Chapter 2

Matrix factorizations / decompositions

Contents (class version)

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.

Source material for this chapter includes [1, §10.1, 5.1, 7.4, 10.2].

Matrix factorizations

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 7, only the SVD accommodates any matrix size and type.

$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$	$M = N$	**LU decomposition** by **Gaussian elimination**, for some (not all) square $\boldsymbol{A}$: $\boldsymbol{L}$ is lower triangular; $\boldsymbol{U}$ is upper triangular. For general $\boldsymbol{A}$, use **pivoting**.
$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}'$	$M = N$	**Cholesky decomposition** (for any **positive-definite** $\boldsymbol{A}$): $\boldsymbol{L}$ is lower triang.
$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$	$M = N$	orthogonal **eigendecomposition** (for any symmetric or normal $\boldsymbol{A}$) $\boldsymbol{Q}$ is unitary; $\boldsymbol{\Lambda}$ is diagonal (and real if $\boldsymbol{A} = \boldsymbol{A}'$)
$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$	$M = N$	diagonalization (possible only for some square $\boldsymbol{A}$): $\boldsymbol{V}$ is (linearly independent) eigenvectors; $\boldsymbol{\Lambda}$ is eigenvalues
$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}\boldsymbol{Q}'$	$M = N$	**Schur decomposition** (for any square $\boldsymbol{A}$): $\boldsymbol{Q}$ is unitary; $\boldsymbol{R}$ is upper triangular
$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$	$M \geq N$	**QR decomposition** via **Gram-Schmidt** orthogonalization (for any "tall" $\boldsymbol{A}$): $\boldsymbol{Q}$ has orthonormal columns; $\boldsymbol{R}$ is upper triangular
$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$	any M, N	**SVD** (for any $\boldsymbol{A}$): $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary, columns are singular vectors $\boldsymbol{\Sigma}$ 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) Hermitian matrices whereas an **SVD** is usually the right tool otherwise.

Square matrices

Recall that any (square) matrix $\boldsymbol{A} \in \mathbb{F}^{N \times N}$ has N (possibly non-distinct) **eigenvalues** $\lambda_1, \ldots, \lambda_N \in \mathbb{C}$.
For each eigenvalue λ_n, the matrix $\boldsymbol{A} - \lambda_n \boldsymbol{I}$ is **singular**, so there must exist a (nonzero) vector $\boldsymbol{v}_n \in \mathbb{F}^N$ (an **eigenvector**) such that

$$(\boldsymbol{A} - \lambda_n \boldsymbol{I})\, \boldsymbol{v}_n = \boldsymbol{0} \Longrightarrow \boldsymbol{A} \boldsymbol{v}_n = \lambda_n \boldsymbol{v}_n, \text{ for } n = 1, \ldots, N. \tag{2.1}$$

In matrix form:

$$\boldsymbol{V} = [\boldsymbol{v}_1, \ \ldots, \ \boldsymbol{v}_N], \quad \boldsymbol{\Lambda} = \mathsf{diag}\{\lambda_1, \ldots, \lambda_N\}\,. \tag{2.2}$$

Because each $\boldsymbol{v}_n$ is nonzero, by convention we always normalize each to have *unit norm* for decompositions.
Despite this normalization, $\boldsymbol{V}$ is never unique because we can always scale each $\boldsymbol{v}_n$ by ± 1 or even $\mathrm{e}^{\imath\phi}$.
For a general square matrix, this is all we can say about an $N \times N$ eigenvector matrix $\boldsymbol{V}$. However, for (Hermitian) symmetric matrices we can say much more, thanks to the spectral theorem discussed next.

2.1 Spectral Theorem (for symmetric matrices)

L§10.1

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

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

$$\tag{2.3}$$

- If $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ is symmetric, then there exists a *real* orthogonal matrix $\boldsymbol{V}$ satisfying (2.3).

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

As mentioned previously, a data matrix $\boldsymbol{X}$ is rarely square, but the Gram matrix $\boldsymbol{X}'\boldsymbol{X}$ and the outer-product matrix $\boldsymbol{X}\boldsymbol{X}'$ 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 $\boldsymbol{X}'\boldsymbol{X}$ or $\boldsymbol{X}\boldsymbol{X}'$.

Proof that eigenvalues are real. If $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{A} = \boldsymbol{A}'$ with eigenvalue λ, then $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \Longrightarrow \boldsymbol{v}'\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}'\boldsymbol{v} \Longrightarrow (\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v})' = \lambda'\boldsymbol{v}'\boldsymbol{v} \Longrightarrow \boldsymbol{v}'\boldsymbol{A}'\boldsymbol{v} = \lambda'\boldsymbol{v}'\boldsymbol{v} \Longrightarrow \boldsymbol{v}'\boldsymbol{A}\boldsymbol{v} = \lambda'\boldsymbol{v}'\boldsymbol{v} \Longrightarrow \lambda = \lambda' = \lambda^*$.

Sketch of proof of orthogonality of eigenvectors for the case of distinct eigenvalues. Suppose $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{A} = \boldsymbol{A}'$ with (real) eigenvalue λ, and $\boldsymbol{u}$ is an eigenvector with different (real) eigenvalue $\beta \neq \lambda$.
$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \Longrightarrow \boldsymbol{u}'\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{u}'\boldsymbol{v} \Longrightarrow (\boldsymbol{u}'\boldsymbol{A}\boldsymbol{v})' = (\lambda\boldsymbol{u}'\boldsymbol{v})' \Longrightarrow \boldsymbol{v}'\boldsymbol{A}'\boldsymbol{u} = \lambda\boldsymbol{v}'\boldsymbol{u} \Longrightarrow \boldsymbol{v}'\boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{v}'\boldsymbol{u}$.
$\boldsymbol{A}\boldsymbol{u} = \beta\boldsymbol{u} \Longrightarrow \boldsymbol{v}'\boldsymbol{A}\boldsymbol{u} = \beta\boldsymbol{v}'\boldsymbol{u} \Longrightarrow \lambda\boldsymbol{v}'\boldsymbol{u} = \beta\boldsymbol{v}'\boldsymbol{u} \Longrightarrow \boldsymbol{v}'\boldsymbol{u} = 0$ because $\beta \neq \lambda$.

(Read)

Example. Consider the symmetric matrix $\boldsymbol{A} = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}$ for which $\det\{\boldsymbol{A} - \lambda\boldsymbol{I}\} = (4-\lambda)(9-\lambda) - 6^2 = \lambda^2 - 13\lambda$. So the eigenvalues of $\boldsymbol{A}$ are $\{13, 0\}$. $\boldsymbol{A} - 13\boldsymbol{I} = \begin{bmatrix} -9 & 6 \\ 6 & -4 \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \begin{bmatrix} -3 & 2 \end{bmatrix}$ so an eigenvector corresponding to eigenvalue 13 is $\boldsymbol{v}_1 = \frac{1}{\sqrt{13}} \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. Similarly an eigenvector corresponding to eigenvalue 0 is $\boldsymbol{v}_2 = \frac{1}{\sqrt{13}} \begin{bmatrix} -3 \\ 2 \end{bmatrix}$. Thus an eigendecomposition is

$$\boldsymbol{A} = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} = \underbrace{\frac{1}{\sqrt{13}} \begin{bmatrix} 2 & -3 \\ 3 & 2 \end{bmatrix}}_{\boldsymbol{V}} \underbrace{\begin{bmatrix} 13 & 0 \\ 0 & 0 \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\frac{1}{\sqrt{13}} \begin{bmatrix} 2 & 3 \\ -3 & 2 \end{bmatrix}}_{\boldsymbol{V}'} = 13\boldsymbol{v}_1\boldsymbol{v}_1' + 0\boldsymbol{v}_2\boldsymbol{v}_2' = \underbrace{\begin{bmatrix} 2 \\ 3 \end{bmatrix}}_{\sqrt{\lambda_1}\boldsymbol{v}_1} \underbrace{\begin{bmatrix} 2 & 3 \end{bmatrix}}_{\sqrt{\lambda_1}\boldsymbol{v}_1'}.$$

Normal matrices

(Read)

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

Define. A square matrix $\boldsymbol{A}$ is a **normal matrix** iff $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}'$.

