Matrix Decompositions

CoE197M/EE298M (Foundations of Machine Learning)

Rowel Atienza, Ph.D.

rowel@eee.upd.edu.ph

Reference: "Mathematics for Machine Learning". Copyright 2020 by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. Published by Cambridge University Press.

Matrix Decompositions

Decomposing a matrix into a product of simpler matrices lets us better understand the data that the matrix represents

For example, by decomposing MNIST images, we understand what makes digits 0 different from 1 to 9 to help us design a better logistic regressor

By decomposing an audio waveform into frequency contents, we understand what makes the sound of “yes” different from “no”

By decomposing connectivity (edges) in a graph neural network, we understand which neurons (nodes) are triggered while an agent is solving a task

Determinant

Given a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, the determinant is $det(\boldsymbol{A}) \in \mathbb{R}$

Use of $det(\boldsymbol{A})$

- Determining the inverse of $\boldsymbol{A}$
- Determining singularity (or invertibility) of $\boldsymbol{A}$

$$det(\boldsymbol{A}) = |\boldsymbol{A}| = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} \in \mathbb{R}$$

Determinant of $\boldsymbol{A} \in \mathbb{R}^{1 \times 1}$

$$det(\boldsymbol{A}) = |a_{11}| = a_{11}$$

Determinant of $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$

$$det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

Determinant of $\boldsymbol{A} \in \mathbb{R}^{3\times 3}$

$$det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

$\boldsymbol{A} \in \mathbb{R}^{3\times 3}$ can be computed by breaking it down into determinants of $\boldsymbol{A}_i \in \mathbb{R}^{2\times 2}$

Color coding shows computation of first term using $a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$

The coefficient of each term is multiplied by $(-1)^{i+j}$. For example, the coefficient of a_{12} is $(-1)^{1+2} = -1$

This method is known as **Laplace Expansion**

Determinant of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ (Laplace Expansion)

For $j = 1,2, \ldots, n$ and $\boldsymbol{A}_{jk}$ is the sub-matrix left after deleting row j and column k

Expansion along column j:

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

Expansion along row j:

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

Exercise: What is $det(\boldsymbol{A})$ if $\boldsymbol{A} \in \mathbb{R}^{4\times 4}$

$$det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

Special Case: Determinant of a Triangular Matrix

Upper Triangular: $\boldsymbol{T} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{(n-1)1} \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}$, $det(\boldsymbol{T}) = \prod_{i=1}^{n} a_{ii}$

Lower Triangular: $\boldsymbol{T} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n1} & \cdots & a_{n(n-1)} & a_{nn} \end{bmatrix}$, $det(\boldsymbol{T}) = \prod_{i=1}^{n} a_{ii}$

Determinant as a Signed Volume

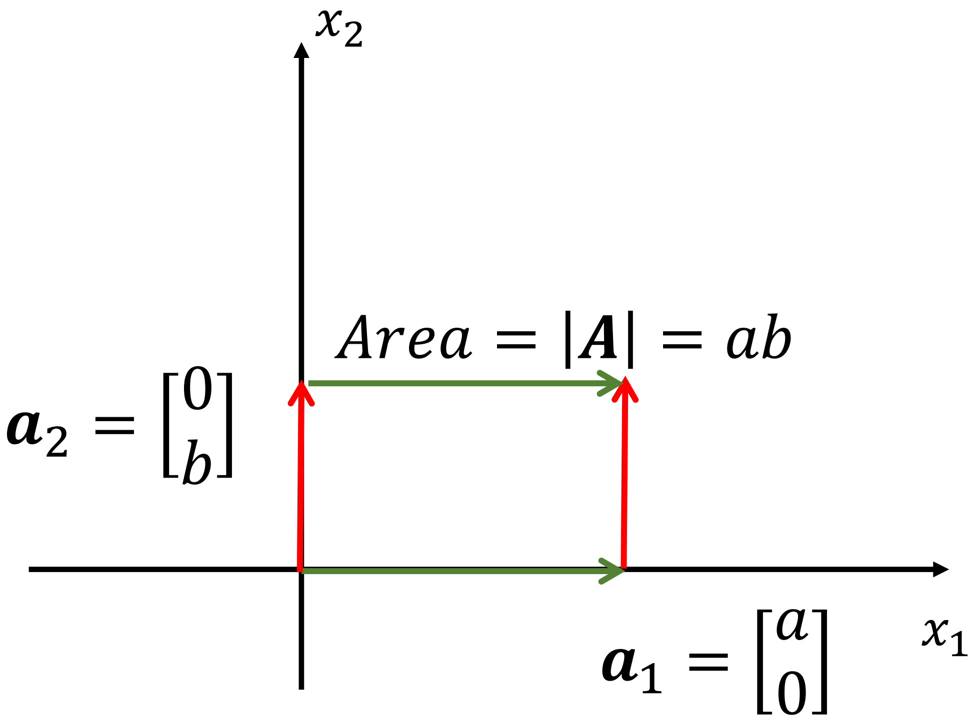
Consider $a > 0$ and $b > 0$:

$$\boldsymbol{A} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} = [\boldsymbol{a}_1 \quad \boldsymbol{a}_2]$$

Determinant:

$$|\boldsymbol{A}| = ab$$

The $Area = |det(\boldsymbol{A})|$ holds true even for non-canonical vectors



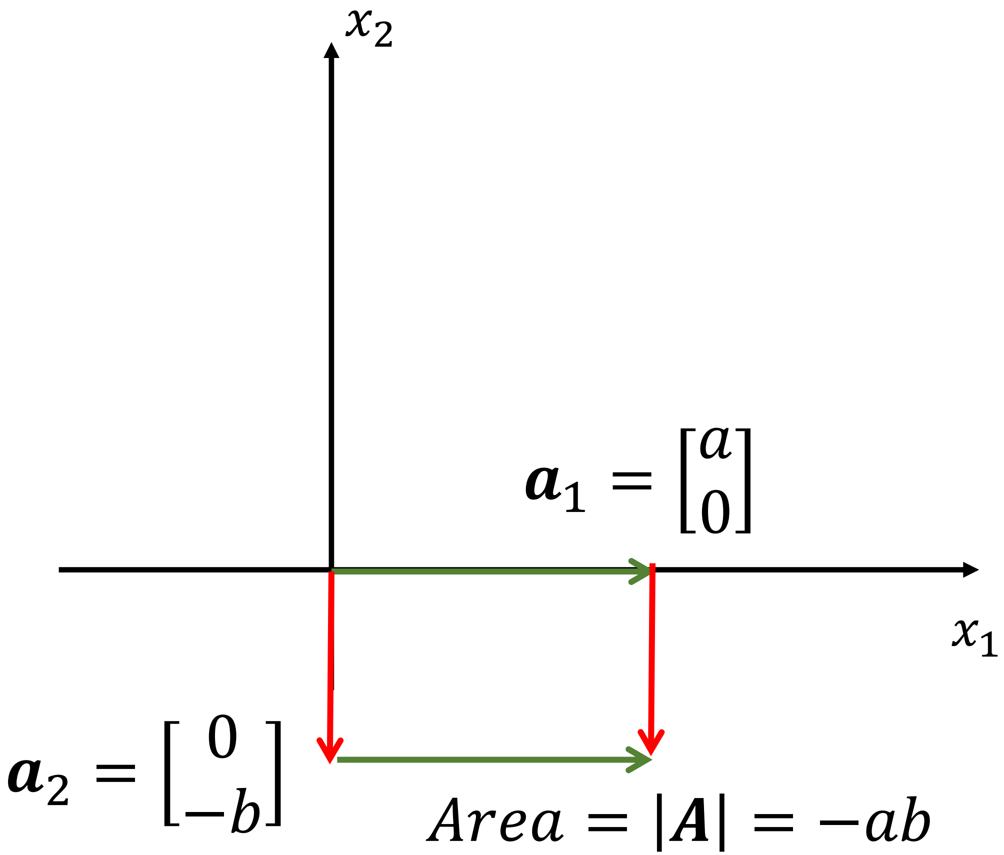
Determinant as a Signed Volume

Consider $a > 0$ and $b > 0$:

$$\boldsymbol{A} = \begin{bmatrix} a & 0 \\ 0 & -b \end{bmatrix} = [\boldsymbol{a}_1 \quad \boldsymbol{a}_2]$$

Determinant:

$$|\boldsymbol{A}| = -ab$$



Determinant as a Signed Volume

Consider:

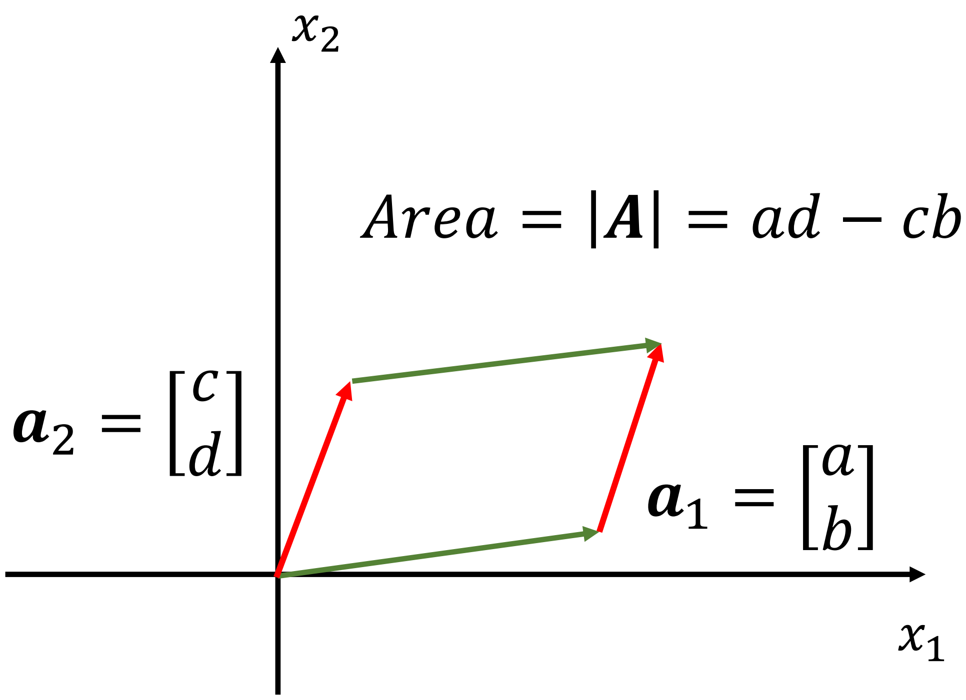
$$\boldsymbol{A} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} = [\boldsymbol{a}_1 \quad \boldsymbol{a}_2]$$

Determinant:

$$|\boldsymbol{A}| = ad - cb$$

Exercise:

Using trigonometric identities, prove that the area of the parallelogram is $|\boldsymbol{A}| = ad - cb$



Determinant as a Signed Volume

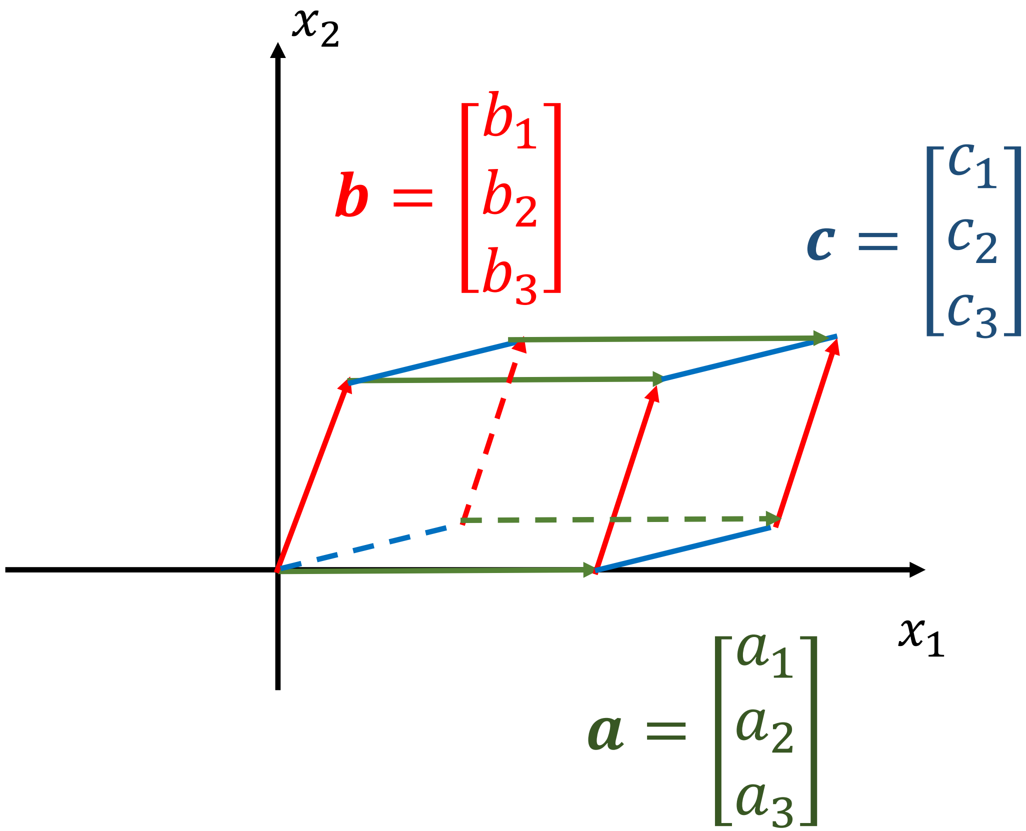
Consider:

$$\boldsymbol{A} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} = [\boldsymbol{a} \quad \boldsymbol{b} \quad \boldsymbol{c}]$$

Determinant:

$$Signed\ Volume = det(\boldsymbol{A}) = |\boldsymbol{A}|$$

The $Volume = |det(\boldsymbol{A})|$ holds true even for non-canonical vectors



$$Vol = |\boldsymbol{A}|$$

Properties of $\boldsymbol{det(A)}$ of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

$det(\boldsymbol{AB}) = det(\boldsymbol{A})\, det(\boldsymbol{B})$

$det(\boldsymbol{A}) = det(\boldsymbol{A}^T)$

If $\boldsymbol{A}$ is invertible:

$$det(\boldsymbol{A}^{-1}) = \frac{1}{det(\boldsymbol{A})}$$

Similar matrices have the same determinant:

$$det\big(\Phi(\boldsymbol{A})\big) = det(\boldsymbol{A})$$

Adding a multiple of a row/col to another does not change the determinant $det(\boldsymbol{A})$

Multiplication of $\boldsymbol{A}$ by $\lambda \in \mathbb{R}$ scales the determinant by λ :

$$det(\lambda\boldsymbol{A}) = \lambda^n det(\boldsymbol{A})$$

Swapping row/col of $\boldsymbol{A}$ changes the sign of $det(\boldsymbol{A})$

Similar Matrices

Two matrices $\boldsymbol{A}, \widetilde{\boldsymbol{A}} \in \mathbb{R}^{n \times n}$, there exists a regular matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ such that:

$$\widetilde{\boldsymbol{A}} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$$

$$det\left(\widetilde{\boldsymbol{A}}\right) = det(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}) = det(\boldsymbol{A})det(\boldsymbol{S}^{-1}\boldsymbol{S}) = det(\boldsymbol{A})$$

Numerical Method in Determining Determinant

Using the properties in the previous slide:

- Use Gaussian Elimination to reduce the matrix into upper triangular form
- Use property $det(\boldsymbol{T}) = \prod_{i=1}^{n} a_{ii}$ to compute the determinant

Example

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & 0 \\ 2 & 0 & 0 & -1 \\ -1 & 1 & 0 & 2 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 2 & -3 \\ 0 & 0 & 0 & -1.5 \end{bmatrix}$$

$$det(\boldsymbol{A}) = 1 \cdot 1 \cdot 2 \cdot -1.5 = -3$$

```
>>> import numpy as np
>>> A = np.array([ [1, 0, -1, 1], [0, 1, 2, 0], [2, 0, 0, -1], [-1, 1, 0, 2]])
>>> A
array([[ 1,  0, -1,  1],
       [ 0,  1,  2,  0],
       [ 2,  0,  0, -1],
       [-1,  1,  0,  2]])
>>> np.linalg.det(A)
-2.9999999999999996
```

Determinant and Rank of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

If $det(\boldsymbol{A}) \neq 0$, then $rank(\boldsymbol{A}) = n$

Trace

Trace of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

Trace of $\boldsymbol{A}$ is the sum of its diagonal elements:

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

Properties of Trace of $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{I} \in \mathbb{R}^{n \times n}$

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

$tr(\alpha \boldsymbol{A}) = \alpha tr(\boldsymbol{A}), \alpha \in \mathbb{R}$

$tr(\boldsymbol{I}) = n$

$\boldsymbol{A} \in \mathbb{R}^{n \times k}, \boldsymbol{B} \in \mathbb{R}^{k \times n}$: $tr(\boldsymbol{AB}) = tr(\boldsymbol{BA})$

$\boldsymbol{A} \in \mathbb{R}^{a \times k}, \boldsymbol{K} \in \mathbb{R}^{k \times l}, \boldsymbol{L} \in \mathbb{R}^{l \times a}$: $tr(\boldsymbol{AKL}) = tr(\boldsymbol{KLA})$

$tr(\boldsymbol{x}\boldsymbol{y}^T) = tr(\boldsymbol{y}^T\boldsymbol{x}) = \boldsymbol{y}^T\boldsymbol{x} \in \mathbb{R}$

Properties of Trace

Let Φ: $V \rightarrow V$ be a linear mapping

If $\boldsymbol{A}$ is used to represent the transformation, then $tr(\Phi) = tr(\boldsymbol{A})$

If $\boldsymbol{B}$ is used to represent the transformation on another basis, then $tr(\Phi) = tr(\boldsymbol{B}) = tr(\boldsymbol{S}^{-1}\boldsymbol{AS}) = tr(\boldsymbol{ASS}^{-1}) = tr(\boldsymbol{A})$

$\therefore$ $tr(\Phi)$ is basis independent

Eigenvalue and Eigenvectors

Eigenvalue and Eigenvectors

Certain vectors respond to certain transformation matrices in such a way that the effect is just a constant scaling

Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\lambda \in \mathbb{R}$. $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ is an eigenvector of $\boldsymbol{A}$ while λ is the corresponding eigenvalue if:

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

Properties of Eigenvalue and Eigenvector

There exists an $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$ such that $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$ for $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\lambda \in \mathbb{R}$ or equivalently $(\boldsymbol{A} - \lambda \boldsymbol{I}_n)\boldsymbol{x} = 0$ can be solved with $\boldsymbol{x} \neq \boldsymbol{0}$

$rank(\boldsymbol{A} - \boldsymbol{I}_n \lambda) < n$

$det(\boldsymbol{A} - \boldsymbol{I}_n \lambda) = 0$

Characteristic Polynomial

Given $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\lambda \in \mathbb{R}$

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

$c_0, c_1, c_2, \dots, c_{n-1}$ are characteristic polynomial of $\boldsymbol{A}$

