Matrix Decompositions

Libin Jiao

Dalian University of Technology

October 9, 2022

Outline

Outline

In machine learning, data is often represented in matrix form as well, e.g., where the rows of the matrix represent different people and the columns describe different features of the people, such as weight, height, and socioeconomic status. In this chapter, we present three aspects of matrices:

- how to summarize matrices,
- how matrices can be decomposed,
- how these decompositions can be used for matrix approximations.

In machine learning, data is often represented in matrix form as well, e.g., where the rows of the matrix represent different people and the columns describe different features of the people, such as weight, height, and socioeconomic status. In this chapter, we present three aspects of matrices:

- how to summarize matrices,
- how matrices can be decomposed,
- how these decompositions can be used for matrix approximations.

In machine learning, data is often represented in matrix form as well, e.g., where the rows of the matrix represent different people and the columns describe different features of the people, such as weight, height, and socioeconomic status. In this chapter, we present three aspects of matrices:

- how to summarize matrices,
- how matrices can be decomposed,
- how these decompositions can be used for matrix approximations.

In machine learning, data is often represented in matrix form as well, e.g., where the rows of the matrix represent different people and the columns describe different features of the people, such as weight, height, and socioeconomic status. In this chapter, we present three aspects of matrices:

- how to summarize matrices,
- how matrices can be decomposed,
- how these decompositions can be used for matrix approximations.

Determinant

Determinants are only defined for square matrices, i.e., matrices with the same number of rows and columns. In this book, we write the determinant as $\det(\boldsymbol{A})$ or sometimes as $|\,\boldsymbol{A}\,|$ so that

$$\det \boldsymbol{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Determinant

- if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $\mid \boldsymbol{A} \mid = a_{11}a_{22} - a_{21}a_{12}$
- in general,

$$\mid \boldsymbol{A} \mid = \sum_{j=1}^{n} a_{ij}\mathrm{cof}(a_{ij}),$$

  - $\mathrm{cof}(a_{ij})$ is the cofactor of element a_{ij} and is defined as the product of $(-1)^{i+j}$ times the determinant of $\boldsymbol{A}$ after deleting its ith row and jth column.

Determinant

- if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $\mid \boldsymbol{A} \mid = a_{11}a_{22} - a_{21}a_{12}$
- in general,

$$\mid \boldsymbol{A} \mid = \sum_{j=1}^{n} a_{ij} \mathrm{cof}(a_{ij}),$$

  - $\mathrm{cof}(a_{ij})$ is the cofactor of element a_{ij} and is defined as the product of $(-1)^{i+j}$ times the determinant of $\boldsymbol{A}$ after deleting its ith row and jth column.

Determinant

- if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then $\mid \boldsymbol{A} \mid = a_{11}a_{22} - a_{21}a_{12}$
- in general,

$$\mid \boldsymbol{A} \mid = \sum_{j=1}^{n} a_{ij}\mathrm{cof}(a_{ij}),$$

  - $\mathrm{cof}(a_{ij})$ is the cofactor of element a_{ij} and is defined as the product of $(-1)^{i+j}$ times the determinant of $\boldsymbol{A}$ after deleting its ith row and jth column.

Determinant

Theorem (Theorem 4.2(Laplace Expansion))

Consider a matrix $\boldsymbol{A} \in (R)^{n\times n}$, then, for all $j = 1, 2, \ldots, n$:

1. *Expansion along column j*

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j}),$$

2. *Expansion along row j*

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A}_{j,k}).$$

Here,$\boldsymbol{A}_{k,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ is the submatrix of $\boldsymbol{A}$ that we obtain deleting row k and column j.

Note: $det(\boldsymbol{A}_{k,j})$ is called minor and $(-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j})$ a cofactor.

Determinant

Theorem (Theorem 4.2(Laplace Expansion))

Consider a matrix $\boldsymbol{A} \in (R)^{n\times n}$*, then, for all* $j = 1, 2, \ldots, n$*:*

1. *Expansion along column* j

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j}),$$

2. *Expansion along row* j

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A}_{j,k}).$$

Here,$\boldsymbol{A}_{k,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ *is the submatrix of* $\boldsymbol{A}$ *that we obtain deleting row* k *and column* j.

Note: $det(\boldsymbol{A}_{k,j})$ is called minor and $(-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j})$ a cofactor.

Determinant

Theorem (Theorem 4.2(Laplace Expansion))

Consider a matrix $\boldsymbol{A} \in (R)^{n\times n}$*, then, for all* $j = 1, 2, \ldots, n$*:*

1. *Expansion along column* j

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j}),$$

2. *Expansion along row* j

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A}_{j,k}).$$

Here, $\boldsymbol{A}_{k,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ *is the submatrix of* $\boldsymbol{A}$ *that we obtain deleting row* k *and column* j*.*

Note: $det(\boldsymbol{A}_{k,j})$ is called minor and $(-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j})$ a cofactor.

Determinant

Theorem (Theorem 4.2(Laplace Expansion))

Consider a matrix $\boldsymbol{A} \in (R)^{n\times n}$, then, for all $j = 1, 2, \dots, n$:

1. *Expansion along column j*

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j}),$$

2. *Expansion along row j*

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A}_{j,k}).$$

Here,$\boldsymbol{A}_{k,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ is the submatrix of $\boldsymbol{A}$ that we obtain deleting row k and column j.

Note: $det(\boldsymbol{A}_{k,j})$ is called minor and $(-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j})$ a cofactor.

Determinant

Theorem (Theorem 4.2(Laplace Expansion))

Consider a matrix $\boldsymbol{A} \in (R)^{n\times n}$*, then, for all* $j = 1, 2, \ldots, n$*:*

1. *Expansion along column* j

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j}),$$

2. *Expansion along row* j

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A}_{j,k}).$$

Here, $\boldsymbol{A}_{k,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ *is the submatrix of* $\boldsymbol{A}$ *that we obtain deleting row* k *and column* j*.*

Note: $det(\boldsymbol{A}_{k,j})$ is called minor and $(-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{k,j})$ a cofactor.

Determinant

Proposition

For a triangular matrix $\boldsymbol{T} \in \mathbb{R}^{n\times n}$, the determinant is the product of the diagonal elements, i.e.,

$$\det(\boldsymbol{T}) = \prod_{i=1}^{n} \boldsymbol{T}_{ii}$$

.

Theorem (Theorem 4.1)

For any square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, it holds that $\boldsymbol{A}$ is invertible if and only if $\det(\boldsymbol{A}) \neq 0$.

Determinant

Proposition

For a triangular matrix $\boldsymbol{T} \in \mathbb{R}^{n \times n}$, the determinant is the product of the diagonal elements, i.e.,

$$\det(\boldsymbol{T}) = \prod_{i=1}^{n} \boldsymbol{T}_{ii}$$

.

Theorem (Theorem 4.1)

For any square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, it holds that $\boldsymbol{A}$ is invertible if and only if $\det(\boldsymbol{A}) \neq 0$.

Determinant

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$
- $|\lambda\boldsymbol{A}| = \lambda^n|\boldsymbol{A}|$

Determinant

Properties:

- determinant is a scalar quantity
- if $\mid \boldsymbol{A} \mid = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $\mid \boldsymbol{A}^T \mid = \mid \boldsymbol{A} \mid$
- $\mid \boldsymbol{AB} \mid = \mid \boldsymbol{BA} \mid = \mid \boldsymbol{A} \mid \mid \boldsymbol{B} \mid$
- $\mid \lambda \boldsymbol{A} \mid = \lambda^n \mid \boldsymbol{A} \mid$

Determinant

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$
- $|\lambda \boldsymbol{A}| = \lambda^n |\boldsymbol{A}|$

Determinant

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$
- $|\lambda\boldsymbol{A}| = \lambda^n |\boldsymbol{A}|$

Determinant

Properties:

- determinant is a scalar quantity
- if $|\boldsymbol{A}| = 0$, then $\boldsymbol{A}$ is singular, otherwise non-singular
- $|\boldsymbol{A}^T| = |\boldsymbol{A}|$
- $|\boldsymbol{AB}| = |\boldsymbol{BA}| = |\boldsymbol{A}||\boldsymbol{B}|$
- $|\lambda\boldsymbol{A}| = \lambda^n |\boldsymbol{A}|$

Determinant

Theorem (Theorem 4.3)

A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ *has* $\det(\boldsymbol{A}) \neq 0$ *if and only if* $\operatorname{rk}(\boldsymbol{A}) = n$. *In other words,* $\boldsymbol{A}$ *is invertible if and only if it is full rank.*

Trace

Definition

The trace of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is defined as

$$tr(\boldsymbol{A}) = \sum_{i}^{n} a_{ii},$$

i.e., the trace is the sum of the diagonal elements of $\boldsymbol{A}$.

Trace

The trace satisfies the following properties:

- $\operatorname{tr}(\boldsymbol{A}+\boldsymbol{B})=\operatorname{tr}(\boldsymbol{A})+\operatorname{tr}(\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$
- $\operatorname{tr}(\alpha \boldsymbol{A})=\alpha \operatorname{tr}(\boldsymbol{A}), \alpha \in \mathbb{R}$ for $\boldsymbol{A} \in \mathbb{R}^{n \times n}$
- $\operatorname{tr}\left(\boldsymbol{I}_n\right)=n$
- $\operatorname{tr}(\boldsymbol{A B})=\operatorname{tr}(\boldsymbol{B A})$ for $\boldsymbol{A} \in \mathbb{R}^{n \times k}, \boldsymbol{B} \in \mathbb{R}^{k \times n}$

Trace

The trace satisfies the following properties:

- $\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$
- $\mathrm{tr}(\alpha\boldsymbol{A}) = \alpha\,\mathrm{tr}(\boldsymbol{A}), \alpha \in \mathbb{R}$ for $\boldsymbol{A} \in \mathbb{R}^{n\times n}$
- $\mathrm{tr}\left(\boldsymbol{I}_n\right) = n$
- $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$ for $\boldsymbol{A} \in \mathbb{R}^{n\times k}, \boldsymbol{B} \in \mathbb{R}^{k\times n}$

Trace

The trace satisfies the following properties:

- $\operatorname{tr}(\boldsymbol{A}+\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$
- $\operatorname{tr}(\alpha\boldsymbol{A}) = \alpha \operatorname{tr}(\boldsymbol{A}), \alpha \in \mathbb{R}$ for $\boldsymbol{A} \in \mathbb{R}^{n\times n}$
- $\operatorname{tr}(\boldsymbol{I}_n) = n$
- $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A})$ for $\boldsymbol{A} \in \mathbb{R}^{n\times k}, \boldsymbol{B} \in \mathbb{R}^{k\times n}$

Trace

The trace satisfies the following properties:

- $\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$
- $\mathrm{tr}(\alpha\boldsymbol{A}) = \alpha\,\mathrm{tr}(\boldsymbol{A}), \alpha \in \mathbb{R}$ for $\boldsymbol{A} \in \mathbb{R}^{n\times n}$
- $\mathrm{tr}\,(\boldsymbol{I}_n) = n$
- $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$ for $\boldsymbol{A} \in \mathbb{R}^{n\times k}, \boldsymbol{B} \in \mathbb{R}^{k\times n}$

Trace

The trace satisfies the following properties:

- $\operatorname{tr}(\boldsymbol{A}+\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B})$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$
- $\operatorname{tr}(\alpha\boldsymbol{A}) = \alpha \operatorname{tr}(\boldsymbol{A}), \alpha \in \mathbb{R}$ for $\boldsymbol{A} \in \mathbb{R}^{n\times n}$
- $\operatorname{tr}(\boldsymbol{I}_n) = n$
- $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A})$ for $\boldsymbol{A} \in \mathbb{R}^{n\times k}, \boldsymbol{B} \in \mathbb{R}^{k\times n}$

Trace

The properties of the trace of matrix products are more general. Specifically, the trace is invariant under cyclic permutations, i.e.,

$$\operatorname{tr}(\boldsymbol{AKL}) = \operatorname{tr}(\boldsymbol{KLA})$$

for matrices $\boldsymbol{A} \in \mathbb{R}^{a\times k}, \boldsymbol{K} \in \mathbb{R}^{k\times l}, \boldsymbol{L} \in \mathbb{R}^{l\times a}$.
This property generalizes to products of an arbitrary number of matrices. As a special case, it follows that for two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

$$\operatorname{tr}\left(\boldsymbol{x}\boldsymbol{y}^\top\right) = \operatorname{tr}\left(\boldsymbol{y}^\top\boldsymbol{x}\right) = \boldsymbol{y}^\top\boldsymbol{x} \in \mathbb{R}$$

Trace

The properties of the trace of matrix products are more general. Specifically, the trace is invariant under cyclic permutations, i.e.,

$$\operatorname{tr}(\boldsymbol{AKL}) = \operatorname{tr}(\boldsymbol{KLA})$$

for matrices $\boldsymbol{A} \in \mathbb{R}^{a\times k}, \boldsymbol{K} \in \mathbb{R}^{k\times l}, \boldsymbol{L} \in \mathbb{R}^{l\times a}$.

This property generalizes to products of an arbitrary number of matrices. As a special case, it follows that for two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

$$\operatorname{tr}\left(\boldsymbol{x}\boldsymbol{y}^\top\right) = \operatorname{tr}\left(\boldsymbol{y}^\top\boldsymbol{x}\right) = \boldsymbol{y}^\top\boldsymbol{x} \in \mathbb{R}$$

Trace

The properties of the trace of matrix products are more general. Specifically, the trace is invariant under cyclic permutations, i.e.,

$$\operatorname{tr}(\boldsymbol{AKL}) = \operatorname{tr}(\boldsymbol{KLA})$$

for matrices $\boldsymbol{A} \in \mathbb{R}^{a \times k}, \boldsymbol{K} \in \mathbb{R}^{k \times l}, \boldsymbol{L} \in \mathbb{R}^{l \times a}$.
This property generalizes to products of an arbitrary number of matrices. As a special case, it follows that for two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

$$\operatorname{tr}\left(\boldsymbol{x}\boldsymbol{y}^{\top}\right) = \operatorname{tr}\left(\boldsymbol{y}^{\top}\boldsymbol{x}\right) = \boldsymbol{y}^{\top}\boldsymbol{x} \in \mathbb{R}$$

Trace

Given a linear mapping $\Phi : V \to V$, where V is a vector space, we define the trace of this map by using the trace of matrix representation of Φ. For a given basis of V, we can describe Φ by means of the transformation matrix $\boldsymbol{A}$. Then the trace of Φ is the trace of $\boldsymbol{A}$. For a different basis of V, it holds that the corresponding transformation matrix $\boldsymbol{B}$ of Φ can be obtained by a basis change of the form $S^{-1}\boldsymbol{A}S$ for suitable S (see Section 2.7.2). For the corresponding trace of Φ, this means

$$\operatorname{tr}(\boldsymbol{B}) = \operatorname{tr}\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}\right) \overset{(4.19)}{=} \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^{-1}\right) = \operatorname{tr}(\boldsymbol{A})$$

Hence, while matrix representations of linear mappings are basis dependent the trace of a linear mapping Φ is independent of the basis.

Trace

Given a linear mapping $\Phi : V \to V$, where V is a vector space, we define the trace of this map by using the trace of matrix representation of Φ. For a given basis of V, we can describe Φ by means of the transformation matrix $\boldsymbol{A}$. Then the trace of Φ is the trace of $\boldsymbol{A}$. For a different basis of V, it holds that the corresponding transformation matrix $\boldsymbol{B}$ of Φ can be obtained by a basis change of the form $S^{-1}\boldsymbol{A}S$ for suitable S (see Section 2.7.2). For the corresponding trace of Φ, this means

$$\operatorname{tr}(\boldsymbol{B}) = \operatorname{tr}\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}\right) \overset{(4.19)}{=} \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^{-1}\right) = \operatorname{tr}(\boldsymbol{A})$$

Hence, while matrix representations of linear mappings are basis dependent the trace of a linear mapping Φ is independent of the basis.

Trace

Given a linear mapping $\Phi : V \to V$, where V is a vector space, we define the trace of this map by using the trace of matrix representation of Φ. For a given basis of V, we can describe Φ by means of the transformation matrix $\boldsymbol{A}$. Then the trace of Φ is the trace of $\boldsymbol{A}$. For a different basis of V, it holds that the corresponding transformation matrix $\boldsymbol{B}$ of Φ can be obtained by a basis change of the form $S^{-1}\boldsymbol{A}S$ for suitable S (see Section 2.7.2). For the corresponding trace of Φ, this means

$$\operatorname{tr}(\boldsymbol{B}) = \operatorname{tr}\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}\right) \overset{(4.19)}{=} \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^{-1}\right) = \operatorname{tr}(\boldsymbol{A})$$

Hence, while matrix representations of linear mappings are basis dependent the trace of a linear mapping Φ is independent of the basis.

