Matrix Decompositions

In this note, we review two common matrix decompositions, the *Singular Value Decomposition* and the *Spectral Decomposition*. Throughout, we are going to use the notation e_i to refer to the vector with a one at the i-th entry, and a zero at all other entries.

1 Eigenvalues and Diagonalization

Let $A \in \mathbb{R}^{d\times d}$ be a square matrix. For now, we will assume that A is *symmetric*, which means that $A = A^\top$. Being symmetric is equivalent to having a spectral decomposition:

Theorem 1 (Spectral Decomposition)**.** *Let $A \in \mathbb{R}^{d\times d}$. Then $A = A^\top$ if and only if A can be written as*

$$A = U\Lambda U^\top,$$

where $\Lambda \in \mathbb{R}^{d\times d}$ is a diagonal matrix, and $U \in \mathbb{R}^{d\times d}$ satisfies $U^\top U = I_{d\times d}$ (U is known as an orthogonal matrix).

We won't prove the spectral theorem here, but you can look it up in your favorite reference on linear algebra. We also remark that it is a special case of the *eigendecomposition*, which applies more generally to many (but not all!) square matrices.

The spectral theorem has the following nice interpretation. Let $u_1, u_2, \dots, u_d \in \mathbb{R}^d$ denote the columns of U, and $\lambda_1, \lambda_2, \dots, \lambda_d$ the diagonals of Λ. Then, we can write out

$$\begin{aligned}
A &= U\Lambda U^\top \\
&= \begin{bmatrix} u_1 & u_2 & \dots & u_d \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & \lambda_d \end{bmatrix} \begin{bmatrix} u_1^\top \\ u_2^\top \\ \dots \\ u_d^\top \end{bmatrix} \\
&= \sum_{i=1}^{d} \lambda_i u_i u_i^\top.
\end{aligned}$$

Since $U^\top U = I$, we can check that $\langle u_i, u_j\rangle = \mathbf{I}(i = j)$. Thus, $\{u_i\}$ are an orthogonal basis for $\mathbb{R}^d$, and the matrix $u_i u_i^\top$ is the rank-one projection onto the span of u_i. Moreover, each u_i is an eigenvector of A, since

$$Au_i = \sum_{j=1}^{d} \lambda_j u_j u_j^\top u_i = \sum_{j=1}^{d} \lambda_j u_j \mathbf{I}(i = j) = \lambda_i u_i.$$

Hence, the spectral decomposition expresses A as a sum of weighted projections onto the span of its eigenvectors.

1.1 Positive Semidefinite matrices

A symmetric matrix $A = A^\top \in \mathbb{R}^{d\times d}$ is said to be *positive semidefinite* or *PSD* if $\langle v, Av\rangle \geq 0$ for all vectors $v \in \mathbb{R}^d$. We write $A \succeq 0$ to denote that A is PSD. Using the spectral decomposition, we have the following necessary and suffient condition that A is PSD:

Proposition 1. *A symmetric matrix A is PSD if and only if $\lambda_{\min}(A) \geq 0$, where $\lambda_{\min}$ denotes the smallest eigenvalue of A.*

Proof. Since A is symmetric, we may let λ_i and u_i be as in the spectral decomposition. Let us order λ so that $\lambda_d = \lambda_{\min}(A)$. Then,

$$v^\top A v = v^\top (\sum_{i=1}^d \lambda_i u_i u_i^\top) v = \sum_{i=1}^d \lambda_i \langle v, u_i\rangle^2.$$

First, suppose $A \succeq 0$. Then, by choosing $v = u_d$, we have

$$0 \leq u_d^\top A u_d = \sum_{i=1}^d \lambda_i \langle u_d, u_i\rangle^2 = \sum_{i=1}^d \lambda_i \mathbf{I}(i = d) = \lambda_d = \lambda_{\min}(A).$$

