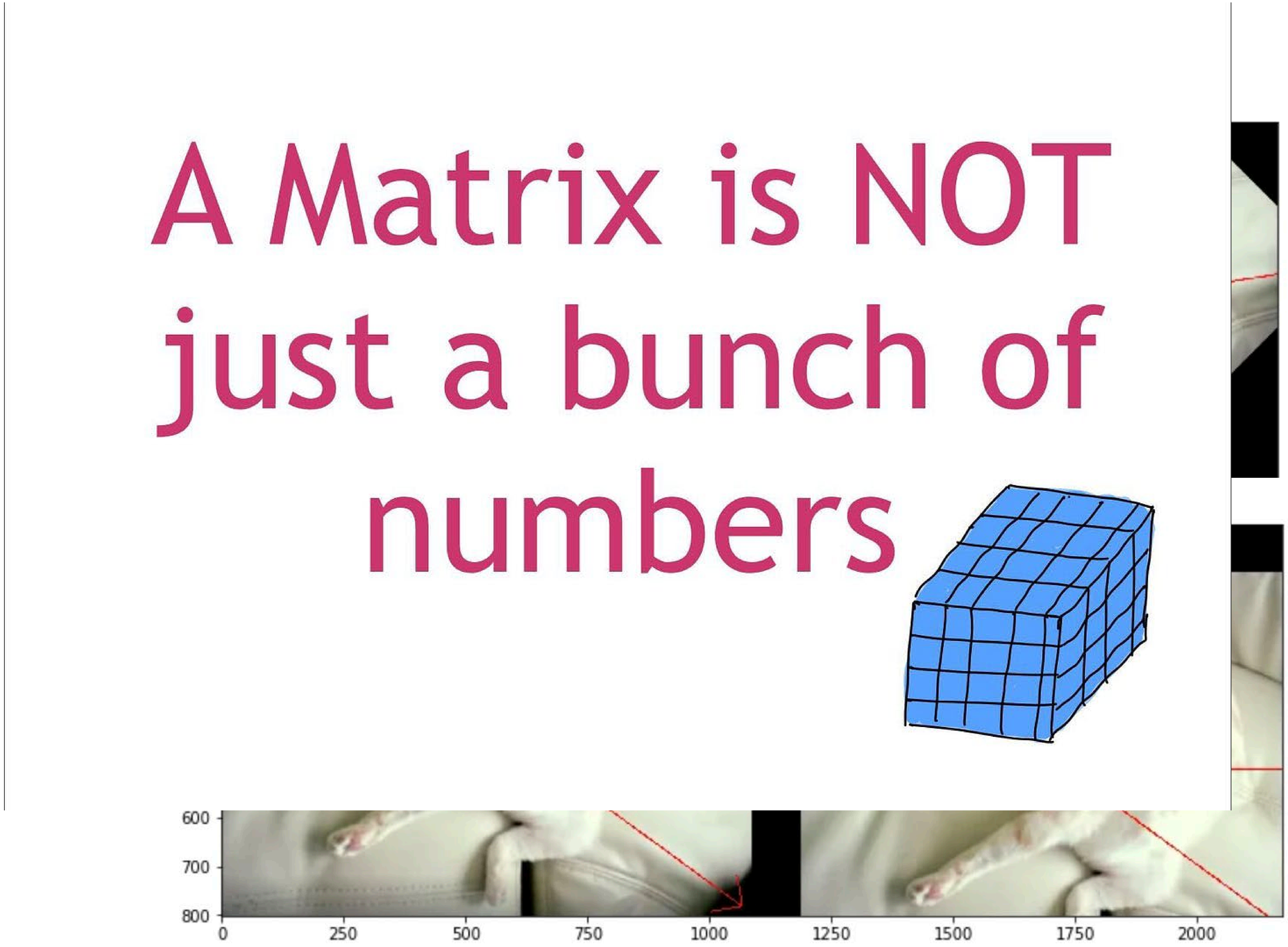
Matrix decomposition

Rahul Shome
based on content by Jo Ciuca and Thang Bui

The essence of a Matrix



The Determinant

- A number associated with a square matrix that essentially “packs” it.