Trace

Given a linear mapping $\Phi : V \to V$, where V is a vector space, we define the trace of this map by using the trace of matrix representation of Φ. For a given basis of V, we can describe Φ by means of the transformation matrix $\boldsymbol{A}$. Then the trace of Φ is the trace of $\boldsymbol{A}$. For a different basis of V, it holds that the corresponding transformation matrix $\boldsymbol{B}$ of Φ can be obtained by a basis change of the form $S^{-1}\boldsymbol{A}S$ for suitable S (see Section 2.7.2). For the corresponding trace of Φ, this means

$$\operatorname{tr}(\boldsymbol{B}) = \operatorname{tr}\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}\right) \overset{(4.19)}{=} \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^{-1}\right) = \operatorname{tr}(\boldsymbol{A})$$

Hence, while matrix representations of linear mappings are basis dependent the trace of a linear mapping Φ is independent of the basis.

Trace

Given a linear mapping $\Phi : V \to V$, where V is a vector space, we define the trace of this map by using the trace of matrix representation of Φ. For a given basis of V, we can describe Φ by means of the transformation matrix $\boldsymbol{A}$. Then the trace of Φ is the trace of $\boldsymbol{A}$. For a different basis of V, it holds that the corresponding transformation matrix $\boldsymbol{B}$ of Φ can be obtained by a basis change of the form $S^{-1}\boldsymbol{A}S$ for suitable S (see Section 2.7.2). For the corresponding trace of Φ, this means

$$\operatorname{tr}(\boldsymbol{B}) = \operatorname{tr}\left(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}\right) \overset{(4.19)}{=} \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{S}\boldsymbol{S}^{-1}\right) = \operatorname{tr}(\boldsymbol{A})$$

Hence, while matrix representations of linear mappings are basis dependent the trace of a linear mapping Φ is independent of the basis.

Trace

In this section, we covered determinants and traces as functions characterizing a square matrix. Taking together our understanding of determinants and traces we can now define an important equation describing a matrix $\boldsymbol{A}$ in terms of a polynomial, which we will use extensively in the following sections.

Definition (Definition 4.5 Characteristic Polynomial)

For $\lambda \in \mathbb{R}$ and a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

$$\begin{aligned} p_{\boldsymbol{A}}(\lambda) &:= \det(\boldsymbol{A} - \lambda \boldsymbol{I}) \\ &= c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_{n-1}\lambda^{n-1} + (-1)^n\lambda^n \end{aligned}$$

$c_0, \ldots, c_{n-1} \in \mathbb{R}$, is the characteristic polynomial of $\boldsymbol{A}$. In particular,

$$\begin{aligned} c_0 &= \det(\boldsymbol{A}) \\ c_{n-1} &= (-1)^{n-1}\operatorname{tr}(\boldsymbol{A}) \end{aligned}$$

Trace

In this section, we covered determinants and traces as functions characterizing a square matrix. Taking together our understanding of determinants and traces we can now define an important equation describing a matrix $\boldsymbol{A}$ in terms of a polynomial, which we will use extensively in the following sections.

Definition (Definition 4.5 Characteristic Polynomial)

For $\lambda \in \mathbb{R}$ and a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

$$\begin{aligned} p_{\boldsymbol{A}}(\lambda) &:= \det(\boldsymbol{A} - \lambda \boldsymbol{I}) \\ &= c_0 + c_1 \lambda + c_2 \lambda^2 + \cdots + c_{n-1} \lambda^{n-1} + (-1)^n \lambda^n \end{aligned}$$

$c_0, \ldots, c_{n-1} \in \mathbb{R}$, is the characteristic polynomial of $\boldsymbol{A}$. In particular,

$$\begin{aligned} c_0 &= \det(\boldsymbol{A}) \\ c_{n-1} &= (-1)^{n-1} \operatorname{tr}(\boldsymbol{A}) \end{aligned}$$

Trace

In this section, we covered determinants and traces as functions characterizing a square matrix. Taking together our understanding of determinants and traces we can now define an important equation describing a matrix $\boldsymbol{A}$ in terms of a polynomial, which we will use extensively in the following sections.

Definition (Definition 4.5 Characteristic Polynomial)

For $\lambda \in \mathbb{R}$ and a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

$$\begin{aligned} p_{\boldsymbol{A}}(\lambda) &:= \det(\boldsymbol{A} - \lambda \boldsymbol{I}) \\ &= c_0 + c_1 \lambda + c_2 \lambda^2 + \cdots + c_{n-1} \lambda^{n-1} + (-1)^n \lambda^n \end{aligned}$$

$c_0, \ldots, c_{n-1} \in \mathbb{R}$, is the characteristic polynomial of $\boldsymbol{A}$. In particular,

$$\begin{aligned} c_0 &= \det(\boldsymbol{A}) \\ c_{n-1} &= (-1)^{n-1} \operatorname{tr}(\boldsymbol{A}) \end{aligned}$$

Outline

Eigenvalues and Eigenvectors

Definition (4.6)

Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ be a square matrix, then, $\lambda \in \mathbb{R}$ is an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{x} \in \mathbb{R}^n \backslash \{0\}$ is the corresponding eigenvector of $\boldsymbol{A}$ if

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}.$$

Remark

In the linear algebra literature and software, it is often a convention that eigenvalues are sorted in descending order, so that the largest eigenvalue and associated eigenvector are called the first eigenvalue and its associated eigenvector, and the second largest called the second eigenvalue and its associated eigenvector, and so on.

Eigenvalues and Eigenvectors

Definition (4.6)

Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be a square matrix, then, $\lambda \in \mathbb{R}$ is an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{x} \in \mathbb{R}^n \backslash \{0\}$ is the corresponding eigenvector of $\boldsymbol{A}$ if

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}.$$

Remark

In the linear algebra literature and software, it is often a convention that eigenvalues are sorted in descending order, so that the largest eigenvalue and associated eigenvector are called the first eigenvalue and its associated eigenvector, and the second largest called the second eigenvalue and its associated eigenvector, and so on.

Eigenvalues and Eigenvectors

Definition (4.7 Collinearity and Codirection)

Two vectors that point in the same direction are called codirected. Two vectors are collinear if they point in the same or the opposite direction.

Remark (Non-uniqueness of eigenvectors)

If $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ associated with eigenvalue λ, then for any $c \in \mathbb{R}\backslash\{0\}$ it holds that $c\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ with the same eigenvalue since

$$\boldsymbol{A}(c\boldsymbol{x}) = c\boldsymbol{A}\boldsymbol{x} = c\lambda\boldsymbol{x} = \lambda(c\boldsymbol{x})$$

Thus, all vectors that are collinear to $\boldsymbol{x}$ are also eigenvectors of $\boldsymbol{A}$.

Eigenvalues and Eigenvectors

Definition (4.7 Collinearity and Codirection)

Two vectors that point in the same direction are called codirected. Two vectors are collinear if they point in the same or the opposite direction.

Remark (Non-uniqueness of eigenvectors)

If $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ associated with eigenvalue λ, then for any $c \in \mathbb{R}\backslash\{0\}$ it holds that $c\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ with the same eigenvalue since

$$\boldsymbol{A}(c\boldsymbol{x}) = c\boldsymbol{A}\boldsymbol{x} = c\lambda\boldsymbol{x} = \lambda(c\boldsymbol{x})$$

Thus, all vectors that are collinear to $\boldsymbol{x}$ are also eigenvectors of $\boldsymbol{A}$.

Eigenvalues and Eigenvectors

Definition (4.7 Collinearity and Codirection)

Two vectors that point in the same direction are called codirected. Two vectors are collinear if they point in the same or the opposite direction.

Remark (Non-uniqueness of eigenvectors)

If $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ associated with eigenvalue λ, then for any $c \in \mathbb{R}\backslash\{0\}$ it holds that $c\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ with the same eigenvalue since

$$\boldsymbol{A}(c\boldsymbol{x}) = c\boldsymbol{A}\boldsymbol{x} = c\lambda\boldsymbol{x} = \lambda(c\boldsymbol{x})$$

Thus, all vectors that are collinear to $\boldsymbol{x}$ are also eigenvectors of $\boldsymbol{A}$.

Eigenvalues and Eigenvectors

Theorem (4.8)

$\lambda \in \mathbb{R}$ is eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ if and only if λ is a root of the characteristic polynomial $p_A(\lambda)$ of $\boldsymbol{A}$.

Definition (4.9)

Let a square matrix $\boldsymbol{A}$ have an eigenvalue λ_i. The algebraic multiplicity of λ_i is the number of times the root appears in the characteristic polynomial.

Definition (4.10 Eigenspace and Eigenspectrum)

For $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, the set of all eigenvectors of $\boldsymbol{A}$ associated with an eigenvalue λ spans a subspace of $\mathbb{R}^n$, which is called the eigenspace of $\boldsymbol{A}$ with respect to λ and is denoted eigenspectrum by E_λ. The set of all eigenvalues of $\boldsymbol{A}$ is called the eigenspectrum, or just spectrum of $\boldsymbol{A}$.

Eigenvalues and Eigenvectors

Theorem (4.8)

$\lambda \in \mathbb{R}$ is eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ if and only if λ is a root of the characteristic polynomial $p_A(\lambda)$ of $\boldsymbol{A}$.

Definition (4.9)

Let a square matrix $\boldsymbol{A}$ have an eigenvalue λ_i. The algebraic multiplicity of λ_i is the number of times the root appears in the characteristic polynomial.

Definition (4.10 Eigenspace and Eigenspectrum)

For $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, the set of all eigenvectors of $\boldsymbol{A}$ associated with an eigenvalue λ spans a subspace of $\mathbb{R}^n$, which is called the eigenspace of $\boldsymbol{A}$ with respect to λ and is denoted eigenspectrum by E_λ. The set of all eigenvalues of $\boldsymbol{A}$ is called the eigenspectrum, or just spectrum of $\boldsymbol{A}$.

Eigenvalues and Eigenvectors

Theorem (4.8)

$\lambda \in \mathbb{R}$ is eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ if and only if λ is a root of the characteristic polynomial $p_A(\lambda)$ of $\boldsymbol{A}$.

Definition (4.9)

Let a square matrix $\boldsymbol{A}$ have an eigenvalue λ_i. The algebraic multiplicity of λ_i is the number of times the root appears in the characteristic polynomial.

Definition (4.10 Eigenspace and Eigenspectrum)

For $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, the set of all eigenvectors of $\boldsymbol{A}$ associated with an eigenvalue λ spans a subspace of $\mathbb{R}^n$, which is called the eigenspace of $\boldsymbol{A}$ with respect to λ and is denoted eigenspectrum by E_λ. The set of all eigenvalues of $\boldsymbol{A}$ is called the eigenspectrum, or just spectrum of $\boldsymbol{A}$.

Properties of Eigenvalues and Eigenvectors

- A matrix A and its transpose $A^\top$ possess the same eigenvalues, but not necessarily the same eigenvectors.
- The eigenspace E_λ is the null space of $A - \lambda I$ since

$$
\begin{aligned}
\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} &\iff \boldsymbol{A}\boldsymbol{x} - \lambda\boldsymbol{x} = \boldsymbol{0} \\
&\iff (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \iff \boldsymbol{x} \in \ker(\boldsymbol{A} - \lambda\boldsymbol{I})
\end{aligned}
$$

- Similar matrices (see Definition 2.22) possess the same eigenvalues. Therefore, a linear mapping Φ has eigenvalues that are independent of the choice of basis of its transformation matrix. This makes eigenvalues, together with the determinant and the trace, key characteristic parameters of a linear mapping as they are all invariant under basis change.
- Symmetric, positive definite matrices always have positive, real eigenvalues.

Properties of Eigenvalues and Eigenvectors

- A matrix A and its transpose $A^\top$ possess the same eigenvalues, but not necessarily the same eigenvectors.
- The eigenspace E_λ is the null space of $A - \lambda \boldsymbol{I}$ since

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} &\iff \boldsymbol{A}\boldsymbol{x} - \lambda \boldsymbol{x} = \boldsymbol{0} \\ &\iff (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \iff \boldsymbol{x} \in \ker(\boldsymbol{A} - \lambda \boldsymbol{I}) \end{aligned}$$

- Similar matrices (see Definition 2.22) possess the same eigenvalues. Therefore, a linear mapping Φ has eigenvalues that are independent of the choice of basis of its transformation matrix. This makes eigenvalues, together with the determinant and the trace, key characteristic parameters of a linear mapping as they are all invariant under basis change.
- Symmetric, positive definite matrices always have positive, real eigenvalues.

Properties of Eigenvalues and Eigenvectors

- A matrix A and its transpose $A^\top$ possess the same eigenvalues, but not necessarily the same eigenvectors.
- The eigenspace E_λ is the null space of $A - \lambda \boldsymbol{I}$ since

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} &\iff \boldsymbol{A}\boldsymbol{x} - \lambda\boldsymbol{x} = \mathbf{0} \\ &\iff (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \mathbf{0} \iff \boldsymbol{x} \in \ker(\boldsymbol{A} - \lambda\boldsymbol{I}) \end{aligned}$$

- Similar matrices (see Definition 2.22) possess the same eigenvalues. Therefore, a linear mapping Φ has eigenvalues that are independent of the choice of basis of its transformation matrix. This makes eigenvalues, together with the determinant and the trace, key characteristic parameters of a linear mapping as they are all invariant under basis change.
- Symmetric, positive definite matrices always have positive, real eigenvalues.

Properties of Eigenvalues and Eigenvectors

- A matrix A and its transpose $A^\top$ possess the same eigenvalues, but not necessarily the same eigenvectors.
- The eigenspace E_λ is the null space of $A - \lambda \boldsymbol{I}$ since

$$\begin{aligned}\boldsymbol{Ax} = \lambda\boldsymbol{x} &\iff \boldsymbol{Ax} - \lambda\boldsymbol{x} = \boldsymbol{0} \\ &\iff (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \iff \boldsymbol{x} \in \ker(\boldsymbol{A} - \lambda\boldsymbol{I})\end{aligned}$$

- Similar matrices (see Definition 2.22) possess the same eigenvalues. Therefore, a linear mapping Φ has eigenvalues that are independent of the choice of basis of its transformation matrix. This makes eigenvalues, together with the determinant and the trace, key characteristic parameters of a linear mapping as they are all invariant under basis change.
- Symmetric, positive definite matrices always have positive, real eigenvalues.

Properties of Eigenvalues and Eigenvectors

- A matrix A and its transpose $A^\top$ possess the same eigenvalues, but not necessarily the same eigenvectors.
- The eigenspace E_λ is the null space of $A - \lambda \boldsymbol{I}$ since

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} &\iff \boldsymbol{A}\boldsymbol{x} - \lambda\boldsymbol{x} = \boldsymbol{0} \\ &\iff (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \iff \boldsymbol{x} \in \ker(\boldsymbol{A} - \lambda\boldsymbol{I}) \end{aligned}$$

- Similar matrices (see Definition 2.22) possess the same eigenvalues. Therefore, a linear mapping Φ has eigenvalues that are independent of the choice of basis of its transformation matrix. This makes eigenvalues, together with the determinant and the trace, key characteristic parameters of a linear mapping as they are all invariant under basis change.
- Symmetric, positive definite matrices always have positive, real eigenvalues.

Example (Example 4.5 Computing Eigenvalues, Eigenvectors, and Eigenspaces)

Let us find the eigenvalues and eigenvectors of the 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$$

Step 1: Characteristic Polynomial.

Step 2: Eigenvalues.

Step 3: Eigenvectors and Eigenspaces

Example (Example 4.5 Computing Eigenvalues, Eigenvectors, and Eigenspaces)

Let us find the eigenvalues and eigenvectors of the 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$$

Step 1: Characteristic Polynomial.

Step 2: Eigenvalues.

Step 3: Eigenvectors and Eigenspaces

Example (Example 4.5 Computing Eigenvalues, Eigenvectors, and Eigenspaces)

Let us find the eigenvalues and eigenvectors of the 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$$

Step 1: Characteristic Polynomial.

Step 2: Eigenvalues.

Step 3: Eigenvectors and Eigenspaces

Example (Example 4.5 Computing Eigenvalues, Eigenvectors, and Eigenspaces)

Let us find the eigenvalues and eigenvectors of the 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$$

Step 1: Characteristic Polynomial.

Step 2: Eigenvalues.

Step 3: Eigenvectors and Eigenspaces

Eigenvalues and Eigenvectors

Definition (Definition 4.11)

Let λ_i be an eigenvalue of a square matrix $\boldsymbol{A}$, then the geometric multiplicity of λ_i is the number of linearly independent eigenvectors associated with λ_i. In other words, it is the dimensionality of the eigenspace spanned by the eigenvectors associated with λ_i.

Remark

A specific eigenvalues geometric multiplicity must be at least one because every eigenvalue has at least one associated eigenvector. An eigenvalues geometric multiplicity cannot exceed its algebraic multiplicity, but it may be lower.

Eigenvalues and Eigenvectors

Definition (Definition 4.11)

Let λ_i be an eigenvalue of a square matrix $\boldsymbol{A}$, then the geometric multiplicity of λ_i is the number of linearly independent eigenvectors associated with λ_i. In other words, it is the dimensionality of the eigenspace spanned by the eigenvectors associated with λ_i.

Remark

A specific eigenvalues geometric multiplicity must be at least one because every eigenvalue has at least one associated eigenvector. An eigenvalues geometric multiplicity cannot exceed its algebraic multiplicity, but it may be lower.