For the other direction, we have

$$v^\top A v = \sum_{i=1}^d \lambda_i \langle u_d, u_i\rangle^2 \geq (\min_{i=1}^d \lambda_i) \sum_{i=1}^d \rangle \langle v, u_i\rangle^2 = \lambda_{\min}(A) \cdot \|v\|_2^2,$$

where the last step uses the fact that $\{u_i\}$ are an orthogonal basis for $\mathbb{R}^d$. Hence, if $\lambda_{\min}(A) \geq 0$, $v^\top A v \geq 0$. $\square$

Here is another fact about PSD matrices:

Proposition 2. *A symmetric matrix $A \in \mathbb{R}^{d\times d}$ is PSD if and only if there exists a matirx $B \in \mathbb{R}^{d\times r}$ such that $A = BB^\top$. Moreover, if A is PSD, then there exists a PSD symmetric matrix, denoted $A^{1/2} \in \mathbb{R}^{d\times d}$, such that $(A^{1/2})^2 = A$.*

Proof. We show that if $A = BB^\top$, then A is PSD. Observe that A is symmetric, since $(BB^\top)^\top = (B^\top)^\top B^\top = BB^\top$. Moreover, for any $v \in \mathbb{R}^d$, $v^\top A v = v^\top B B^\top v = (B^\top v)^\top (B^\top v) = \|B^\top v\|_2^2 \geq 0$. The rest is left as an exercise. (*Hint:* Use the spectral decomp to show $A^{1/2}$ exists, and choose $B = A^{1/2}$)). $\square$

2 Singular Value Decomposition

Recall that the spectral decomposition gives us a way of expressing symmetric, square matrices. We can use the spectral decomposition to derive a more general decomposition for *arbitrary, rectangular matrices:*

Theorem 2 (SVD). *Any rectangular matrix $X \in \mathbb{R}^{n\times d}$ can be written as*

$$X = U\Sigma V^\top,$$

where $U \in \mathbb{R}^{n\times n}$, $V \in \mathbb{R}^{d\times d}$, and $\Sigma \in \mathbb{R}^{n\times d}$. Moreover, $U^\top U = I_{n\times n}$, $V^\top V = I_{d\times d}$, and Σ has nonnegative entries *along its first* $\min\{n,d\}$ *diagonals, and zeros elsewhere. The decomposition $X = U\Sigma V^\top$, with the dimensions given above, is known as the (full) singular vale decomposition.*

To make things clearer, we will assume that $n \geq d$ throughout, we can always get that $d \geq n$ case by taking transposes. First, let's write out X as a sum of rank one terms, just like in the spectral decomposition. Then,

$$X = U\Sigma V^\top = \sum_{i=1}^{d} \sigma_i u_i v_i^\top, \tag{1}$$

where u_i are the columns of U, v_i are the columns of V, and σ_i the diagonals. We call u_i the left singular vectors, v_i the right-singular vectors, and σ_i the singular values. Typically, we order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d$.

Recall that since $U^\top U = I_{n\times n}$, and $V^\top V = I_{d\times d}$, we see that $\{u_i\}_{i=1}^n$ and $\{v_i\}_{i=1}^d$ are orthonormal bases for $\mathbb{R}^n$ and $\mathbb{R}^d$, respectively. Observe that since right and left inverses are the same, we have $V^\top V = VV^\top = I_{d\times d}$ and $V^\top = V^{-1}$, and $UU^\top = U^\top U = I_{n\times n}$ and $U^\top = U^{-1}$.

Thus, the SVD of X has the interpretation that ...