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

For a normal matrix, $\boldsymbol{\Lambda}$ need not be real, whereas for a (Hermitian) symmetric matrix, $\boldsymbol{\Lambda}$ is real.

Every unitary matrix $\boldsymbol{U}$ is a normal matrix. (?)

A: True B: False $\boxed{\textbf{??}}$

Permutation matrix

Example. An 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 $\boldsymbol{P}$ is a permutation matrix, then $\boldsymbol{P}^{-1} = \boldsymbol{P}'$ so $\boldsymbol{P}$ is an orthogonal matrix and is **normal**. Thus $\boldsymbol{P}$ 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

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

This permutation matrix happens to be a **circulant matrix** (see Ch. 7), so its eigenvalues are given by the 3-point **DFT** of the first column: $\left\{ \mathrm{e}^{-\imath 2\pi k/3} : k = 0, 1, 2 \right\} = \left\{ 1, \mathrm{e}^{\pm \imath 2\pi/3} \right\}$. (See HW.)

Example. The following **rotation matrix** is asymmetric unless ϕ is a multiple of π: $\boldsymbol{R}_\phi = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}$.
Yet this matrix must have two eigenvalues with corresponding eigenvectors. The eigenvalues satisfy $(\cos\phi - \lambda)^2 + \sin^2\phi = 0$ leading to $\lambda_\pm = \mathrm{e}^{\pm\imath\phi}$. Corresponding eigenvectors are $\boldsymbol{v}_\pm = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ \pm\imath \end{bmatrix}$ because

$$\boldsymbol{R}_\phi \begin{bmatrix} 1 \\ \pm\imath \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} 1 \\ \pm\imath \end{bmatrix} = \begin{bmatrix} \cos\phi \pm \imath\sin\phi \\ -\sin\phi \pm \imath\cos\phi \end{bmatrix} = (\cos\phi \pm \imath\sin\phi) \begin{bmatrix} 1 \\ \pm\imath \end{bmatrix} = \mathrm{e}^{\pm\imath\phi} \begin{bmatrix} 1 \\ \pm\imath \end{bmatrix}.$$

This rotation matrix is **normal** because $\boldsymbol{R}_{-\phi} = \boldsymbol{R}'_\phi = \boldsymbol{R}_\phi^{-1}$, *i.e.*, multiplying by $\boldsymbol{R}$ and then $\boldsymbol{R}'$ corresponds to rotating by ϕ and then rotating back, and $\boldsymbol{R}'\boldsymbol{R} = \boldsymbol{R}\boldsymbol{R}' = \boldsymbol{I}_2$. (In fact $\boldsymbol{R}$ is unitary.)
$\boldsymbol{R}$ is diagonalizable by the unitary matrix $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_+ & \boldsymbol{v}_- \end{bmatrix}$ and here is a **unitary eigendecomposition**:

$$\boldsymbol{R}_\phi = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ \imath & -\imath \end{bmatrix} \right) \begin{bmatrix} \mathrm{e}^{\imath\phi} & 0 \\ 0 & \mathrm{e}^{-\imath\phi} \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -\imath \\ 1 & \imath \end{bmatrix} \right).$$

$\boldsymbol{R}$ is *not* diagonalizable by a real matrix $\boldsymbol{V}$ because rotation leads to a vector pointing in a different direction, unless ϕ is a multiple of π. But $\boldsymbol{R}$ *is* diagonalizable by the above unitary matrix $\boldsymbol{V}$.

What is the best way to think about the rotation matrix $\boldsymbol{R}$?
A: A data matrix. B: An operator matrix. C: Neither. $\boxed{??}$

Square asymmetric and non-normal matrices

(Read)

Some, but not all, square *asymmetric* matrices that are not **normal** 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 $\boldsymbol{V}$ such that $\boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V}$ is diagonal.

Specifically, if $\boldsymbol{A} \in \mathbb{F}^{N \times N}$ has N **linearly independent** eigenvectors $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_N \end{bmatrix}$, then

$$\boldsymbol{A} =$$

- If $\boldsymbol{A}$ is asymmetric and its eigenvalues are all real, then $\boldsymbol{A}$ cannot have a unitary eigendecomposition. (If it did, then we would have $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}' = \boldsymbol{A}'$, contradicting asymmetry.)
- If $\boldsymbol{A}$ has N distinct eigenvalues (no repeated roots of characteristic equation), then $\boldsymbol{A}$ is diagonalizable. (But having distinct eigenvalues is not a necessary condition for being diagonalizable.)
- If $\boldsymbol{A}$ is both invertible and diagonalizable then $\boldsymbol{A}^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}^{-1}$.
- 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 $\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ has repeated eigenvalue $\lambda = 3$, and $\boldsymbol{A} - 3\boldsymbol{I} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ so $\boldsymbol{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.
Here $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda}$, where $\boldsymbol{V} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and $\boldsymbol{\Lambda} = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$ contains both eigenvalues of $\boldsymbol{A}$, but the columns of $\boldsymbol{V}$ are linearly dependent so $\boldsymbol{A}$ is not diagonalizable.

Readers interested in general **eigendecompositions** and **diagonalizable** matrices should study **minimal polynomials** and the **Jordan normal form**.

Define. We say a set of matrices $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots$ is **simultaneously diagonalizable** iff there exists an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}_k\boldsymbol{P}$ is **diagonal** for every $\boldsymbol{A}_k$ in the set.

Challenge. When $\boldsymbol{A}$ is 3×3 and symmetric, it has 6 **degrees of freedom** (**DoF**) (the upper triangular elements). Now $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$, where $\boldsymbol{V}$ and $\boldsymbol{\Lambda}$ are both 3×3 matrices. Explain the DoF in terms of $\boldsymbol{V}$ and $\boldsymbol{\Lambda}$. ♦♦

??

Every permutation matrix has a linearly independent set of eigenvectors. (?)

A: True B: False $\boxed{??}$

Every permutation matrix has real eigenvalues. (?)

A: True B: False $\boxed{??}$

Geometry of matrix diagonalization

Let $\boldsymbol{A} \in \mathbb{F}^{N \times N}$ be a (Hermitian) symmetric matrix, so $\boldsymbol{A} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}'$.

Consider the linear transform $\boldsymbol{x} \mapsto \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}' \boldsymbol{x}$.

We can think of $\boldsymbol{A}\boldsymbol{x}$ as a cascade of three linear transforms:

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

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

It is useful to understand these three operations geometrically.

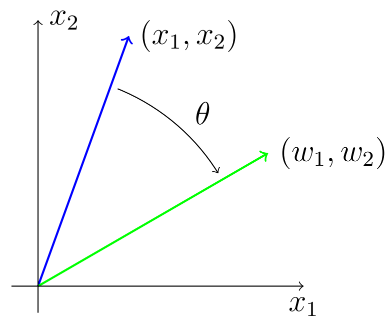
Example. We illustrate for the case of a symmetric 2×2 matrix.

Fact. Every $\boldsymbol{V} \in \mathbb{R}^{2\times 2}$ with $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_2$ (*i.e.*, **orthogonal**) has the following form for some rotation angle θ:

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

Exercise. Verify that $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_2$.
For simplicity we focus on the case of rotation matrices hereafter where $q = +1$.