Graphical Intuition in Two Dimensions

Let us gain some intuition for determinants, eigenvectors, and eigenvalues using different linear mappings. Figure 4.4 depicts five transformation matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_5$ and their impact on a square grid of points, centered at the origin.

Eigenvalues and Eigenvectors

Theorem (Theorem 4.12)

The eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ with n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ are linearly independent.

This theorem states that eigenvectors of a matrix with n distinct eigenvalues form a basis of $\mathbb{R}^n$.

Eigenvalues and Eigenvectors

Theorem (Theorem 4.12)

The eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ *of a matrix* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *with* n *distinct eigenvalues* $\lambda_1, \ldots, \lambda_n$ *are linearly independent.*

This theorem states that eigenvectors of a matrix with n distinct eigenvalues form a basis of $\mathbb{R}^n$.

Eigenvalues and Eigenvectors

Definition (Definition 4.13)

A square matrix $A \in \mathbb{R}^{n \times n}$ is defective if it possesses fewer than n linearly independent eigenvectors.

A non-defective matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ does not necessarily require n distinct eigenvalues, but it does require that the eigenvectors form a basis of $\mathbb{R}^n$. Looking at the eigenspaces of a defective matrix, it follows that the sum of the dimensions of the eigenspaces is less than n. Specifically, a defective matrix has at least one eigenvalue λ_i with an algebraic multiplicity $m > 1$ and a geometric multiplicity of less than m.

Remark

A defective matrix cannot have n distinct eigenvalues, as distinct eigenvalues have linearly independent eigenvectors (Theorem 4.12).

Eigenvalues and Eigenvectors

Definition (Definition 4.13)

A square matrix $A \in \mathbb{R}^{n \times n}$ is defective if it possesses fewer than n linearly independent eigenvectors.

A non-defective matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ does not necessarily require n distinct eigenvalues, but it does require that the eigenvectors form a basis of $\mathbb{R}^n$. Looking at the eigenspaces of a defective matrix, it follows that the sum of the dimensions of the eigenspaces is less than n. Specifically, a defective matrix has at least one eigenvalue λ_i with an algebraic multiplicity $m > 1$ and a geometric multiplicity of less than m.

Remark

A defective matrix cannot have n distinct eigenvalues, as distinct eigenvalues have linearly independent eigenvectors (Theorem 4.12).

Eigenvalues and Eigenvectors

Definition (Definition 4.13)

A square matrix $A \in \mathbb{R}^{n\times n}$ is defective if it possesses fewer than n linearly independent eigenvectors.

A non-defective matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ does not necessarily require n distinct eigenvalues, but it does require that the eigenvectors form a basis of $\mathbb{R}^n$. Looking at the eigenspaces of a defective matrix, it follows that the sum of the dimensions of the eigenspaces is less than n. Specifically, a defective matrix has at least one eigenvalue λ_i with an algebraic multiplicity $m > 1$ and a geometric multiplicity of less than m.

Remark

A defective matrix cannot have n distinct eigenvalues, as distinct eigenvalues have linearly independent eigenvectors (Theorem 4.12).

Eigenvalues and Eigenvectors

Theorem (Theorem 4.14)

Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, we can always obtain a symmetric, positive semi-definite matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ by defining

$$\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}.$$

Remark

If $\mathrm{rk}(\boldsymbol{A}) = n$, then $\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}$ is symmetric, positive definite.

Eigenvalues and Eigenvectors

Theorem (Theorem 4.14)

Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, we can always obtain a symmetric, positive semi-definite matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ by defining

$$\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}.$$

Remark

If $\operatorname{rk}(\boldsymbol{A}) = n$, then $\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}$ is symmetric, positive definite.

Eigenvalues and Eigenvectors

Understanding why Theorem 4.14 holds is insightful for how we can use symmetrized matrices: Symmetry requires $S = S^\top$, and by inserting (4.36), we obtain $\boldsymbol{S} = \boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}^\top (\boldsymbol{A}^\top)^\top = (\boldsymbol{A}^\top \boldsymbol{A})^\top = \boldsymbol{S}^\top$.
Moreover, positive semi-definiteness (Section 3.2.3) requires that $\boldsymbol{x}^\top \boldsymbol{S} \boldsymbol{x} \geqslant 0$ and inserting (4.36) we obtain $\boldsymbol{x}^\top \boldsymbol{S} \boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{x} = (\boldsymbol{x}^\top \boldsymbol{A}^\top)(\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x})^\top (\boldsymbol{A}\boldsymbol{x}) \geqslant 0$, because the dot product computes a sum of squares (which are themselves non-negative).

Eigenvalues and Eigenvectors

Understanding why Theorem 4.14 holds is insightful for how we can use symmetrized matrices: Symmetry requires $S = S^\top$, and by inserting (4.36), we obtain $\boldsymbol{S} = \boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}^\top (\boldsymbol{A}^\top)^\top = (\boldsymbol{A}^\top \boldsymbol{A})^\top = \boldsymbol{S}^\top$.
Moreover, positive semi-definiteness (Section 3.2.3) requires that $\boldsymbol{x}^\top \boldsymbol{S}\boldsymbol{x} \geqslant 0$ and inserting (4.36) we obtain $\boldsymbol{x}^\top \boldsymbol{S}\boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = (\boldsymbol{x}^\top \boldsymbol{A}^\top)(\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x})^\top(\boldsymbol{A}\boldsymbol{x}) \geqslant 0$, because the dot product computes a sum of squares (which are themselves non-negative).

Eigenvalues and Eigenvectors

Theorem (Theorem 4.15 Spectral Theorem)

If $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is symmetric, there exists an orthonormal basis of the corresponding vector space $\boldsymbol{V}$ consisting of eigenvectors of $\boldsymbol{A}$, and each eigenvalue is real.

A direct implication of the spectral theorem is that the eigendecomposition of a symmetric matrix $\boldsymbol{A}$ exists (with real eigenvalues), and that we can find an ONB of eigenvectors so that $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top$, where $\boldsymbol{D}$ is diagonal and the columns of $\boldsymbol{P}$ contain the eigenvectors.

Eigenvalues and Eigenvectors

Theorem (Theorem 4.15 Spectral Theorem)

If $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is symmetric, there exists an orthonormal basis of the corresponding vector space $\boldsymbol{V}$ consisting of eigenvectors of $\boldsymbol{A}$, and each eigenvalue is real.

A direct implication of the spectral theorem is that the eigendecomposition of a symmetric matrix $\boldsymbol{A}$ exists (with real eigenvalues), and that we can find an ONB of eigenvectors so that $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\top}$, where $\boldsymbol{D}$ is diagonal and the columns of $\boldsymbol{P}$ contain the eigenvectors.

Example (Example 4.8)

Consider the matrix

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$$

Eigenvalues and Eigenvectors

Theorem (Theorem 4.16)

The determinant of a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is the product of its eigenvalues, i.e.,

$$\det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i$$

where λ_i are (possibly repeated) eigenvalues of $\boldsymbol{A}$.

Theorem (Theorem 4.17)

The trace of a matrix $A \in \mathbb{R}^{n\times n}$ is the sum of its eigenvalues, i.e.,

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i,$$

where λ_i are (possibly repeated) eigenvalues of $\boldsymbol{A}$.

Eigenvalues and Eigenvectors

Theorem (Theorem 4.16)

The determinant of a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is the product of its eigenvalues, i.e.,

$$\det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i$$

where λ_i are (possibly repeated) eigenvalues of $\boldsymbol{A}$.

Theorem (Theorem 4.17)

The trace of a matrix $A \in \mathbb{R}^{n-n}$ is the sum of its eigenvalues, i.e.,

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i,$$

where λ_i are (possibly repeated) eigenvalues of $\boldsymbol{A}$.

Geometric intuition of Theorem 4.16 and 4.17

Let us provide a geometric intuition of these two theorems. Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ that possesses two linearly independent eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2$. For this example, we assume $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ are an ONB of $\mathbb{R}^2$ so that they are orthogonal and the area of the square they span is 1 ; see Figure 4.6. From Section 4.1, we know that the determinant computes the change of area of unit square under the transformation $\boldsymbol{A}$. In this example, we can compute the change of area explicitly: Mapping the eigenvectors using $\boldsymbol{A}$ gives us vectors $\boldsymbol{v}_1 = \boldsymbol{A}\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1$ and $\boldsymbol{v}_2 = \boldsymbol{A}\boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_2$, i.e., the new vectors $\boldsymbol{v}_i$ are scaled versions of the eigenvectors $\boldsymbol{x}_i$, and the scaling factors are the corresponding eigenvalues λ_i.$\boldsymbol{v}_1, \boldsymbol{v}_2$ are still orthogonal, and the area of the rectangle they span is $|\lambda_1 \lambda_2|$.

Geometric intuition of Theorem 4.16 and 4.17

Let us provide a geometric intuition of these two theorems. Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ that possesses two linearly independent eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2$. For this example, we assume $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ are an ONB of $\mathbb{R}^2$ so that they are orthogonal and the area of the square they span is 1 ; see Figure 4.6. From Section 4.1, we know that the determinant computes the change of area of unit square under the transformation $\boldsymbol{A}$. In this example, we can compute the change of area explicitly: Mapping the eigenvectors using $\boldsymbol{A}$ gives us vectors $\boldsymbol{v}_1 = \boldsymbol{A}\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1$ and $\boldsymbol{v}_2 = \boldsymbol{A}\boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_2$, i.e., the new vectors $\boldsymbol{v}_i$ are scaled versions of the eigenvectors $\boldsymbol{x}_i$, and the scaling factors are the corresponding eigenvalues $\lambda_i . \boldsymbol{v}_1, \boldsymbol{v}_2$ are still orthogonal, and the area of the rectangle they span is $|\lambda_1 \lambda_2|$.

Geometric intuition of Theorem 4.16 and 4.17

Let us provide a geometric intuition of these two theorems. Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ that possesses two linearly independent eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2$. For this example, we assume $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ are an ONB of $\mathbb{R}^2$ so that they are orthogonal and the area of the square they span is 1 ; see Figure 4.6. From Section 4.1, we know that the determinant computes the change of area of unit square under the transformation $\boldsymbol{A}$. In this example, we can compute the change of area explicitly: Mapping the eigenvectors using $\boldsymbol{A}$ gives us vectors $\boldsymbol{v}_1 = \boldsymbol{A}\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1$ and $\boldsymbol{v}_2 = \boldsymbol{A}\boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_2$, i.e., the new vectors $\boldsymbol{v}_i$ are scaled versions of the eigenvectors $\boldsymbol{x}_i$, and the scaling factors are the corresponding eigenvalues λ_i.$\boldsymbol{v}_1, \boldsymbol{v}_2$ are still orthogonal, and the area of the rectangle they span is $|\lambda_1 \lambda_2|$.

Geometric intuition of Theorem 4.16 and 4.17

Let us provide a geometric intuition of these two theorems. Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ that possesses two linearly independent eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2$. For this example, we assume $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ are an ONB of $\mathbb{R}^2$ so that they are orthogonal and the area of the square they span is 1 ; see Figure 4.6. From Section 4.1, we know that the determinant computes the change of area of unit square under the transformation $\boldsymbol{A}$. In this example, we can compute the change of area explicitly: Mapping the eigenvectors using $\boldsymbol{A}$ gives us vectors $\boldsymbol{v}_1 = \boldsymbol{A}\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1$ and $\boldsymbol{v}_2 = \boldsymbol{A}\boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_2$, i.e., the new vectors $\boldsymbol{v}_i$ are scaled versions of the eigenvectors $\boldsymbol{x}_i$, and the scaling factors are the corresponding eigenvalues λ_i.$\boldsymbol{v}_1, \boldsymbol{v}_2$ are still orthogonal, and the area of the rectangle they span is $|\lambda_1 \lambda_2|$.

Geometric intuition of Theorem 4.16 and 4.17

Let us provide a geometric intuition of these two theorems. Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ that possesses two linearly independent eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2$. For this example, we assume $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ are an ONB of $\mathbb{R}^2$ so that they are orthogonal and the area of the square they span is 1 ; see Figure 4.6. From Section 4.1, we know that the determinant computes the change of area of unit square under the transformation $\boldsymbol{A}$. In this example, we can compute the change of area explicitly: Mapping the eigenvectors using $\boldsymbol{A}$ gives us vectors $\boldsymbol{v}_1 = \boldsymbol{A}\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1$ and $\boldsymbol{v}_2 = \boldsymbol{A}\boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_2$, i.e., the new vectors $\boldsymbol{v}_i$ are scaled versions of the eigenvectors $\boldsymbol{x}_i$, and the scaling factors are the corresponding eigenvalues $\lambda_i . \boldsymbol{v}_1, \boldsymbol{v}_2$ are still orthogonal, and the area of the rectangle they span is $|\lambda_1 \lambda_2|$.

Geometric intuition of Theorem 4.16 and 4.17

Given that $\boldsymbol{x}_1, \boldsymbol{x}_2$ (in our example) are orthonormal, we can directly compute the circumference of the unit square as $2(1+1)$. Mapping the eigenvectors using $\boldsymbol{A}$ creates a rectangle whose circumference is $2(|\lambda_1| + |\lambda_2|)$. Therefore, the sum of the absolute values of the eigenvalues tells us how the circumference of the unit square changes under the transformation matrix $\boldsymbol{A}$.

Geometric intuition of Theorem 4.16 and 4.17

Given that $\boldsymbol{x}_1, \boldsymbol{x}_2$ (in our example) are orthonormal, we can directly compute the circumference of the unit square as $2(1+1)$. Mapping the eigenvectors using $\boldsymbol{A}$ creates a rectangle whose circumference is $2(|\lambda_1| + |\lambda_2|)$. Therefore, the sum of the absolute values of the eigenvalues tells us how the circumference of the unit square changes under the transformation matrix $\boldsymbol{A}$.

Outline

Cholesky Decomposition

Theorem (Theorem 4.18 Cholesky Decomposition)

A symmetric, positive definite matrix $\boldsymbol{A}$ can be factorized into a product $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$, where $\boldsymbol{L}$ is a lower-triangular matrix with positive diagonal elements:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & \cdots & l_{n1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_{nn} \end{bmatrix}$$

Cholesky factor $\boldsymbol{L}$ is called the Cholesky factor of $\boldsymbol{A}$, and $\boldsymbol{L}$ is unique.

Cholesky Decomposition

Theorem (Theorem 4.18 Cholesky Decomposition)

A symmetric, positive definite matrix $\boldsymbol{A}$ can be factorized into a product $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$, where $\boldsymbol{L}$ is a lower-triangular matrix with positive diagonal elements:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & \cdots & l_{n1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_{nn} \end{bmatrix}$$

Cholesky factor $\boldsymbol{L}$ is called the Cholesky factor of $\boldsymbol{A}$, and $\boldsymbol{L}$ is unique.

