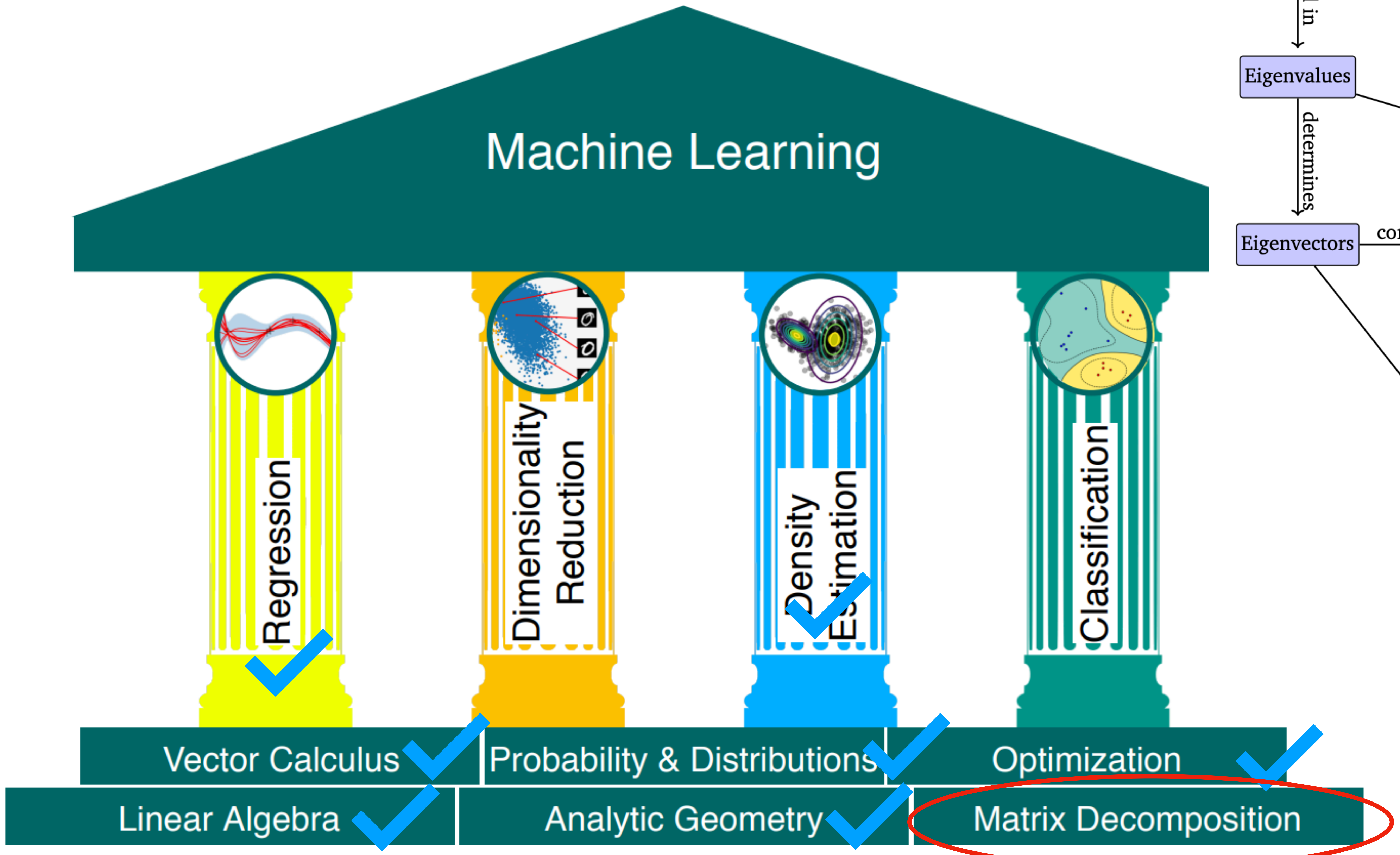
Matrix decomposition [part 2]