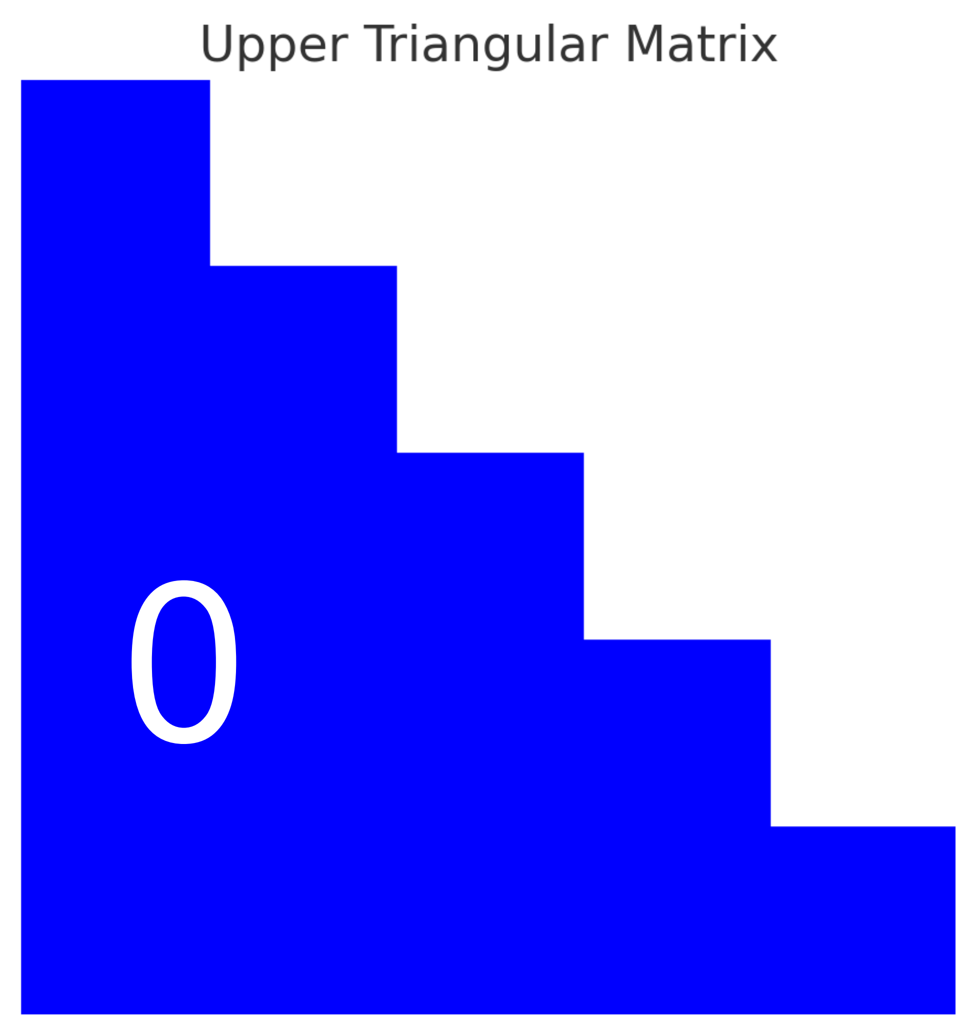
- We write the determinant as $\det(\boldsymbol{A})$ or sometimes as $|\boldsymbol{A}|$ so that

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

- The determinant of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is a function that maps $\boldsymbol{A}$ onto a real number.

Triangular Matrices

- We call a square matrix T an upper-triangular matrix if T_{ij} for $i > j$, i.e., the matrix is zero below its diagonal.
- Analogously, we define a lower-triangular matrix as a matrix with zeros above its diagonal.



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

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

Properties of the determinant

1. $\det(I_n) = 1$

2. Exchanging two rows of a matrix reverses the sign of the determinant.

3. The determinant is a linear function.

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

$$\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$$

The Trace and its properties

The **trace** of a square matrix $A \in \mathbb{R}^{n \times n}$ is the sum of the diagonal elements of A.

$$\mathrm{tr}(A) := \sum_{i=1}^{n} a_{ii}$$

Properties:

$$\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$$

$$\mathrm{tr}(\alpha A) = \alpha \mathrm{tr}(A)$$

$$\mathrm{tr}(I_n) = n$$

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

Eigenvalues and Eigenvectors

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

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

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

- The characteristic polynomial $p_{\boldsymbol{A}}(\lambda) := \det(\boldsymbol{A} - \lambda \boldsymbol{I})$ will allow us to compute eigenvalues and eigenvectors.