Cholesky Decomposition

Theorem (Theorem 4.18 Cholesky Decomposition)

A symmetric, positive definite matrix $\boldsymbol{A}$ *can be factorized into a product* $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^{\top}$*, where* $\boldsymbol{L}$ *is a lower-triangular matrix with positive diagonal elements:*

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & \cdots & l_{n1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_{nn} \end{bmatrix}$$

Cholesky factor $\boldsymbol{L}$ *is called the Cholesky factor of* $\boldsymbol{A}$*, and* $\boldsymbol{L}$ *is unique.*

Cholesky Decomposition

Example (Example 4.10 Cholesky Factorization)

Consider a symmetric, positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{3\times3}$. We are interested in finding its Cholesky factorization $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$, i.e.,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{L}\boldsymbol{L}^\top = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

Multiplying out the right-hand side yields

$$\boldsymbol{A} = \begin{bmatrix} l_{11}^2 & l_{21}l_{11} & l_{31}l_{11} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{31}l_{21} + l_{32}l_{22} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Cholesky Decomposition

Example (Example 4.10 Cholesky Factorization)

Consider a symmetric, positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{3\times 3}$. We are interested in finding its Cholesky factorization $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$, i.e.,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{L}\boldsymbol{L}^\top = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

Multiplying out the right-hand side yields

$$\boldsymbol{A} = \begin{bmatrix} l_{11}^2 & l_{21}l_{11} & l_{31}l_{11} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{31}l_{21} + l_{32}l_{22} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Cholesky Decomposition

Example (Example 4.10 Cholesky Factorization)

Consider a symmetric, positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{3\times3}$. We are interested in finding its Cholesky factorization $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$, i.e.,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{L}\boldsymbol{L}^\top = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

Multiplying out the right-hand side yields

$$\boldsymbol{A} = \begin{bmatrix} l_{11}^2 & l_{21}l_{11} & l_{31}l_{11} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 & l_{31}l_{21} + l_{32}l_{22} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Cholesky Decomposition

Comparing the left-hand side of (4.45) and the right-hand side of (4.46) shows that there is a simple pattern in the diagonal elements l_{ii} :

$$l_{11} = \sqrt{a_{11}}, \quad l_{22} = \sqrt{a_{22} - l_{21}^2}, \quad l_{33} = \sqrt{a_{33} - \left(l_{31}^2 + l_{32}^2\right)}$$

Similarly for the elements below the diagonal (l_{ij}, where $i > j$), there is also a repeating pattern:

$$l_{21} = \frac{1}{l_{11}} a_{21}, \quad l_{31} = \frac{1}{l_{11}} a_{31}, \quad l_{32} = \frac{1}{l_{22}} \left(a_{32} - l_{31} l_{21}\right)$$

Thus, we constructed the Cholesky decomposition for any symmetric, positive definite 3×3 matrix. The key realization is that we can backward calculate what the components l_{ij} for the $\boldsymbol{L}$ should be, given the values a_{ij} for $\boldsymbol{A}$ and previously computed values of l_{ij}

Cholesky Decomposition

Comparing the left-hand side of (4.45) and the right-hand side of (4.46) shows that there is a simple pattern in the diagonal elements l_{ii} :

$$l_{11} = \sqrt{a_{11}}, \quad l_{22} = \sqrt{a_{22} - l_{21}^2}, \quad l_{33} = \sqrt{a_{33} - \left(l_{31}^2 + l_{32}^2\right)}$$

Similarly for the elements below the diagonal (l_{ij}, where $i > j$), there is also a repeating pattern:

$$l_{21} = \frac{1}{l_{11}} a_{21}, \quad l_{31} = \frac{1}{l_{11}} a_{31}, \quad l_{32} = \frac{1}{l_{22}} \left(a_{32} - l_{31} l_{21}\right)$$

Thus, we constructed the Cholesky decomposition for any symmetric, positive definite 3×3 matrix. The key realization is that we can backward calculate what the components l_{ij} for the $\boldsymbol{L}$ should be, given the values a_{ij} for $\boldsymbol{A}$ and previously computed values of l_{ij}

Cholesky Decomposition

Comparing the left-hand side of (4.45) and the right-hand side of (4.46) shows that there is a simple pattern in the diagonal elements l_{ii} :

$$l_{11} = \sqrt{a_{11}}, \quad l_{22} = \sqrt{a_{22} - l_{21}^2}, \quad l_{33} = \sqrt{a_{33} - \left(l_{31}^2 + l_{32}^2\right)}$$

Similarly for the elements below the diagonal (l_{ij}, where $i > j$), there is also a repeating pattern:

$$l_{21} = \frac{1}{l_{11}} a_{21}, \quad l_{31} = \frac{1}{l_{11}} a_{31}, \quad l_{32} = \frac{1}{l_{22}} \left(a_{32} - l_{31} l_{21}\right)$$

Thus, we constructed the Cholesky decomposition for any symmetric, positive definite 3×3 matrix. The key realization is that we can backward calculate what the components l_{ij} for the $\boldsymbol{L}$ should be, given the values a_{ij} for $\boldsymbol{A}$ and previously computed values of l_{ij}

Cholesky Decomposition

The Cholesky decomposition is an important tool for the numerical computations underlying machine learning. Here, symmetric positive definite matrices require frequent manipulation, e.g., the covariance matrix of a multivariate Gaussian variable (see Section 6.5) is symmetric, positive definite. The Cholesky factorization of this covariance matrix allows us to generate samples from a Gaussian distribution. It also allows us to perform a linear transformation of random variables, which is heavily exploited when computing gradients in deep stochastic models, such as the variational auto-encoder (Jimenez Rezende et al., 2014; Kingma and Welling, 2014).

Cholesky Decomposition

The Cholesky decomposition is an important tool for the numerical computations underlying machine learning. Here, symmetric positive definite matrices require frequent manipulation, e.g., the covariance matrix of a multivariate Gaussian variable (see Section 6.5) is symmetric, positive definite. The Cholesky factorization of this covariance matrix allows us to generate samples from a Gaussian distribution. It also allows us to perform a linear transformation of random variables, which is heavily exploited when computing gradients in deep stochastic models, such as the variational auto-encoder (Jimenez Rezende et al., 2014; Kingma and Welling, 2014).

Cholesky Decomposition

The Cholesky decomposition also allows us to compute determinants very efficiently. Given the Cholesky decomposition $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$, we know that $\det(\boldsymbol{A}) = \det(\boldsymbol{L})\det(\boldsymbol{L}^\top) = \det(\boldsymbol{L})^2$. Since $\boldsymbol{L}$ is a triangular matrix, the determinant is simply the product of its diagonal entries so that $\det(\boldsymbol{A}) = \prod_i L_{ii}^2$. Thus, many numerical software packages use the Cholesky decomposition to make computations more efficient.

Cholesky Decomposition

The Cholesky decomposition also allows us to compute determinants very efficiently. Given the Cholesky decomposition $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^\top$, we know that $\det(\boldsymbol{A}) = \det(\boldsymbol{L})\det(\boldsymbol{L}^\top){=}\det(\boldsymbol{L})^2$. Since $\boldsymbol{L}$ is a triangular matrix, the determinant is simply the product of its diagonal entries so that $\det(\boldsymbol{A}) = \prod_i \boldsymbol{L}_{ii}^2$. Thus, many numerical software packages use the Cholesky decomposition to make computations more efficient.

Outline

Eigendecomposition and Diagonalization

Definition (Definition 4.19 Diagonalizable)

A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is diagonalizable if it is similar to a diagonal matrix, i.e., if there exists an invertible matrix $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ such that $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$.

In the following, we will see that diagonalizing a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is a way of expressing the same linear mapping but in another basis (see Section 2.6.1), which will turn out to be a basis that consists of the eigenvectors of $\boldsymbol{A}$.

Eigendecomposition and Diagonalization

Definition (Definition 4.19 Diagonalizable)

A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is diagonalizable if it is similar to a diagonal matrix, i.e., if there exists an invertible matrix $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ such that $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$.

In the following, we will see that diagonalizing a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is a way of expressing the same linear mapping but in another basis (see Section 2.6.1), which will turn out to be a basis that consists of the eigenvectors of $\boldsymbol{A}$.

Eigendecomposition and Diagonalization

Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, let $\lambda_1, \dots, \lambda_n$ be a set of scalars, and let $\boldsymbol{p}_1, \dots, \boldsymbol{p}_n$ be a set of vectors in $\mathbb{R}^n$. We define $\boldsymbol{P} := [\boldsymbol{p}_1, \dots, \boldsymbol{p}_n]$ and let $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ be a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$. Then we can show that

$$\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{D},$$

if and only if $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{p}_1, \dots, \boldsymbol{p}_n$ are corresponding eigenvectors of $\boldsymbol{A}$.

Eigendecomposition and Diagonalization

Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, let $\lambda_1, \ldots, \lambda_n$ be a set of scalars, and let $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ be a set of vectors in $\mathbb{R}^n$. We define $\boldsymbol{P} := [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n]$ and let $\boldsymbol{D} \in \mathbb{R}^{n \times n}$ be a diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Then we can show that

$$\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{D},$$

if and only if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$ are corresponding eigenvectors of $\boldsymbol{A}$.

Eigendecomposition and Diagonalization

We can see that this statement holds because

$$\boldsymbol{A}\boldsymbol{P} = \boldsymbol{A}[\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n] = [\boldsymbol{A}\boldsymbol{p}_1, \ldots, \boldsymbol{A}\boldsymbol{p}_n]$$

$$\boldsymbol{P}\boldsymbol{D} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n]\begin{bmatrix} \lambda_1 & 0 \\ & \ddots \\ 0 & \lambda_n \end{bmatrix} = [\lambda_1\boldsymbol{p}_1, \ldots, \lambda_n\boldsymbol{p}_n]$$

Thus, $\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{D}$ implies that

$$\begin{gathered} \boldsymbol{A}\boldsymbol{p}_1 = \lambda_1\boldsymbol{p}_1 \\ \vdots \\ \boldsymbol{A}\boldsymbol{p}_n = \lambda_n\boldsymbol{p}_n \end{gathered}$$

Therefore, the columns of $\boldsymbol{P}$ must be eigenvectors of $\boldsymbol{A}$.

Eigendecomposition and Diagonalization

Our definition of diagonalization requires that $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ is invertible, i.e., $\boldsymbol{P}$ has full rank (Theorem 4.3). This requires us to have n linearly independent eigenvectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$, i.e., the $\boldsymbol{p}_i$ form a basis of $\mathbb{R}^n$.

Theorem (Theorem 4.20 Eigendecomposition)

A square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ can be factored into

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$$

where $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{D}$ is a diagonal matrix whose diagonal entries are the eigenvalues of $\boldsymbol{A}$, if and only if the eigenvectors of $\boldsymbol{A}$ form a basis of $\mathbb{R}^n$.

Eigendecomposition and Diagonalization

Our definition of diagonalization requires that $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ is invertible, i.e., $\boldsymbol{P}$ has full rank (Theorem 4.3). This requires us to have n linearly independent eigenvectors $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n$, i.e., the $\boldsymbol{p}_i$ form a basis of $\mathbb{R}^n$.

Theorem (Theorem 4.20 Eigendecomposition)

A square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ can be factored into

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$$

where $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{D}$ is a diagonal matrix whose diagonal entries are the eigenvalues of $\boldsymbol{A}$, if and only if the eigenvectors of $\boldsymbol{A}$ form a basis of $\mathbb{R}^n$.

Eigendecomposition and Diagonalization

Theorem (Theorem 4.21)

A symmetric matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ can always be diagonalized.

Theorem 4.21 follows directly from the spectral Theorem 4.15. Moreover, the spectral theorem states that we can find an ONB of eigenvectors of $\mathbb{R}^n$. This makes $\boldsymbol{P}$ an orthogonal matrix so that $\boldsymbol{D} = \boldsymbol{P}^\top \boldsymbol{A} \boldsymbol{P}$.

Eigendecomposition and Diagonalization

Theorem (Theorem 4.21)

A symmetric matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ can always be diagonalized.

Theorem 4.21 follows directly from the spectral Theorem 4.15. Moreover, the spectral theorem states that we can find an ONB of eigenvectors of $\mathbb{R}^n$. This makes $\boldsymbol{P}$ an orthogonal matrix so that $\boldsymbol{D} = \boldsymbol{P}^\top \boldsymbol{A}\boldsymbol{P}$.

Outline

Singular Value Decomposition

The singular value decomposition (SVD) of a matrix is a central matrix decomposition method in linear algebra. It has been referred to as the "fundamental theorem of linear algebra" (Strang, 1993) because it can be applied to all matrices, not only to square matrices, and it always exists. Moreover, as we will explore in the following, the SVD of a matrix $\boldsymbol{A}$, which represents a linear mapping $\Phi : V \to W$, quantifies the change between the underlying geometry of these two vector spaces.

Singular Value Decomposition

The singular value decomposition (SVD) of a matrix is a central matrix decomposition method in linear algebra. It has been referred to as the "fundamental theorem of linear algebra" (Strang, 1993) because it can be applied to all matrices, not only to square matrices, and it always exists. Moreover, as we will explore in the following, the SVD of a matrix $\boldsymbol{A}$, which represents a linear mapping $\Phi : V \to W$, quantifies the change between the underlying geometry of these two vector spaces.

Singular Value Decomposition

The singular value decomposition (SVD) of a matrix is a central matrix decomposition method in linear algebra. It has been referred to as the ”fundamental theorem of linear algebra” (Strang, 1993) because it can be applied to all matrices, not only to square matrices, and it always exists. Moreover, as we will explore in the following, the SVD of a matrix $\boldsymbol{A}$, which represents a linear mapping $\Phi : V \to W$, quantifies the change between the underlying geometry of these two vector spaces.

Singular Value Decomposition

Theorem (Theorem 4.22 SVD Theorem)

Let $\boldsymbol{A}^{m\times n}$ be a rectangular matrix of rank $r \in [0, \min(m,n)]$. The SVD of $\boldsymbol{A}$ is a decomposition of the form

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$

with an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ with column vectors $\boldsymbol{u}_i, i = 1, \ldots, m$, and an orthogonal matrix $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ with column vectors $\boldsymbol{v}_j, j = 1, \ldots, n$. Moreover, $\boldsymbol{\Sigma}$ is an $m \times n$ matrix with $\Sigma_{ii} = \sigma_i \geqslant 0$ and $\Sigma_{ij} = 0, i \neq j$.

Singular Value Decomposition

Theorem (Theorem 4.22 SVD Theorem)

Let $\boldsymbol{A}^{m\times n}$ *be a rectangular matrix of rank* $r \in [0, \min(m,n)]$. *The SVD of* $\boldsymbol{A}$ *is a decomposition of the form*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$

with an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ *with column vectors* $\boldsymbol{u}_i, i = 1, \ldots, m$, *and an orthogonal matrix* $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ *with column vectors* $\boldsymbol{v}_j, j = 1, \ldots, n$. *Moreover,* Σ *is an* $m \times n$ *matrix with* $\Sigma_{ii} = \sigma_i \geqslant 0$ *and* $\Sigma_{ij} = 0, i \neq j$.

Singular Value Decomposition

Theorem (Theorem 4.22 SVD Theorem)

Let $\boldsymbol{A}^{m\times n}$ *be a rectangular matrix of rank* $r \in [0, \min(m,n)]$. *The SVD of* $\boldsymbol{A}$ *is a decomposition of the form*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$

with an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ *with column vectors* $\boldsymbol{u}_i, i = 1, \ldots, m$, *and an orthogonal matrix* $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ *with column vectors* $\boldsymbol{v}_j, j = 1, \ldots, n$. *Moreover,* $\boldsymbol{\Sigma}$ *is an* $m \times n$ *matrix with* $\Sigma_{ii} = \sigma_i \geqslant 0$ *and* $\Sigma_{ij} = 0, i \neq j$.

Singular Value Decomposition

The diagonal entries $\sigma_i, i = 1, \ldots, r$, of $\boldsymbol{\Sigma}$ are called the singular values, $\boldsymbol{u}_i$ are called the left-singular vectors, and $\boldsymbol{v}_j$ are called the right-singular vectors. By convention, the singular values are ordered, i.e., $\sigma_1 \geqslant \sigma_2 \geqslant \sigma_r \geqslant 0$.

Singular Value Decomposition

The diagonal entries $\sigma_i, i = 1, \ldots, r$, of $\boldsymbol{\Sigma}$ are called the singular values, $\boldsymbol{u}_i$ are called the left-singular vectors, and $\boldsymbol{v}_j$ are called the right-singular vectors. By convention, the singular values are ordered, i.e., $\sigma_1 \geqslant \sigma_2 \geqslant \sigma_r \geqslant 0$.

Singular Value Decomposition

The diagonal entries $\sigma_i, i = 1, \ldots, r$, of $\boldsymbol{\Sigma}$ are called the singular values, $\boldsymbol{u}_i$ are called the left-singular vectors, and $\boldsymbol{v}_j$ are called the right-singular vectors. By convention, the singular values are ordered, i.e., $\sigma_1 \geqslant \sigma_2 \geqslant \sigma_r \geqslant 0$.

Singular Value Decomposition

The singular value matrix $\boldsymbol{\Sigma}$ is unique, but it requires some attention. Observe that the $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ is rectangular. In particular, $\boldsymbol{\Sigma}$ is of the same size as $\boldsymbol{A}$. This means that $\boldsymbol{\Sigma}$ has a diagonal submatrix that contains the singular values and needs additional zero padding. Specifically, if $m > n$, then the matrix $\boldsymbol{\Sigma}$ has diagonal structure up to row n and then consists of $\mathbf{0}^\top$ row vectors from $n+1$ to m below so that

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

Singular Value Decomposition

The singular value matrix $\boldsymbol{\Sigma}$ is unique, but it requires some attention. Observe that the $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is rectangular. In particular, $\boldsymbol{\Sigma}$ is of the same size as $\boldsymbol{A}$. This means that $\boldsymbol{\Sigma}$ has a diagonal submatrix that contains the singular values and needs additional zero padding. Specifically, if $m > n$, then the matrix $\boldsymbol{\Sigma}$ has diagonal structure up to row n and then consists of $\boldsymbol{0}^\top$ row vectors from $n+1$ to m below so that

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

Singular Value Decomposition

The singular value matrix $\boldsymbol{\Sigma}$ is unique, but it requires some attention. Observe that the $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ is rectangular. In particular, $\boldsymbol{\Sigma}$ is of the same size as $\boldsymbol{A}$. This means that $\boldsymbol{\Sigma}$ has a diagonal submatrix that contains the singular values and needs additional zero padding. Specifically, if $m > n$, then the matrix $\boldsymbol{\Sigma}$ has diagonal structure up to row n and then consists of $\mathbf{0}^\top$ row vectors from $n+1$ to m below so that

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

Singular Value Decomposition

The singular value matrix $\mathbf{\Sigma}$ is unique, but it requires some attention. Observe that the $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ is rectangular. In particular, $\mathbf{\Sigma}$ is of the same size as $\boldsymbol{A}$. This means that $\mathbf{\Sigma}$ has a diagonal submatrix that contains the singular values and needs additional zero padding. Specifically, if $m > n$, then the matrix $\mathbf{\Sigma}$ has diagonal structure up to row n and then consists of $\mathbf{0}^\top$ row vectors from $n+1$ to m below so that

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{bmatrix}$$

Singular Value Decomposition

If $m < n$, the matrix $\boldsymbol{\Sigma}$ has a diagonal structure up to column m and columns that consist of $\mathbf{0}$ from $m+1$ to n :

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 & & 0 \\ 0 & 0 & \sigma_m & 0 & \dots & 0 \end{bmatrix}$$

Remark. The SVD exists for any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.