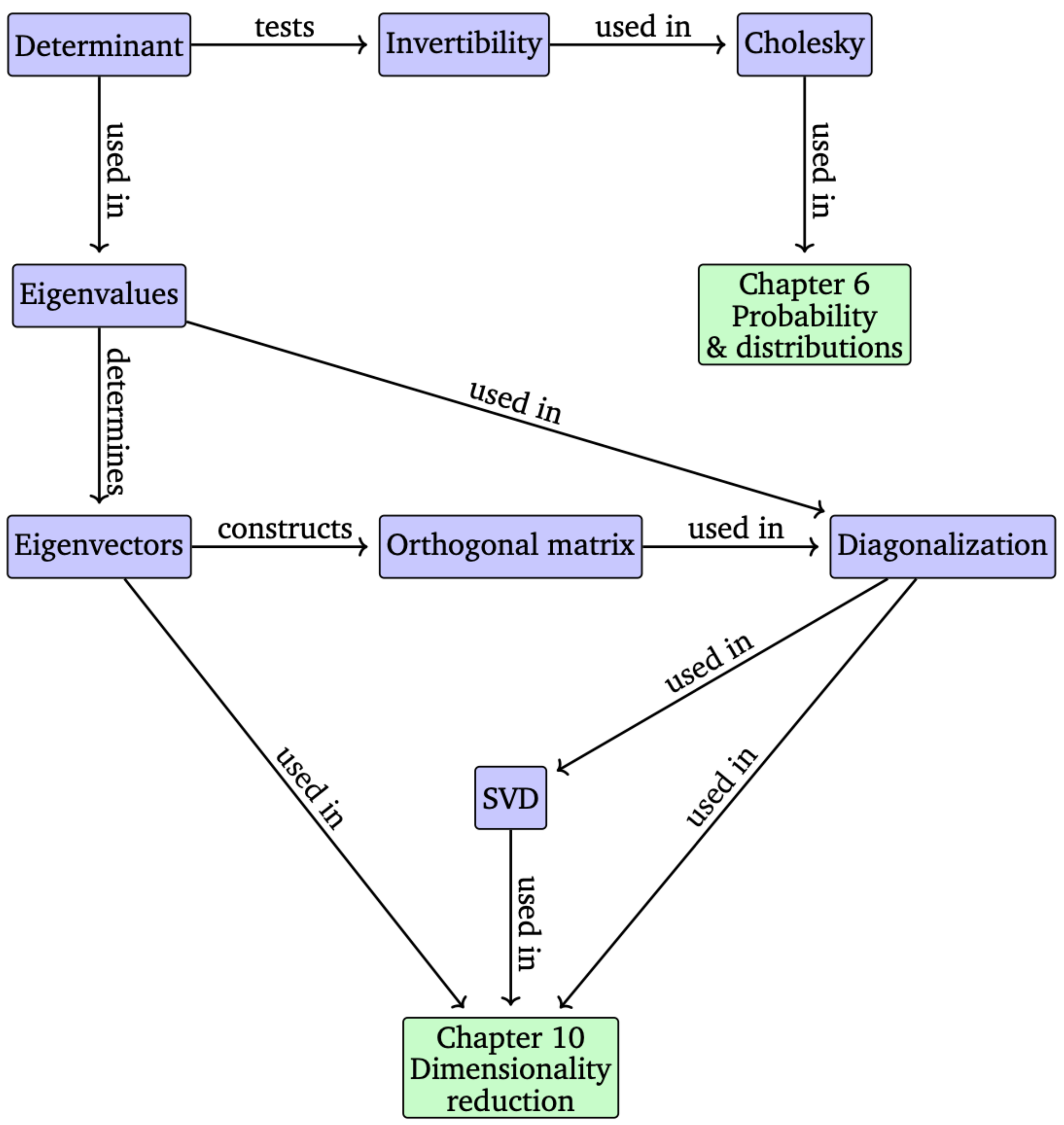
Week 9 - Introduction to ML / Thang Bui / ANU / 2023 S2

Housekeeping

- Assignment 4 will be released later today or tomorrow
- [Centrally invigilated, written] exam timetable is now available
- Calculators must **not** have any of the following features:
  - alpha-numeric keypad [full alphabet on the face];
  - dictionaries;
  - language translators;
  - retrieval or manipulation of text;
  - graphic or word display;
  - sound;
  - pocket organisers; or
  - external communications