Theorem

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

Definition:

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 $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$ is the corresponding eigenvector of $\boldsymbol{A}$ if

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

We call this equation the **eigenvalue equation**.

Matrix we are finding the eigenvector/eigenvalue of

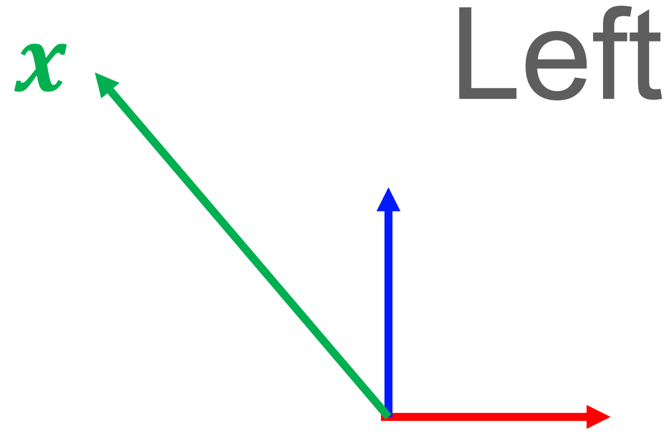
eigenvalue

$$Ax = \lambda x$$

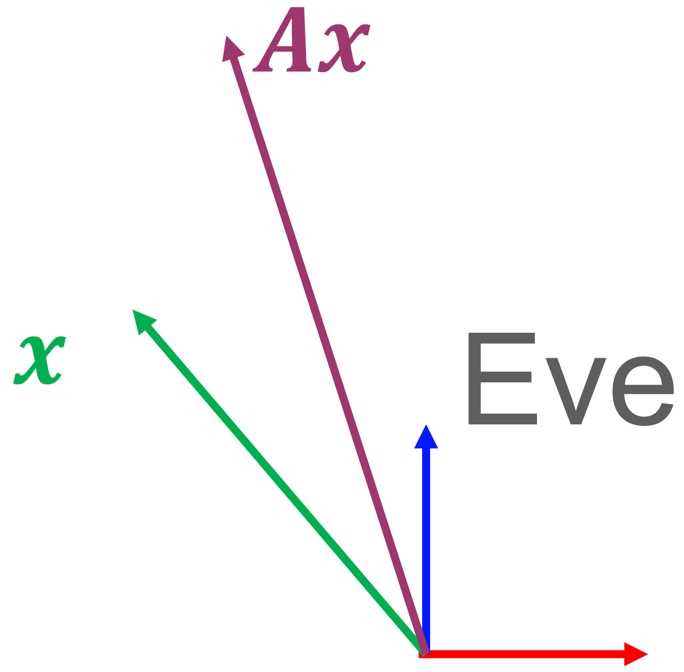
$$(A - \lambda I)x = 0$$

identity matrix

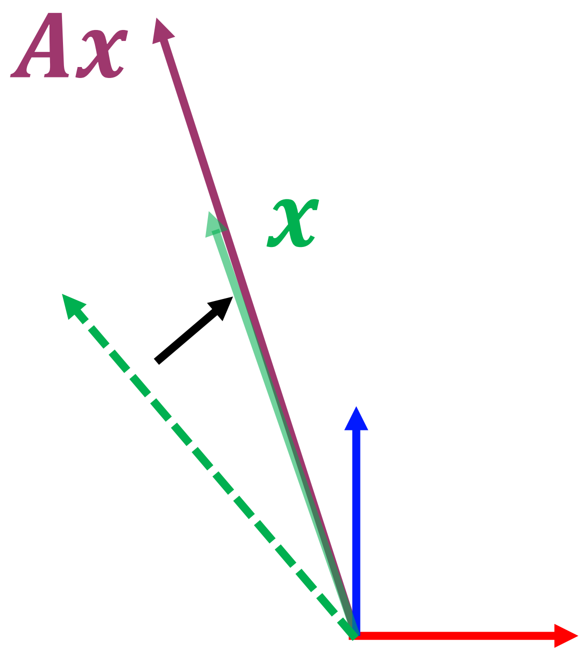
Eigenvalues and Eigenvectors



Left multiplied by a matrix $\boldsymbol{A}$

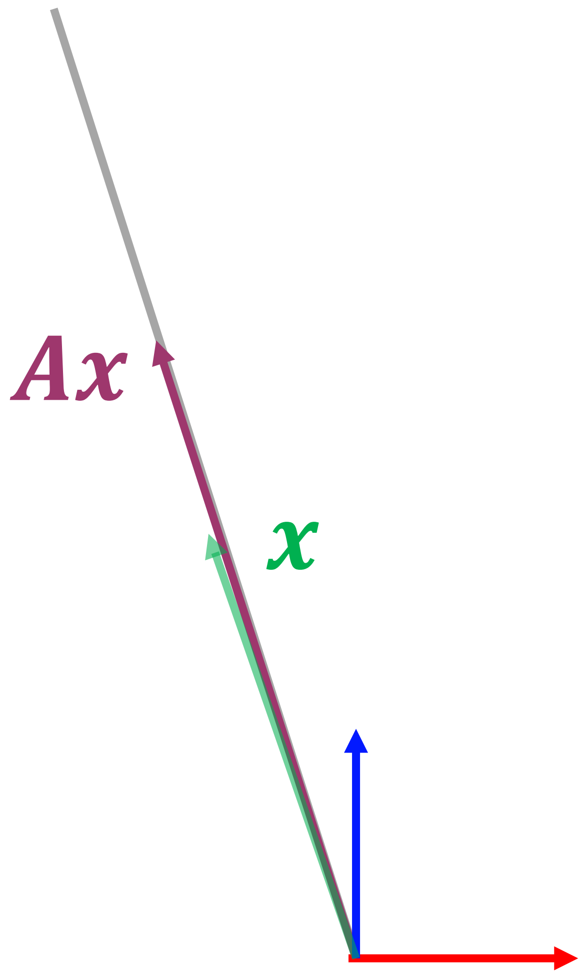


Everything is normal



Now, $\boldsymbol{x}$ and $\boldsymbol{Ax}$ are of the same line. The length of $\boldsymbol{Ax}$ is greater than $\boldsymbol{x}$

$$\boldsymbol{Ax} = \lambda \boldsymbol{x}$$



The grey line is the eigenspace of $\boldsymbol{A}$ with respect to λ

Every vector on this grey line is an eigenvector of $\boldsymbol{A}$, and they all correspond to the eigenvalue λ

Definition:

Let a square matrix 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:

For $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, the union of the $\mathbf{0}$ vector and the set of all eigenvectors of $\boldsymbol{A}$ associated with an eigenvalue λ is a subspace of $\mathbb{R}^{n}$, which is called the eigenspace of $\boldsymbol{A}$ with respect to λ and is denoted by E_{λ}.

