Chapter 2
Matrix factorizations / decompositions

Contents (final version)

2.0 Introduction (s,intro) ......................................................... 2.2
2.1 Spectral Theorem (for symmetric matrices) (s,spectral) ................. 2.4
  Normal matrices (s,normal) ...................................................... 2.6
  Square asymmetric matrices (s,asym) ........................................ 2.8
  Geometry of matrix diagonalization (s,geom0) ................................ 2.9
2.2 SVD (s,svd1) ................................................................. 2.15
  Existence ................................................................. 2.15
  Geometry ................................................................. 2.16
  The matrix 2-norm or spectral norm (s,2norm) ................................ 2.20
  When does $U = V$? (s,svd,uv) ............................................. 2.24
  Positive semidefinite matrices (s,posdef) .................................... 2.26
  Summary (s,summ) ............................................................. 2.28
  SVD computation using eigendecomposition (s,svd,tall) ................... 2.30

2.1
**2.0 Introduction**

One of the main topics in the previous chapter was matrix multiplication. This chapter focuses on matrix factorizations, which is perhaps somewhat like the reverse of matrix multiplication. The reading for this chapter is [1, §10.1, 5.1, 7.4, 10.2].

There are many factorizations used in linear algebra and numerical linear algebra. Here are 7 important ones. The first 5 are for square matrices only. Of these, only the SVD accommodates any matrix size and type.

\[ A = LU \quad M = N \]

**LU decomposition by Gaussian elimination**

- \(L\) is lower triangular; \(U\) is upper triangular

\[ A = LL' \quad M = N \]

**Cholesky decomposition** (when \(A\) is symmetric): \(L\) is lower triangular

\[ A = Q\Lambda Q' \quad M = N \]

orthogonal **eigendecomposition** (when \(A\) is symmetric or normal)

- \(Q\) is unitary; \(\Lambda\) is diagonal (and real if \(A = A'\))

\[ A = V\Lambda V^{-1} \quad M = N \]

diagonalization (when possible)

- \(V\) is (linearly independent) eigenvectors; \(\Lambda\) is eigenvalues

\[ A = QRQ' \quad M = N \]

**Schur decomposition**

- \(Q\) is unitary; \(R\) is upper triangular

\[ A = QR \quad M \geq N \]

**QR decomposition** via **Gram-Schmidt** orthogonalization

- \(Q\) has orthonormal columns; \(R\) is upper triangular

\[ A = U\Sigma V' \quad \text{any } M, N \]

**SVD**: \(U\) and \(V\) are unitary, columns are singular vectors

- \(\Sigma\) is (rectangular) diagonal with real, nonnegative singular values
The LU, QR and Cholesky decompositions are important for solving systems of equations, but are less helpful for analysis than the other forms. This chapter focuses on the two particularly important matrix decompositions: the eigendecomposition and the SVD.

There are no applications in this chapter but the tools are the foundation for most of the applications that appear in later chapters. Sometimes we use these tools for mathematical analysis (on paper only, especially for very large problems) but often we use them numerically.

A short summary of this chapter is: an eigendecomposition is usually the right tool for (square) symmetric matrices whereas an SVD is the right tool otherwise.

**Square matrices**

Recall that any (square) matrix $A \in \mathbb{F}^{N \times N}$ has $N$ (possibly non-distinct) eigenvalues $\lambda_1, \ldots, \lambda_N \in \mathbb{C}$.

For each eigenvalue $\lambda_n$, the matrix $A - \lambda_n I$ is singular, so there must exist a (nonzero) vector $v_n \in \mathbb{F}^N$ (an eigenvector) such that

$$(A - \lambda_n I) v_n = 0 \implies Av_n = \lambda_n v_n, \text{ for } n = 1, \ldots, N. \quad (2.1)$$

In matrix form:

$$AV = V \Lambda, \quad V = [v_1, \ldots, v_N], \quad \Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_N\}. \quad (2.2)$$

Because each $v_n$ is nonzero, by convention we always normalize each to have unit norm for decompositions. Despite this normalization, $V$ is never unique because we can always scale each $v_n$ by $\pm 1$ or even $e^{i\phi}$.

For a general square matrix, this is all we can say about an $N \times N$ eigenvector matrix $V$. However, for (Hermitian) symmetric matrices we can say much more, thanks to the spectral theorem discussed next.
2.1 Spectral Theorem (for symmetric matrices)

If \( A \in \mathbb{F}^{N \times N} \) is (Hermitian) symmetric, i.e., \( A = A' \), then the spectral theorem says the following.