Foundations of ML





Today:
Eigendecomposition and SVD

Source: Maths for ML Textbook

Overview

Last lecture:

1. **Trace** and **Determinant**
2. **Eigenvectors** and **eigenvalues**
3. **Symmetric** matrices

This lecture: Decompose/factorise a matrix into a product of matrices

1. **Eigen-decomposition**: using eigenvalues and eigenvectors, for square matrices
2. **Singular Value Decomposition (SVD)**: using singular values and singular vectors, for general matrices

Review and some examples

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 8 & 3 \\ 2 & 7 \end{bmatrix}$

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

Overview

Last lecture:

1. **Trace** and **Determinant**
2. **Eigenvectors** and **eigenvalues**
3. **Symmetric** matrices

This lecture: Decompose/factorise a matrix into a product of matrices

1. **Eigen-decomposition**: using eigenvalues and eigenvectors, for square matrices
2. **Singular Value Decomposition (SVD)**: using singular values and singular vectors, for general matrices

Some background: *similar* matrices

Two matrices A, B are *similar* if there exists an invertible matrix P, such that $B = P^{-1}AP$.

Property: Similar matrices have the same eigenvalues

Some background: *similar* matrices

Two matrices A, B are *similar* if there exists an invertible matrix P, such that $B = P^{-1}AP$.

Property: Similar matrices have the same eigenvalues

Proof: If $Ax = \lambda x$ then $P^{-1}A(PP^{-1})x = P^{-1}\lambda x$, or $B(P^{-1}x) = \lambda(P^{-1}x)$ or $By = \lambda y$

Eigendecomposition: diagonalisable matrices

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

Eigendecomposition: diagonalisable matrices

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

Consider a matrix $P = [p_1, p_2, \ldots, p_n]$, p_i is the i-th column and a diagonal matrix D whose diagonal is $[\lambda_1, \lambda_2, \ldots, \lambda_n]$. We can show that the columns of P are *eigenvectors* of A, and the diagonal of D contains the corresponding *eigenvalues*. See handwritten notes.

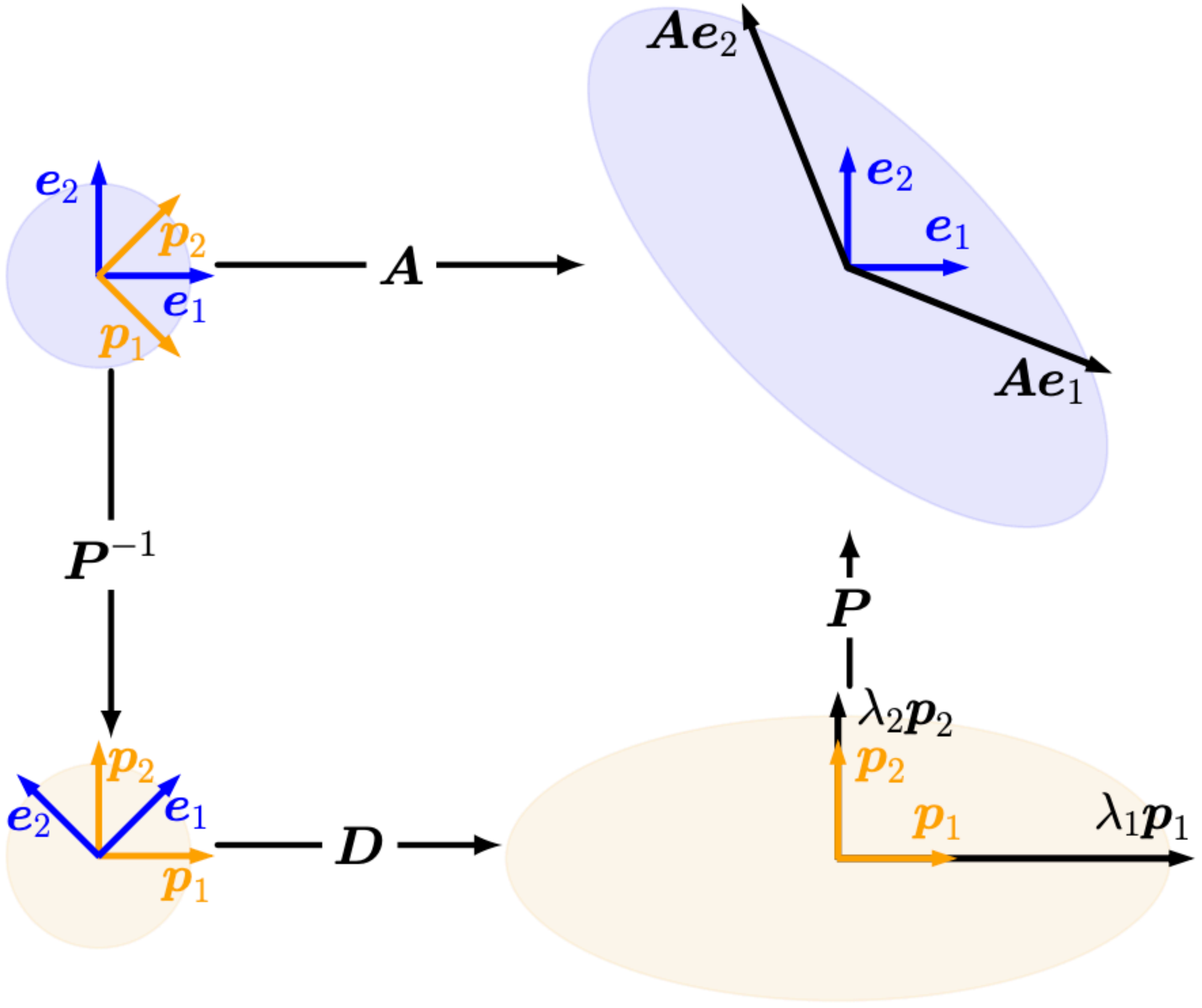
Eigendecomposition: diagonalisable matrices