Singular Value Decomposition

If $m < n$, the matrix $\boldsymbol{\Sigma}$ has a diagonal structure up to column m and columns that consist of $\mathbf{0}$ from $m+1$ to n :

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 & & 0 \\ 0 & 0 & \sigma_m & 0 & \dots & 0 \end{bmatrix}$$

Remark. The SVD exists for any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.

Singular Value Decomposition

If $m < n$, the matrix $\mathbf{\Sigma}$ has a diagonal structure up to column m and columns that consist of $\mathbf{0}$ from $m+1$ to n :

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & 0 & & 0 \\ 0 & 0 & \sigma_m & 0 & \dots & 0 \end{bmatrix}$$

Remark. The SVD exists for any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.

Geometric Intuitions for the SVD

The SVD offers geometric intuitions to describe a transformation matrix A. In the following, we will discuss the SVD as sequential linear transformations performed on the bases. In Example 4.12, we will then apply transformation matrices of the SVD to a set of vectors in $\mathbb{R}^2$, which allows us to visualize the effect of each transformation more clearly.

Geometric Intuitions for the SVD

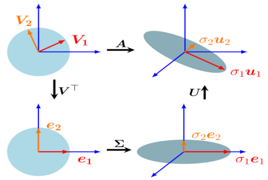
The SVD offers geometric intuitions to describe a transformation matrix A. In the following, we will discuss the SVD as sequential linear transformations performed on the bases. In Example 4.12, we will then apply transformation matrices of the SVD to a set of vectors in $\mathbb{R}^2$, which allows us to visualize the effect of each transformation more clearly.

Geometric Intuitions for the SVD

The SVD offers geometric intuitions to describe a transformation matrix A. In the following, we will discuss the SVD as sequential linear transformations performed on the bases. In Example 4.12, we will then apply transformation matrices of the SVD to a set of vectors in $\mathbb{R}^2$, which allows us to visualize the effect of each transformation more clearly.

Geometric Intuitions for the SVD

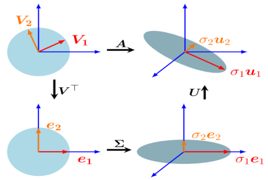
The SVD of a matrix can be interpreted as a decomposition of a corresponding linear mapping (recall Section 2.7.1) $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ into three operations; see Figure 4.8.



The SVD intuition follows superficially a similar structure to our eigendecomposition intuition.

Geometric Intuitions for the SVD

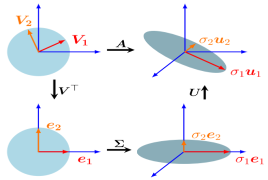
The SVD of a matrix can be interpreted as a decomposition of a corresponding linear mapping (recall Section 2.7.1) $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ into three operations; see Figure 4.8.



The SVD intuition follows superficially a similar structure to our eigendecomposition intuition.

Geometric Intuitions for the SVD

The SVD of a matrix can be interpreted as a decomposition of a corresponding linear mapping (recall Section 2.7.1) $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ into three operations; see Figure 4.8.



The SVD intuition follows superficially a similar structure to our eigendecomposition intuition.

Geometric Intuitions for the SVD

Assume we are given a transformation matrix of a linear mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ with respect to the standard bases B and C of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Moreover, assume a second basis $\tilde{B}$ of $\mathbb{R}^n$ and $\tilde{C}$ of $\mathbb{R}^m$. Then

1. The matrix $\boldsymbol{V}$ performs a basis change in the domain $\mathbb{R}^n$ from $\tilde{B}$ (represented by the red and orange vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ in the top-left of Figure 4.8) to the standard basis $B.\boldsymbol{V}^\top = \boldsymbol{V}^{-1}$ performs a basis change from B to $\tilde{B}$. The red and orange vectors are now aligned with the canonical basis in the bottom-left of Figure 4.8.

Geometric Intuitions for the SVD

Assume we are given a transformation matrix of a linear mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ with respect to the standard bases B and C of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Moreover, assume a second basis $\tilde{B}$ of $\mathbb{R}^n$ and $\tilde{C}$ of $\mathbb{R}^m$. Then

1. The matrix $\boldsymbol{V}$ performs a basis change in the domain $\mathbb{R}^n$ from $\tilde{B}$ (represented by the red and orange vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ in the top-left of Figure 4.8) to the standard basis $B. \boldsymbol{V}^\top = \boldsymbol{V}^{-1}$ performs a basis change from B to $\tilde{B}$. The red and orange vectors are now aligned with the canonical basis in the bottom-left of Figure 4.8.

Geometric Intuitions for the SVD

Assume we are given a transformation matrix of a linear mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ with respect to the standard bases B and C of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Moreover, assume a second basis $\tilde{B}$ of $\mathbb{R}^n$ and $\tilde{C}$ of $\mathbb{R}^m$. Then

1. The matrix $\boldsymbol{V}$ performs a basis change in the domain $\mathbb{R}^n$ from $\tilde{B}$ (represented by the red and orange vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ in the top-left of Figure 4.8) to the standard basis B.$\boldsymbol{V}^\top = \boldsymbol{V}^{-1}$ performs a basis change from B to $\tilde{B}$. The red and orange vectors are now aligned with the canonical basis in the bottom-left of Figure 4.8.

Geometric Intuitions for the SVD

2. Having changed the coordinate system to $\tilde{B}$, $\boldsymbol{\Sigma}$ scales the new coordinates by the singular values σ_i (and adds or deletes dimensions), i.e., Σ is the transformation matrix of Φ with respect to $\tilde{B}$ and $\tilde{C}$, represented by the red and orange vectors being stretched and lying in the $e_1 - e_2$ plane, which is now embedded in a third dimension in the bottom-right of Figure 4.8.
3. $\boldsymbol{U}$ performs a basis change in the codomain $\mathbb{R}^m$ from $\tilde{C}$ into the canonical basis of $\mathbb{R}^m$, represented by a rotation of the red and orange vectors out of the $e_1 - e_2$ plane. This is shown in the top-right of Figure 4.8.

Geometric Intuitions for the SVD

2. Having changed the coordinate system to $\tilde{B}$, $\boldsymbol{\Sigma}$ scales the new coordinates by the singular values σ_i (and adds or deletes dimensions), i.e., Σ is the transformation matrix of Φ with respect to $\tilde{B}$ and $\tilde{C}$, represented by the red and orange vectors being stretched and lying in the $e_1 - e_2$ plane, which is now embedded in a third dimension in the bottom-right of Figure 4.8.
3. $\boldsymbol{U}$ performs a basis change in the codomain $\mathbb{R}^m$ from $\tilde{C}$ into the canonical basis of $\mathbb{R}^m$, represented by a rotation of the red and orange vectors out of the $e_1 - e_2$ plane. This is shown in the top-right of Figure 4.8.

Geometric Intuitions for the SVD

The SVD expresses a change of basis in both the domain and codomain. This is in contrast with the eigendecomposition that operates within the same vector space, where the same basis change is applied and then undone. What makes the SVD special is that these two different bases are simultaneously linked by the singular value matrix $\boldsymbol{\Sigma}$.

Geometric Intuitions for the SVD

The SVD expresses a change of basis in both the domain and codomain. This is in contrast with the eigendecomposition that operates within the same vector space, where the same basis change is applied and then undone. What makes the SVD special is that these two different bases are simultaneously linked by the singular value matrix Σ.

Geometric Intuitions for the SVD

The SVD expresses a change of basis in both the domain and codomain. This is in contrast with the eigendecomposition that operates within the same vector space, where the same basis change is applied and then undone. What makes the SVD special is that these two different bases are simultaneously linked by the singular value matrix $\boldsymbol{\Sigma}$.

Construction of the SVD

In the following, we will explore why Theorem 4.22 holds and how the SVD is constructed. Computing the SVD of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is equivalent to finding two sets of orthonormal bases $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m)$ and $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ of the codomain $\mathbb{R}^m$ and the domain $\mathbb{R}^n$, respectively. From these ordered bases, we will construct the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$.

Construction of the SVD

In the following, we will explore why Theorem 4.22 holds and how the SVD is constructed. Computing the SVD of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is equivalent to finding two sets of orthonormal bases $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m)$ and $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ of the codomain $\mathbb{R}^m$ and the domain $\mathbb{R}^n$, respectively. From these ordered bases, we will construct the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$.

Construction of the SVD

In the following, we will explore why Theorem 4.22 holds and how the SVD is constructed. Computing the SVD of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is equivalent to finding two sets of orthonormal bases $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m)$ and $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$ of the codomain $\mathbb{R}^m$ and the domain $\mathbb{R}^n$, respectively. From these ordered bases, we will construct the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$.

Construction of the SVD

Our plan is to start with constructing the orthonormal set of rightsingular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{R}^n$. We then construct the orthonormal set of left-singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \mathbb{R}^m$. Thereafter, we will link the two and require that the orthogonality of the $\boldsymbol{v}_i$ is preserved under the transformation of $\boldsymbol{A}$. This is important because we know that the images $\boldsymbol{A}\boldsymbol{v}_i$ form a set of orthogonal vectors. We will then normalize these images by scalar factors, which will turn out to be the singular values.

Construction of the SVD

Our plan is to start with constructing the orthonormal set of rightsingular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{R}^n$. We then construct the orthonormal set of left-singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \mathbb{R}^m$. Thereafter, we will link the two and require that the orthogonality of the $\boldsymbol{v}_i$ is preserved under the transformation of $\boldsymbol{A}$. This is important because we know that the images $\boldsymbol{A}\boldsymbol{v}_i$ form a set of orthogonal vectors. We will then normalize these images by scalar factors, which will turn out to be the singular values.

Construction of the SVD

Our plan is to start with constructing the orthonormal set of rightsingular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{R}^n$. We then construct the orthonormal set of left-singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \mathbb{R}^m$. Thereafter, we will link the two and require that the orthogonality of the $\boldsymbol{v}_i$ is preserved under the transformation of $\boldsymbol{A}$. This is important because we know that the images $\boldsymbol{A}\boldsymbol{v}_i$ form a set of orthogonal vectors. We will then normalize these images by scalar factors, which will turn out to be the singular values.

Construction of the SVD

Our plan is to start with constructing the orthonormal set of rightsingular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{R}^n$. We then construct the orthonormal set of left-singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \mathbb{R}^m$. Thereafter, we will link the two and require that the orthogonality of the $\boldsymbol{v}_i$ is preserved under the transformation of $\boldsymbol{A}$. This is important because we know that the images $\boldsymbol{A}\boldsymbol{v}_i$ form a set of orthogonal vectors. We will then normalize these images by scalar factors, which will turn out to be the singular values.

Construction of the SVD

Our plan is to start with constructing the orthonormal set of rightsingular vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{R}^n$. We then construct the orthonormal set of left-singular vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \mathbb{R}^m$. Thereafter, we will link the two and require that the orthogonality of the $\boldsymbol{v}_i$ is preserved under the transformation of $\boldsymbol{A}$. This is important because we know that the images $\boldsymbol{A}\boldsymbol{v}_i$ form a set of orthogonal vectors. We will then normalize these images by scalar factors, which will turn out to be the singular values.

Construction of the SVD

Let us begin with constructing the right-singular vectors. The spectral theorem (Theorem 4.15) tells us that a symmetric matrix possesses an ONB of eigenvectors, which also means it can be diagonalized.

Moreover, from Theorem 4.14 we can always construct a symmetric, positive semidefinite matrix $\boldsymbol{A}^\top \boldsymbol{A} \in \mathbb{R}^{n\times n}$ from any rectangular matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Thus, we can always diagonalize $\boldsymbol{A}^\top \boldsymbol{A}$ and obtain

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top = \boldsymbol{P}\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}\boldsymbol{P}^\top \tag{4.71}$$

where $\boldsymbol{P}$ is an orthogonal matrix, which is composed of the orthonormal eigenbasis. The $\lambda_i \geqslant 0$ are the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$.

Construction of the SVD

Let us begin with constructing the right-singular vectors. The spectral theorem (Theorem 4.15) tells us that a symmetric matrix possesses an ONB of eigenvectors, which also means it can be diagonalized. Moreover, from Theorem 4.14 we can always construct a symmetric, positive semidefinite matrix $\boldsymbol{A}^\top \boldsymbol{A} \in \mathbb{R}^{n\times n}$ from any rectangular matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Thus, we can always diagonalize $\boldsymbol{A}^\top \boldsymbol{A}$ and obtain

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top = \boldsymbol{P}\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}\boldsymbol{P}^\top \tag{4.71}$$

where $\boldsymbol{P}$ is an orthogonal matrix, which is composed of the orthonormal eigenbasis. The $\lambda_i \geqslant 0$ are the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$.

Construction of the SVD

Let us begin with constructing the right-singular vectors. The spectral theorem (Theorem 4.15) tells us that a symmetric matrix possesses an ONB of eigenvectors, which also means it can be diagonalized. Moreover, from Theorem 4.14 we can always construct a symmetric, positive semidefinite matrix $\boldsymbol{A}^\top\boldsymbol{A} \in \mathbb{R}^{n\times n}$ from any rectangular matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$. Thus, we can always diagonalize $\boldsymbol{A}^\top\boldsymbol{A}$ and obtain

$$\boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top = \boldsymbol{P}\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}\boldsymbol{P}^\top \tag{4.71}$$

where $\boldsymbol{P}$ is an orthogonal matrix, which is composed of the orthonormal eigenbasis. The $\lambda_i \geqslant 0$ are the eigenvalues of $\boldsymbol{A}^\top\boldsymbol{A}$.

Construction of the SVD

Let us begin with constructing the right-singular vectors. The spectral theorem (Theorem 4.15) tells us that a symmetric matrix possesses an ONB of eigenvectors, which also means it can be diagonalized. Moreover, from Theorem 4.14 we can always construct a symmetric, positive semidefinite matrix $\boldsymbol{A}^\top \boldsymbol{A} \in \mathbb{R}^{n \times n}$ from any rectangular matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Thus, we can always diagonalize $\boldsymbol{A}^\top \boldsymbol{A}$ and obtain

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top = \boldsymbol{P}\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}\boldsymbol{P}^\top \tag{4.71}$$

where $\boldsymbol{P}$ is an orthogonal matrix, which is composed of the orthonormal eigenbasis. The $\lambda_i \geqslant 0$ are the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$.

Construction of the SVD

Let us assume the SVD of $\boldsymbol{A}$ exists and inject $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ into (4.71). This yields

$$\boldsymbol{A}^\top\boldsymbol{A} = \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)^\top\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right) = \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$

where $\boldsymbol{U}, \boldsymbol{V}$ are orthogonal matrices. Therefore, with $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}$ we obtain

$$\boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{V}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}\boldsymbol{V}^\top \tag{4.73}$$

Comparing now (4.71) and (4.73), we identify

$$\boldsymbol{V}^\top = \boldsymbol{P}^\top \tag{4.74}$$

$$\sigma_i^2 = \lambda_i \tag{4.75}$$

Construction of the SVD

Let us assume the SVD of $\boldsymbol{A}$ exists and inject $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ into (4.71). This yields

$$\boldsymbol{A}^\top \boldsymbol{A} = \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)^\top \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right) = \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$

where $\boldsymbol{U}, \boldsymbol{V}$ are orthogonal matrices. Therefore, with $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}$ we obtain

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{V}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}\boldsymbol{V}^\top \tag{4.73}$$

Comparing now (4.71) and (4.73), we identify

$$\boldsymbol{V}^\top = \boldsymbol{P}^\top \tag{4.74}$$

$$\sigma_i^2 = \lambda_i \tag{4.75}$$

Construction of the SVD

Let us assume the SVD of $\boldsymbol{A}$ exists and inject $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ into (4.71). This yields

$$\boldsymbol{A}^\top\boldsymbol{A} = \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)^\top\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right) = \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$

where $\boldsymbol{U}, \boldsymbol{V}$ are orthogonal matrices. Therefore, with $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}$ we obtain

$$\boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{V}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}\boldsymbol{V}^\top \tag{4.73}$$

Comparing now (4.71) and (4.73), we identify

$$\boldsymbol{V}^\top = \boldsymbol{P}^\top \tag{4.74}$$

$$\sigma_i^2 = \lambda_i \tag{4.75}$$

Construction of the SVD

Let us assume the SVD of $\boldsymbol{A}$ exists and inject $\boldsymbol{A} = \boldsymbol{U\Sigma V}^\top$ into (4.71). This yields

$$\boldsymbol{A}^\top\boldsymbol{A} = \left(\boldsymbol{U\Sigma V}^\top\right)^\top\left(\boldsymbol{U\Sigma V}^\top\right) = \boldsymbol{V\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{U\Sigma V}^\top$$

where $\boldsymbol{U}, \boldsymbol{V}$ are orthogonal matrices. Therefore, with $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}$ we obtain

$$\boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{V\Sigma}^\top\boldsymbol{\Sigma V}^\top = \boldsymbol{V}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}\boldsymbol{V}^\top \tag{4.73}$$

Comparing now (4.71) and (4.73), we identify

$$\boldsymbol{V}^\top = \boldsymbol{P}^\top \tag{4.74}$$

$$\sigma_i^2 = \lambda_i \tag{4.75}$$

Construction of the SVD

Therefore, the eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$ that compose $\boldsymbol{P}$ are the right-singular vectors $\boldsymbol{V}$ of $\boldsymbol{A}$ (see (4.74)). The eigenvalues of $\boldsymbol{A}^\top\boldsymbol{A}$ are the squared singular values of $\boldsymbol{\Sigma}$ (see (4.75)).