- The eigenvalues of \( A \) are all real.
- Remarkably, there is an orthonormal basis for \( \mathbb{F}^N \) consisting of eigenvectors of \( A \), i.e., there exists \( V \) in (2.2) that is an orthogonal (or unitary) matrix, i.e., \( V'V = VV' = I \), so \( V^{-1} = V' \).
- Multiplying (2.2) on the right by \( V' \) yields an eigendecomposition (aka a matrix factorization):

\[
A = V \Lambda V'.
\]  

(2.3)

In words, every symmetric (hence square) matrix has an orthogonal (or unitary) eigendecomposition. This form is very useful for analysis and sometimes for computation.

As mentioned previously, a data matrix \( X \) is rarely square, but the Gram matrix \( X'X \) and the outer-product matrix \( XX' \) are always square. Furthermore, the Gram matrix and the outer-product matrix are always (Hermitian) symmetric, so the spectral theorem applies.

In signal processing, we usually need an eigendecomposition only for matrices of the form \( X'X \) or \( XX' \).

Proof that eigenvalues are real. If \( v \) is an eigenvector of \( A = A' \) with eigenvalue \( \lambda \), then \( Av = \lambda v \implies v'Av = \lambda v'v \implies (v'Av)' = \lambda v'v \implies v'A'v = \lambda v'v \implies v'Av = \lambda v'v \implies \lambda = \lambda' = \lambda^* \).

Sketch of proof of orthogonality of eigenvectors for the case of distinct eigenvalues. Suppose \( v \) is an eigen-
vector of $A = A'$ with (real) eigenvalue $\lambda$, and $u$ is an eigenvector with different (real) eigenvalue $\beta \neq \lambda$.

$Av = \lambda v \implies u'Av = \lambda u'v \implies (u'Av)' = (\lambda u)v' \implies v'A'u = \lambda v'u \implies v'Au = \lambda v'u$.

$Au = \beta u \implies v'Au = \beta v'u \implies \lambda v'u = \beta v'u \implies v'u = 0$ because $\beta \neq \lambda$.

Example. Consider the symmetric matrix $A = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}$ for which $\det\{A - \lambda I\} = (4 - \lambda)(9 - \lambda) - 6^2 = \lambda^2 - 13\lambda$. So the eigenvalues of $A$ are $\{13, 0\}$. $A - 13I = \begin{bmatrix} -9 & 6 \\ 6 & -4 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \begin{bmatrix} -3 & 2 \end{bmatrix}$ so an eigenvector corresponding to eigenvalue 13 is $v_1 = \frac{1}{\sqrt{13}} \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. Similarly an eigenvector corresponding to eigenvalue 0 is $v_2 = \frac{1}{\sqrt{13}} \begin{bmatrix} -3 \\ 2 \end{bmatrix}$. Thus an eigendecomposition is

$$A = \frac{1}{\sqrt{13}} \begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 13 & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{\sqrt{13}} \begin{bmatrix} 2 & 3 \\ -3 & 2 \end{bmatrix} = 13v_1v_1' + 0v_2v_2' = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \sqrt{13}v_1 + \sqrt{13}v_1'$$.
Normal matrices

Hermitian symmetry is a sufficient but not necessary condition for existence of a unitary eigendecomposition.

Define. A square matrix $A$ is a normal matrix iff $A' A = AA'$.

The spectral theorem says: A square matrix $A$ is diagonalizable by a unitary matrix, i.e., $A = V \Lambda V'$, iff it is a normal matrix.

For a normal matrix, $\Lambda$ need not be real, whereas for a symmetric matrix, $\Lambda$ is real.

Example. One important type of normal matrix is a permutation matrix.

Define. A $N \times N$ permutation matrix has exactly one 1 in each row and one 1 in each column and all other elements are zero.

If $P$ is a permutation matrix, then $P^{-1} = P'$ so $P$ is an orthogonal matrix and is normal. Thus $P$ has a unitary eigendecomposition and typically its eigenvalues and eigenvectors are complex [wiki].

Example. The permutation matrix that (circularly) shifts each element of a vector in $\mathbb{F}_3$ by one index is

$$
P = \begin{bmatrix}
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix}.
$$

This permutation matrix happens to be a circulant matrix so its eigenvalues are given by the 3-point DFT of the first column: $\{e^{-i2\pi k/3} : k = 0, 1, 2\} = \{1, e^{ \pm i2\pi / 3}\}$. (See HW.)
Example. The following rotation matrix is asymmetric unless $\phi$ is a multiple of $\pi$: $R_\phi = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix}$.