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

Consider a matrix $P = [p_1, p_2, \ldots, p_n]$, p_i is the i-th column and a diagonal matrix D whose diagonal is $[\lambda_1, \lambda_2, \ldots, \lambda_n]$. We can show that the columns of P are *eigenvectors* of A, and the diagonal of D contains the corresponding *eigenvalues*. See handwritten notes.

Theorem (eigendecomposition) A square matrix $A \in \mathbb{R}^{n \times n}$ can be factored into $A = PDP^{-1}$ where $P \in \mathbb{R}^{n \times n}$ and D is a diagonal matrix whose diagonal entries are the eigenvalues of A, *if and only if* the eigenvectors of A form a basis of $\mathbb{R}^n$ [A has a full set of n linearly independent eigenvectors].

Eigendecomposition: geometric intuition



Special case: symmetric matrices

A square, symmetric matrix $S \in \mathbb{R}^{n \times n}$ is always *diagonalisable.*

The eigenvectors can be chosen orthogonal, and re-scaled to be unit vector so they are orthonormal: $P^\intercal P = I$ and $P^\intercal = P^{-1}$. That is S = $S = PDP^\intercal$ or $D = P^\intercal SP$

Examples

Compute the eigendecomposition of $A = \begin{bmatrix} 8 & 3 \\ 2 & 7 \end{bmatrix}$

Compute the eigendecomposition of symmetric matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

But why eigendecomposition?

Some operations can be performed more efficiently: matrix power A^k, determinant $\det(A)$, matrix exponential (in differential equations)….. See handwritten notes.

Overview

Last lecture:

1. **Trace** and **Determinant**
2. **Eigenvectors** and **eigenvalues**
3. **Symmetric** matrices

This lecture: Decompose/factorise a matrix into a product of matrices

1. **Eigen-decomposition:** using eigenvalues and eigenvectors, for square matrices
2. **Singular Value Decomposition (SVD)**: using singular values and singular vectors, for general matrices

Singular Value Decomposition

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

$$A = U\Sigma V^\intercal = \begin{bmatrix} \vdots & \vdots & & \vdots \\ u_1 & u_2 & \cdots & u_m \\ \vdots & \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & \sigma_r & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \vdots & \vdots & & \vdots \\ v_1 & v_2 & \cdots & v_n \\ \vdots & \vdots & & \vdots \end{bmatrix}^\intercal$$

$U \in \mathbb{R}^{m\times m}$

left singular vectors

$\Sigma \in \mathbb{R}^{m\times n}$

singular values

$V \in \mathbb{R}^{n\times n}$

right singular vectors

U and V are orthogonal matrices, $U^\intercal = U^{-1}, V^\intercal = V^{-1}$. Columns are orthonormal.