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



Importantly, the length of $\boldsymbol{x}$ and $\boldsymbol{w}$ are the same:

$$\boldsymbol{w} = \boldsymbol{V}'\boldsymbol{x} \Longrightarrow \|\boldsymbol{w}\|_2^2 = \boldsymbol{w}'\boldsymbol{w} = (\boldsymbol{V}'\boldsymbol{x})'\boldsymbol{V}'\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{V}\boldsymbol{V}'\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{x} = \|\boldsymbol{x}\|_2^2.$$

The next mapping is

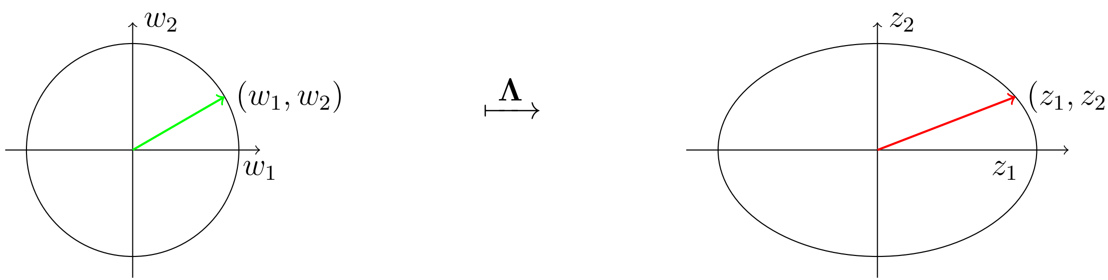
$$\boldsymbol{w} \mapsto \boldsymbol{z} = \boldsymbol{\Lambda w}, \text{ where } \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \Longrightarrow \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 w_1 \\ \lambda_2 w_2 \end{bmatrix}.$$

For interpretation, assume $\lambda_1, \lambda_2 \neq 0$ and suppose $\|\boldsymbol{x}\|_2 = 1 \Longrightarrow \boldsymbol{x}'\boldsymbol{x} = 1 \Longrightarrow x_1^2 + x_2^2 = 1$, which in turn implies $w_1^2 + w_2^2 = 1$, *i.e.*, $\boldsymbol{x}$ and $\boldsymbol{w}$ 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., $\boldsymbol{z}$ lies on an ellipse with axes governed by λ_1 and λ_2.

Graphically:



Typically $\boldsymbol{z}$ is *not* collinear with $\boldsymbol{w}$.
The exception is when $\lambda_1 = \lambda_2$, in which case $\boldsymbol{A} = \lambda_1 \boldsymbol{I}_2$, which is a trivial case.

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

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

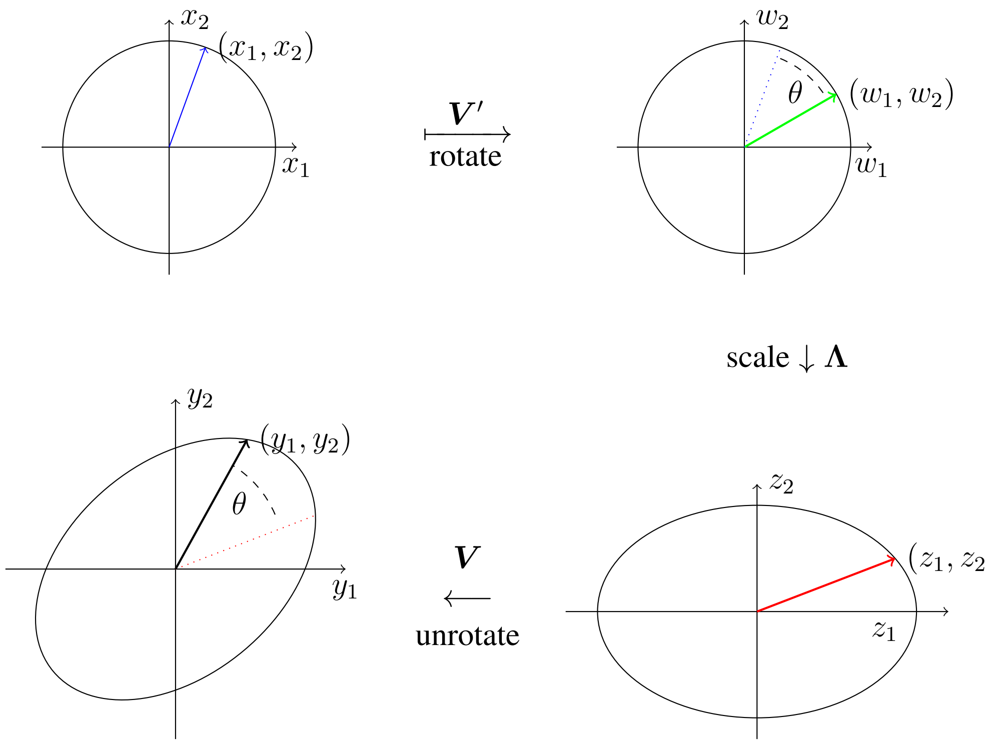
If $\boldsymbol{V}$ includes a sign flip, *e.g.*, $\boldsymbol{V} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, then $\boldsymbol{V}^{-1} = \boldsymbol{V}'$ also undoes that sign flip.

- 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.

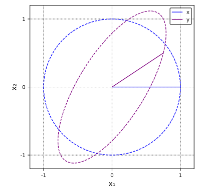
The next page gives a graphical summary in 2×2 case of $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{V} \overbrace{\boldsymbol{\Lambda} \underbrace{\boldsymbol{V}'\boldsymbol{x}}_{\boldsymbol{w}}}^{\boldsymbol{z}}$.

(The same principles apply in higher dimensions.)


x_2
(x_1, x_2)
x_1
$\boldsymbol{V}'$
rotate
w_2
θ
(w_1, w_2)
w_1
scale $\downarrow$ $\boldsymbol{\Lambda}$
y_2
(y_1, y_2)
θ
y_1
$\boldsymbol{V}$
unrotate
z_2
(z_1, z_2)
z_1

Demo

https://web.eecs.umich.edu/~fessler/course/551/julia/demo/02_eigshow1.html
https://web.eecs.umich.edu/~fessler/course/551/julia/demo/02_eigshow1.ipynb



Practical implementation

`d = eigvals(A)` returns 1D array of eigenvalues
`V = eigvecs(A)` returns matrix of eigenvectors
`obj = eigen(A)` returns an "object" (akin to a dictionary) of type `Eigen` with components:
`d = obj.values` or `d = eigen(A).values` (vector containing eigenvalues)
`V = obj.vectors` or `V = eigen(A).vectors` (matrix with eigenvectors)
Note the use of argument/index chaining: `f(arg1).arg2` `f(arg1)[1]`
To extract both parts in one line use: `d,V = eigen(A)`
If $\boldsymbol{A}$ is **diagonalizable**, then $\boldsymbol{A} = \boldsymbol{V} \operatorname{diag}\{\boldsymbol{d}\} \boldsymbol{V}^{-1}$, *i.e.*, `A = V * Diagonal(d) * inv(V)`

2.2 SVD

Singular values and singular vectors

Define. A non-negative real number σ is called a **singular value** of a $M \times N$ matrix $\boldsymbol{A}$ iff there exists unit norm vectors $\boldsymbol{u} \in \mathbb{C}^M$ and $\boldsymbol{v} \in \mathbb{C}^N$ for which

- $\boldsymbol{A}\boldsymbol{v} = \sigma \boldsymbol{u}$, and
- $\boldsymbol{A}'\boldsymbol{u} = \sigma \boldsymbol{v}$.