The set of all eigenvalues of $\boldsymbol{A}$ is called the eigenspectrum, or just spectrum, of $\boldsymbol{A}$.

If λ is an eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, then the corresponding eigenspace E_{λ} is the solution space of the homogeneous system of linear equations $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = \mathbf{0}$

Example 6: The case of the Identity Matrix

The identity matrix $\boldsymbol{I} \in \mathbb{R}^{n \times n}$ has characteristic polynomial $p_{\boldsymbol{I}}(\lambda) = \det(\boldsymbol{I} - \lambda \boldsymbol{I}) = (1 - \lambda)^n = 0$. It has only one eigenvalue $\lambda = 1$ that occurs n times.

- Moreover, $\boldsymbol{I}\boldsymbol{x} = \lambda \boldsymbol{x}$ holds for all vectors $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$.

- Therefore, the sole eigenspace E_1 of the identity matrix spans n dimensions, and all n standard basis vectors of $\mathbb{R}^n$ are eigenvectors of $\boldsymbol{I}$.

Definition:

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.

- **In our previous example, the geometric multiplicity of $\lambda = 5$ and $\lambda = 2$ is 1.**

- In another example, the matrix $\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ has two repeated eigenvalues $\lambda_1 = \lambda_2 = 2$. The algebraic multiplicity of λ_1 $and\, \lambda_2$ is 2.

- The eigenvalue has only one distinct unit eigenvector $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and thus geometric multiplicity is 1.