2.1 The compact SVD

You might have noticed that even those U has n-columns, only $d \leq n$ of those columns show up in rank-one SVD in Equation 1. One way to see this that when $n \geq d$, Σ has the form

$$\Sigma = \begin{bmatrix} \Sigma_{1:d} \\ \mathbf{0}_{(n-d)\times d} \end{bmatrix}$$

where $\Sigma_{1:d}$ is the matrix with $\sigma_1, \ldots, \sigma_n$ on the diagonal, and thus partition U into the first d and last $(n-d)$ columns, we have

$$X = U\Sigma V^\top = \begin{bmatrix} U_{1:d} & U_{d+1:n} \end{bmatrix} \begin{bmatrix} \Sigma_{1:d} \\ \mathbf{0}_{(n-d)\times d} \end{bmatrix} V^\top = U_{1:d}\Sigma_{1:d}V^\top.$$

The expression $X = U_{1:d}\Sigma_{1:d}V^\top$, where $U_{1:d} \in \mathbb{R}^{n\times d}$ and $\Sigma_{1:d} \in \mathbb{R}^{d\times d}$ is known as an *compact* SVD, because we have removed the extra zero packing from our original Σ. Now, $\Sigma_{1:d} \in \mathbb{R}^{d\times d}$ is a true diagonal matrix, where $\Sigma \in \mathbb{R}^{n\times d}$ was almost diagonal, but rectangular. For example,

$$\Sigma_{1:d}^2 = \Sigma_{1:d}^\top \Sigma_{1:d} = \Sigma_{1:d}\Sigma_{1:d}^\top,$$

but if we used the full Σ, then

$$\Sigma^\top\Sigma = \Sigma_{1:d}^2 \in \mathbb{R}^{d\times d} \quad \text{but} \quad \Sigma\Sigma^\top = \begin{bmatrix} \Sigma_{1:d} & \mathbf{0}_{(n-d)\times d} \\ \mathbf{0}_{(n-d)\times d} & \mathbf{0}_{(n-d)\times(n-d)} \end{bmatrix} \mathbb{R}^{n\times n}.$$

Moreover, when $d < n$, we can show $U_{1:d}^\top U_{1:d} = I_{d\times d}$. However, $U_{1:d}U_{1:d}^\top \in \mathbb{R}^{n\times n}$, and cannot be equal to the identity since it has rank at most $d < n$. In fact, from the homework, we see that $U_{1:d}U_{1:d}^\top = \sum_{i=1}^d u_i u_i^\top$ is the *orthogonal projection* onto the first d singular values.

2.2 A super-compact SVD

We can make the SVD even more compact by removing the 0-singular values. Indeed, we have the following lemma:

Lemma 1. $\operatorname{rank}(X) = r$ *if and only if* X *has exactly* r *nonzero singular vectors.*

There are many ways to prove this lemma. For example, if we use the non-compact SVD, U and V are full rank, so $\operatorname{rank}(X) = \operatorname{rank}(\Sigma)$, which is just the number of non-zero entries on the diagonal of Σ. For rank-r matrices X, we can write

$$X = \sum_{i=1}^{r} \sigma_i u_i v_i^\top,$$

where we ordered the nonzero singular values $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_r$. As above, we can write an even more compact SVD

$$X = U_{1:r} \Sigma_{1:r} V_{1:r}^\top.$$

Note that when $r = d$ (X is full rank), then this "more compact" SVD coincides with the regular compact SVD.

2.3 Which SVD should I to use?

Many textbooks in China and India teach this "more compact" SVD as the cannonical (standard) SVD, whereas textbooks in the US (and Wikipedia) often start with the full SVD. Because we have students from diverse backgrounds, in the course we shall request that

> When writing $X = U\Sigma V^\top$, **you must specify that you are taking an SVD, and give the dimensions of U, Σ, and V.**

This is good practice more generally, because it ensures that the reader understands what you are writing. Moreover, it will help the grader undestand your train of logic, and avoiding taking off points. To keep things clear in this note, we will take $d \leq n$, use $U\Sigma V^\top$ for the full SVD, and $U_{1:d}\Sigma_{1:d}V^\top$ for the compact SVD, and $U_{1:r}\Sigma_{1:r}V_{1:r}^\top$ for the more compct SVD. Here are some examples of where confusion may arise.