Construction of the SVD

Therefore, the eigenvectors of $\boldsymbol{A}^\top \boldsymbol{A}$ that compose $\boldsymbol{P}$ are the right-singular vectors $\boldsymbol{V}$ of $\boldsymbol{A}$ (see (4.74)). The eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$ are the squared singular values of $\boldsymbol{\Sigma}$ (see (4.75)).

Construction of the SVD

To obtain the left-singular vectors $\boldsymbol{U}$, we follow a similar procedure. We start by computing the SVD of the symmetric matrix $\boldsymbol{A}\boldsymbol{A}^\top \in \mathbb{R}^{m \times m}$ (instead of the previous $\boldsymbol{A}^\top\boldsymbol{A} \in \mathbb{R}^{n \times n}$). The SVD of $\boldsymbol{A}$ yields

$$\boldsymbol{A}\boldsymbol{A}^\top = \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)^\top = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top \quad (4.76a)$$

$$= \boldsymbol{U}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m^2 \end{bmatrix}\boldsymbol{U}^\top \quad (4.76b)$$

Construction of the SVD

To obtain the left-singular vectors $\boldsymbol{U}$, we follow a similar procedure. We start by computing the SVD of the symmetric matrix $\boldsymbol{A}\boldsymbol{A}^\top \in \mathbb{R}^{m\times m}$ (instead of the previous $\boldsymbol{A}^\top\boldsymbol{A} \in \mathbb{R}^{n\times n}$). The SVD of $\boldsymbol{A}$ yields

$$\boldsymbol{A}\boldsymbol{A}^\top = \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)^\top = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top \tag{4.76a}$$

$$= \boldsymbol{U}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m^2 \end{bmatrix}\boldsymbol{U}^\top \tag{4.76b}$$

Construction of the SVD

To obtain the left-singular vectors $\boldsymbol{U}$, we follow a similar procedure. We start by computing the SVD of the symmetric matrix $\boldsymbol{A}\boldsymbol{A}^\top \in \mathbb{R}^{m\times m}$ (instead of the previous $\boldsymbol{A}^\top\boldsymbol{A} \in \mathbb{R}^{n\times n}$). The SVD of $\boldsymbol{A}$ yields

$$\boldsymbol{A}\boldsymbol{A}^\top = \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\right)^\top = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top \tag{4.76a}$$

$$= \boldsymbol{U}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m^2 \end{bmatrix}\boldsymbol{U}^\top \tag{4.76b}$$

Construction of the SVD

The spectral theorem tells us that $AA^\top = SDS^\top$ can be diagonalized and we can find an ONB of eigenvectors of $AA^\top$, which are collected in S. The orthonormal eigenvectors of $AA^\top$ are the left-singular vectors U and form an orthonormal basis set in the codomain of the SVD.

Construction of the SVD

The spectral theorem tells us that $A\boldsymbol{A}^\top = \boldsymbol{SDS}^\top$ can be diagonalized and we can find an ONB of eigenvectors of $\boldsymbol{AA}^\top$, which are collected in S. The orthonormal eigenvectors of $\boldsymbol{AA}^\top$ are the left-singular vectors $\boldsymbol{U}$ and form an orthonormal basis set in the codomain of the SVD.

Construction of the SVD

This leaves the question of the structure of the matrix $\boldsymbol{\Sigma}$. Since $\boldsymbol{A}\boldsymbol{A}^{\top}$ and $A^{\top}\boldsymbol{A}$ have the same nonzero eigenvalues (see page 106) the nonzero entries of the Σ matrices in the SVD for both cases have to be the same. The last step is to link up all the parts we touched upon so far. We have an orthonormal set of right-singular vectors in V. To finish the construction of the SVD, we connect them with the orthonormal vectors $\boldsymbol{U}$. To reach this goal, we use the fact the images of the $\boldsymbol{v}_i$ under $\boldsymbol{A}$ have to be orthogonal, too. We can show this by using the results from Section 3.4.

Construction of the SVD

This leaves the question of the structure of the matrix $\boldsymbol{\Sigma}$. Since $\boldsymbol{A}\boldsymbol{A}^\top$ and $A^\top\boldsymbol{A}$ have the same nonzero eigenvalues (see page 106) the nonzero entries of the Σ matrices in the SVD for both cases have to be the same. The last step is to link up all the parts we touched upon so far. We have an orthonormal set of right-singular vectors in V. To finish the construction of the SVD, we connect them with the orthonormal vectors $\boldsymbol{U}$. To reach this goal, we use the fact the images of the $\boldsymbol{v}_i$ under $\boldsymbol{A}$ have to be orthogonal, too. We can show this by using the results from Section 3.4.

Construction of the SVD

This leaves the question of the structure of the matrix $\boldsymbol{\Sigma}$. Since $\boldsymbol{A}\boldsymbol{A}^{\top}$ and $A^{\top}\boldsymbol{A}$ have the same nonzero eigenvalues (see page 106) the nonzero entries of the Σ matrices in the SVD for both cases have to be the same. The last step is to link up all the parts we touched upon so far. We have an orthonormal set of right-singular vectors in V. To finish the construction of the SVD, we connect them with the orthonormal vectors $\boldsymbol{U}$. To reach this goal, we use the fact the images of the $\boldsymbol{v}_i$ under $\boldsymbol{A}$ have to be orthogonal, too. We can show this by using the results from Section 3.4.

Construction of the SVD

This leaves the question of the structure of the matrix $\boldsymbol{\Sigma}$. Since $\boldsymbol{A}\boldsymbol{A}^{\top}$ and $A^{\top}\boldsymbol{A}$ have the same nonzero eigenvalues (see page 106) the nonzero entries of the Σ matrices in the SVD for both cases have to be the same. The last step is to link up all the parts we touched upon so far. We have an orthonormal set of right-singular vectors in V. To finish the construction of the SVD, we connect them with the orthonormal vectors $\boldsymbol{U}$. To reach this goal, we use the fact the images of the $\boldsymbol{v}_i$ under $\boldsymbol{A}$ have to be orthogonal, too. We can show this by using the results from Section 3.4.

Construction of the SVD

We require that the inner product between $\boldsymbol{A}\boldsymbol{v}_i$ and $\boldsymbol{A}\boldsymbol{v}_j$ must be 0 for $i \neq j$. For any two orthogonal eigenvectors $\boldsymbol{v}_i, \boldsymbol{v}_j, i \neq j$, it holds that

$$(\boldsymbol{A}\boldsymbol{v}_i)^\top (\boldsymbol{A}\boldsymbol{v}_j) = \boldsymbol{v}_i^\top \left(\boldsymbol{A}^\top \boldsymbol{A}\right) \boldsymbol{v}_j = \boldsymbol{v}_i^\top (\lambda_j \boldsymbol{v}_j) = \lambda_j \boldsymbol{v}_i^\top \boldsymbol{v}_j = 0 \tag{4.77}$$

Construction of the SVD

We require that the inner product between $\boldsymbol{A}\boldsymbol{v}_i$ and $\boldsymbol{A}\boldsymbol{v}_j$ must be 0 for $i \neq j$. For any two orthogonal eigenvectors $\boldsymbol{v}_i, \boldsymbol{v}_j, i \neq j$, it holds that

$$(\boldsymbol{A}\boldsymbol{v}_i)^\top (\boldsymbol{A}\boldsymbol{v}_j) = \boldsymbol{v}_i^\top \left(\boldsymbol{A}^\top \boldsymbol{A}\right) \boldsymbol{v}_j = \boldsymbol{v}_i^\top (\lambda_j \boldsymbol{v}_j) = \lambda_j \boldsymbol{v}_i^\top \boldsymbol{v}_j = 0 \tag{4.77}$$

Construction of the SVD

For the case $m \geqslant r$, it holds that $\{\boldsymbol{A}\boldsymbol{v}_1, \ldots, \boldsymbol{A}\boldsymbol{v}_r\}$ is a basis of an r dimensional subspace of $\mathbb{R}^m$. To complete the SVD construction, we need left-singular vectors that are orthonormal: We normalize the images of the right-singular vectors $\boldsymbol{A}\boldsymbol{v}_i$ and obtain

$$\boldsymbol{u}_i := \frac{\boldsymbol{A}\boldsymbol{v}_i}{\|\boldsymbol{A}\boldsymbol{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{A}\boldsymbol{v}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i \tag{4.78}$$

where the last equality was obtained from (4.75) and (4.76b), showing us that the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\top$ are such that $\sigma_i^2 = \lambda_i$. Therefore, the eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$, which we know are the rightsingular vectors $\boldsymbol{v}_i$, and their normalized images under $\boldsymbol{A}$, the left-singular vectors $\boldsymbol{u}_i$, form two self-consistent ONBs that are connected through the singular value matrix $\boldsymbol{\Sigma}$.

Construction of the SVD

For the case $m \geqslant r$, it holds that $\{\boldsymbol{Av}_1, \ldots, \boldsymbol{Av}_r\}$ is a basis of an r dimensional subspace of $\mathbb{R}^m$. To complete the SVD construction, we need left-singular vectors that are orthonormal: We normalize the images of the right-singular vectors $\boldsymbol{Av}_i$ and obtain

$$\boldsymbol{u}_i := \frac{\boldsymbol{Av}_i}{\|\boldsymbol{Av}_i\|} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{Av}_i = \frac{1}{\sigma_i}\boldsymbol{Av}_i \tag{4.78}$$

where the last equality was obtained from (4.75) and (4.76b), showing us that the eigenvalues of $\boldsymbol{AA}^\top$ are such that $\sigma_i^2 = \lambda_i$.

Therefore, the eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$, which we know are the rightsingular vectors $\boldsymbol{v}_i$, and their normalized images under $\boldsymbol{A}$, the left-singular vectors $\boldsymbol{u}_i$, form two self-consistent ONBs that are connected through the singular value matrix $\boldsymbol{\Sigma}$.

Construction of the SVD

For the case $m \geqslant r$, it holds that $\{\boldsymbol{A}\boldsymbol{v}_1, \ldots, \boldsymbol{A}\boldsymbol{v}_r\}$ is a basis of an r dimensional subspace of $\mathbb{R}^m$. To complete the SVD construction, we need left-singular vectors that are orthonormal: We normalize the images of the right-singular vectors $\boldsymbol{A}\boldsymbol{v}_i$ and obtain

$$\boldsymbol{u}_i := \frac{\boldsymbol{A}\boldsymbol{v}_i}{\|\boldsymbol{A}\boldsymbol{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{A}\boldsymbol{v}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i \tag{4.78}$$

where the last equality was obtained from (4.75) and (4.76b), showing us that the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\top$ are such that $\sigma_i^2 = \lambda_i$. Therefore, the eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$, which we know are the rightsingular vectors $\boldsymbol{v}_i$, and their normalized images under $\boldsymbol{A}$, the left-singular vectors $\boldsymbol{u}_i$, form two self-consistent ONBs that are connected through the singular value matrix $\boldsymbol{\Sigma}$.

Construction of the SVD

For the case $m \geqslant r$, it holds that $\{\boldsymbol{A}\boldsymbol{v}_1, \ldots, \boldsymbol{A}\boldsymbol{v}_r\}$ is a basis of an r dimensional subspace of $\mathbb{R}^m$. To complete the SVD construction, we need left-singular vectors that are orthonormal: We normalize the images of the right-singular vectors $\boldsymbol{A}\boldsymbol{v}_i$ and obtain

$$\boldsymbol{u}_i := \frac{\boldsymbol{A}\boldsymbol{v}_i}{\|\boldsymbol{A}\boldsymbol{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{A}\boldsymbol{v}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i \tag{4.78}$$

where the last equality was obtained from (4.75) and (4.76b), showing us that the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\top$ are such that $\sigma_i^2 = \lambda_i$.

Therefore, the eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$, which we know are the rightsingular vectors $\boldsymbol{v}_i$, and their normalized images under $\boldsymbol{A}$, the left-singular vectors $\boldsymbol{u}_i$, form two self-consistent ONBs that are connected through the singular value matrix $\boldsymbol{\Sigma}$.

Construction of the SVD

For the case $m \geqslant r$, it holds that $\{\boldsymbol{A}\boldsymbol{v}_1, \ldots, \boldsymbol{A}\boldsymbol{v}_r\}$ is a basis of an r dimensional subspace of $\mathbb{R}^m$. To complete the SVD construction, we need left-singular vectors that are orthonormal: We normalize the images of the right-singular vectors $\boldsymbol{A}\boldsymbol{v}_i$ and obtain

$$\boldsymbol{u}_i := \frac{\boldsymbol{A}\boldsymbol{v}_i}{\|\boldsymbol{A}\boldsymbol{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{A}\boldsymbol{v}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i \tag{4.78}$$

where the last equality was obtained from (4.75) and (4.76b), showing us that the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\top$ are such that $\sigma_i^2 = \lambda_i$.
Therefore, the eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$, which we know are the rightsingular vectors $\boldsymbol{v}_i$, and their normalized images under $\boldsymbol{A}$, the left-singular vectors $\boldsymbol{u}_i$, form two self-consistent ONBs that are connected through the singular value matrix $\boldsymbol{\Sigma}$.

Construction of the SVD

For the case $m \geqslant r$, it holds that $\{\boldsymbol{A}\boldsymbol{v}_1, \ldots, \boldsymbol{A}\boldsymbol{v}_r\}$ is a basis of an r dimensional subspace of $\mathbb{R}^m$. To complete the SVD construction, we need left-singular vectors that are orthonormal: We normalize the images of the right-singular vectors $\boldsymbol{A}\boldsymbol{v}_i$ and obtain

$$\boldsymbol{u}_i := \frac{\boldsymbol{A}\boldsymbol{v}_i}{\|\boldsymbol{A}\boldsymbol{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}\boldsymbol{A}\boldsymbol{v}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i \tag{4.78}$$

where the last equality was obtained from (4.75) and (4.76b), showing us that the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\top$ are such that $\sigma_i^2 = \lambda_i$.
Therefore, the eigenvectors of $\boldsymbol{A}^\top\boldsymbol{A}$, which we know are the rightsingular vectors $\boldsymbol{v}_i$, and their normalized images under $\boldsymbol{A}$, the left-singular vectors $\boldsymbol{u}_i$, form two self-consistent ONBs that are connected through the singular value matrix $\boldsymbol{\Sigma}$.

Construction of the SVD

Let us rearrange (4.78) to obtain the singular value equation

$$\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i, \quad i = 1, \ldots, r \tag{4.79}$$

This equation closely resembles the eigenvalue equation (4.25), but the vectors on the left- and the right-hand sides are not the same.

Construction of the SVD

Let us rearrange (4.78) to obtain the singular value equation

$$\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i, \quad i = 1, \ldots, r \tag{4.79}$$

This equation closely resembles the eigenvalue equation (4.25), but the vectors on the left- and the right-hand sides are not the same.

Construction of the SVD

For $n > m$, (4.79) holds only for $i \leqslant m$ and (4.79) says nothing about the $\boldsymbol{u}_i$ for $i > m$. However, we know by construction that they are orthonormal. Conversely, for $m > n$, (4.79) holds only for $i \leqslant n$. For $i > n$, we have $\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{0}$ and we still know that the $\boldsymbol{v}_i$ form an orthonormal set. This means that the SVD also supplies an orthonormal basis of the kernel (null space) of $\boldsymbol{A}$, the set of vectors $\boldsymbol{x}$ with $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ (see Section 2.7.3).

Construction of the SVD

For $n > m$, (4.79) holds only for $i \leqslant m$ and (4.79) says nothing about the $\boldsymbol{u}_i$ for $i > m$. However, we know by construction that they are orthonormal. Conversely, for $m > n$, (4.79) holds only for $i \leqslant n$. For $i > n$, we have $\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{0}$ and we still know that the $\boldsymbol{v}_i$ form an orthonormal set. This means that the SVD also supplies an orthonormal basis of the kernel (null space) of $\boldsymbol{A}$, the set of vectors $\boldsymbol{x}$ with $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ (see Section 2.7.3).

