) and eq. (62),

$$\text{Null}(A) = \text{Col}(V_{n-r}), \tag{63}$$

which is what we wanted to show.

- $\text{Col}(V_r) \perp \text{Null}(A)$.
  - Since V has orthonormal columns, we know that every vector in $\vec{v}_1, \ldots, \vec{v}_r$ is orthogonal to every vector in $\vec{v}_{r+1}, \ldots, \vec{v}_n$. Thus each of $\vec{v}_1, \ldots, \vec{v}_r$ is orthogonal to $\text{span}(\vec{v}_{r+1}, \ldots, \vec{v}_n) = \text{Col}(V_{n-r})$. Thus any linear combination of $\vec{v}_1, \ldots, \vec{v}_r$ is too, so

$$\text{span}(\vec{v}_1, \ldots, \vec{v}_r) \perp \text{Col}(V_{n-r}). \tag{64}$$

But the first term is just $\text{Col}(V_r)$ by definition, and we just proved that $\text{Col}(V_{n-r}) = \text{Null}(A)$. So

$$\text{Col}(V_r) \perp \text{Null}(A). \tag{65}$$

- $\text{Col}(U_r) = \text{Col}(A)$.
  - Remember that the columns of U_r are $\vec{u}_1, \ldots, \vec{u}_r$. From eq. (32) we obtain $\vec{u}_i = \frac{A\vec{v}_i}{\sigma_i}$, which is proportional to $A\vec{v}_i$.

We already showed that $\vec{u}_1, \ldots, \vec{u}_m$ are orthonormal and hence $\vec{u}_1, \ldots, \vec{u}_r$ are orthonormal. This takes care of the proportionality, so all we need to show is that $\text{span}(A\vec{v}_1, \ldots, A\vec{v}_r) = \text{Col}(A)$. But $A\vec{v}_1, \ldots, A\vec{v}_r$ are exactly the columns of AV_r, so the left hand side is $\text{Col}(AV_r)$.

First, we want to show $\text{Col}(A) \subseteq \text{Col}(AV_r)$. Let $\vec{v} \in \text{Col}(A)$; we want to show that $\vec{v} \in \text{Col}(AV_r)$. Indeed, let $\vec{w}$ be such that $\vec{v} = A\vec{w}$; we know $\vec{w}$ exists by the definition of $\text{Col}(A)$. Then since $\vec{v}_1, \ldots, \vec{v}_n$ is an orthonormal basis for $\mathbb{R}^n$ and also the columns of V, there is a $\vec{y}$ such that $\vec{w} = V\vec{y}$. Then

$$\vec{v} = A\vec{w} = AV\vec{y}. \tag{66}$$

Writing V in terms of V_r and V_{n-r},

$$\vec{v} = AV\vec{y} = A\begin{bmatrix} V_r & V_{n-r} \end{bmatrix}\vec{y} = \begin{bmatrix} AV_r & AV_{n-r} \end{bmatrix}\vec{y}. \tag{67}$$

But we already proved that

$$\mathrm{Col}(V_{n-r}) = \mathrm{Null}(A) \tag{68}$$

so

$$AV_{n-r} = 0_{m\times(n-r)}. \tag{69}$$

Thus

$$\vec{v} = \begin{bmatrix} AV_r & AV_{n-r} \end{bmatrix} \vec{y} = \begin{bmatrix} AV_r & 0_{m\times(n-r)} \end{bmatrix} \vec{y}. \tag{70}$$

Thus $\vec{v}$ is a linear combination of the columns of AV_r, so $\vec{v} \in \mathrm{Col}(AV_r)$. Since this is true for arbitrary $\vec{v}$,

$$\mathrm{Col}(A) \subseteq \mathrm{Col}(AV_r). \tag{71}$$

The reverse direction is easier. Take $\vec{v} \in \mathrm{Col}(AV_r)$. Then there is $\vec{w}$ such that

$$\vec{v} = AV_r\vec{w} = A(V_r\vec{w}). \tag{72}$$

So in the end $\vec{v}$ is a linear combination of columns of A, hence $\vec{v} \in \mathrm{Col}(A)$. Since this is true for arbitrary $\vec{v}$,

$$\mathrm{Col}(AV_r) \subseteq \mathrm{Col}(A). \tag{73}$$

Hence by eq. (71) and eq. (73),

$$\mathrm{Col}(AV_r) = \mathrm{Col}(A). \tag{74}$$

To wrap up the proof, recall that we showed at the beginning that the columns of AV_r, i.e., the vectors $A\vec{v}_1, \dots, A\vec{v}_r$ are scaled versions of $\vec{u}_1, \dots, \vec{u}_r$. Thus they span the same set, which is $\mathrm{Col}(A)$, as we desired.