$c_0 = det(\boldsymbol{A})$

$c_{n-1} = (-1)^{n-1} tr(\boldsymbol{A})$

Non-uniqueness of Eigenvector

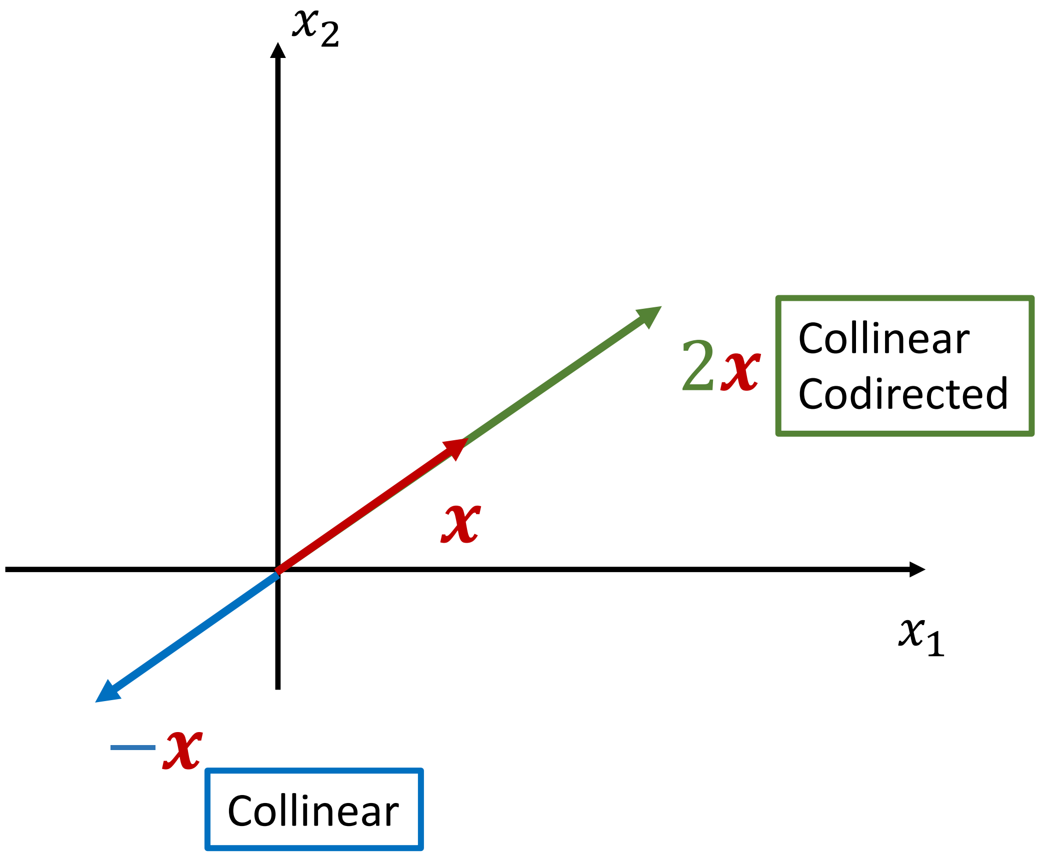
Collinear – 2 vectors are on the same or opposite direction

Codirected – 2 vectors are on the same direction

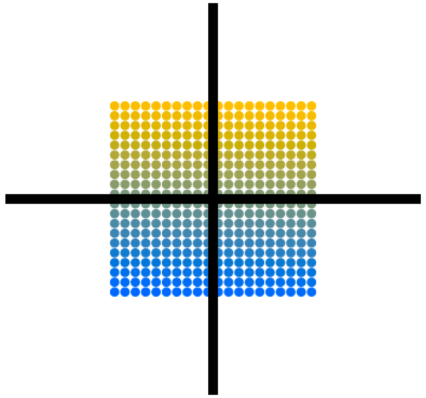
For a given $c \in \mathbb{R}\backslash\{0\}$:

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

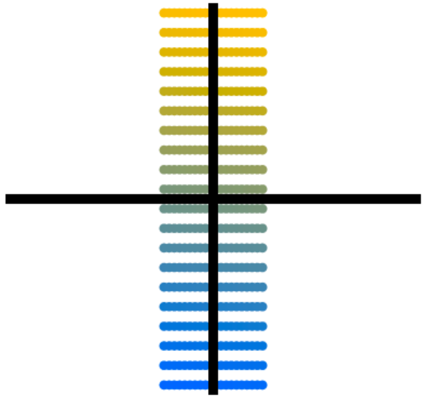
∴ all vectors collinear to $\boldsymbol{x}$ are also eigenvectors; same eigenvalue, different eigenvectors



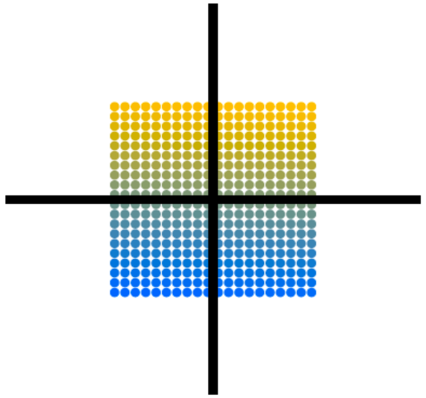
$$A = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}$$



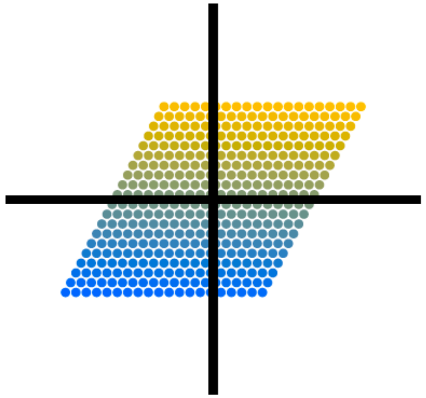
$\lambda_1 = 2.0$
$\lambda_2 = 0.5$
$\det(\boldsymbol{A}) = 1.0$



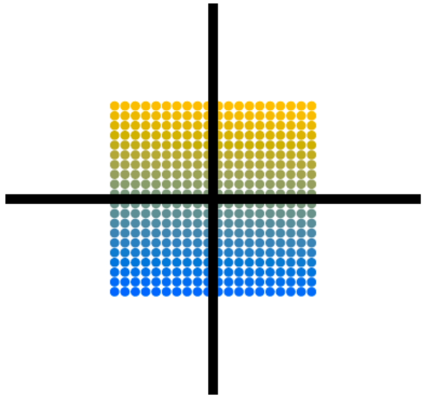
$$A = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$$