1. If Σ is from the full SVD, then Σ is rectangular, and thus Σ^{-1} is not defined for $d \neq n$. If we use $\Sigma_{1:d}$, then the inverse may not exists because Σ may not have full rank. But if we use $\Sigma_{1:r}$, then $\Sigma_{1:r}$ is square and full rank, and thus invertible. If you want to invert Σ or $\Sigma_{1:d}$, you will need the pseuodinverse.

2. If we use U form the full SVD, then $UU^\top = U^\top U = I_{n\times n}$. But if we use $U_{1:d}$, then $U_{1:d}^\top U_{1:d} = I_{d\times d}$, but $U_{1:d}U_{1:d}^\top$ is a projection matrix that may not have a clear interpretation. And, if we use $U_{1:r}$, then $U_{1:r}^\top U_{1:r} = I_{r\times r}$, and $U_{1:r}U_{1:r}^\top$ is the projection onto the span of the non-zero left singular vectors u_i, which we can show (see Discussion Sheet 2), is the *rowspace of* X.

At the end of the day, which SVD you use is a matter of personal preference. Some people like having $\Sigma_{1:r}$ be square and invertible, others enjoy have U and V be square, orthogonal matrices. It's up to you, *as long as you are clear*.

3 Relating the SVD to the Spectral Decomposition

In general, the SVD and the Spectral Decomposition are not the same. Here are some key differences:

1. The spectral decomposition applies to symmetric (square) matrices, whereas the SVD applies to all matrices.

2. In the spectral decomposition, $A = U\Lambda U^\top \in \mathbb{R}^{d\times d}$, that is, we have a U on both sides of the diagonal matrix. In the SVD, we have $X = U\Sigma V^\top \in \mathbb{R}^{n\times d}$, where U and V may be different (and have different dimensions when $n \neq d$)

3. Given, $A = U\Lambda U^\top$, the diagonals of Λ are *eigenvalues* of A, that is, there exists $v \in \mathbb{R}^{d\times d}$ for which $Av = \Lambda_{ii} v$ for each $i \in \{1, \ldots, d\}$. For $X = U\Sigma V^\top$, the diagonals of Σ are the *singular values*, which in general are unrelated to eigenvalues. Indeed, if X is rectangular, it cannot have eigenvalues.

4. Similarly, if $A = U\Lambda U^\top$, the columns of U are the *eigenvectors*, whereas if $X = U\Sigma V^\top$, U and V contain left- and right *singular vectors*.

5. Eigenvalues - even of a symmetric matrix - can be positive or negative. Singular values are *always nonnegative*.

Though the singular vectors/values of X have *nothing* to do with eigenvalues/vectors of X in general, they have *everything* to do with the eigenvectors and value of $X^\top X$ and $XX^\top$. Indeed, let $X \in \mathbb{R}^{n\times d}$, for $d \le n$, and let us take the *full* SVD:

$$X = U\Sigma V^\top, \quad U \in \mathbb{R}^{n\times n}, \Sigma \in \mathbb{R}^{n\times d}, V \in \mathbb{R}^{d\times d}.$$

Then,

$$X^\top X = (U\Sigma V^\top)^\top (U\Sigma V^\top) = V\Sigma^\top U^\top U\Sigma V^\top = V\Sigma^\top \Sigma V^\top$$

since $U^\top U = I_{n\times n}$.

$$XX^\top = (U\Sigma V^\top)(U\Sigma V^\top)^\top = U\Sigma V^\top V\Sigma U^\top = U\Sigma\Sigma^\top U^\top.$$

We can check that $\Sigma\Sigma^\top$ and $\Sigma^\top\Sigma$ are diagonal matrices,

$$\Sigma^\top\Sigma = \begin{bmatrix} \Sigma_{1:d}^\top & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Sigma_{1:d} \\ \mathbf{0} \end{bmatrix} = \Sigma_{1:d}^\top \Sigma_{1:d} = \Sigma_{1:d}^2 \overset{\text{def}}{=} \Lambda_{1:d}. \tag{2}$$