- $\mathrm{Col}(U_{m-r}) \perp \mathrm{Col}(A)$.
  - Since U has orthonormal columns, every vector in $\vec{u}_1, \dots, \vec{u}_r$ is orthogonal to every vector in $\vec{u}_{r+1}, \dots, \vec{u}_m$. Thus each of $\vec{u}_{r+1}, \dots, \vec{u}_m$ is orthogonal to $\mathrm{span}(\vec{u}_1, \dots, \vec{u}_r) = \mathrm{Col}(U_r)$. Thus any linear combination of $\vec{u}_{r+1}, \dots, \vec{u}_m$ is too, so

    $$\mathrm{span}(\vec{u}_{r+1}, \dots, \vec{u}_m) \perp \mathrm{Col}(U_r). \tag{75}$$

    But the first term is just $\mathrm{Col}(U_{m-r})$ by definition, and we just proved that $\mathrm{Col}(U_r) = \mathrm{Col}(A)$. So

    $$\mathrm{Col}(U_{m-r}) \perp \mathrm{Col}(A) \tag{76}$$

    as desired.
- The columns of V are eigenvectors of $A^\top A$.
  - In the second step of the construction of Method 1, one sets the columns of V to the $\vec{v}_1, \dots, \vec{v}_n$ which are orthonormal eigenvectors of $A^\top A$.
- The columns of U are eigenvectors of $AA^\top$.
  - We first show that $\vec{u}_1, \dots, \vec{u}_r$ are eigenvectors of $AA^\top$. To do so, we may as well compute $AA^\top\vec{u}_i$ and see what we get. More specifically, for $1 \le i \le r$,

    $$AA^\top\vec{u}_i = AA^\top\left(\frac{A\vec{v}_i}{\sigma_i}\right) \tag{77}$$

$$
\begin{aligned}
&= \frac{1}{\sigma_i} AA^\top A\vec{v}_i && (78)\\
&= \frac{1}{\sigma_i} A(A^\top A\vec{v}_i) && (79)\\
&= \frac{1}{\sigma_i} A(\lambda_i\vec{v}_i) && (80)\\
&= \frac{\lambda_i}{\sigma_i} A\vec{v}_i && (81)\\
&= \lambda_i \frac{A\vec{v}_i}{\sigma_i} && (82)\\
&= \lambda_i\vec{u}_i. && (83)
\end{aligned}
$$

Hence $\vec{u}_i$ is an eigenvector of $AA^\top$.

We then show that $\vec{u}_{r+1}, \ldots, \vec{u}_m$ are eigenvectors of $AA^\top$. Since $\operatorname{rank} A = r$, we know $\operatorname{rank}(AA^\top) = r$, so $AA^\top$ has r nonzero eigenvalues, corresponding to $\vec{u}_1, \ldots, \vec{u}_r$. Then the remaining $m - r$ eigenvalues of $AA^\top$ are equal to 0. Since $\operatorname{span}(\vec{u}_1, \ldots, \vec{u}_m) = \mathbb{R}^m$, the remaining $m-r$ eigenvectors of $AA^\top$ span exactly the same space as $\operatorname{span}(\vec{u}_{r+1}, \ldots, \vec{u}_m)$. Since each eigenvector corresponding to 0 of $AA^\top$ has the same eigenvalue, any linear combination of them is also an eigenvector of $AA^\top$ corresponding to 0. Thus $\vec{u}_{r+1}, \ldots, \vec{u}_m$ are eigenvectors of $AA^\top$ corresponding to 0.

- The diagonal entries of Σ are eigenvalues of $A^\top A$ or $AA^\top$.
  - In the third step of Method 1 algorithm, we calculate the diagonal entries of Σ by taking the square roots of the eigenvalues of $A^\top A$. So we just want to show that the eigenvalues of $A^\top A$ are exactly the eigenvalues of $AA^\top$. We will show this in the following way.

    Suppose $(\vec{v}, \lambda)$ is an eigenvector-eigenvalue pair for $A^\top A$.

    To show that λ is an eigenvalue of $AA^\top$, the easiest thing we could do is to find a vector $\vec{u}$ such that $AA^\top\vec{u} = \lambda\vec{u}$, and thus conclude that $(\vec{u}, \lambda)$ is an eigenvector-eigenvalue pair for $AA^\top$.