Any time this pair of relationships holds, we call $\boldsymbol{u}$ and $\boldsymbol{v}$ a pair of left and right **singular vectors** of $\boldsymbol{A}$.

Fact.

- A $M \times N$ matrix has at most $\min(M, N)$ distinct **singular values**.
- In contrast, the set of all possible **singular vectors** of $\boldsymbol{A}$ is uncountably infinite because if $\boldsymbol{u}$ and $\boldsymbol{v}$ are a pair of left and right singular vectors, then $-\boldsymbol{u}$ and $-\boldsymbol{v}$ are also a pair of left and right singular vectors. More generally, $\mathrm{e}^{\imath\phi}\, \boldsymbol{u}$ and $\mathrm{e}^{\imath\phi}\, \boldsymbol{v}$ are also a pair of left and right singular vectors.

Example. Consider the 1×2 matrix $\boldsymbol{A} = \begin{bmatrix} 3 & 4 \end{bmatrix}$ and let $\boldsymbol{u} = [1]$ and $\boldsymbol{v} = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix}$. (Read)

Then $\boldsymbol{A}\boldsymbol{v} = 5 = 5\boldsymbol{u}$ and $\boldsymbol{A}'\boldsymbol{u} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 5\boldsymbol{v}$ so $\boldsymbol{u}$ and $\boldsymbol{v}$ are a pair of left and right singular vectors corresponding to the singular value $\sigma = 5$. This matrix has no other singular values because $\min(M, N) = 1$. This matrix is not square so it does not have an eigendecomposition.

Existence of SVD

L§5.1

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

$$\boldsymbol{X} = \qquad (2.4)$$

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

- $\boldsymbol{U}$ is $M \times M$ and unitary: $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}' = \boldsymbol{I}_M$, and its columns consist of M **left singular vectors** of $\boldsymbol{X}$.
- $\boldsymbol{V}$ is $N \times N$ and unitary: $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}' = \boldsymbol{I}_N$ and its columns consist of N **right singular vectors** of $\boldsymbol{X}$.
- $\boldsymbol{\Sigma}$ is a $M \times N$ **rectangular diagonal matrix** containing *all* the $\min(M, N)$ **singular values** of $\boldsymbol{X}$.
- The $M \times N$ matrix $\boldsymbol{\Sigma}$ looks like one of:

$$\underbrace{\boldsymbol{\Sigma} = \left[\begin{array}{ccc} \sigma_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma_N \\ \hline & \mathbf{0}_{M-N,N} & \end{array}\right]}_{M > N \text{ (tall)}} \text{ or } \underbrace{\boldsymbol{\Sigma} = \left[\begin{array}{ccc|c} \sigma_1 & & \mathbf{0} & \\ & \ddots & & \mathbf{0}_{M,N-M} \\ \mathbf{0} & & \sigma_M & \end{array}\right]}_{N > M \text{ (wide)}} \text{ or } \underbrace{\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_M \end{bmatrix}}_{M = N \text{ (square)}}.$$

- These possible shapes of $\boldsymbol{\Sigma}$ are why the sum in (2.4) has the $\min(M, N)$ limit.
- The **singular values** $\sigma_1, \ldots, \sigma_{\min\{M,N\}}$ are real and nonnegative.

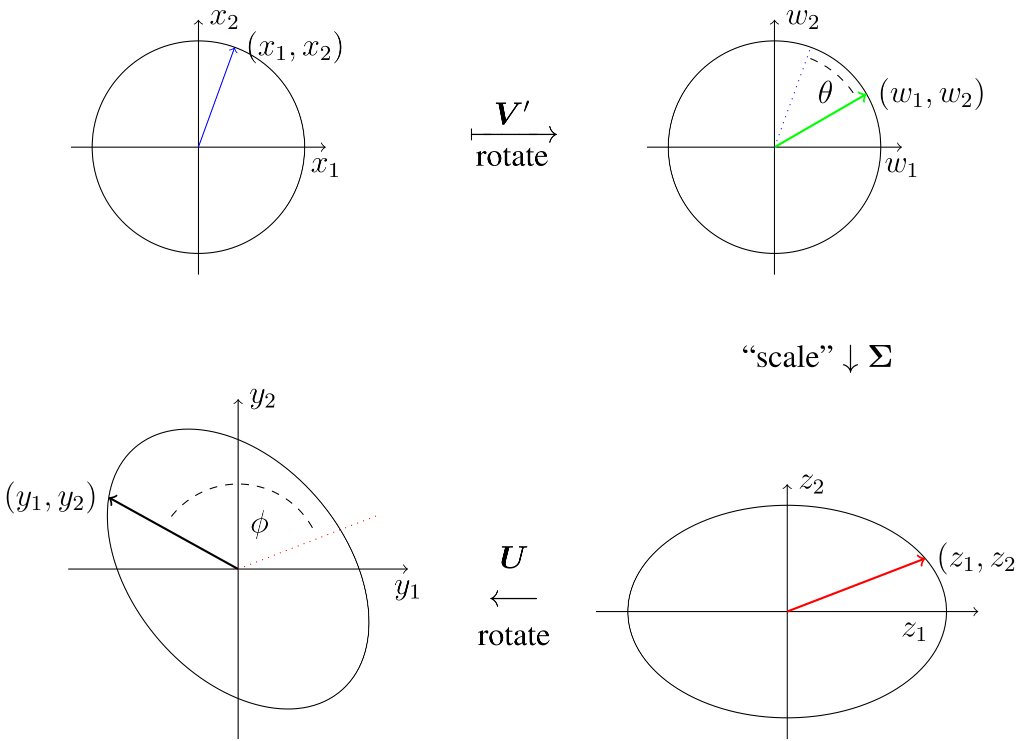
- By convention we always use descending order: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(M,N)}$.
- The first r singular values are positive, where $0 \leq r \leq \min\{M, N\}$ is the **rank** of the matrix (later).
- We often casually call $\boldsymbol{U}$ and $\boldsymbol{V}$ "the" left and right singular vectors of $\boldsymbol{X}$, but this wording is imprecise.
- For a history of the SVD, see [2].
- A subtle point is that the sum form (2.4) does not use all columns of $\boldsymbol{U}$ when $M > N$, nor all columns of $\boldsymbol{V}$ when $N > M$. More on this later when we discuss the **compact SVD**.

Geometry

Example. If $\boldsymbol{A}$ is any real 2×2 matrix, then its SVD looks like:

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

where $\theta \neq \phi$ in general. (In fact when $M \neq N$, $\boldsymbol{U}$ and $\boldsymbol{V}$ even have different sizes!)
Consider the case $q_1 = q_2 = +1$ for simplicity. The next page illustrates the 2×2 geometry is graphically:


x_2
(x_1, x_2)
x_1
$\boldsymbol{V}'$
rotate
w_2
θ
(w_1, w_2)
w_1
"scale" $\downarrow \boldsymbol{\Sigma}$
y_2
(y_1, y_2)
ϕ
y_1
$\boldsymbol{U}$
rotate
z_2
(z_1, z_2)
z_1

What are the differences here (for SVD) from before (eigendecomposition?

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

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

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

By inspection we can choose $\boldsymbol{U} = \boldsymbol{R}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_2$ and $\boldsymbol{V} = \boldsymbol{I}_2$.

This matrix $\boldsymbol{R}$ has a unique SVD. (?)