Yet this matrix must have two eigenvalues with corresponding eigenvectors. The eigenvalues satisfy 

$(\cos \phi - \lambda)^2 + \sin^2 \phi = 0$ leading to $\lambda_\pm = e^{\pm i\phi}$. The corresponding eigenvectors are $v_\pm = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ \pm i \end{bmatrix}$,

because $R_\phi \begin{bmatrix} 1 \\ \pm i \end{bmatrix} = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} 1 \\ \pm i \end{bmatrix} = \begin{bmatrix} \cos \phi \pm i \sin \phi \\ -\sin \phi \pm i \cos \phi \end{bmatrix} = (\cos \phi \pm i \sin \phi) \begin{bmatrix} 1 \\ \pm i \end{bmatrix} = e^{\pm i\phi} \begin{bmatrix} 1 \\ \pm i \end{bmatrix}$.

This rotation matrix is normal because $R_{-\phi} = R_\phi^t = R_\phi^{-1}$, i.e., multiplying by $R$ and then $R'$ corresponds to rotating by $\phi$ and then rotating back, and $R'R = RR' = I_2$. So $R$ is diagonalizable by $V = [v_+ \ v_-]$:

$$R_\phi = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix} = \left( \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \right) e^{i\phi} \begin{bmatrix} 0 & e^{-i\phi} \\ \frac{1}{\sqrt{2}} & 1 \end{bmatrix} \left( \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ 1 & i \end{bmatrix} \right).$$

$R$ is not diagonalizable by a real matrix $V$ because rotation leads to a vector pointing in a different direction, unless $\phi$ is a multiple of $\pi$. But $R$ is diagonalizable by the above unitary matrix $V$.

What is the best way to think about the rotation matrix $R$?

A: A data matrix.  
B: An operator matrix.  
C: Neither.  

??
**Square asymmetric matrices**

What about *asymmetric* matrices? Some, but not all, square and asymmetric matrices are **diagonalizable**.

**Define.** A square matrix is **diagonalizable** iff it is **similar** to a diagonal matrix, *i.e.*, iff there exists an invertible matrix $V$ such that $V^{-1}AV$ is diagonal.

Specifically, if $A$ has **linearly independent** eigenvectors $V$, then

$$A = V \Lambda V^{-1}.$$  

- If $A$ is asymmetric and its eigenvalues are all real, then $A$ cannot have a unitary eigendecomposition. (If it did, then we would have $A = V \Lambda V' = A'$, contradicting asymmetry.)
- If $A$ has $N$ distinct eigenvalues (no repeated roots of characteristic equation), then $A$ is diagonalizable. (But this is not a necessary condition for being diagonalizable.)
- If $A$ is both invertible and diagonalizable then $A^{-1} = V \Lambda^{-1}V^{-1}$.
- Being diagonalizable does not imply invertibility because some eigenvalues can be 0.

Some square matrices are not diagonalizable (see example below). Such matrices might arise when trying to solve a system of $N$ equations in $N$ unknowns so they are a major topic in a linear algebra course, but they are much less important in signal processing so we do not dwell on them further here.

Anyway, every matrix, even if non-square, has a **singular value decomposition** (*SVD*) so that will be the primary tool we use for data matrices.
Example. The matrix $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ has repeated eigenvalue $\lambda = 3$, and $A - 3I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ so $v = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Here $AV = V\Lambda$, where $V = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and $\Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ contains both eigenvalues of $A$, but the columns of $V$ are linearly dependent so $A$ is not diagonalizable.

---

**Geometry of matrix diagonalization**

Let $A \in \mathbb{F}^{N \times N}$ be a (Hermitian) symmetric matrix, so $A = V\Lambda V'$. Consider the linear transform $x \mapsto y = Ax = V\Lambda V'x$.

We can think of $Ax$ as a cascade of three linear transforms:

$$x \xrightarrow{V'} w \xrightarrow{\Lambda} z \xrightarrow{V} y.$$

- $x \mapsto w = V'x$ is a coordinate change (a rotation in fact, possibly with a sign flip for one axis). $w$ denotes the coefficient vector for $x$ in the basis $V$, because $x = Vw$.
- $w \mapsto z = \Lambda x$ is scaling of each coordinate by the diagonal elements of $\Lambda$.
- $z \mapsto y = Vz$ is going back to the original coordinate system.