    But we don't have any vector $\vec{u}$. We know that it should depend on A in some way, and also that maybe it could depend on $\vec{v}$. To figure out exactly what we can try, let's check the dimensions. More precisely, we know that $AA^\top$ is an $m \times m$ matrix, while $A^\top A$ is an $n \times n$ matrix. This means that $\vec{v}$ is a length-n vector, but we want $\vec{u}$ to be a length-m vector. The easiest way we can get a length m vector from a length n vector is just to multiply by A, which is an $m \times n$ matrix and thus multiplies length n vectors to length m vectors. So we will guess $\vec{u} = A\vec{v}$ and see if it works.

    Indeed, let us test our guess in the eigenvalue equation. In other words, we will multiply $\vec{u}$ by $AA^\top$ and see if we get something involving $\vec{u}$ on the right-hand side.

$$
\begin{aligned}
AA^\top\vec{u} &= AA^\top(A\vec{v}) && (84)\\
&= A(A^\top A\vec{v}) && (85)\\
&= A(\lambda\vec{v}) && (86)\\
&= \lambda A\vec{v} && (87)\\
&= \lambda\vec{u}. && (88)
\end{aligned}
$$

    Great, so it all works out! Indeed, $(\vec{u}, \lambda)$ is an eigenvector-eigenvalue pair for $AA^\top$. And so λ is an eigenvalue of $AA^\top$, as desired.

To show that if λ is an eigenvalue of $AA^\top$ then it is an eigenvector of $A^\top A$, one can show it in the same way; namely, if $(\vec{u}, \lambda)$ is an eigenvector-eigenvalue pair for $AA^\top$, and if we define $\vec{v} = A^\top \vec{u}$, then we can show in the same way that $(\vec{v}, \lambda)$ is an eigenvector-eigenvalue pair for $A^\top A$. And so if λ is an eigenvalue of $AA^\top$, then it is an eigenvector for $A^\top A$.
So the diagonal entries of Σ are the eigenvalues of $AA^\top$ and $A^\top A$, since they are exactly the same eigenvalues.

We proved all the properties we wanted to prove, for Method 1 of finding the SVD. The corresponding properties for Method 2 can be similarly verified.

6 Geometric Interpretation

To reiterate, let $A = U\Sigma V^\top$ be the full SVD.

Note that multiplying a vector $\vec{x}$ by an orthonormal matrix U does not change its norm. This follows because $U^\top U = I$, which implies

$$\|U\vec{x}\|^2 = (U\vec{x})^\top (U\vec{x}) = \vec{x}^\top U^\top U\vec{x} = \vec{x}^\top \vec{x} = \|\vec{x}\|^2 . \tag{89}$$

Thus we can interpret multiplication by an orthonormal matrix as a combination of operations that don't change length, such as rotations, and reflections.

Since Σ_r is diagonal with entries $\sigma_1, \ldots, \sigma_r$, multiplying a vector by Σ stretches the first entry of the vector by σ_1, the second entry by σ_2, and so on.

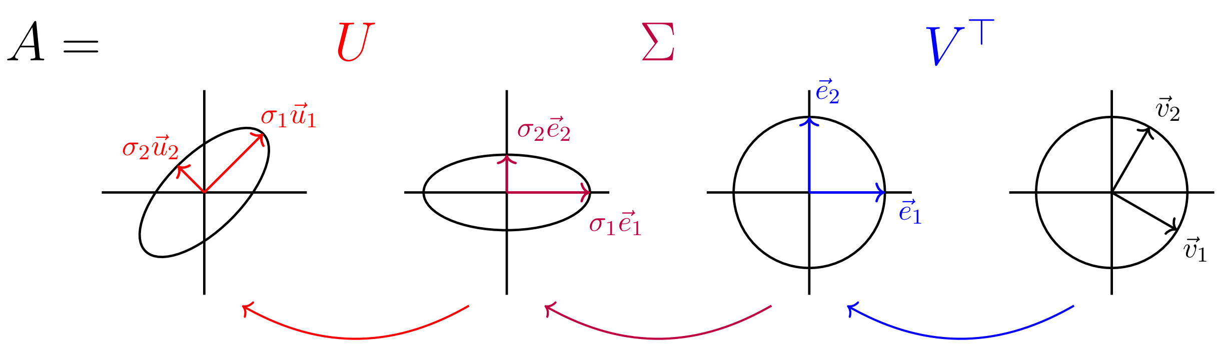
Combining these observations, we interpret $A\vec{x}$ as the composition of three operations:

(a) $V^\top \vec{x}$ which rotates $\vec{x}$ without changing its length.

(b) $\Sigma V^\top \vec{x}$ which stretches the resulting vector along each axis with the corresponding singular value,

(c) $U\Sigma V^\top \vec{x}$ which again rotates the resulting vector without changing its length.