A: True B: False **??**

More geometry

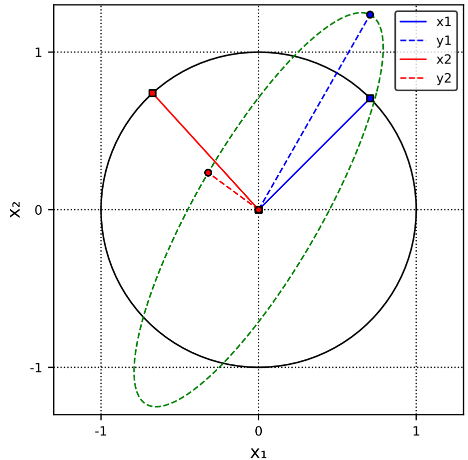
For eigendecomposition we asked "when is $\boldsymbol{A}\boldsymbol{x}$ aligned with $\boldsymbol{x}$?"

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

What choice of vectors $\boldsymbol{x}$ and $\boldsymbol{z}$ makes $\boldsymbol{A}\boldsymbol{x}$ perpendicular to $\boldsymbol{A}\boldsymbol{z}$?

Demo

https://web.eecs.umich.edu/~fessler/course/551/julia/demo/02_eigshow2.html
https://web.eecs.umich.edu/~fessler/course/551/julia/demo/02_eigshow2.ipynb



In general, SVD maps "circles" (precisely: the set of vectors with unit norm) to "rotated ellipses."

- Major and minor axes directions correspond to the left singular vectors in $\boldsymbol{U}$.
- Eccentricity depends on the singular values.

Practical implementation (Read)

`obj = svd(A)` returns an "object" (akin to a dictionary) of type `SVD` with components:
`U = obj.U` has size $M \times \min(M, N)$.
Caution: `U` differs from $\boldsymbol{U} \in \mathbb{F}^{M \times M}$ in the typical **tall** case where $M > N$
`s = obj.S` for vector $\boldsymbol{s}$ of length $\min(M, N)$ such that $\operatorname{diag}\{\boldsymbol{s}\}$ is the core of $\boldsymbol{\Sigma}$
`Vt = obj.Vt` for $\boldsymbol{V}'$ has size $\min(M, N) \times N$
`V = obj.V` (internally this is an `Adjoint` array to avoid a transpose!)
Caution: `V` differs from $\boldsymbol{V} \in \mathbb{F}^{N \times N}$ in the **wide** case where $M < N$.

Another option is `U,s,V = svd(A)` or `U,s,V = (svd(A)...,)`

One can rebuild $\boldsymbol{A}$ (to within numerical precision) using `U * Diagonal(s) * V'`

If one wants a **full SVD** where $\boldsymbol{U}$ is $M \times M$ and $\boldsymbol{V}$ is $N \times N$ then use `svd(A, full=true)`
but we rarely need that in practice!
The default `svd(A)` is equivalent to `svd(A, full=false)` because this is the usual use case.

The non-full default in JULIA is equivalent to the **economy SVD** option `svd(A,'econ')` in MATLAB, also sometimes called the **thin SVD**.

Ch. 3 describes a **compact SVD** that differs further from the **economy SVD** and **full SVD**.
In terms of sizes: **compact** $\leq$ **economy** $=$ **thin** $\leq$ **full**. See Ch. 3 for an example.

See [3] for a recent survey of methods for computing an SVD.

2.3 The matrix 2-norm or spectral norm

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

Note the "systems" language here. Very often when we talk about $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ we also think about a system block diagram like

$$\underset{\text{input } \mathbb{F}^M}{\boldsymbol{x}} \rightarrow \boxed{\underset{\text{(linear) system}}{\boldsymbol{A}}} \rightarrow \underset{\text{output } \mathbb{F}^N}{\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}}.$$

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

Expressing the question mathematically: L§7.4

$$\boldsymbol{x}_\star = \qquad\qquad \text{where } \|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{\boldsymbol{x}'\boldsymbol{x}}.$$

Claim: the first right singular vector (having the largest singular value σ_1).

Another maximizer is

Proof. Because $\boldsymbol{V}$ is orthogonal (or unitary), we can write $\boldsymbol{x}$ in terms of the $\boldsymbol{V}$ basis as $\boldsymbol{x} = \boldsymbol{V}\boldsymbol{z}$ where $\boldsymbol{z} = \boldsymbol{V}'\boldsymbol{x}$ are the coefficients. Note that $\|\boldsymbol{x}\| = 1 \iff \|\boldsymbol{z}\| = 1$.

Thus $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'\boldsymbol{x} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'\boldsymbol{V}\boldsymbol{z} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{z}$ so $\|\boldsymbol{A}\boldsymbol{x}\| = \sqrt{(\boldsymbol{A}\boldsymbol{x})'(\boldsymbol{A}\boldsymbol{x})} = \sqrt{\boldsymbol{z}'\boldsymbol{\Sigma}'\boldsymbol{U}'\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{z}} = \sqrt{\boldsymbol{z}'\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\boldsymbol{z}}$ $= \sqrt{\sum_{k=1}^r \sigma_k^2 \left|z_k\right|^2} \leq \sqrt{\sigma_1^2 \sum_{k=1}^r \left|z_k\right|^2} = \sigma_1 \|\boldsymbol{z}\| = \sigma_1$, where $r = \min(M, N)$. So $\|\boldsymbol{x}\| = 1 \Longrightarrow \|\boldsymbol{A}\boldsymbol{x}\| \leq \sigma_1$. This gives an upper bound on the norm of the output. But we can achieve that upper bound by choosing $\boldsymbol{x} = \boldsymbol{v}_1$ because then $\boldsymbol{z} = (1, 0, 0, \ldots, 0)$ and $\boldsymbol{A}\boldsymbol{x} = \sigma_1 \boldsymbol{u}_1$ so $\|\boldsymbol{A}\boldsymbol{x}\| = \|\sigma_1 \boldsymbol{u}_1\| = \sigma_1$. □

Fact. The solution $\boldsymbol{x}_\star = \boldsymbol{v}_1$ is unique to within a phase factor $\mathrm{e}^{\imath\phi}$ iff $\sigma_1 > \sigma_2$.

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

$$\|\boldsymbol{A}\|_2 \triangleq \max_{\boldsymbol{x} \,:\, \|\boldsymbol{x}\|_2 = 1} \|\boldsymbol{A}\boldsymbol{x}\|_2 = \max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} = \sigma_1. \tag{2.5}$$

By definition the 2-norm of a matrix $\boldsymbol{A}$ gives a **tight upper bound** on how much the 2-norm of a vector can be amplified when multiplying by $\boldsymbol{A}$:

$$\|\boldsymbol{A}\boldsymbol{x}\|_2 \leq$$

It is tight because the upper bound is achieved when $\boldsymbol{x} = \boldsymbol{v}_1$.

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 jth transmit antenna and the ith receive antenna.
(Matrix $\boldsymbol{A}$ depends on amplifier and receiver properties and the multi-path wave propagation between them.)

M Receive			N Transmit
		$>-$	1
1	$-<$	$>-$	2
$\vdots$	$-<$	$>-$	$\vdots$
M	$-<$	$>-$	N

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: $\|\boldsymbol{x}\| \leq 1$.

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

Exercise. The reader should verify the second equality in (2.5) and that

$$\underset{\boldsymbol{x} : \|\boldsymbol{x}\| \leq 1}{\arg\max} \|\boldsymbol{A}\boldsymbol{x}\| = \underset{\boldsymbol{x} : \|\boldsymbol{x}\| = 1}{\arg\max} \|\boldsymbol{A}\boldsymbol{x}\| .$$

In practice, the designer of a wifi system does not know the $\boldsymbol{A}$ for your house, and in fact $\boldsymbol{A}$ changes if you rearrange the furniture or relocate your computer. So the wifi router must determine $\boldsymbol{A}$ "on the fly" and this process is called **channel estimation**. The basic idea is that the transmitter first sends N orthonormal training waveforms $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{C}^N$, called **pilot signals**, to the receiver. The receiver records the corresponding outputs $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \in \mathbb{C}^M$, where $\boldsymbol{y}_n = \boldsymbol{A}\boldsymbol{x}_n$. Writing as a matrix:

$$[\boldsymbol{y}_1 \ \ldots \ \boldsymbol{y}_N] = \boldsymbol{A}[\boldsymbol{x}_1 \ \ldots \ \boldsymbol{x}_N] \Longrightarrow \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}.$$

Because $\boldsymbol{X}$ is unitary by design, we can estimate $\boldsymbol{A}$ as

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

The receiver must know what pilot signals $\boldsymbol{X}$ are transmitted; this specification is part of the protocol.