It is useful to understand these three operations geometrically.
Example. We illustrate for the case of a symmetric $2 \times 2$ matrix.

Fact. Every $V \in \mathbb{R}^{2 \times 2}$ with $V'V = I_2$ (i.e., orthogonal) has the following form for some rotation angle $\theta$:

$$V = \begin{bmatrix} \cos \theta & -q \sin \theta \\ \sin \theta & q \cos \theta \end{bmatrix}, \quad q \in \{\pm 1\}.$$

Exercise. Verify that $V'V = I_2$.

For simplicity we focus on the case of rotation matrices hereafter where $q = +1$.

Consider the linear transformation $x \mapsto w = V'x = \begin{bmatrix} x_1 \cos \theta + x_2 \sin \theta \\ -x_1 \sin \theta + x_2 \cos \theta \end{bmatrix}$. Graphically:

Importantly, the length of $x$ and $w$ are the same:

$$w = V'x \implies \|w\|^2 = w'w = (V'z)'V'x = x'VV'x = x'Ix = x'x = \|x\|^2.$$

\[
\begin{align*}
\begin{bmatrix}
\cos \theta & -q \sin \theta \\
\sin \theta & q \cos \theta
\end{bmatrix}, \quad q \in \{\pm 1\}
\end{align*}
\]
The next mapping is

\[ w \mapsto z = \Lambda w, \text{ where } \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \Rightarrow \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 w_1 \\ \lambda_2 w_2 \end{bmatrix}. \]

For interpretation, assume \( \lambda_1, \lambda_2 \neq 0 \) and suppose \( \|x\|_2 = 1 \Rightarrow x'x = 1 \Rightarrow x_1^2 + x_2^2 = 1 \), which in turn implies \( w_1^2 + w_2^2 = 1 \), i.e., \( x \) and \( w \) lie on the unit circle. Because \( w_1 = z_1/\lambda_1 \) and \( w_2 = z_2/\lambda_2 \), we have

\[ \left( \frac{z_1}{\lambda_1} \right)^2 + \left( \frac{z_2}{\lambda_2} \right)^2 = 1, \]

i.e., \( z \) lies on an ellipse with axes governed by \( \lambda_1 \) and \( \lambda_2 \).

Graphically:
Typically $z$ is not collinear with $w$.
The exception is when $\lambda_1 = \lambda_2$, in which case $A = \lambda_1 I_2$, which is a trivial case.

Q. When does $y = Ax$ produce $y$ that is collinear with $x$?
A. When $x$ is eigenvector of $A$, because then $Ax = \lambda x$.

For the third and final mapping, we return to geometry.
If $x \mapsto w$ represents counter-clockwise rotation, then $w \mapsto x$ must represent clockwise rotation, because $VV' = I$ so $V(V'x) = (VV')x = Ix = x$.

If $V$ includes a sign flip, e.g., $V = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, then $V^{-1} = V'$ also undoes that sign flip.

- An orthogonal matrix with determinant equal to $+1$ is called a rotation.
- If the determinant is $-1$ then it is an improper rotation.

We ignore the possibility of a sign flip in the graphical illustrations, for simplicity.

The next page gives a graphical summary in $2 \times 2$ case of $y = Ax = V \begin{bmatrix} \hat{z} \\ \hat{w} \end{bmatrix}$.

(The same principles apply in higher dimensions.)
\[ x_2 \rightarrow (x_1, x_2) \quad V' \rightarrow \text{rotate} \]

\[ w_2 \rightarrow (w_1, w_2) \quad \theta \quad \text{scale} \downarrow \Lambda \]

\[ y_2 \rightarrow (y_1, y_2) \quad \theta \quad \text{unrotate} \]

\[ z_2 \rightarrow (z_1, z_2) \]
2.2 SVD

Existence

If $X \in \mathbb{F}^{M \times N}$ then there exists (for proof, see [1, Theorem 5.1] or [wiki]) $U$ and $V$ and $\Sigma$ such that

$$X = U \Sigma V' = \sum_{k=1}^{\min(M,N)} \sigma_k u_k v_k'.$$

(2.4)

This factorization is called the singular value decomposition (SVD), where:

- $U$ is $M \times M$ and unitary: $U'U = UU' = I_M$, and consists of the left singular vectors
- $V$ is $N \times N$ and unitary: $V'V = VV' = I_N$ and consists of the right singular vectors
- $\Sigma$ is a $M \times N$ rectangular diagonal matrix consisting of the singular values
• The $M \times N$ matrix $\Sigma$ looks like one of:

$$
\Sigma = \begin{bmatrix}
\sigma_1 & \cdots & 0 \\
0 & \ddots & \vdots \\
0_{M-N,N} & \cdots & \sigma_N
\end{bmatrix}
$$

or

$$
\Sigma = \begin{bmatrix}
\sigma_1 & \cdots & 0 \\
0 & \ddots & \vdots \\
0_{M-N,N} & \cdots & \sigma_N
\end{bmatrix}
$$

or

$$
\Sigma = \begin{bmatrix}
\sigma_1 & \cdots & 0 \\
0 & \ddots & \vdots \\
0 & \cdots & \sigma_M
\end{bmatrix}
$$

$M > N$ (tall) \quad N > M$ (wide) \quad $M = N$ (square)

• These possible shapes of $\Sigma$ are why the sum in (2.4) has $\min(M, N)$ terms.

• The **singular values** $\sigma_1, \ldots, \sigma_{\min\{M,N\}}$ are real and nonnegative.

• By convention we use descending order: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(M,N)}$.

• The first $r$ singular values are positive, where $0 \leq r \leq \min\{M, N\}$ is the **rank** of the matrix (later).

• A subtle point is that the sum form does not use all columns of $U$ when $M > N$, nor all columns of $V$ when $N > M$. More on this later when we discuss the **thin SVD**.

---

**Geometry**

Example. If $A$ is any real $2 \times 2$ matrix, then its SVD looks like:

$$
A = U \Sigma V' = \begin{bmatrix}
\cos \phi & q_1 \sin \phi \\
-q_1 \sin \phi & \cos \phi
\end{bmatrix}
\begin{bmatrix}
\sigma_1 & 0 \\
0 & \sigma_2
\end{bmatrix}
\begin{bmatrix}
\cos \theta & -q_2 \sin \theta \\
\sin \theta & q_2 \cos \theta
\end{bmatrix}, \quad q_1, q_2 \in \{\pm 1\},
$$

where $\theta \neq \phi$ in general. (In fact when $M \neq N$, $U$ and $V$ even have different sizes!)

Consider the case $q_1 = q_2 = +1$ for simplicity. The next page illustrates the $2 \times 2$ geometry is graphically:
\[ (x_1, x_2) \xrightarrow{\text{rotate}} (w_1, w_2) \]

\[ \begin{align*}
V' & \quad \text{rotate} \\
\theta & \\
\Sigma & \quad \text{“scale”} \\
\phi & \quad \text{rotate}
\end{align*} \]
What are the differences here (for SVD) from before (for Eigenvalue decomposition)?

- Final rotation angle differs: $\phi \neq \theta$ in general, i.e., $U \neq V$ in general.
- For non-square matrices, $\Sigma$ is non-square, so the interpretation that it “scales” is incomplete.
- The SVD always exists, even for (asymmetric) $2 \times 2$ matrices that have no eigendecomposition.

Example. Determine the SVD of the rotation matrix $R = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix}$.

(Recall that no eigendecomposition exists for this important matrix in general.)

By inspection we can choose $U = R$ and $\Sigma = I_2$ and $V = I_2$.

Is the SVD unique here?

A: Yes  B: No  C: Neither
More geometry

For eigendecomposition we asked “when is $Ax$ aligned with $x$?”

Should we ask that question for the SVD? No, $A$ is non-square in general!

What choice of vectors $x$ and $z$ makes $Ax$ perpendicular to $Az$?

Choose the right singular vectors: $x = v_1$ and $z = v_2$.

Proof: $A = U \Sigma V' = \sigma_1 u_1 v_1' + \sigma_2 u_2 v_2'$ (a sum of outer products) so $Ax = Av_1 = \sigma_1 u_1$ and $Az = Av_2 = \sigma_2 u_2$ but $u_1 \perp u_2$. so $Ax \perp Az$. □
The matrix 2-norm or spectral norm

Moving towards one (of many) applications of the SVD, we now ask: What unit norm “input vector” $x$ produces an “output vector” $Ax$ having the “largest” output (as defined by norm, aka energy)? In other words, find unit norm $x_\star$ such that $\|Ax_\star\| \geq \|Ax\|$ for all unit norm $x$.