By convention, the singular values are ordered $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$

SVD - singular value matrix

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

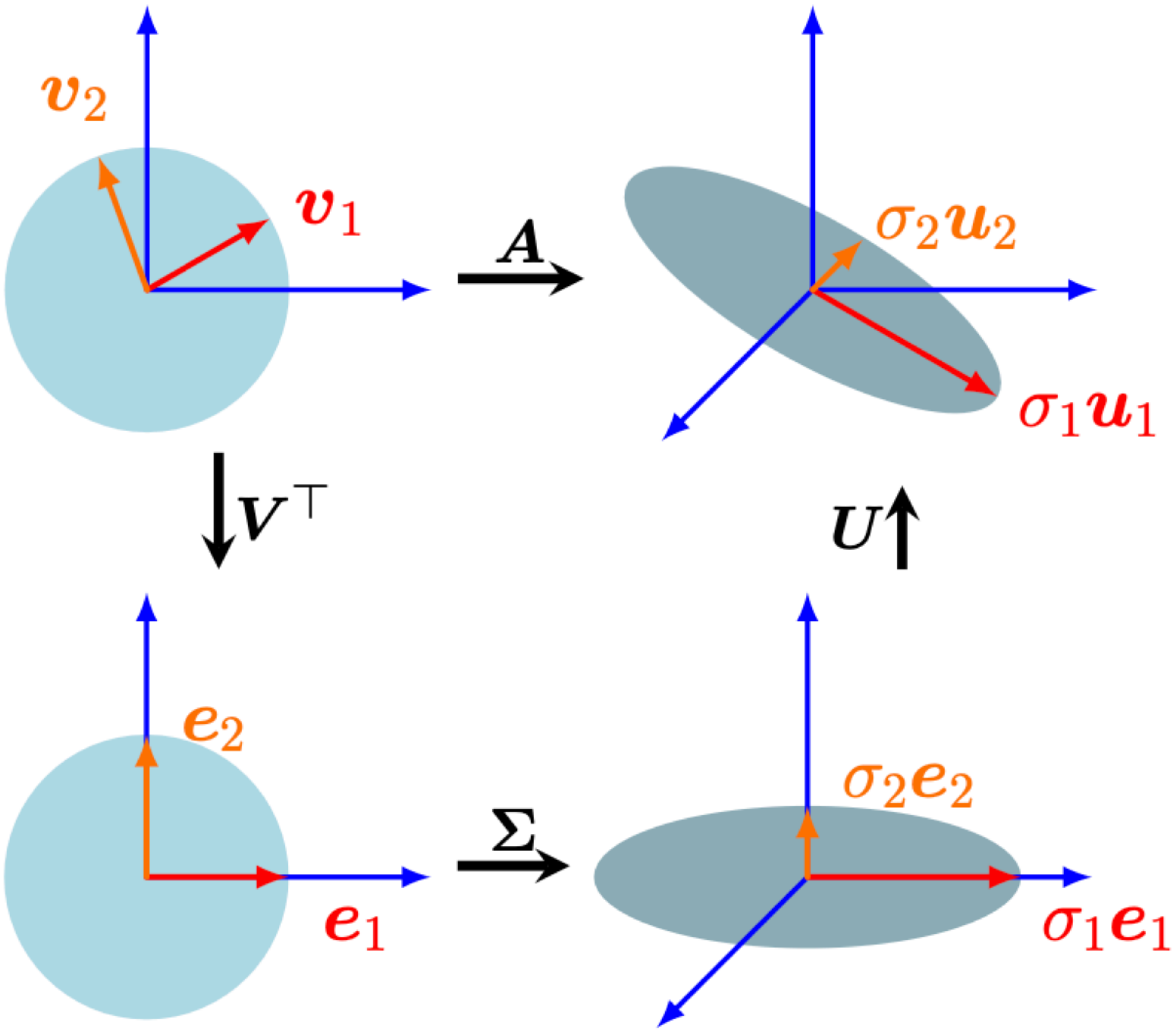
$$n < m$$

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

$$n > m$$

The singular value matrix is unique, and the SVD exists for any matrix $A \in \mathbb{R}^{m \times n}$