Construction of the SVD

For $n > m$, (4.79) holds only for $i \leqslant m$ and (4.79) says nothing about the $\boldsymbol{u}_i$ for $i > m$. However, we know by construction that they are orthonormal. Conversely, for $m > n$, (4.79) holds only for $i \leqslant n$. For $i > n$, we have $\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{0}$ and we still know that the $\boldsymbol{v}_i$ form an orthonormal set. This means that the SVD also supplies an orthonormal basis of the kernel (null space) of $\boldsymbol{A}$, the set of vectors $\boldsymbol{x}$ with $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ (see Section 2.7.3).

Construction of the SVD

For $n > m$, (4.79) holds only for $i \leqslant m$ and (4.79) says nothing about the $\boldsymbol{u}_i$ for $i > m$. However, we know by construction that they are orthonormal. Conversely, for $m > n$, (4.79) holds only for $i \leqslant n$. For $i > n$, we have $\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{0}$ and we still know that the $\boldsymbol{v}_i$ form an orthonormal set. This means that the SVD also supplies an orthonormal basis of the kernel (null space) of $\boldsymbol{A}$, the set of vectors $\boldsymbol{x}$ with $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ (see Section 2.7.3).

Construction of the SVD

Moreover, concatenating the $\boldsymbol{v}_i$ as the columns of $\boldsymbol{V}$ and the $\boldsymbol{u}_i$ as the columns of $\boldsymbol{U}$ yields

$$\boldsymbol{AV} = \boldsymbol{U\Sigma}$$

where $\boldsymbol{\Sigma}$ has the same dimensions as $\boldsymbol{A}$ and a diagonal structure for rows $1, \ldots, r$. Hence, right-multiplying with $\boldsymbol{V}^\top$ yields $\boldsymbol{A} = \boldsymbol{U\Sigma V}^\top$, which is the SVD of $\boldsymbol{A}$.

Construction of the SVD

Moreover, concatenating the $\boldsymbol{v}_i$ as the columns of $\boldsymbol{V}$ and the $\boldsymbol{u}_i$ as the columns of $\boldsymbol{U}$ yields

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{\Sigma}$$

where $\boldsymbol{\Sigma}$ has the same dimensions as $\boldsymbol{A}$ and a diagonal structure for rows $1, \ldots, r$. Hence, right-multiplying with $\boldsymbol{V}^\top$ yields $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$, which is the SVD of $\boldsymbol{A}$.

Construction of the SVD

Example (Example 4.13 Computing the SVD)

Let us find the singular value decomposition of

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$$

The SVD requires us to compute the right-singular vectors $\boldsymbol{v}_j$, the singular values σ_k, and the left-singular vectors $\boldsymbol{u}_i$.

Construction of the SVD

Example (Example 4.13 Computing the SVD)

Let us find the singular value decomposition of

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$$

The SVD requires us to compute the right-singular vectors $\boldsymbol{v}_j$, the singular values σ_k, and the left-singular vectors $\boldsymbol{u}_i$.

Construction of the SVD

Step 1: Right-singular vectors as the eigenbasis of $A^\top A$.

We start by computing

$$\boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

We compute the singular values and right-singular vectors $\boldsymbol{v}_j$ through the eigenvalue decomposition of $\boldsymbol{A}^\top \boldsymbol{A}$, which is given as

$$\boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top$$

and we obtain the right-singular vectors as the columns of $\boldsymbol{P}$ so that

$$\boldsymbol{V} = \boldsymbol{P} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}.$$

Construction of the SVD

Step 1: Right-singular vectors as the eigenbasis of $A^\top A$.

We start by computing

$$\boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

We compute the singular values and right-singular vectors $\boldsymbol{v}_j$ through the eigenvalue decomposition of $\boldsymbol{A}^\top \boldsymbol{A}$, which is given as

$$\boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top$$

and we obtain the right-singular vectors as the columns of $\boldsymbol{P}$ so that

$$\boldsymbol{V} = \boldsymbol{P} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}.$$

Construction of the SVD

Step 1: Right-singular vectors as the eigenbasis of $A^\top A$.

We start by computing

$$\boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

We compute the singular values and right-singular vectors $\boldsymbol{v}_j$ through the eigenvalue decomposition of $\boldsymbol{A}^\top \boldsymbol{A}$, which is given as

$$\boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top$$

and we obtain the right-singular vectors as the columns of $\boldsymbol{P}$ so that

$$\boldsymbol{V} = \boldsymbol{P} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}.$$

Construction of the SVD

Step 1: Right-singular vectors as the eigenbasis of $A^\top A$.

We start by computing

$$\boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

We compute the singular values and right-singular vectors $\boldsymbol{v}_j$ through the eigenvalue decomposition of $\boldsymbol{A}^\top \boldsymbol{A}$, which is given as

$$\boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^\top$$

and we obtain the right-singular vectors as the columns of $\boldsymbol{P}$ so that

$$\boldsymbol{V} = \boldsymbol{P} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}.$$

Construction of the SVD

Step 2: Singular-value matrix.

As the singular values σ_i are the square roots of the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$, we obtain them straight from $\boldsymbol{D}$. Since $\mathrm{rk}(\boldsymbol{A}) = 2$, there are only two nonzero singular values: $\sigma_1 = \sqrt{6}$ and $\sigma_2 = 1$. The singular value matrix must be the same size as A, and we obtain

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Construction of the SVD

Step 2: Singular-value matrix.

As the singular values σ_i are the square roots of the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$, we obtain them straight from $\boldsymbol{D}$. Since $\text{rk}(\boldsymbol{A}) = 2$, there are only two nonzero singular values: $\sigma_1 = \sqrt{6}$ and $\sigma_2 = 1$. The singular value matrix must be the same size as A, and we obtain

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Construction of the SVD

Step 2: Singular-value matrix.
As the singular values σ_i are the square roots of the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$, we obtain them straight from $\boldsymbol{D}$. Since $\mathrm{rk}(\boldsymbol{A}) = 2$, there are only two nonzero singular values: $\sigma_1 = \sqrt{6}$ and $\sigma_2 = 1$. The singular value matrix must be the same size as A, and we obtain

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Construction of the SVD

Step 2: Singular-value matrix.
As the singular values σ_i are the square roots of the eigenvalues of $\boldsymbol{A}^\top \boldsymbol{A}$, we obtain them straight from $\boldsymbol{D}$. Since $\mathrm{rk}(\boldsymbol{A}) = 2$, there are only two nonzero singular values: $\sigma_1 = \sqrt{6}$ and $\sigma_2 = 1$. The singular value matrix must be the same size as A, and we obtain

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Construction of the SVD

Step 3: Left-singular vectors as the normalized image of the rightsingular vectors.

We find the left-singular vectors by computing the image of the rightsingular vectors under $\boldsymbol{A}$ and normalizing them by dividing them by their corresponding singular value. We obtain

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1}\boldsymbol{A}\boldsymbol{v}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{5}{\sqrt{30}} \\ \frac{-2}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix}$$

$$\boldsymbol{u}_2 = \frac{1}{\sigma_2}\boldsymbol{A}\boldsymbol{v}_2 = \frac{1}{1}\begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$$

$$\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2] = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

Construction of the SVD

Step 3: Left-singular vectors as the normalized image of the rightsingular vectors.

We find the left-singular vectors by computing the image of the rightsingular vectors under $\boldsymbol{A}$ and normalizing them by dividing them by their corresponding singular value. We obtain

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1}\boldsymbol{A}\boldsymbol{v}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{5}{\sqrt{30}} \\ \frac{-2}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix}$$

$$\boldsymbol{u}_2 = \frac{1}{\sigma_2}\boldsymbol{A}\boldsymbol{v}_2 = \frac{1}{1}\begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$$

$$\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2] = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

Construction of the SVD

Note that on a computer the approach illustrated here has poor numerical behavior, and the SVD of A is normally computed without resorting to the eigenvalue decomposition of $\boldsymbol{A}^\top \boldsymbol{A}$.

Eigenvalue Decomposition vs. Singular Value Decomposition

Let us consider the eigendecomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ and the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}$ and review the core elements of the past sections.

- The SVD always exists for any matrix $\mathbb{R}^{m\times n}$. The eigendecomposition is only defined for square matrices $\mathbb{R}^{n\times n}$ and only exists if we can find a basis of eigenvectors of $\mathbb{R}^{n}$.
- The vectors in the eigendecomposition matrix $\boldsymbol{P}$ are not necessarily orthogonal, i.e., the change of basis is not a simple rotation and scaling. On the other hand, the vectors in the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in the SVD are orthonormal, so they do represent rotations.

Eigenvalue Decomposition vs. Singular Value Decomposition

Let us consider the eigendecomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ and the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}$ and review the core elements of the past sections.

- The SVD always exists for any matrix $\mathbb{R}^{m\times n}$. The eigendecomposition is only defined for square matrices $\mathbb{R}^{n\times n}$ and only exists if we can find a basis of eigenvectors of $\mathbb{R}^{n}$.
- The vectors in the eigendecomposition matrix $\boldsymbol{P}$ are not necessarily orthogonal, i.e., the change of basis is not a simple rotation and scaling. On the other hand, the vectors in the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in the SVD are orthonormal, so they do represent rotations.

Eigenvalue Decomposition vs. Singular Value Decomposition

Let us consider the eigendecomposition $\boldsymbol{A} = \boldsymbol{PDP}^{-1}$ and the SVD $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\top}$ and review the core elements of the past sections.

- The SVD always exists for any matrix $\mathbb{R}^{m\times n}$. The eigendecomposition is only defined for square matrices $\mathbb{R}^{n\times n}$ and only exists if we can find a basis of eigenvectors of $\mathbb{R}^n$.
- The vectors in the eigendecomposition matrix $\boldsymbol{P}$ are not necessarily orthogonal, i.e., the change of basis is not a simple rotation and scaling. On the other hand, the vectors in the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in the SVD are orthonormal, so they do represent rotations.

Eigenvalue Decomposition vs. Singular Value Decomposition

Both the eigendecomposition and the SVD are compositions of three linear mappings:

- Change of basis in the domain
- Independent scaling of each new basis vector and mapping from domain to codomain
- Change of basis in the codomain

Eigenvalue Decomposition vs. Singular Value Decomposition

Both the eigendecomposition and the SVD are compositions of three linear mappings:

- Change of basis in the domain
- Independent scaling of each new basis vector and mapping from domain to codomain
- Change of basis in the codomain

Eigenvalue Decomposition vs. Singular Value Decomposition

Both the eigendecomposition and the SVD are compositions of three linear mappings:

- Change of basis in the domain
- Independent scaling of each new basis vector and mapping from domain to codomain
- Change of basis in the codomain

Outline

Matrix Approximation

We considered the SVD as a way to factorize $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \in \mathbb{R}^{m\times n}$ into the product of three matrices, where $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ are orthogonal and $\boldsymbol{\Sigma}$ contains the singular values on its main diagonal. Instead of doing the full SVD factorization, we will now investigate how the SVD allows us to represent a matrix $\boldsymbol{A}$ as a sum of simpler (low-rank) matrices A_i, which lends itself to a matrix approximation scheme that is cheaper to compute than the full SVD.

Matrix Approximation

We considered the SVD as a way to factorize $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \in \mathbb{R}^{m\times n}$ into the product of three matrices, where $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ are orthogonal and $\boldsymbol{\Sigma}$ contains the singular values on its main diagonal. Instead of doing the full SVD factorization, we will now investigate how the SVD allows us to represent a matrix $\boldsymbol{A}$ as a sum of simpler (low-rank) matrices A_i, which lends itself to a matrix approximation scheme that is cheaper to compute than the full SVD.

Matrix Approximation

We considered the SVD as a way to factorize $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \in \mathbb{R}^{m\times n}$ into the product of three matrices, where $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ are orthogonal and $\boldsymbol{\Sigma}$ contains the singular values on its main diagonal. Instead of doing the full SVD factorization, we will now investigate how the SVD allows us to represent a matrix $\boldsymbol{A}$ as a sum of simpler (low-rank) matrices A_i, which lends itself to a matrix approximation scheme that is cheaper to compute than the full SVD.

Matrix Approximation

We construct a rank-1 matrix $\boldsymbol{A}_i \in \mathbb{R}^{m\times n}$ as

$$\boldsymbol{A}_i := \boldsymbol{u}_i \boldsymbol{v}_i^\top$$

which is formed by the outer product of the i th orthogonal column vector of $\boldsymbol{U}$ and $\boldsymbol{V}$.

Matrix Approximation

A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ of rank r can be written as a sum of rank-1 matrices $\boldsymbol{A}_i$ so that

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{r} \sigma_i \boldsymbol{A}_i \tag{4.91}$$

where the outer-product matrices $\boldsymbol{A}_i$ are weighted by the i th singular value σ_i. We can see why (4.91) holds: The diagonal structure of the singular value matrix $\boldsymbol{\Sigma}$ multiplies only matching left- and right-singular vectors $\boldsymbol{u}_i \boldsymbol{v}_i^\top$ and scales them by the corresponding singular value σ_i. All terms $\sum_{ij} \boldsymbol{u}_i \boldsymbol{v}_j^\top$ vanish for $i \neq j$ because $\boldsymbol{\Sigma}$ is a diagonal matrix. Any terms $i > r$ vanish because the corresponding singular values are 0.

Matrix Approximation

A matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ of rank r can be written as a sum of rank-1 matrices $\boldsymbol{A}_i$ so that

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{r} \sigma_i \boldsymbol{A}_i \tag{4.91}$$

where the outer-product matrices $\boldsymbol{A}_i$ are weighted by the i th singular value σ_i. We can see why (4.91) holds: The diagonal structure of the singular value matrix $\boldsymbol{\Sigma}$ multiplies only matching left- and right-singular vectors $\boldsymbol{u}_i \boldsymbol{v}_i^\top$ and scales them by the corresponding singular value σ_i. All terms $\sum_{ij} \boldsymbol{u}_i \boldsymbol{v}_j^\top$ vanish for $i \neq j$ because $\boldsymbol{\Sigma}$ is a diagonal matrix. Any terms $i > r$ vanish because the corresponding singular values are 0.

Matrix Approximation

A matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ of rank r can be written as a sum of rank-1 matrices $\boldsymbol{A}_i$ so that

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{r} \sigma_i \boldsymbol{A}_i \tag{4.91}$$

where the outer-product matrices $\boldsymbol{A}_i$ are weighted by the i th singular value σ_i. We can see why (4.91) holds: The diagonal structure of the singular value matrix $\boldsymbol{\Sigma}$ multiplies only matching left- and right-singular vectors $\boldsymbol{u}_i \boldsymbol{v}_i^\top$ and scales them by the corresponding singular value σ_i. All terms $\sum_{ij} \boldsymbol{u}_i \boldsymbol{v}_j^\top$ vanish for $i \neq j$ because $\boldsymbol{\Sigma}$ is a diagonal matrix. Any terms $i > r$ vanish because the corresponding singular values are 0.

Matrix Approximation

A matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ of rank r can be written as a sum of rank-1 matrices $\boldsymbol{A}_i$ so that

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{r} \sigma_i \boldsymbol{A}_i \tag{4.91}$$

where the outer-product matrices $\boldsymbol{A}_i$ are weighted by the i th singular value σ_i. We can see why (4.91) holds: The diagonal structure of the singular value matrix $\boldsymbol{\Sigma}$ multiplies only matching left- and right-singular vectors $\boldsymbol{u}_i \boldsymbol{v}_i^\top$ and scales them by the corresponding singular value σ_i. All terms $\sum_{ij} \boldsymbol{u}_i \boldsymbol{v}_j^\top$ vanish for $i \neq j$ because $\boldsymbol{\Sigma}$ is a diagonal matrix. Any terms $i > r$ vanish because the corresponding singular values are 0.

Matrix Approximation

A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ of rank r can be written as a sum of rank-1 matrices $\boldsymbol{A}_i$ so that

$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{r} \sigma_i \boldsymbol{A}_i \tag{4.91}$$

where the outer-product matrices $\boldsymbol{A}_i$ are weighted by the i th singular value σ_i. We can see why (4.91) holds: The diagonal structure of the singular value matrix $\boldsymbol{\Sigma}$ multiplies only matching left- and right-singular vectors $\boldsymbol{u}_i \boldsymbol{v}_i^\top$ and scales them by the corresponding singular value σ_i. All terms $\sum_{ij} \boldsymbol{u}_i \boldsymbol{v}_j^\top$ vanish for $i \neq j$ because $\boldsymbol{\Sigma}$ is a diagonal matrix. Any terms $i > r$ vanish because the corresponding singular values are 0.

Matrix Approximation

We summed up the r individual rank-1 matrices to obtain a rank-r matrix $\boldsymbol{A}$; see (4.91). If the sum does not run over all matrices $\boldsymbol{A}_i, i = 1, \ldots, r$, but only up to an intermediate value $k < r$, we obtain a rank-k approximation

$$\widehat{\boldsymbol{A}}(k) := \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{k} \sigma_i \boldsymbol{A}_i$$

of $\boldsymbol{A}$ with $\mathrm{rk}(\widehat{\boldsymbol{A}}(k)) = k$.