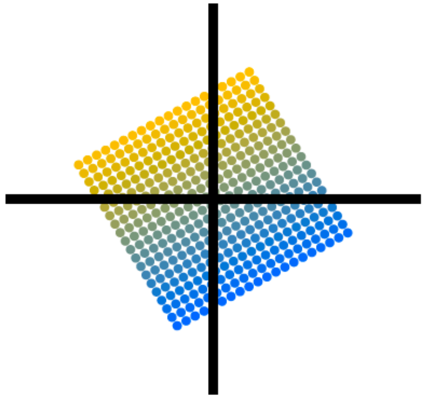
$\lambda_1 = 1.0$
$\lambda_2 = 1.0$
$\det(\boldsymbol{A}) = 1.0$



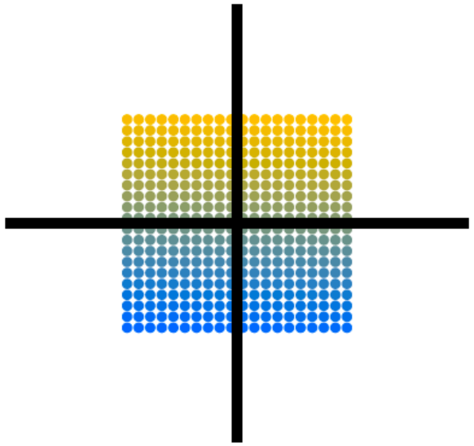
$$A = \begin{bmatrix} \cos\frac{\pi}{6} & -\sin\frac{\pi}{6} \\ \sin\frac{\pi}{6} & \cos\frac{\pi}{6} \end{bmatrix}$$

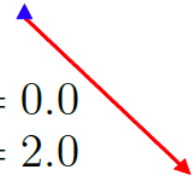


$\lambda_1 = (0.87\text{-}0.5\text{j})$
$\lambda_2 = (0.87+0.5\text{j})$
$\det(\boldsymbol{A}) = 1.0$

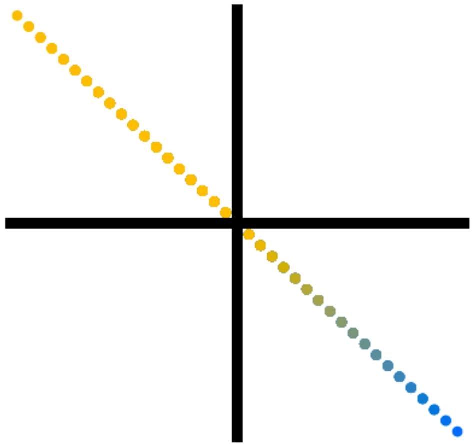


$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

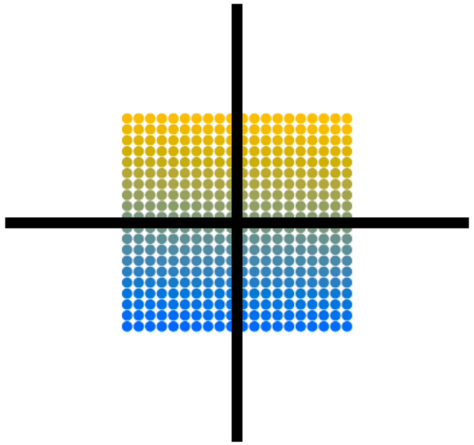


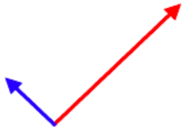


$\lambda_1 = 0.0$
$\lambda_2 = 2.0$
$\det(\boldsymbol{A}) = 0.0$

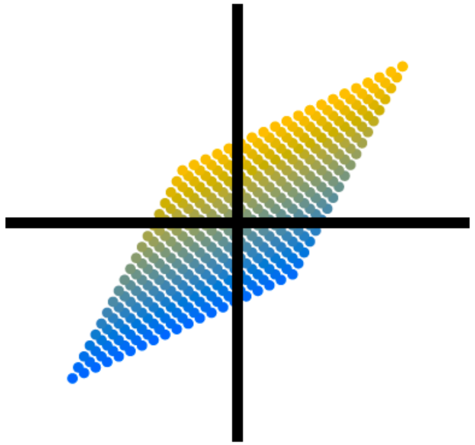


$$A = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$





$\lambda_1 = 0.5$
$\lambda_2 = 1.5$
$\det(\boldsymbol{A}) = 0.75$



Properties of Eigenvalue and Eigenvector

λ is an eigenvalue of $\boldsymbol{A}$ if and only if λ is a root of characteristic polynomial of $\boldsymbol{A}$:

$$p_{\boldsymbol{A}}(\lambda) = det(\boldsymbol{A} - \lambda \boldsymbol{I}) = c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_{n-1}\lambda^{n-1} + (-1)^n\lambda^n = 0$$

The **algebraic multiplicity** of λ as an eigenvalue of $\boldsymbol{A}$ is the number of times it appears as a root in $p_{\boldsymbol{A}}(\lambda)$

Eigenspace and Eigenspectrum

Eigenspace: The set of all eigenvectors, $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m\}$ of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ for a corresponding eigenvalue $\lambda \in \mathbb{R}$ spans a subspace of $\mathbb{R}^n$ called Eigenspace of E_λ

Eigenspectrum: The set of all eigenvalues, $\{\lambda_1, \lambda_2, \dots, \lambda_m\}$, of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

Identity Matrix $\boldsymbol{I} \in \mathbb{R}^{n \times n}$ has $p_{\boldsymbol{A}}(\lambda) = det(\boldsymbol{I} - \lambda \boldsymbol{I}) = (1 - \lambda)^n = 0$

- Solution is λ repeated n times resulting to Eigenspectrum of $\{1\}$ and Eigenspace $E_\lambda = \boldsymbol{x} \in \mathbb{R}^n$

Other Properties

The matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and its transpose $\boldsymbol{A}^T \in \mathbb{R}^{n \times n}$ have the same eigenvalues but not necessarily the same eigenvectors

Null space or Kernel: The Eigenspace is the Null space or Kernel of $(\boldsymbol{A} - \lambda\boldsymbol{I})$ since $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$ or $\boldsymbol{x} \in kernel(\boldsymbol{A} - \lambda\boldsymbol{I})$