SVD: geometric intuition

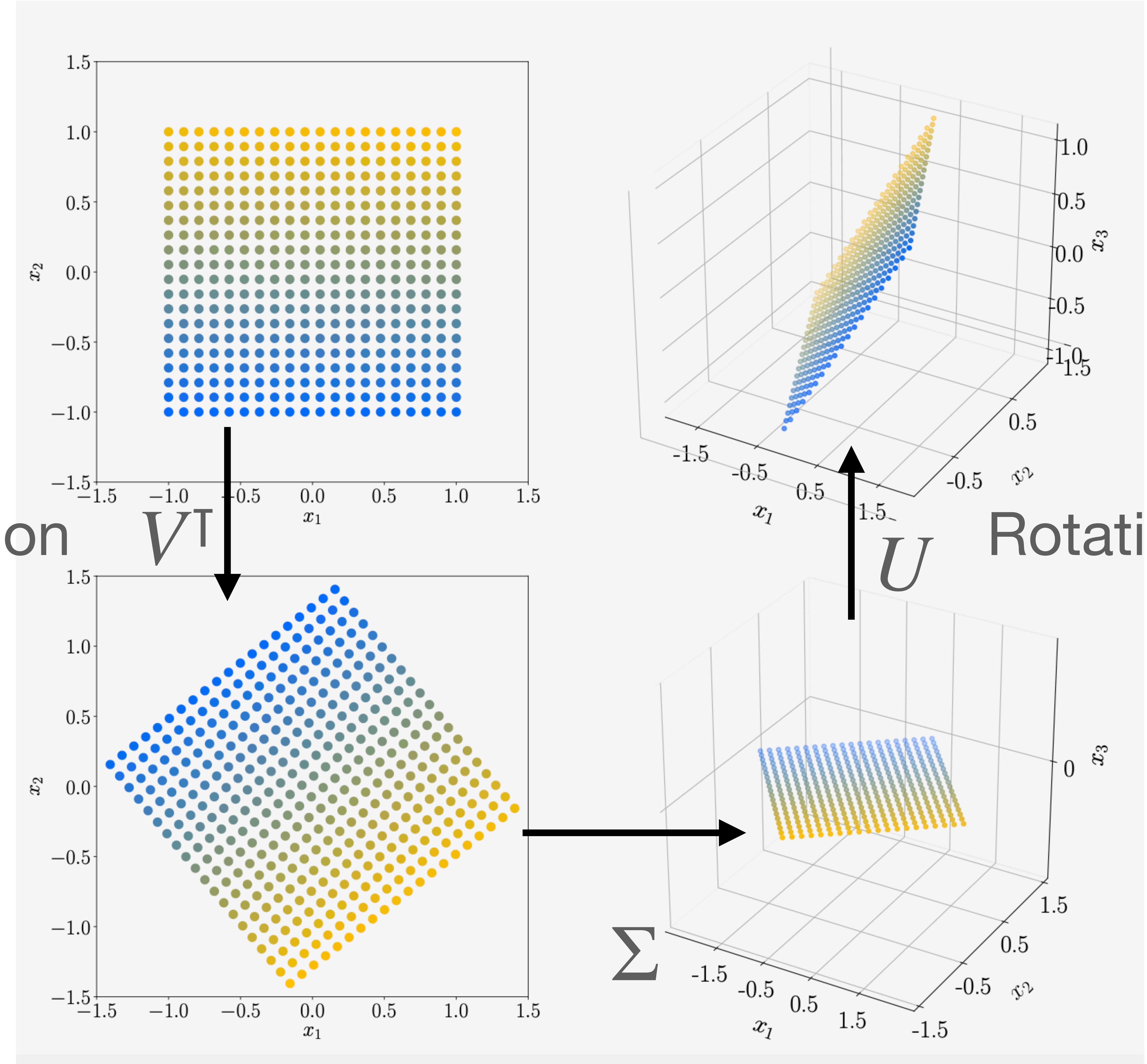


SVD: geometric intuition

Consider a mapping of a square grid of vectors $\mathcal{X} \in \mathbb{R}^2$ that fit in a box of size 2×2 centered at the origin. Using the standard basis, we map these vectors using