And

$$\Sigma\Sigma^\top = \begin{bmatrix} \Sigma_{1:d} \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \Sigma_{1:d} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \Sigma_{1:d}\Sigma_{1:d}^\top & \mathbf{0}_{d\times(n-d)} \\ \mathbf{0}_{(n-d)\times d} & \mathbf{0}_{(n-d)\times(n-d)} \end{bmatrix} \overset{\text{def}}{=} \Lambda_{1:n}. \tag{3}$$

Hence, since $\Lambda_{1:n}$ and $\Lambda_{1:d}$ are diagonal, and $V^\top V = I_{d\times d}$ and $U^\top U = I_{n\times n}$, we see that

$$X^\top X = V\Lambda_{1:d}V^\top \quad \text{and} \quad XX^\top = U\Lambda_{1:n}U^\top$$

give the spectral decomposition of $X^\top X$ and $XX^\top$. We can summarize our findings with the following theorem:

Theorem 3. *Let $X \in \mathbb{R}^{n \times d}$ have full SVD, given by $U\Sigma V^\top$. Then*

1. *The spectral decomposition of $X^\top X$ can be expressed as $V \cdot (\Sigma^\top \Sigma) \cdot V^\top$, and the spectral decomposition of $XX^\top$ can be expressed as $U \cdot (\Sigma\Sigma^\top) \cdot U^\top$*

2. *The left-signular vectors of X are the eigenvectors of $XX^\top$, and the right-singular vectors of X are the eigenvectors of $X^\top X$.*

3. *$X^\top X$ and $XX^\top$ have the same non-zero eigenvalues, which are precisely the squares of the nonzero singular vectors of X*

3.1 The SVD ofsymmetric matrices

In the previous section, we stated how the SVD differs from the spectral decomposition, and focused a lot on dimnesion mismatch. It turns out that that even for symmetric matrices, the SVD and the spectral decomposition are different. Consider the identity matrix $-I$ in $\mathbb{R}^{d \times d}$. Then, we can write, somewhat pedantically, that

$$-I = (I) \cdot (-I) \cdot (I^\top).$$

This is actually an *valid* spectral decomposition of I - $-I$ is a diagonal matrix, and I is an orthogonal matrix. However, it is *not* a valid SVD of $-I$, because $-I$ has negative entries on its diagonal.

On the other hand, we can write

$$-I = (-I) \cdot (I) \cdot (I^\top).$$

Now, this is a valid SVD of I, because the "U" and "V" matrices are orthogonal, and the Σ matrix is diagonal and has non-negative entries. However, it is *not* a valid spectral decomposition, because $(-I)^\top \neq I$.

This example may seem silly, but it illustrates a key difference between eigenvalues and singular values. Let A be an arbitrary symemtric matrix. Then, like $-I$, A may have negative eigenvalues. But, the singular values of A are the square roots eigenvalues of $A^\top A = A^2$ (note A is symmetric). In particuar, one can show the following:

Proposition 3 (SVD for Symmetric Matrices)**.** *Let A be symmetric. Then the singular values of A are the square roots of the eigenvalues of A^2, which are the* absolute values *of the eigenvalue sof A. Hence, the SVD and the spectral decomposition of A can be the same if and only if A is symmetric and has nonnegative eigenvalues; that is, if and only if $A \succeq 0$.*

4 Optional Topics

4.1 SVD for Square Matrices (Optional)

In the previous section, we saw a nice relationship between the eigenvalues of a symmetric matrix and its singular values. For arbitrary square matrices, this may not be the case. For one, square matrices may not even have eigenvalues, like

$$X = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

which is known as a *Jordan block*. Other real square matrix may have *complex* eigenvalues, like the matrix