Similar matrices have the same eigenvalues. Therefore, under basis change, the following are invariant:

- Determinant
- Trace
- Eigenvalues

Positive definite matrices always have positive real eigenvalues

Multiplicity

The **algebraic multiplicity** of λ as an eigenvalue of $\boldsymbol{A}$ is the number of times it appears as a root in $p_{\boldsymbol{A}}(\lambda)$

The **geometric multiplicity** of λ is the number of independent eigenvectors associated with λ

The dimensionality of space spanned by the eigenvectors of λ

Example

$\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$, solve for the eigenvalues and eigenvectors

$det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \begin{bmatrix} 2-\lambda & 1 \\ 0 & 2-\lambda \end{bmatrix} = 0 \Rightarrow p_{\boldsymbol{A}}(\lambda) = (2-\lambda)(2-\lambda) = 0$

$det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 4 - 4\lambda + \lambda^2, det(\boldsymbol{A}) = 4, -4 = -1^1 Tr(A),$

Eigenvalues: $\lambda_1 = \lambda_2 = 2$, the algebraic multiplicity is 2

Eigenvectors for : $\lambda_1 = \lambda_2 = 2$

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

$\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, the geometric multiplicity is 1

Eigenvalues and Eigenvectors

```
>>> import numpy as np
>>> a = np.array([[2, 1],[0, 2]])
>>> a.shape
(2, 2)
>>> np.linalg.det(a)
4.0
>>> np.linalg.eig(a)
(array([2., 2.]), array([[ 1.0000000e+00, -1.0000000e+00],
       [ 0.0000000e+00,  4.4408921e-16]]))
>>>
```

Recall: Symmetric Positive Definite Matrix

Symmetric Positive Definite Matrix implies:

$$\forall \boldsymbol{x} \in V \backslash \{\mathbf{0}\}, \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > \mathbf{0}$$

Symmetric Positive Semi-Definite Matrix if:

$$\forall \boldsymbol{x} \in V \backslash \{\mathbf{0}\}, \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} \geq \mathbf{0}$$

More Properties

Theorem: The eigenvectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ of matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with distinct eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ are linearly independent

> *Definition*: If there are fewer than n linearly independent eigenvectors, the matrix is called defective

Theorem (SPSD): For a given matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, we can always obtain a symmetric positive semi-definite matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$: $\boldsymbol{S} = \boldsymbol{A}^T\boldsymbol{A}$

> If $rank(\boldsymbol{A}) = n$, then $\boldsymbol{S}$ is a symmetric positive definite (SPD) matrix

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

Determinant and Eigenvalues

Theorem: The determinant of 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 $\lambda_i \in \mathbb{C}$ (complex) and may be repeated eigenvalues of $\boldsymbol{A}$

Trace and Eigenvalues

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

$$Tr(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i$$

where $\lambda_i \in \mathbb{C}$ (complex) and may be repeated eigenvalues of $\boldsymbol{A}$

Cholesky Decomposition

Decomposition as Product of 2 Numbers

In positive real numbers, square root of a number is a useful decomposition. A number is expressed as a product of 2 identical numbers.

- The square root of area of square is the length of its side: $A = s^2$
- The square root of a number greater than 1 is always greater than 1:
  - $n = m^2, n, m > 1$
- The square root of a number less than 1 is always less than 1
  - $n = m^2, n, m < 1$

In positive integers, factorization determines if a number is prime.

- A number is prime if it has only 2 factors: itself and 1

Can we factor $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ as a product of 2 or more matrices?

Cholesky Decomposition

Theorem: A **symmetric positive definite** matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ can be decomposed into a product of 2 matrices $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^T$ where $\boldsymbol{L}$ is a lower triangular matrix with positive diagonal elements:

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

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

Example Cholesky Decomposition

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

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

Example Cholesky Decomposition

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

$$l_{11} = \sqrt{a_{11}}$$

$$l_{21} = \frac{a_{21}}{l_{11}}$$

$$l_{31} = \frac{a_{31}}{l_{11}}$$

$$l_{22} = \sqrt{a_{22} - l_{21}^2}$$

$$l_{32} = \frac{a_{32} - l_{21}l_{31}}{l_{22}}$$

$$l_{33} = \sqrt{a_{33} - l_{31}^2 - l_{32}^2}$$

Cholesky Decomposition: Applications

Numerical computation of determinant of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$:

$$det(\boldsymbol{A}) = det(\boldsymbol{L})det(\boldsymbol{L}^T) = \prod_{i=1}^{n} l_{ii}^2$$

Modelling of covariance matrix of a multi-variate Gaussian which is symmetric and positive definite

Eigendecomposition and Diagonalization

Diagonal Matrices

Determinant, Inverse, and Power are easy to compute for diagonal matrices

Definition: Diagonal Matrix

$$\boldsymbol{D} = \begin{bmatrix} d_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_{nn} \end{bmatrix}$$

Determinant

$$det(\boldsymbol{D}) = \prod_{i=1}^{n} d_{ii}$$

Inverse

$$\boldsymbol{D}^{-1} = \begin{bmatrix} \frac{1}{d_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{d_{nn}} \end{bmatrix}$$

Power

$$\boldsymbol{D}^{k} = \begin{bmatrix} d_{11}^{k} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_{nn}^{k} \end{bmatrix}$$

Diagonal Matrices

Trick: Transform a matrix into a diagonal matrix using change of basis

Definition (Diagonalizable): A matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is diagonalizable if it is similar to a diagonal matrix. There exists an invertible matrix $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$

Eigendecomposition

Theorem (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

$\boldsymbol{D} = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}$ and $\boldsymbol{P} = [\boldsymbol{p}_1 \quad \cdots \quad \boldsymbol{p}_n]$ are linearly independent eigenvectors of $\boldsymbol{A}$ or the basis of $\mathbb{R}^n$

The Matrix $\boldsymbol{P} \in \mathbb{R}^{n\times n}$

Let $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ be the eigenvalues of matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$

Let $\boldsymbol{D} = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}$

Let $\boldsymbol{P} = [\boldsymbol{p}_1 \quad \cdots \quad \boldsymbol{p}_n]$

Then $\boldsymbol{PD} = \boldsymbol{AP} \Longrightarrow [\boldsymbol{p}_1 \quad \cdots \quad \boldsymbol{p}_n]\begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix} = \boldsymbol{A}[\boldsymbol{p}_1 \quad \cdots \quad \boldsymbol{p}_n]$