Theorem

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

- Eigenvectors of a matrix with n distinct eigenvalues form a basis of $\mathbb{R}^n$.

Definition:

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

- Looking at the eigenspaces of a defective matrix, it follows that the sum of the dimensions of the eigenspaces is less than n.
- A defective matrix cannot have n distinct eigenvalues, as distinct eigenvalues have linearly independent eigenvectors.

Theorem

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

$\boldsymbol{S} \coloneqq \boldsymbol{A}^{\mathrm{T}} \boldsymbol{A}$

The Spectral Theorem

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

Theorems

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

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

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

The trace of a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is the sum of its eigenvalues:

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

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

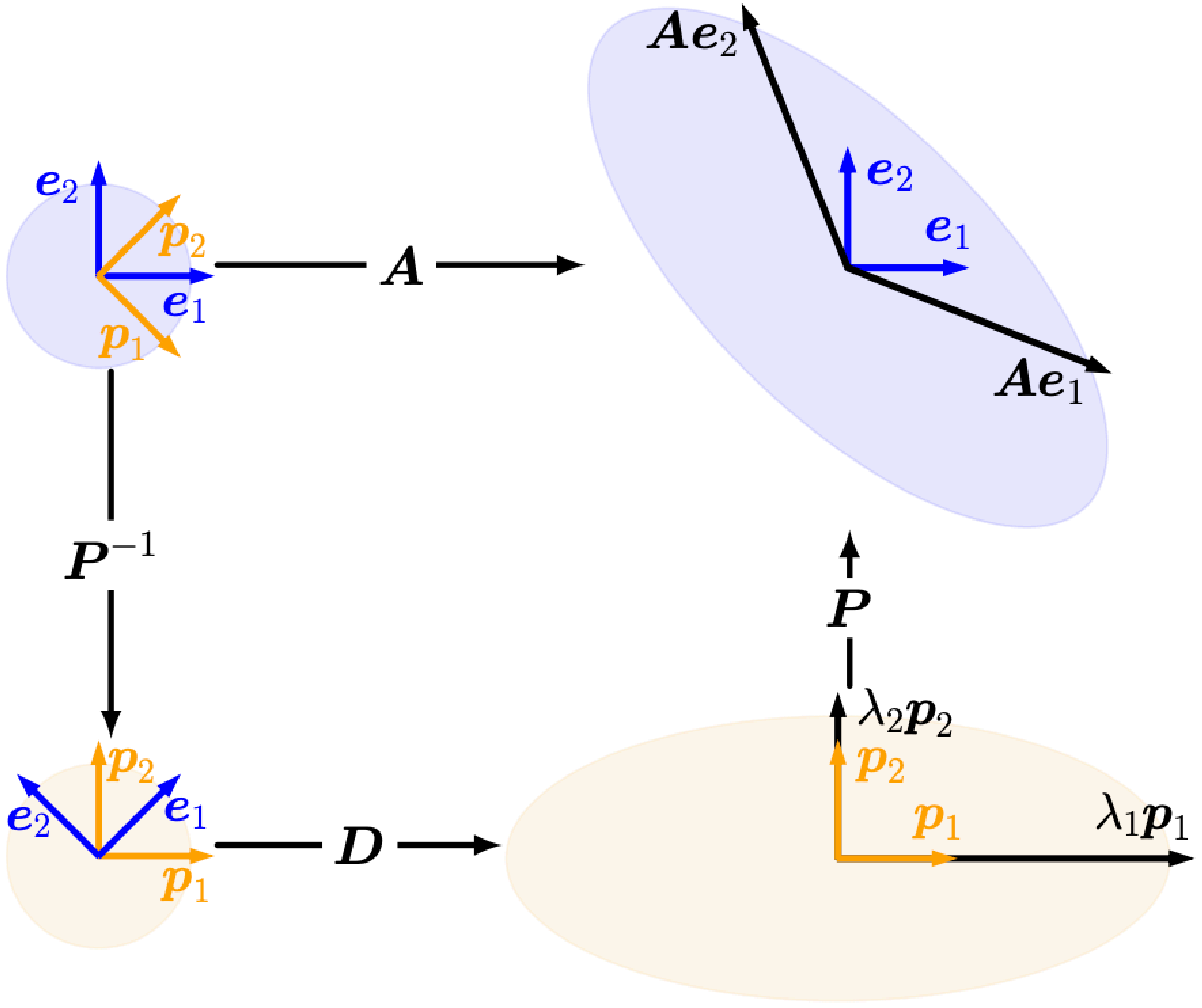
Eigendecomposition: diagonalisable matrices

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 $D = P^{-1}AP$.