Note the “systems” language here. Very often when we talk about $y = Ax$ we also think about a system block diagram like

$\begin{align*}
  x &\quad \text{input } \mathbb{F}^M \\
  &\quad \rightarrow \quad A \\
  &\quad \quad \text{(linear) system} \\
  &\quad \rightarrow \quad y = Ax \\
  &\quad \text{output } \mathbb{F}^N.
\end{align*}$

In this setting, we are thinking of $A$ as an operation, not as data.

Expressing the question mathematically:

$L\S7.4$

$$x_\star = \arg \max_{x: \|x\|=1} \|Ax\| = \arg \max_{x \neq 0} \frac{\|Ax\|}{\|x\|},$$

where $\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{x'x}$.

Claim: $x_\star = v_1$, the first right singular vector (having the largest singular value $\sigma_1$).
Proof. Because $V$ is orthogonal (or unitary), we can write $x$ in terms of the $V$ basis as $x = Vz$ where $z = V'x$ are the coefficients. Note that $\|x\| = 1 \iff \|z\| = 1$.

Thus $Ax = U\Sigma V'x = U\Sigma V'Vz = U\Sigma z$ so $\|Ax\| = \sqrt{(Ax)'(Ax)} = \sqrt{z'\Sigma'U'U\Sigma z} = \sqrt{z'\Sigma'\Sigma z}$

$= \sqrt{\sum_{k=1}^{r} \sigma_k^2 |z_k|^2} \leq \sqrt{\sigma_1^2 \sum_{k=1}^{r} |z_k|^2} = \sigma_1 \|z\| = \sigma_1$, where $r = \min(M, N)$. So $\|x\| = 1 \implies \|Ax\| \leq \sigma_1$. This gives an upper bound on the norm of the output. But we can achieve that upper bound by choosing $x = v_1$ because then $z = (1, 0, 0, \ldots, 0)$ and $Ax = \sigma_1 u_1$ so $\|Ax\| = \|\sigma_1 u_1\| = \sigma_1$.

Fact. The solution $x^* = v_1$ is unique iff $\sigma_1 > \sigma_2$.

Define. This property of a matrix is very important; it is called the matrix 2-norm or spectral norm:

$$\|A\|_2 \triangleq \max_{x: \|x\|_2 = 1} \|Ax\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1.$$  \hspace{1cm} (2.5)

By definition the 2-norm of a matrix $A$ gives a (tight) upper bound on how much the 2-norm of a vector can be amplified when multiplying by $A$:

$$\|Ax\|_2 \leq \|A\|_2 \|x\|_2 = \sigma_1 \|x\|_2.$$  

This important SVD property and has practical applications for power maximization, shown next.
Example. Multi-input multi-output (MIMO) communications with multi-antenna systems

Consider a system with $N$ transmit antennas and $M$ receiving antennas, such as in 802.11n wifi.

Let $a_{ij}$ denote the (complex) gain between the $j$th transmit antenna and the $i$th receive antenna.

(Matrix $A$ depends on amplifier and receiver properties and the multi-path wave propagation between them.)

Goal: Design transmit amplitudes so that the received signal has largest possible signal to noise ratio (SNR).

For some signal we want to transmit, we use amplitudes $x_1, \ldots, x_N$ for the $N$ transmit antennas.

There are transmit power limits (amplifier hardware and allowable interference) so we constrain the input amplitudes: $\|x\| \leq 1$.

After transmission, the signals received by the $M$ antennas will have amplitudes $y = Ax$. To maximize SNR, we want to design $x$ to make the received signal energy $\|y\|$ as large as possible. (The background noise power is independent of $x$.) This problem is exactly of the form $\arg \max_x : \|x\| = 1 \|Ax\|$ so the solution is

$$x = v_1,$$

the first right singular vector.
In practice, the designer of a wifi system does not know the $A$ for your house, and in fact $A$ changes if you rearrange the furniture or move your computer to a different position. So the wifi router must determine $A$ “on the fly” and this process is called **channel estimation**. The basic idea is that the transmitter first sends $N$ orthonormal training waveforms $x_1, \ldots, x_N \in \mathbb{C}^N$, called **pilot signals**, to the receiver. The receiver records the corresponding outputs $y_1, \ldots, y_N \in \mathbb{C}^M$, where $y_n = Ax_n$. Writing as a matrix:

$$[y_1 \  \ldots \ y_N] = A[x_1 \  \ldots \ x_N] \implies Y = AX.$$  

Because $X$ is unitary by design, we can estimate $A$ as

$$A = Y X^{-1} = Y X'.$$

Note that the receiver must know what pilot signals $X$ are transmitted.