$\lambda_1\boldsymbol{p}_1 = \boldsymbol{A}\boldsymbol{p}_1, \dots, \lambda_n\boldsymbol{p}_n = \boldsymbol{A}\boldsymbol{p}_n$

$\therefore \boldsymbol{p}_1, \quad \cdots, \quad \boldsymbol{p}_n$ are eigenvectors of $\boldsymbol{A}$ that must be linearly independent so that $\boldsymbol{P}$ is invertible

Example

Given $\boldsymbol{A} = \frac{1}{2}\begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$, perform Eigendecomposition

Let $\boldsymbol{D} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$

Let $\boldsymbol{P} = [\boldsymbol{p}_1 \quad \boldsymbol{p}_2]$

To find the eigenvalues: $det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \begin{vmatrix} \frac{5}{2} - \lambda & -1 \\ -1 & \frac{5}{2} - \lambda \end{vmatrix} = \boldsymbol{0}$

The characteristic polynomial:

$$\left(\frac{5}{2} - \lambda\right)^2 - 1 = 0 \;\; or \;\; \lambda = \frac{7}{2}, \frac{3}{2}$$

$$\therefore \boldsymbol{D} = \begin{bmatrix} \frac{7}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}$$

$$\boldsymbol{A}\boldsymbol{p}_1 = \lambda_1 \boldsymbol{p}_1 \Rightarrow \frac{1}{2}\begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}\boldsymbol{p}_1 = \frac{7}{2}\boldsymbol{p}_1 \Rightarrow \begin{bmatrix} 5-7 & -2 \\ -2 & 5-7 \end{bmatrix}\boldsymbol{p}_1$$

$$\Rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\boldsymbol{p}_1 \Rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\boldsymbol{p}_1 \ by\ GE\ or \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}\boldsymbol{p}_1\ by\ -1\ Trick$$

$\boldsymbol{p}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, similarly $\boldsymbol{p}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\ or\ \boldsymbol{P} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \boldsymbol{P}^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

Check

$$\boldsymbol{A} = \boldsymbol{PDP}^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{7}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{7}{2} & \frac{3}{2} \\ -\frac{7}{2} & \frac{3}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$\boldsymbol{A} = \frac{1}{2} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$$

Symmetric Matrix

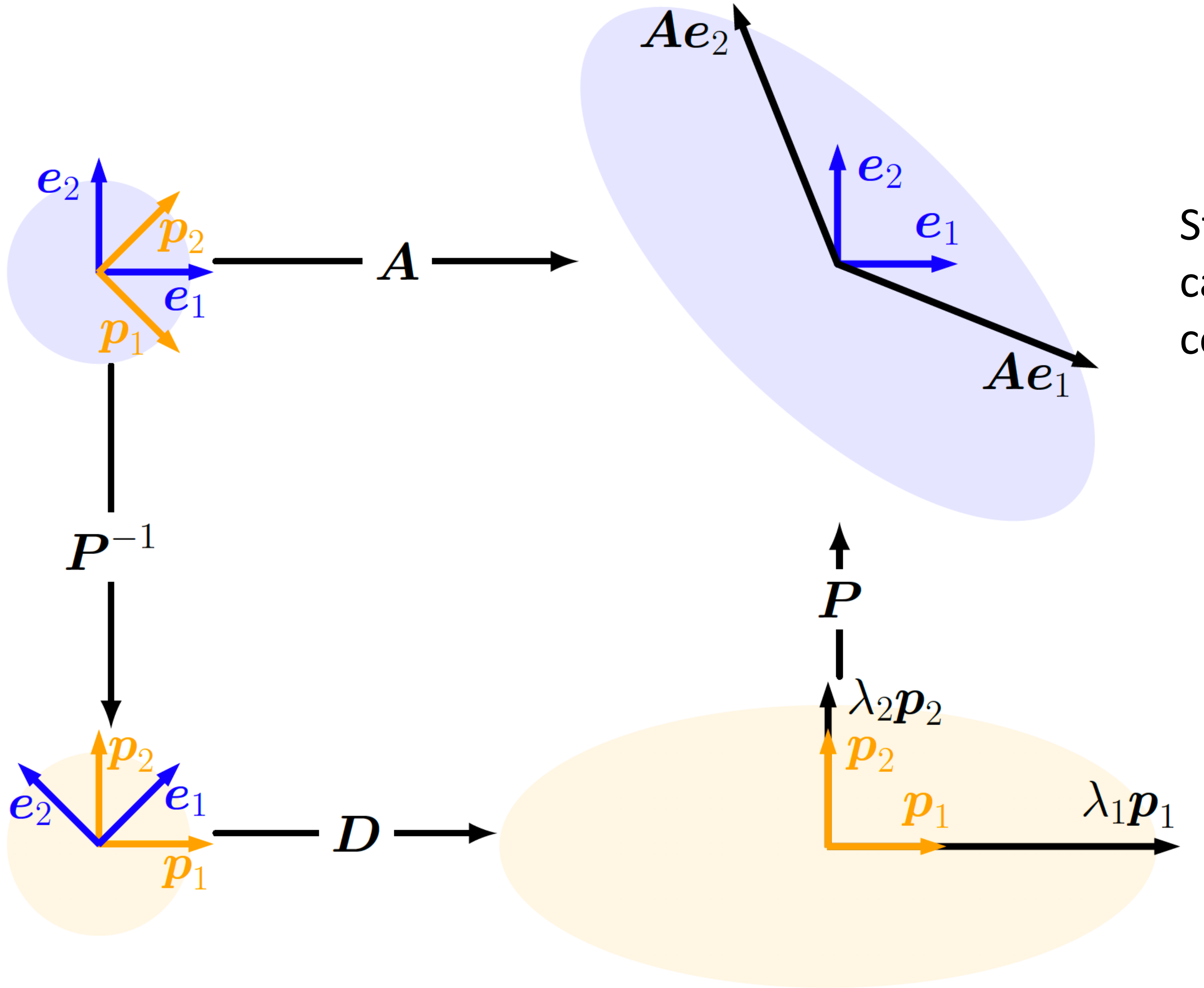
Theorem (Eigendecomposition of Symmetric Matrix): A symmetric matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ can always be diagonalized

Proof: A symmetric matrix has an orthonormal eigenvectors

Implication: $\boldsymbol{P}$ is an orthogonal matrix

Geometric Interpretation

$\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are standard basis vectors
$\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are eigenvectors