The following figure illustrates these three operations moving from the right to the left.



Here as usual $\vec{e}_1, \vec{e}_2$ are the first and second standard basis vectors.

The geometric interpretation above reveals that σ_1 is the largest amplification factor a vector can experience upon multiplication by A. More specifically, if $\|\vec{x}\| \leq 1$ then $\|A\vec{x}\| \leq \sigma_1$. We achieve equality at $\vec{x} = \vec{v}_1$, because then $V^\top \vec{v}_1$ is the first unit vector, which gets magnified by σ_1 when multiplied by Σ to end up with a total length of σ_1.

7 Examples

Example SVD Interpretation.

Suppose we have an $m \times n$ matrix A, of rank r, that contains the ratings of m viewers for n movies. Write

$$A = U\Sigma V^\top = \sum_{i=1}^{r} \sigma_i \vec{u}_i \vec{v}_i^\top. \tag{90}$$

We can interpret each rank 1 matrix $\sigma_i \vec{u}_i \vec{v}_i^\top$ to be due to a particular attribute, e.g., comedy, action, sci-fi, or romance content. Then σ_i determines how strongly the ratings depend on the i^{th} attribute; the entries of $\vec{v}_i^\top$ score each movie with respect to this attribute, and the entries of $\vec{u}_i$ evaluate how much each viewer cares about this particular attribute. Interestingly, the $(r+1)^{\text{th}}$ attributes onwards don't influence the ratings, according to our analysis.

Numerical Example 1.

Let's find the SVD for

$$A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}. \tag{91}$$

We use **Method 2**. We calculate

$$AA^\top = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}. \tag{92}$$

This happens to be diagonal, so we can read off the eigenvalues:

$$\lambda_1 = 32 \qquad \lambda_2 = 18 \tag{93}$$

We can select the orthonormal eigenvectors:

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \vec{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{94}$$

The singular values are

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{32} = 4\sqrt{2} \qquad \sigma_2 = \sqrt{\lambda_2} = \sqrt{18} = 3\sqrt{2}. \tag{95}$$

Then to find $\vec{v}_1, \vec{v}_2$, we do

$$\vec{v}_1 = \frac{A^\top \vec{u}_1}{\sigma_1} = \frac{1}{4\sqrt{2}} \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \tag{96}$$

$$\vec{v}_2 = \frac{A^\top \vec{u}_2}{\sigma_2} = \frac{1}{3\sqrt{2}} \begin{bmatrix} -3 \\ 3 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}. \tag{97}$$

Thus our SVD is

$$A = U\Sigma V^\top = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}. \tag{98}$$

Note that we can change the signs of $\vec{u}_1, \vec{u}_2$ and they are still orthonormal eigenvectors, and produce a valid SVD. However, changing the sign of $\vec{u}_i$ requires us to change the sign of $\vec{v}_i = A^\top \vec{u}_i$, so therefore the product of $\vec{u}_i \vec{v}_i^\top$ remains unchanged.

Another source of non-uniqueness arises when we have repeated singular values, as seen in the next example.

Numerical Example 2.

We want to find an SVD for

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{99}$$

Again, we use **Method 2**. Note that $AA^\top = I_2$, which has repeated eigenvalues at $\lambda_1 = \lambda_2 = 1$. In particular, *any* pair of orthonormal vectors is a set of orthonormal eigenvectors for $I_2 = AA^\top$. We can parameterize all such pairs as

$$\vec{u}_1 = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} \qquad \vec{u}_2 = \begin{bmatrix} -\sin(\theta) \\ \cos(\theta) \end{bmatrix} \tag{100}$$

where θ is a free parameter. Since $\sigma_1 = \sigma_2 = 1$, we obtain

$$\vec{v}_1 = \frac{A^\top \vec{u}_1}{\sigma_1} = \begin{bmatrix} \cos(\theta) \\ -\sin(\theta) \end{bmatrix} \qquad \vec{v}_2 = \frac{A^\top \vec{u}_2}{\sigma_2} = \begin{bmatrix} -\sin(\theta) \\ -\cos(\theta) \end{bmatrix}. \tag{101}$$

Thus an SVD is

$$A = U\Sigma V^\top = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ -\sin(\theta) & -\cos(\theta) \end{bmatrix} \tag{102}$$

for any value of θ. Thus this matrix has *infinite* valid SVDs, one for each value of θ in the interval $[0, 2\pi)$.

Contributors:

- Druv Pai.
- Gireeja Ranade.
- Murat Arcak.