Once we have $A$ we can compute its SVD and use $v_1$ as the best transmit amplitude vector.

This is related to the topic known as **beamforming**. See [this illustration of an improved constellation using SVD-based transmit beamforming](#).

One might ask: why estimate *all* of $A$ when we need only $v_1$? Later we will see how to estimate $v_1$ with fewer measurements (at least when $N$ is large).
When does $U = V$?

For the SVD $A = U \Sigma V'$, when does $U = V$?

First, if $N \neq M$ then $U$ and $V$ have different sizes.
So focus here on the square case where $N = M$.

Answer: When $A$ is square, $U = V$ iff $A = A'$ and all eigenvalues of $A$ are nonnegative.

Recall from (2.3) that we can write any (Hermitian) symmetric matrix as $A = V \Lambda V'$ where $\Lambda$ is the diagonal of the eigenvalues. This looks a lot like an SVD with $U = V$. However, for the SVD we always have $\sigma_k \geq 0$ whereas the eigenvalues of (symmetric) $A$ are real but not necessarily nonnegative.

So to have $U = V$ we need $A$ to be symmetric and to have nonnegative eigenvalues.
Example. Consider the eigendecomposition and the following three forms of the (not unique!) SVD of the following matrix:

\[
A = \begin{bmatrix}
0 & 2 \\
2 & 0 \\
\end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix}
1 & -1 \\
1 & 1 \\
\end{bmatrix} \begin{bmatrix}
2 & 0 \\
0 & -2 \\
\end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix}
1 & 1 \\
-1 & 1 \\
\end{bmatrix} \quad \text{(eigendecomposition)}
\]

\[
= \frac{1}{\sqrt{2}} \begin{bmatrix}
1 & 1 \\
1 & -1 \\
\end{bmatrix} \begin{bmatrix}
2 & 0 \\
0 & 2 \\
\end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix}
1 & 1 \\
-1 & 1 \\
\end{bmatrix} \quad \text{(SVD version 1)}
\]

\[
= \begin{bmatrix}
0 & 1 \\
1 & 0 \\
\end{bmatrix} \begin{bmatrix}
2 & 0 \\
0 & 2 \\
\end{bmatrix} \begin{bmatrix}
1 & 0 \\
0 & 1 \\
\end{bmatrix} \quad \text{(SVD version 2)}
\]

\[
= \begin{bmatrix}
1 & 0 \\
0 & 1 \\
\end{bmatrix} \begin{bmatrix}
2 & 0 \\
0 & 2 \\
\end{bmatrix} \begin{bmatrix}
0 & 1 \\
1 & 0 \\
\end{bmatrix} \quad \text{(SVD version 3)}.
\]

Note that \( \mathbf{U} \neq \mathbf{V} \) here even though \( \mathbf{A} \) is symmetric because \( \mathbf{A} \) has a negative eigenvalue:

\[
\det\{\mathbf{A} - z\mathbf{I}\} = z^2 - 4 \implies z = \pm 2.
\]
Positive semidefinite matrices

The preceding example is the exception, not the rule. Most of the time we use the SVD for non-square matrices anyway. As mentioned previously, when we do consider a square matrix, it is usually a Gram matrix $X'X$ or an outer-product matrix $XX'$. These matrices are not only symmetric, they are positive semi-definite.

**Define.** A (square) Hermitian matrix $A$ is **positive semi-definite** iff $x'Ax \geq 0$ for all $x$.

**Define.** A (square) Hermitian matrix $A$ is **positive definite** iff $x'Ax > 0$ for all $x \neq 0$.

**Lemma.** If $A = BB'$ for any matrix $B$, then $A$ is positive semi-definite.

**Proof.** $x'Ax = x'BB'x = \| B'x \|^2 \geq 0$.

Which of the following statements is true?

- A: All positive-semidefinite matrices are positive definite.
- B: All positive-definite matrices are positive semi-definite.
- C: Neither statement is always true.

In words, any Gram matrix $X'X$ or outer-product matrix $XX'$ is positive semi-definite.
Theorem. If $A = BB'$ for any matrix $B$, then $A = UΣU'$ with $Σ_{ii} ≥ 0$. In words, the SVD and the eigendecomposition are the same for such matrices.

Proof. Let $B = UΣ_BV'$ denote the SVD of $B$. Recall $Σ_B$ is real and nonnegative. Then $A = BB' = UΣ_BV'VΣ_BU' = UΣ_BΣ_B^*U' = U \text{diag}\{σ_k^2\}U'$.