Once the router has determined $\boldsymbol{A}$, it can compute its SVD and use $\boldsymbol{v}_1$ as the best transmit amplitude vector.

This process 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 $\boldsymbol{A}$ when we need only $\boldsymbol{v}_1$?
Later we will see how to estimate $\boldsymbol{v}_1$ with fewer measurements (at least when N is large).

Eigenvalues as optimization problems

(Read)

The equality (2.5) expresses the first **singular value** σ_1 of an arbitrary matrix $\boldsymbol{A}$ as the solution of an optimization problem:

$$\sigma_1 = \max_{\boldsymbol{x} \,:\, \|\boldsymbol{x}\|_2 = 1} \|\boldsymbol{A}\boldsymbol{x}\|_2 = \max_{\boldsymbol{x} \,:\, \|\boldsymbol{x}\|_2 = 1} \sqrt{\boldsymbol{x}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x}}.$$

Can we also express any **eigenvalue** of a square matrix as some optimization problem?
The answer is yes, at least for Hermitian matrices.

Fact. If $\boldsymbol{A}$ is a $N \times N$ **Hermitian matrix**, with eigenvalues ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$, then

$$\lambda_1 = \max_{\boldsymbol{x} \,:\, \|\boldsymbol{x}\|_2 = 1} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}, \qquad \lambda_N = \min_{\boldsymbol{x} \,:\, \|\boldsymbol{x}\|_2 = 1} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}.$$

A generalization (under the same assumptions) is [4]: ♦♦

$$\sum_{k=1}^{K} \lambda_k = \max_{(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_K) \in \mathcal{X}_K} \sum_{k=1}^{K} \boldsymbol{x}_k'\boldsymbol{A}\boldsymbol{x}_k, \qquad \sum_{k=N-K+1}^{N} \lambda_k = \min_{(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_K) \in \mathcal{X}_K} \sum_{k=N-K+1}^{N} \boldsymbol{x}_k'\boldsymbol{A}\boldsymbol{x}_k,$$

where $\mathcal{X}_K$ denotes the collection of all possible sets of K **orthonormal vectors** in $\mathbb{F}^N$.

For other eigenvalue and **generalized eigenvalue** problems put in optimization form see [5].

2.4 Relating SVDs and eigendecompositions

The two big topics of this chapter are SVDs and eigendecompositions. Are they related? Every matrix, even if rectangular, has an SVD, whereas only some square matrices have an eigendecomposition. Nevertheless, we can find some relationships.

- A **right singular vector** of $\boldsymbol{A}$ is an eigenvector of $\boldsymbol{A}'\boldsymbol{A}$.
- A **left singular vector** of $\boldsymbol{A}$ is an eigenvector of $\boldsymbol{A}\boldsymbol{A}'$.
- The $\min(M, N)$ singular values of a matrix $\boldsymbol{A} \in \mathbb{F}^{M \times N}$ are the square roots of the eigenvalues of $\boldsymbol{A}'\boldsymbol{A}$ or $\boldsymbol{A}\boldsymbol{A}'$, whichever is the smaller matrix.

Example. For the case where $\boldsymbol{A}$ is tall and $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$, then

$$\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{V} \underbrace{\boldsymbol{\Sigma}'\boldsymbol{\Sigma}}_{\boldsymbol{\Lambda}} \boldsymbol{V}' = \boldsymbol{V} \underbrace{\begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_N^2 \end{bmatrix}}_{\boldsymbol{\Lambda}} \boldsymbol{V}',$$

which has eigenvalues $\{\sigma_k^2\}$. So the square roots of those eigenvalues of $\boldsymbol{A}'\boldsymbol{A}$ (properly sorted) are the singular values of $\boldsymbol{A}$.

This is about all we can say to relate SVDs and eigendecompositions for general (rectangular) matrices, so next we turn to square matrices.

In general, if $\boldsymbol{A}$ is an arbitrary square matrix, there is not too much one can say to relate its eigenvalues and its singular values. One known relationship is **Weyl's inequality**: $|\lambda_1(\boldsymbol{A}) \cdots \lambda_K(\boldsymbol{A})| \leq \sigma_1(\boldsymbol{A}) \cdots \sigma_K(\boldsymbol{A})$ for $1 \leq K \leq N$, assuming that we order the eigenvalues of $\boldsymbol{A}$ so that $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_N|$.
Probably the most useful inequality is the case $K = 1$, saying $|\lambda_1| \leq \sigma_1$.

In general, if $\boldsymbol{A}$ is an arbitrary **diagonalizable** square matrix, then its eigendecomposition $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$ is *unrelated* to any SVD of $\boldsymbol{A}$.

However, for **normal** matrices, we can relate any eigendecomposition of $\boldsymbol{A}$ to an SVD of $\boldsymbol{A}$.

If $\boldsymbol{A}$ is a **normal** $N \times N$ matrix (*e.g.*, Hermitian symmetric), then it has a unitary eigendecomposition

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}' = \textstyle\sum_{n=1}^{N} \lambda_n \boldsymbol{v}_n \boldsymbol{v}_n'.$$

Without loss of generality, we can order the eigenvalues with decreasing magnitudes, *i.e.*, $|\lambda_1| \geq \ldots \geq |\lambda_N|$. However, even after doing so, $\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$ is still not an SVD of $\boldsymbol{A}$ in general because some of the eigenvalues λ_n can be negative or even complex, so $\boldsymbol{\Sigma} \neq \boldsymbol{\Lambda}$. To construct an SVD of $\boldsymbol{A}$ in terms of $\boldsymbol{V}$ and $\boldsymbol{\Lambda}$, we use the fact that $\lambda_n = \operatorname{sign}(\lambda_n) |\lambda_n|$, where $\operatorname{sign}\big(|z| \, \mathrm{e}^{\imath \angle z}\big) = \mathrm{e}^{\imath \angle z}$ for $z \in \mathbb{C}$, as follows:

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}' = \sum_{n=1}^{N} \lambda_n \boldsymbol{v}_n \boldsymbol{v}_n' = \qquad (2.6)$$

So when $\boldsymbol{A}$ is normal with eigenvectors $\boldsymbol{V}$ and eigenvalues $\boldsymbol{\Lambda}$ with descending magnitudes, an SVD of $\boldsymbol{A}$ is:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}', \quad \boldsymbol{U} = \boldsymbol{V}\boldsymbol{S}, \quad \boldsymbol{S} \triangleq \mathsf{diag}\{\mathrm{sign}(\lambda_n)\}, \quad \boldsymbol{\Sigma} = \mathsf{diag}\{|\lambda_n|\}\,. \tag{2.7}$$

This SVD is not unique; we could have associated some or all of the sign values with $\boldsymbol{V}$, for example.

The construction (2.6) shows that if $\boldsymbol{A}$ is **normal**, then for any unitary eigendecomposition of $\boldsymbol{A}$ of the form $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$, we can construct an SVD of the form (2.7) where the matrix $\boldsymbol{V}$ of right singular vectors consists entirely of eigenvectors of $\boldsymbol{A}$, and where the matrix $\boldsymbol{U}$ of left singular vectors also consists of entirely of eigenvectors of $\boldsymbol{A}$, because in (2.6) we have $\boldsymbol{u}_n = \mathrm{sign}(\lambda_n)\,\boldsymbol{v}_n$ so $\boldsymbol{u}_n$ is also an eigenvector of $\boldsymbol{A}$.