$$\boldsymbol{A} = \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top \tag{4.67a}$$

$$= \begin{bmatrix} -0.79 & 0 & -0.62 \\ 0.38 & -0.78 & -0.49 \\ -0.48 & -0.62 & 0.62 \end{bmatrix} \begin{bmatrix} 1.62 & 0 \\ 0 & 1.0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.78 & 0.62 \\ -0.62 & -0.78 \end{bmatrix}. \tag{4.67b}$$



Rotation $V^\top$

Rotation U

Σ Map to codomain + scaling/stretching

SVD construction: finding V and Σ

SVD construction: finding V and Σ

We can always eigen-decompose $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and obtain

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

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

SVD construction: finding V and Σ

We can always eigen-decompose $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and obtain

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

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

Let us assume the SVD of $\boldsymbol{A}$ exists and takes the form of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$

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

SVD construction: finding V and Σ

We can always eigen-decompose $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and obtain

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

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

Let us assume the SVD of $\boldsymbol{A}$ exists and takes the form of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$

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

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}\boldsymbol{V}^{\mathrm{T}}$$

Leading to

$$\boldsymbol{V} = \boldsymbol{P}$$
$$\sigma_i^2 = \lambda_i$$

SVD construction: finding U

Note: $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} \Leftrightarrow \boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{\Sigma}$ which means

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

where r is the rank of $\boldsymbol{A}$. So, we can calculate

$$\boldsymbol{u}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i, i = 1,\ldots,r \quad (1)$$

We look at matrices with full rank, i.e., $r = \min(m, \ n)$. Remember that $\boldsymbol{U}$ is an $m \times m$ matrix.

If $m \leq n$, $\boldsymbol{U} = \left[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\right]$; All the $\boldsymbol{u}_i$ have been calculated through (1)

If $m > n$, $\boldsymbol{U} = \left[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n, \ \ldots, \boldsymbol{u}_m\right]$;

$\boldsymbol{u}_1, \ldots, \ \boldsymbol{u}_n$ have been calculate through (1)

In order to calculate $\boldsymbol{u}_{n+1}, \ldots, \boldsymbol{u}_m$, you use the fact that $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n, \ \ldots, \boldsymbol{u}_m$ are orthonormal vectors.

Examples

Find the SVD of $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$

Find the SVD of $A = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$

Eigendecomposition and SVD [1]

The SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ always exists for any matrix $\mathbb{R}^{m \times n}$. The eigendecomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ 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. On the other hand, the vectors in the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in the SVD are orthonormal, so they represent rotations.

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

A key difference between the eigendecomposition and the SVD is that in the SVD, domain and codomain can be vector spaces of different dimensions

Eigendecomposition and SVD [2]

In the SVD, the left- and right-singular vector matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are generally not inverse of each other (they perform basis changes in different vector spaces). In the eigendecomposition, the basis change matrices $\boldsymbol{P}$ and $\boldsymbol{P}^{-1}$ are inverses of each other.

In the SVD, the entries in the diagonal matrix $\boldsymbol{\Sigma}$ are all real and nonnegative, which is not generally true for the diagonal matrix in the eigendecomposition.

The SVD and the eigendecomposition are closely related through their projections

- The right-singular vectors of $\boldsymbol{A}$ are eigenvectors of $\boldsymbol{A}^{\top}\boldsymbol{A}$.
- The nonzero singular values of $\boldsymbol{A}$ are the square roots of the nonzero eigenvalues of $\boldsymbol{A}^{\top}\boldsymbol{A}$.

For symmetric matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, the eigenvalue decomposition and the SVD are one and the same, which follows from the spectral theorem.

Overview

Last lecture:

1. **Trace** and **Determinant**
2. **Eigenvectors** and **eigenvalues**
3. **Symmetric** matrices

This lecture: Decompose/factorise a matrix into a product of matrices

1. **Eigen-decomposition**: using eigenvalues and eigenvectors, for square matrices
2. **Singular Value Decomposition (SVD)**: using singular values and singular vectors, for general matrices