Matrix Approximation

We summed up the r individual rank-1 matrices to obtain a rank-r matrix $\boldsymbol{A}$; see (4.91). If the sum does not run over all matrices $\boldsymbol{A}_i, i = 1, \ldots, r$, but only up to an intermediate value $k < r$, we obtain a rank-k approximation

$$\widehat{\boldsymbol{A}}(k) := \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{k} \sigma_i \boldsymbol{A}_i$$

of $\boldsymbol{A}$ with $\mathrm{rk}(\widehat{\boldsymbol{A}}(k)) = k$.

Matrix Approximation

We summed up the r individual rank-1 matrices to obtain a rank-r matrix $\boldsymbol{A}$; see (4.91). If the sum does not run over all matrices $\boldsymbol{A}_i, i = 1, \ldots, r$, but only up to an intermediate value $k < r$, we obtain a rank-k approximation

$$\widehat{\boldsymbol{A}}(k) := \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top = \sum_{i=1}^{k} \sigma_i \boldsymbol{A}_i$$

of $\boldsymbol{A}$ with $\mathrm{rk}(\widehat{\boldsymbol{A}}(k)) = k$.

Matrix Approximation

To measure the difference (error) between $\boldsymbol{A}$ and its rank-k k approximation $\widehat{\boldsymbol{A}}(k)$, we need the notion of a norm. In Section 3.1, we already used norms on vectors that measure the length of a vector. By analogy we can also define norms on matrices.

Matrix Approximation

To measure the difference (error) between $\boldsymbol{A}$ and its rank-k k approximation $\widehat{\boldsymbol{A}}(k)$, we need the notion of a norm. In Section 3.1, we already used norms on vectors that measure the length of a vector. By analogy we can also define norms on matrices.

Matrix Approximation

Definition (Definition 4.23 Spectral Norm of a Matrix)

For $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, the spectral norm of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is defined as

$$\|\boldsymbol{A}\|_2 := \max_{\boldsymbol{x}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \tag{4.93}$$

We introduce the notation of a subscript in the matrix norm (left-hand side), similar to the Euclidean norm for vectors (right-hand side), which has subscript 2. The spectral norm (4.93) determines how long any vector $\boldsymbol{x}$ can at most become when multiplied by $\boldsymbol{A}$.

Matrix Approximation

Definition (Definition 4.23 Spectral Norm of a Matrix)

For $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$, the spectral norm of a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is defined as

$$\|\boldsymbol{A}\|_2 := \max_{\boldsymbol{x}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \tag{4.93}$$

We introduce the notation of a subscript in the matrix norm (left-hand side), similar to the Euclidean norm for vectors (right-hand side), which has subscript 2. The spectral norm (4.93) determines how long any vector $\boldsymbol{x}$ can at most become when multiplied by $\boldsymbol{A}$.

Matrix Approximation

Definition (Definition 4.23 Spectral Norm of a Matrix)

For $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$, the spectral norm of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is defined as

$$\|\boldsymbol{A}\|_2 := \max_{\boldsymbol{x}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \tag{4.93}$$

We introduce the notation of a subscript in the matrix norm (left-hand side), similar to the Euclidean norm for vectors (right-hand side), which has subscript 2. The spectral norm (4.93) determines how long any vector $\boldsymbol{x}$ can at most become when multiplied by $\boldsymbol{A}$.

Matrix Approximation

Theorem (Theorem 4.24)

The spectral norm of $\boldsymbol{A}$ is its largest singular value σ_1.

We leave the proof of this theorem as an exercise.

Matrix Approximation

Theorem (Theorem 4.24)

The spectral norm of $\boldsymbol{A}$ is its largest singular value σ_1.

We leave the proof of this theorem as an exercise.

Matrix Approximation

Theorem (Theorem 4.25 Eckart-Young Theorem (Eckart and Young, 1936))

Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ *of rank* r *and let* $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ *be a matrix of rank* k. *For any* $k \leqslant r$ *with* $\widehat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ *it holds that*

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\operatorname{rk}(\boldsymbol{B})=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2, \tag{4.94}$$

$$\|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2 = \sigma_{k+1}. \tag{4.95}$$

The Eckart-Young theorem states explicitly how much error we introduce by approximating $\boldsymbol{A}$ using a rank-k approximation. We can interpret the rank-k approximation obtained with the SVD as a projection of the full-rank matrix $\boldsymbol{A}$ onto a lower-dimensional space of rank-at-most-k matrices. Of all possible projections, the SVD minimizes the error (with respect to the spectral norm) between $\boldsymbol{A}$ and any rank- k approximation.

Matrix Approximation

Theorem (Theorem 4.25 Eckart-Young Theorem (Eckart and Young, 1936))

Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ of rank r and let $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ be a matrix of rank k. For any $k \leqslant r$ with $\widehat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ it holds that

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\operatorname{rk}(\boldsymbol{B})=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2, \tag{4.94}$$

$$\|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2 = \sigma_{k+1}. \tag{4.95}$$

The Eckart-Young theorem states explicitly how much error we introduce by approximating $\boldsymbol{A}$ using a rank-k approximation. We can interpret the rank-k approximation obtained with the SVD as a projection of the full-rank matrix $\boldsymbol{A}$ onto a lower-dimensional space of rank-at-most-k matrices. Of all possible projections, the SVD minimizes the error (with respect to the spectral norm) between $\boldsymbol{A}$ and any rank- k approximation.

Matrix Approximation

Theorem (Theorem 4.25 Eckart-Young Theorem (Eckart and Young, 1936))

Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ *of rank* r *and let* $\boldsymbol{B} \in \mathbb{R}^{m\times n}$ *be a matrix of rank* k. *For any* $k \leqslant r$ *with* $\widehat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ *it holds that*

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\operatorname{rk}(B)=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2, \tag{4.94}$$

$$\|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2 = \sigma_{k+1}. \tag{4.95}$$

The Eckart-Young theorem states explicitly how much error we introduce by approximating $\boldsymbol{A}$ using a rank-k approximation. We can interpret the rank-k approximation obtained with the SVD as a projection of the full-rank matrix $\boldsymbol{A}$ onto a lower-dimensional space of rank-at-most-k matrices. Of all possible projections, the SVD minimizes the error (with respect to the spectral norm) between $\boldsymbol{A}$ and any rank- k approximation.

Matrix Approximation

Theorem (Theorem 4.25 Eckart-Young Theorem (Eckart and Young, 1936))

Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ *of rank* r *and let* $\boldsymbol{B} \in \mathbb{R}^{m\times n}$ *be a matrix of rank* k*. For any* $k \leqslant r$ *with* $\widehat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ *it holds that*

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\operatorname{rk}(B)=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2, \tag{4.94}$$

$$\|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2 = \sigma_{k+1}. \tag{4.95}$$

The Eckart-Young theorem states explicitly how much error we introduce by approximating $\boldsymbol{A}$ using a rank-k approximation. We can interpret the rank-k approximation obtained with the SVD as a projection of the full-rank matrix $\boldsymbol{A}$ onto a lower-dimensional space of rank-at-most-k matrices. Of all possible projections, the SVD minimizes the error (with respect to the spectral norm) between $\boldsymbol{A}$ and any rank- k approximation.

Matrix Approximation

Theorem (Theorem 4.25 Eckart-Young Theorem (Eckart and Young, 1936))

Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ *of rank* r *and let* $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ *be a matrix of rank* k*. For any* $k \leqslant r$ *with* $\widehat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ *it holds that*

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\operatorname{rk}(B)=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2, \tag{4.94}$$

$$\|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2 = \sigma_{k+1}. \tag{4.95}$$

The Eckart-Young theorem states explicitly how much error we introduce by approximating $\boldsymbol{A}$ using a rank-k approximation. We can interpret the rank-k approximation obtained with the SVD as a projection of the full-rank matrix $\boldsymbol{A}$ onto a lower-dimensional space of rank-at-most-k matrices. Of all possible projections, the SVD minimizes the error (with respect to the spectral norm) between $\boldsymbol{A}$ and any rank- k approximation.

Matrix Approximation

Theorem (Theorem 4.25 Eckart-Young Theorem (Eckart and Young, 1936))

Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ *of rank* r *and let* $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ *be a matrix of rank* k*. For any* $k \leqslant r$ *with* $\widehat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\top}$ *it holds that*

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\operatorname{rk}(B)=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2, \tag{4.94}$$

$$\|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2 = \sigma_{k+1}. \tag{4.95}$$

The Eckart-Young theorem states explicitly how much error we introduce by approximating $\boldsymbol{A}$ using a rank-k approximation. We can interpret the rank-k approximation obtained with the SVD as a projection of the full-rank matrix $\boldsymbol{A}$ onto a lower-dimensional space of rank-at-most-k matrices. Of all possible projections, the SVD minimizes the error (with respect to the spectral norm) between $\boldsymbol{A}$ and any rank- k approximation.

Matrix Approximation

Theorem (Theorem 4.25 Eckart-Young Theorem (Eckart and Young, 1936))

Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ of rank r and let $\boldsymbol{B} \in \mathbb{R}^{m\times n}$ be a matrix of rank k. For any $k \leqslant r$ with $\widehat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ it holds that

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\operatorname{rk}(B)=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2, \tag{4.94}$$

$$\|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2 = \sigma_{k+1}. \tag{4.95}$$

The Eckart-Young theorem states explicitly how much error we introduce by approximating $\boldsymbol{A}$ using a rank-k approximation. We can interpret the rank-k approximation obtained with the SVD as a projection of the full-rank matrix $\boldsymbol{A}$ onto a lower-dimensional space of rank-at-most-k matrices. Of all possible projections, the SVD minimizes the error (with respect to the spectral norm) between $\boldsymbol{A}$ and any rank- k approximation.

Matrix Approximation

We can retrace some of the steps to understand why (4.95) should hold. We observe that the difference between $\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)$ is a matrix containing the sum of the remaining rank-1 matrices

$$\boldsymbol{A} - \widehat{\boldsymbol{A}}(k) = \sum_{i=k+1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$$

Matrix Approximation

By Theorem 4.24, we immediately obtain σ_{k+1} as the spectral norm of the difference matrix. Let us have a closer look at (4.94),

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\operatorname{rk}(B)=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2 .$$

If we assume that there is another matrix $\boldsymbol{B}$ with $\operatorname{rk}(\boldsymbol{B}) \leqslant k$, such that

$$\|\boldsymbol{A} - \boldsymbol{B}\|_2 < \|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2$$

then there exists an at least $(n-k)$-dimensional null space $Z \subseteq \mathbb{R}^n$, such that $\boldsymbol{x} \in Z$ implies that $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$.

Matrix Approximation

By Theorem 4.24, we immediately obtain σ_{k+1} as the spectral norm of the difference matrix. Let us have a closer look at (4.94),

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\mathrm{rk}(B)=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2.$$

If we assume that there is another matrix $\boldsymbol{B}$ with $\mathrm{rk}(\boldsymbol{B}) \leqslant k$, such that

$$\|\boldsymbol{A} - \boldsymbol{B}\|_2 < \|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2$$

then there exists an at least $(n-k)$-dimensional null space $Z \subseteq \mathbb{R}^n$, such that $\boldsymbol{x} \in Z$ implies that $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$.

Matrix Approximation

By Theorem 4.24, we immediately obtain σ_{k+1} as the spectral norm of the difference matrix. Let us have a closer look at (4.94),

$$\widehat{\boldsymbol{A}}(k) = \operatorname{argmin}_{\operatorname{rk}(B)=k} \|\boldsymbol{A} - \boldsymbol{B}\|_2 .$$

If we assume that there is another matrix $\boldsymbol{B}$ with $\operatorname{rk}(\boldsymbol{B}) \leqslant k$, such that

$$\|\boldsymbol{A} - \boldsymbol{B}\|_2 < \|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\|_2$$

then there exists an at least $(n-k)$-dimensional null space $Z \subseteq \mathbb{R}^n$, such that $\boldsymbol{x} \in Z$ implies that $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$.

Matrix Approximation

Then it follows that

$$\|\boldsymbol{A}\boldsymbol{x}\|_2 = \|(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{x}\|_2$$

and by using a version of the Cauchy-Schwartz inequality (3.17) that encompasses norms of matrices, we obtain

$$\|\boldsymbol{A}\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{A} - \boldsymbol{B}\|_2 \|\boldsymbol{x}\|_2 < \sigma_{k+1} \|\boldsymbol{x}\|_2$$

Matrix Approximation

Then it follows that

$$\|\boldsymbol{A}\boldsymbol{x}\|_2 = \|(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{x}\|_2$$

and by using a version of the Cauchy-Schwartz inequality (3.17) that encompasses norms of matrices, we obtain

$$\|\boldsymbol{A}\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{A} - \boldsymbol{B}\|_2 \|\boldsymbol{x}\|_2 < \sigma_{k+1} \|\boldsymbol{x}\|_2$$

Matrix Approximation

However, there exists a $(k+1)$-dimensional subspace where $\|\boldsymbol{Ax}\|_2 \geqslant \sigma_{k+1}\|\boldsymbol{x}\|_2$, which is spanned by the right-singular vectors $\boldsymbol{v}_j, j \leqslant k+1$ of A. Adding up dimensions of these two spaces yields a number greater than n, as there must be a nonzero vector in both spaces. This is a contradiction of the rank-nullity theorem (Theorem 2.24) in Section 2.7.3.

Matrix Approximation

However, there exists a $(k+1)$-dimensional subspace where $\|\boldsymbol{Ax}\|_2 \geqslant \sigma_{k+1}\|\boldsymbol{x}\|_2$, which is spanned by the right-singular vectors $\boldsymbol{v}_j, j \leqslant k+1$ of A. Adding up dimensions of these two spaces yields a number greater than n, as there must be a nonzero vector in both spaces. This is a contradiction of the rank-nullity theorem (Theorem 2.24) in Section 2.7.3.

Matrix Approximation

However, there exists a $(k+1)$-dimensional subspace where $\|\boldsymbol{A}\boldsymbol{x}\|_2 \geqslant \sigma_{k+1}\|\boldsymbol{x}\|_2$, which is spanned by the right-singular vectors $\boldsymbol{v}_j, j \leqslant k+1$ of A. Adding up dimensions of these two spaces yields a number greater than n, as there must be a nonzero vector in both spaces. This is a contradiction of the rank-nullity theorem (Theorem 2.24) in Section 2.7.3.

Matrix Approximation

However, there exists a $(k+1)$-dimensional subspace where $\|\boldsymbol{Ax}\|_2 \geqslant \sigma_{k+1}\|\boldsymbol{x}\|_2$, which is spanned by the right-singular vectors $\boldsymbol{v}_j, j \leqslant k+1$ of A. Adding up dimensions of these two spaces yields a number greater than n, as there must be a nonzero vector in both spaces. This is a contradiction of the rank-nullity theorem (Theorem 2.24) in Section 2.7.3.

Matrix Approximation

The Eckart-Young theorem implies that we can use SVD to reduce a rank-r matrix $\boldsymbol{A}$ to a rank-k matrix $\widehat{\boldsymbol{A}}$ in a principled, optimal (in the spectral norm sense) manner. We can interpret the approximation of $\boldsymbol{A}$ by a rank-k matrix as a form of lossy compression. Therefore, the low-rank approximation of a matrix appears in many machine learning applications, e.g., image processing, noise filtering, and regularization of ill-posed problems. Furthermore, it plays a key role in dimensionality reduction and principal component analysis, as we will see in Chapter 10 .

Matrix Approximation

The Eckart-Young theorem implies that we can use SVD to reduce a rank-r matrix $\boldsymbol{A}$ to a rank-k matrix $\widehat{\boldsymbol{A}}$ in a principled, optimal (in the spectral norm sense) manner. We can interpret the approximation of $\boldsymbol{A}$ by a rank-k matrix as a form of lossy compression. Therefore, the low-rank approximation of a matrix appears in many machine learning applications, e.g., image processing, noise filtering, and regularization of ill-posed problems. Furthermore, it plays a key role in dimensionality reduction and principal component analysis, as we will see in Chapter 10 .

Matrix Approximation

The Eckart-Young theorem implies that we can use SVD to reduce a rank-r matrix $\boldsymbol{A}$ to a rank-k matrix $\widehat{\boldsymbol{A}}$ in a principled, optimal (in the spectral norm sense) manner. We can interpret the approximation of $\boldsymbol{A}$ by a rank-k matrix as a form of lossy compression. Therefore, the low-rank approximation of a matrix appears in many machine learning applications, e.g., image processing, noise filtering, and regularization of ill-posed problems. Furthermore, it plays a key role in dimensionality reduction and principal component analysis, as we will see in Chapter 10 .

Matrix Approximation

The Eckart-Young theorem implies that we can use SVD to reduce a rank-r matrix $\boldsymbol{A}$ to a rank-k matrix $\widehat{\boldsymbol{A}}$ in a principled, optimal (in the spectral norm sense) manner. We can interpret the approximation of $\boldsymbol{A}$ by a rank-k matrix as a form of lossy compression. Therefore, the low-rank approximation of a matrix appears in many machine learning applications, e.g., image processing, noise filtering, and regularization of ill-posed problems. Furthermore, it plays a key role in dimensionality reduction and principal component analysis, as we will see in Chapter 10 .

Matrix Approximation

Example (Example 4.15)

Finding Structure in Movie Ratings and Consumers (continued).

Thanks for your attention!