In short, if $\boldsymbol{A}$ is **normal** than we can construct an SVD of $\boldsymbol{A}$ using any orthonormal set of eigenvectors of $\boldsymbol{A}$. However, it does *not* follow that all singular vectors of a **normal** matrix $\boldsymbol{A}$ are eigenvectors!

<u>Example</u>. Consider $\boldsymbol{A} = \begin{bmatrix} 3 & 0 \\ 0 & -3 \end{bmatrix}$ and $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. This $\boldsymbol{x}$ is a right singular vector because $\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = 9\boldsymbol{x}$. However, this $\boldsymbol{x}$ is not an eigenvector of $\boldsymbol{A}$. To elaborate, here are two different SVDs of $\boldsymbol{A}$, one that is constructed from a set of eigenvectors of $\boldsymbol{A}$, and one that is not:

$$\boldsymbol{A} = \begin{bmatrix} 3 & 0 \\ 0 & -3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{V}} \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & -3 \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{V}'} \qquad \text{(an elementary eigendecomposition)}$$

$$= \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}}_{\boldsymbol{U}} \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}}_{\boldsymbol{\Sigma}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{V}'} \qquad \text{(SVD vers. 1, using } \boldsymbol{V} \text{ and } \boldsymbol{u}_n = \operatorname{sign}(\lambda_n)\, \boldsymbol{v}_n\text{)}$$

$$= \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{\tilde{\boldsymbol{U}}} \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}}_{\boldsymbol{\Sigma}} \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{\tilde{\boldsymbol{V}}'} \qquad \text{(SVD vers. 2, unrelated to eigenvectors of } \boldsymbol{A}\text{).}$$

Fact. If $\boldsymbol{A}$ is **normal** and has eigenvalues with distinct magnitudes, or if eigenvalues with equal magnitudes have equal values, then for every SVD of $\boldsymbol{A}$, the left and right singular vectors are all eigenvectors of $\boldsymbol{A}$. (Proof in a HW problem.)

When does $\boldsymbol{U} = \boldsymbol{V}$?

A question about the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$ that sometimes arises is: When does $\boldsymbol{U} = \boldsymbol{V}$?
This is an imprecise question because $\boldsymbol{U}$ and $\boldsymbol{V}$ are not unique. A more precise version is this:
For what class of matrices does there exist an SVD in which $\boldsymbol{U} = \boldsymbol{V}$?

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

If $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}'$ then clearly $\boldsymbol{A}$ is Hermitian symmetric.
But is symmetry a sufficient condition for possibly having $\boldsymbol{U} = \boldsymbol{V}$? No.

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

So to have $\boldsymbol{U} = \boldsymbol{V}$ we need $\boldsymbol{A}$ to be Hermitian symmetric *and* to have nonnegative eigenvalues.

An SVD of $\boldsymbol{A}$ can have $\boldsymbol{U} = \boldsymbol{V}$ iff

Refer to (2.6) for the key idea.

Example. Consider the eigendecomposition of the following matrix and the following three (not unique!) SVD forms:

$$\begin{aligned}
\boldsymbol{A} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix} &= \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_{\boldsymbol{V}} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{\boldsymbol{V}'} && \text{(eigendecomposition)} \\
&= \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{\boldsymbol{U}} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}}_{\boldsymbol{\Sigma}} \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{\boldsymbol{V}'} && \text{(SVD version 1)} \\
&= \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{\tilde{\boldsymbol{U}}} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}}_{\boldsymbol{\Sigma}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\tilde{\boldsymbol{V}}'} && \text{(SVD version 2)} \\
&= \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\tilde{\boldsymbol{U}}} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}}_{\boldsymbol{\Sigma}} \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{\tilde{\boldsymbol{V}}'} && \text{(SVD version 3).}
\end{aligned}$$

Even though $\boldsymbol{A}$ is symmetric, and we have exhibited three different SVDs for it, we cannot find an SVD here where $\boldsymbol{U}$ and $\boldsymbol{V}$ are the same because $\boldsymbol{A}$ has a negative eigenvalue:
$\det\{\boldsymbol{A} - z\boldsymbol{I}\} = z^2 - 4 \Longrightarrow z = \pm 2.$

Which of the above forms corresponds to (2.6)? SVD version 1, where $\boldsymbol{U} = \boldsymbol{V} \operatorname{diag}\{\operatorname{sign}(\lambda_i)\}$.

SVD clicker questions

Every permutation matrix has an SVD. (?)

A: True B: False $\boxed{??}$

If $\boldsymbol{A} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, then $\max_{\boldsymbol{x} \in \mathbb{R}^3 \,:\, \|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\| = 5$. (?)

A: True B: False $\boxed{??}$

The singular values of a permutation matrix all equal to 1. (?)

A: True B: False $\boxed{??}$

(A uniqueness question.) If $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$ and $\boldsymbol{A} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}'$ are both SVDs of $\boldsymbol{A}$, then $\boldsymbol{\Sigma} = \tilde{\boldsymbol{\Sigma}}$. (?)

A: True B: False $\boxed{??}$

$\boxed{??}$

By definition, to determine if $\boldsymbol{x}$ is an **eigenvector** of a square matrix $\boldsymbol{A}$, simply compute $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ and see if $\boldsymbol{y} = \alpha \boldsymbol{x}$ for some $\alpha \in \mathbb{F}$.

If $\boldsymbol{x}$ is an **eigenvector** of a **normal** matrix $\boldsymbol{A}$ having unitary eigendecomposition $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$, then $\boldsymbol{x} = \alpha \boldsymbol{v}_j$ for some $\alpha \in \mathbb{F}$, where $\boldsymbol{v}_j$ is one of the columns of $\boldsymbol{V}$. (?)

A: True B: False

$\boxed{\textbf{??}}$

For a $M \times N$ matrix $\boldsymbol{A}$, how do we test if is $\boldsymbol{v} \in \mathbb{F}^N$ is a right singular vector or $\boldsymbol{u} \in \mathbb{F}^M$ is a left singular vector?

- A **right singular vector** of $\boldsymbol{A}$ is an eigenvector of $\boldsymbol{A}'\boldsymbol{A}$.
- A **left singular vector** of $\boldsymbol{A}$ is an eigenvector of $\boldsymbol{A}\boldsymbol{A}'$.

If $\boldsymbol{x}$ is a **right singular vector** of a matrix $\boldsymbol{A}$ having SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$, then $\boldsymbol{x} = \alpha \boldsymbol{v}_j$ for some $\alpha \in \mathbb{F}$, where $\boldsymbol{v}_j$ is one of the columns of $\boldsymbol{V}$. (?)

A: True B: False $\boxed{\textbf{??}}$

2.5 Positive semidefinite matrices

L§10.2

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 $\boldsymbol{X}'\boldsymbol{X}$ or an outer-product matrix $\boldsymbol{X}\boldsymbol{X}'$. These matrices are not only symmetric, they are **positive semi-definite**.

Define. A $N \times N$ (square) Hermitian matrix $\boldsymbol{A}$ is **positive semi-definite** iff

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

Lemma. If $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}'$ for any matrix $\boldsymbol{B}$, then $\boldsymbol{A}$ is positive

Proof. $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{B}\boldsymbol{B}'\boldsymbol{x} = \|\boldsymbol{B}'\boldsymbol{x}\|_2^2 \geq 0$.

We write to denote positive definite and to denote positive semi-definite.