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

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 $D = P^\intercal SP$

But why eigendecomposition?

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

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

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

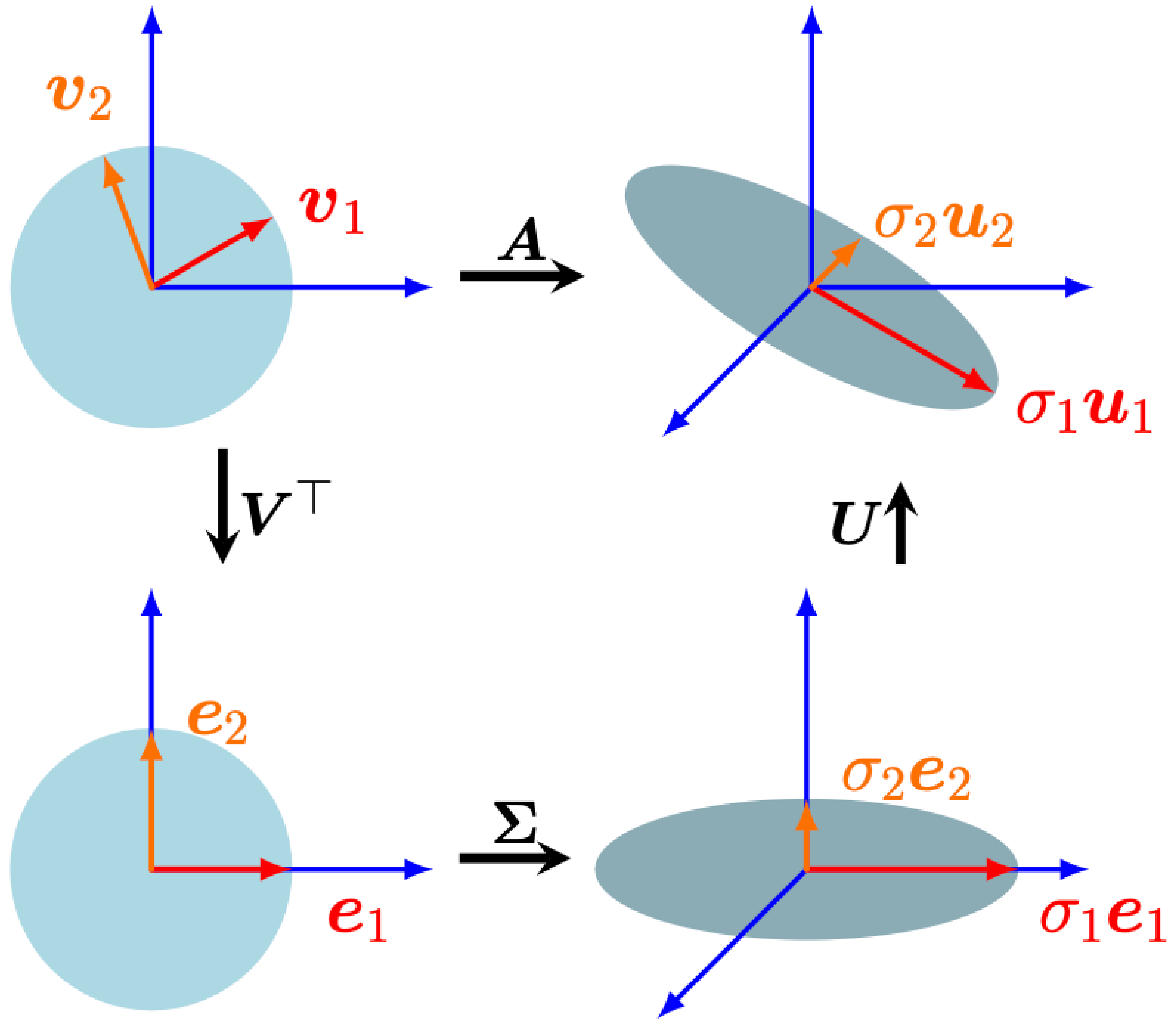
$n < m$

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

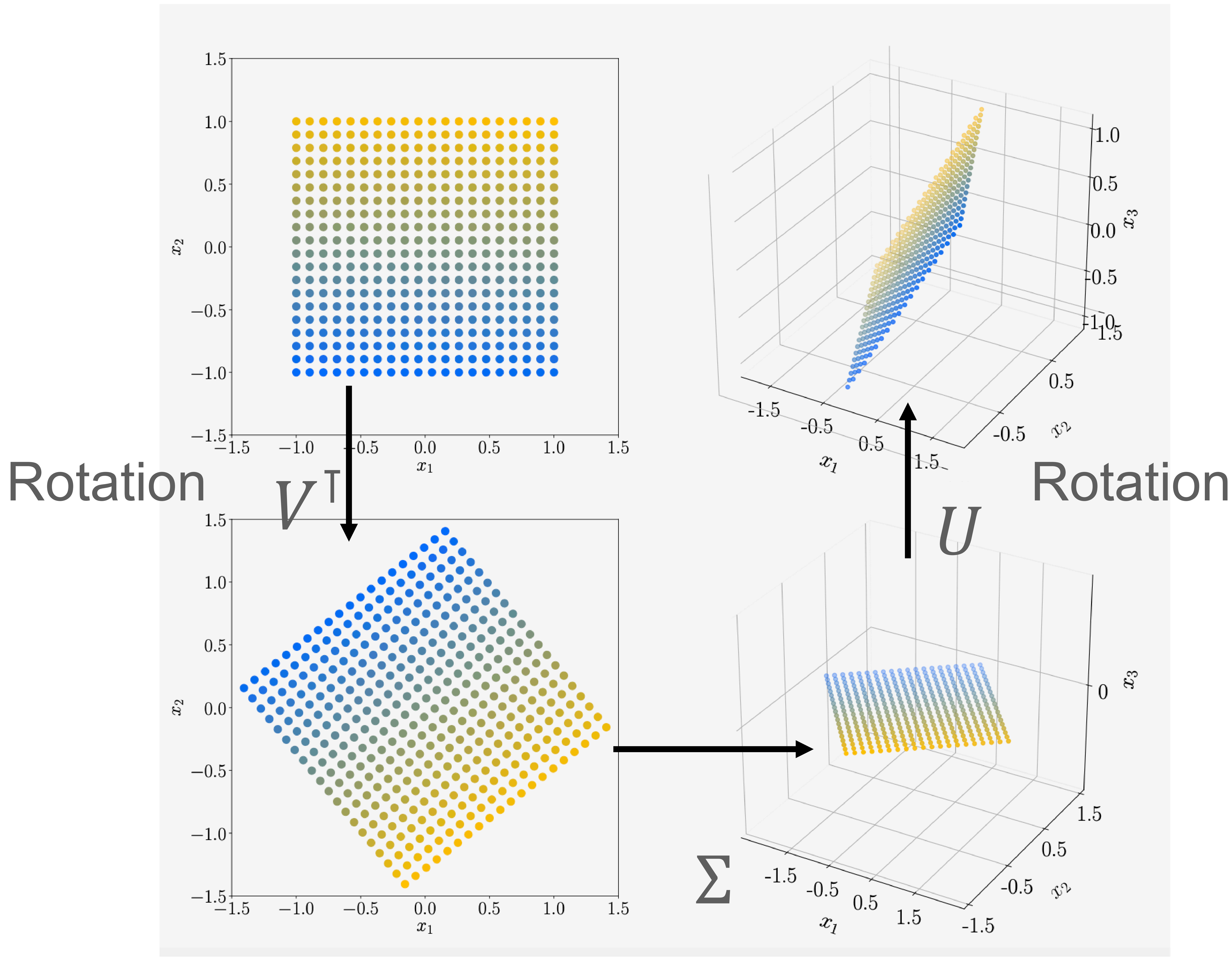
$n > m$

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



Map to codomain + scaling/stretching

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

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

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

SVD construction: finding $V and \Sigma$

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 t

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

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

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

Leading to

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

SVD construction: finding U

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

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

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

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

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

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

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

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

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

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 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.
- The nonzero singular values of $\boldsymbol{A}$ are the square roots of the nonzero eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

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.