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

And, even if X has real eigenvalues, they may be totally different form the singular values of X. Indeed, suppose that

$$X = \begin{bmatrix} 1 & K \\ 0 & 2 \end{bmatrix}$$

Then, $\det(\lambda I_{2\times 2} - X) = (\lambda - 1)(\lambda - 2)$, so the eigenvalues of X are $\{1, 2\}$. However,

$$XX^\top = \begin{bmatrix} K^2+1 & 2K \\ 2K & 4 \end{bmatrix}$$

The eigenvalues of this matrix are the roots of

$$\begin{aligned}(K^2+1-\lambda)(4-\lambda) - 4K^2 &= 4K^2 - \lambda K^2 + 4 - \lambda - 4\lambda + \lambda^2 - 4K^2 \\ &= \lambda^2 - \lambda(5+K^2) + 1.\end{aligned}$$

Solving the quadratic equation, we find that we have one eigenvalue of

$$\lambda_+ := \frac{5 + K^2 + \sqrt{K^2+4}}{2}.$$

Not ony is λ_+ not equal to either 0 or 1, but by making K arbitrarily large (positive or negative), we can drive λ_+ to ∞.

4.2 Reduced Spectral Decomposition (Optional)

Let $A = U\Lambda U^\top$ be a symmetric matrix, as in the previous section. If $\mathrm{rank}(\Lambda) = r < d$, then we know that $d - r$ of the eigenvalues of A must be zero, which means that $d - r$ entries of Λ must be zero. Because we can always relabel the entries, we can write Λ so that $\lambda_{r+1}, \ldots, \lambda_d = 0$. Therefore, we may write

$$A = \sum_{i=1}^{r} \lambda_i u_i u_i^\top.$$

Then we can write A is an *reduced* spectral decomposition

$$
\begin{aligned}
A &= \widetilde{U}\widetilde{\Lambda}\widetilde{U}^\top \\
&= \begin{bmatrix} u_1 & u_2 & \dots & u_r \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & \lambda_r \end{bmatrix} \begin{bmatrix} u_1^\top \\ u_2^\top \\ \dots \\ u_r^\top \end{bmatrix},
\end{aligned}
$$

where $\widetilde{U} \in \mathbb{R}^{d\times r}$, and $\widetilde{\Lambda} \in \mathbb{R}^{r\times r}$. This decomposition is called *reduced* because we have removed the extra zero eigenvalues. Keep in mind:

> Whenever you invoke the spectral theorem, it is standard to talk about the non-reduced spectral decomposition with $U \in \mathbb{R}^{d\times d}$ and $\Lambda \in \mathbb{R}^{d\times d}$. If you want to use the *reduced* decomposition in a problem set or exam, **you should make it clear that this is what you are doing!**. The best way to do this is to make sure to specify dimensions of all matrices you write out.

The reduced and non-reduced decompositions have some subtle differences. Because we still have that $\langle u_i, u_j\rangle = \mathbf{I}(i = j)$, we can check that

$$
\widetilde{U}^\top\widetilde{U} = I_{r\times r}.
$$

However, $\widetilde{U}\widetilde{U}^\top \neq I_{d\times d}$; this is impossible because $\mathrm{rank}(\widetilde{U}\widetilde{U}^\top) \le \mathrm{rank}(\widetilde{U}) \le \min\{d, r\} = r < d$. In fact, as we can check on the homework, we see that that since u_i are orthogonal,

$$
\widetilde{U}\widetilde{U}^\top = \sum_{i=1}^{r} u_i u_i^\top
$$

is the *orthogonal projection* onto the span of $(u_1, \dots, u_r)$. From the work in Discussion 2, we can verify that $(u_1, \dots, u_r)$ is a basis for the rowspace of A, and thereofre,

Proposition 4. *If $\widetilde{U} = [u_1|u_2|\dots|u_r]$ is the matrix whose columns are the r-eigenvectors associated with nonzero eigenvalues, or equivalent, the matrix from the reduced spectral decomposition, then $\widetilde{U}\widetilde{U}^\top$ is a rank-r projection matrix, and projects onto the rowspace of A. Since $A = A^\top$, it is also a projection onto the columnspace of A.*