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 $\boldsymbol{X}'\boldsymbol{X}$ or outer-product matrix $\boldsymbol{X}\boldsymbol{X}'$ is positive semi-definite, *i.e.*, $\boldsymbol{X}'\boldsymbol{X} \succeq \boldsymbol{0}$.

Theorem. If $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}'$ for any matrix $\boldsymbol{B}$, then $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}'$ with $\Sigma_{ii} \geq 0$.
In words, for such matrices an eigendecomposition with (real, nonnegative) eigenvalues in descending order is also an SVD.

Proof. Let $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}_B\boldsymbol{V}'$ denote an SVD of $\boldsymbol{B}$. Recall $\boldsymbol{\Sigma}_B$ is real and nonnegative.
Then $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}' = \boldsymbol{U}\boldsymbol{\Sigma}_B\boldsymbol{V}'\boldsymbol{V}\boldsymbol{\Sigma}_B'\boldsymbol{U}' = \boldsymbol{U}\boldsymbol{\Sigma}_B\boldsymbol{\Sigma}_B'\boldsymbol{U}' = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}'$, where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_B\boldsymbol{\Sigma}_B'$ is diagonal with entries σ_k^2 (and some zeros if $\boldsymbol{B}$ is tall).

So when $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}'$ the *eigenvalues* of $\boldsymbol{A}$ are the *square of the singular values* of $\boldsymbol{B}$ (and some zeros if $\boldsymbol{B}$ is tall), and hence real and nonnegative.

Example. Consider $\boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \\ 1 & 0 \end{bmatrix} \Longrightarrow \boldsymbol{A} = \boldsymbol{B}\boldsymbol{B}' = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 9 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

`eigvals(A)` returns $(0, 2, 9)$
`svdvals(B)` returns $\left(3, \sqrt{2}\right)$

Note that `eigen` and `eigvals` do not return eigenvalues in descending magnitude order in general, whereas `svd` and `svdvals` always return singular values in descending order.

Relating positive (semi)definiteness to eigenvalues

Let $\boldsymbol{A}$ be any $N \times N$ Hermitian matrix, then $\boldsymbol{A}$ has an eigendecomposition of the form $\boldsymbol{A} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}'$.
Now $\boldsymbol{x}' \boldsymbol{A} \boldsymbol{x} = \boldsymbol{x}' \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}' \boldsymbol{x} = \boldsymbol{z}' \boldsymbol{\Lambda} \boldsymbol{z} = \sum_i \lambda_i \left|z_i\right|^2$ where $\boldsymbol{z} = \boldsymbol{V}' \boldsymbol{x}$.
Thus if (Hermitian) $\boldsymbol{A}$ has all nonnegative eigenvalues, then $\boldsymbol{x}' \boldsymbol{A} \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in \mathbb{C}^N$, so $\boldsymbol{A} \succeq \boldsymbol{0}$.
The converse is also true: If $\boldsymbol{A}$ is symmetric positive semi-definite, then $\boldsymbol{A}$ has nonnegative eigenvalues.

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

To summarize, any positive semi-definite matrix has real and nonnegative eigenvalues,
and it has an SVD that matches its eigendecomposition (with descending eigenvalue order), with $\boldsymbol{U} = \boldsymbol{V}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}$, *i.e.*, $\sigma_n(\boldsymbol{A}) = \lambda_n(\boldsymbol{A}) \geq 0$.

Venn diagram of matrices and decompositions

Rectangular SVD: $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}'$	Square $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda}$	Diagonalizable $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$ iff $\boldsymbol{V}$ linearly ind.	Normal $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$ $\boldsymbol{V}$ unitary $\sigma_n = \left|\lambda_{(n)}\right|$	Hermitian $\boldsymbol{\Lambda}$ real	PSD: $\boldsymbol{A} \succeq \boldsymbol{0}$ $\boldsymbol{\Sigma} = \boldsymbol{\Lambda} \succeq \boldsymbol{0}$ $\boldsymbol{U} = \boldsymbol{V}$

where $\left|\lambda_{(n)}\right|$ denotes the nth largest magnitude eigenvalue.

2.6 Summary

In practice, we usually end up using:
- an eigendecomposition, $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$, when working with positive semi-definite matrices (like a Gram matrix or outer-product matrix),
- the SVD, $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$, for most other matrices.

A curious note about terminology:
- columns of $\boldsymbol{V}$ are called the **right eigenvectors** for the eigendecomposition, because $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Lambda}$ even though $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$.
- columns of $\boldsymbol{U}$ are called the **left singular vectors** for the SVD, because $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$.

SVD computation using eigendecomposition (Read)

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 `eigen` 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 $\boldsymbol{A}\boldsymbol{A}'$ to find $\boldsymbol{U}$ and $\boldsymbol{\Sigma}$:

$$\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}' = \boldsymbol{U} \underbrace{\boldsymbol{\Sigma}\boldsymbol{\Sigma}'}_{M \times M} \boldsymbol{U}'.$$

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

$$\begin{aligned}
\mathsf{diag}\{1/\sigma_n\}\, \boldsymbol{U}[:, 1:N]'\boldsymbol{A} &= \mathsf{diag}\{1/\sigma_n\}\, \boldsymbol{U}[:, 1:N]'(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}') = \mathsf{diag}\{1/\sigma_n\} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \end{bmatrix} \boldsymbol{\Sigma}\boldsymbol{V}' \\
&= \mathsf{diag}\{1/\sigma_n\}\, \mathsf{diag}\{\sigma_n\}\, \boldsymbol{V}' = \boldsymbol{V}'.
\end{aligned}$$

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

Example. (Read)

Exercise. Find an **SVD** of $\boldsymbol{A} = \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix}$.

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

$$\boldsymbol{A}'\boldsymbol{A} = \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 9 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}' \Longrightarrow \boldsymbol{V} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}.$$

$$\boldsymbol{A}\boldsymbol{A}' = \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} 9 & 0 \\ 0 & 0 \end{bmatrix} = \boldsymbol{I}_2 \begin{bmatrix} 9 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{I}_2 = \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}' \Longrightarrow \boldsymbol{U} = \boldsymbol{I}_2.$$

Note that to find $\boldsymbol{V}$ properly, we applied a permutation to have the eigenvalues of $\boldsymbol{A}'\boldsymbol{A}$ in descending order. And in general one would need to do more work to properly match the $\boldsymbol{U}$ and $\boldsymbol{V}$ ordering.

Thus an SVD is $\boldsymbol{A} = \underbrace{\boldsymbol{I}_2}_{\boldsymbol{U}} \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}}_{\boldsymbol{\Sigma}} \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{\boldsymbol{V}'}$.

Bibliography

[1] A. J. Laub. *Matrix analysis for scientists and engineers*. Soc. Indust. Appl. Math., 2005 (cit. on pp. 2.2, 2.19).

[2] G. W. Stewart. "On the early history of the singular value decomposition". In: *SIAM Review* 35.4 (1993), 551–66 (cit. on p. 2.20).

[3] J. Dongarra, M. Gates, A. Haidar, J. Kurzak, P. Luszczek, S. Tomov, and I. Yamazaki. "The singular value decomposition: anatomy of optimizing an algorithm for extreme scale". In: *SIAM Review* 60.4 (2018), 808–65 (cit. on p. 2.24).

[4] K. Fan. "On a theorem of Weyl concerning eigenvalues of linear transformations I". In: *Proc. Natl. Acad. Sci.* 35.11 (Nov. 1949), 652–5 (cit. on p. 2.29).

[5] B. Ghojogh, F. Karray, and M. Crowley. *Eigenvalue and generalized eigenvalue problems: Tutorial*. 2019 (cit. on p. 2.29).