So when $A = BB'$ the eigenvalues of $A$ are the square of the singular values of $B$, and hence real and nonnegative.

Let $A$ be any Hermitian matrix, then $A$ has an eigendecomposition of the form $A = VΛV'$. Now $x'Ax = x'VΛV'x = z'Λz = \sum_i λ_i |z_i|^2$ where $z = V'x$.

Thus if (Hermitian) $A$ has all nonnegative eigenvalues, then $x'Ax ≥ 0$ for all $x$ so $A$ is positive semi-definite.

The converse is also true: If $A$ is symmetric positive semi-definite, then $A$ has nonnegative eigenvalues.

Combining, a Hermitian matrix $A$ is positive semi-definite iff all of its eigenvalues are nonnegative. Similarly, a Hermitian matrix $A$ is positive definite iff all of its eigenvalues are positive.

We write $A ≻ 0$ to denote positive definite and $A ⪰ 0$ to denote positive semi-definite.

To summarize, any positive semi-definite matrix has real and nonnegative eigenvalues, and its eigendecomposition matches its SVD, with $U = V$ and $Σ = Λ$, i.e., $σ_n(A) = λ_n(A) ≥ 0$. 
Summary

In practice, we usually end up using:

- an eigendecomposition, $A = V \Lambda V'$, when working with positive semi-definite matrices (like a Gram matrix or outer-product matrix),
- the SVD, $A = U \Sigma V'$, for most other matrices.

A curious note about terminology:

- columns of $V$ are called the right eigenvectors for the eigendecomposition, because $AV = V \Lambda$ even though $A = V \Lambda V'$.
- columns of $U$ are called the left singular vectors for the SVD, because $A = U \Sigma V'$. 
## Venn diagram of matrices

<table>
<thead>
<tr>
<th>Rectangular SVD: ( A = U \Sigma V' )</th>
<th>Square Diagonalizable: ( A = V \Lambda V^{-1} ) V linearly indep.</th>
<th>Normal: ( A = V \Lambda V' ) V unitary</th>
<th>Hermitian: ( \Lambda ) real</th>
<th>PSD: ( A \succeq 0 ) ( \Sigma = \Lambda \succeq 0 ) ( U = V )</th>
</tr>
</thead>
</table>

### Example.

(For student reading only.)

**Exercise.** Determine the SVD of \( A = \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} \).

Recall that this (square but asymmetric) matrix does not have an orthogonal eigendecomposition.

\[
A' A = \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 9 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = V \Sigma^2 V' \implies V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}.
\]

\[
A A' = \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 9 & 0 \\ 0 & 0 \end{bmatrix} = I_2 \begin{bmatrix} 9 & 0 \\ 0 & 0 \end{bmatrix} I_2 = U \Sigma^2 U' \implies U = I_2.
\]

Note that to find \( V \) properly, we applied a permutation to have the eigenvalues of \( A' A \) in descending order.

Thus a SVD of \( A \) is \( A = \underbrace{I_2}_{U} \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{V'} \).
**SVD computation using eigendecomposition**  
(For student reading only.)

There is a concise overview of practical computation of the SVD here: [wiki].

To perform an SVD by hand using an eigendecomposition, or if you were stuck on a desert island with a computer that had an `eig` command but no `svd` command, here is how you could do it for the case of a tall $M \times N$ matrix with $M \geq N$ and rank $N$.

First use the eigendecomposition of $AA'$ to find $U$ and $\Sigma$:

$$AA' = U \Lambda U' = U \Sigma \Sigma' U', \quad M \times M$$

Here $U$ will be the left singular vectors of $A$ and the $N \leq M$ singular values will be $\sigma_n = \sqrt{\lambda_n}, \quad n = 1, \ldots, N$. Now obtain $V'$ by multiplying $A$ on the left by $\text{diag}\{1/\sigma_n\} U[:, 1 : N]'$ as follows:

$$\text{diag}\{1/\sigma_n\} U[:, 1 : N]' A = \text{diag}\{1/\sigma_n\} U[:, 1 : N]' (U \Sigma V') = \text{diag}\{1/\sigma_n\} [I, 0] \Sigma V'$$

$$= \text{diag}\{1/\sigma_n\} \text{diag}\{1/\sigma_n\} V' = V'.$$

Unfortunately this process does not work when $A$ is not full rank, and it requires an SVD of the “large” size $M \times M$ so it is impractical.

**Bibliography**