Standard canonical coordinates

Eigenvectors are now Eigenbasis

Scaled Eigenbasis by Eigenvalues

Benefits of EigenDecomposition

Power: $\boldsymbol{A}^k = (\boldsymbol{PDP}^{-1})^k = \boldsymbol{PD}^k\boldsymbol{P}^{-1}$

$det(\boldsymbol{A}) = det(\boldsymbol{PDP}^{-1}) = det(\boldsymbol{P})det(\boldsymbol{D})det(\boldsymbol{P}^{-1})$

$$\Rightarrow det(\boldsymbol{P}) \prod_{i=1}^{n} d_{ii} det(\boldsymbol{P}^{-1}) = \prod_{i=1}^{n} d_{ii}\, det(\boldsymbol{P})det(\boldsymbol{P}^{-1})$$

$$\Rightarrow \prod_{i=1}^{n} d_{ii}\, det(\boldsymbol{PP}^{-1}) = \prod_{i=1}^{n} d_{ii} = \prod_{i=1}^{n} \lambda_i$$

Singular Value Decomposition

Singular Value Decomposition (SVD)

Eigendecomposition and Cholesky Decomposition are limited to square matrices

SVD exists for all matrices

SVD

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

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

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{mn} \end{bmatrix} \begin{bmatrix} v_{11} & \dots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \dots & v_{nn} \end{bmatrix}^T$$

$$[\boldsymbol{a}_1 \quad \cdots \quad \boldsymbol{a}_n] = [\boldsymbol{u}_1 \quad \cdots \quad \boldsymbol{u}_m][\boldsymbol{\sigma}_1 \quad \cdots \quad \boldsymbol{\sigma}_n][\boldsymbol{v}_1 \quad \cdots \quad \boldsymbol{v}_n]^T$$

Orthogonal matrix: $\boldsymbol{U} \in \mathbb{R}^{m\times m}$, Diagonal Matrix: $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$, Orthogonal matrix: $\boldsymbol{V} \in \mathbb{R}^{n\times n}$

$\sigma_{ii} \geq 0$ (*singular values*), $\sigma_{ij} = 0$ for $i \neq j$

SVD

$\boldsymbol{u}_i$- left singular vectors

$\boldsymbol{v}_i$- right singular vectors

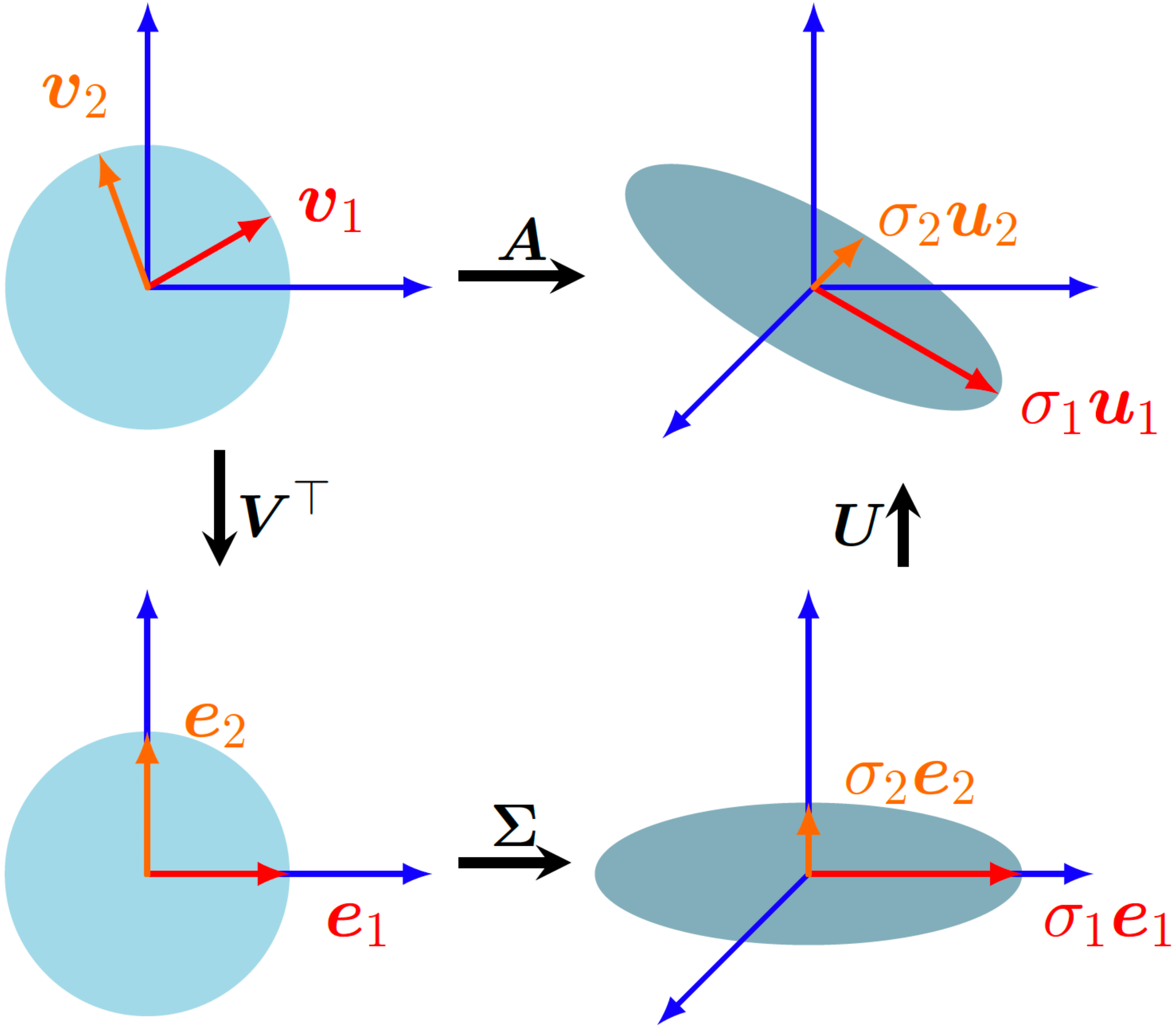
$\boldsymbol{\Sigma}$ – singular matrix

$$\begin{cases} \begin{bmatrix} \sigma_{11} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{nn} \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix} & if\ m > n \\ \begin{bmatrix} \sigma_{11} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{mm} & 0 & \dots & 0 \end{bmatrix} & if\ m < n \end{cases}$$

Geometric Interpretation

$\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are basis vectors in $\mathbb{R}^2$

$\sigma_1\boldsymbol{u}_1$ and $\sigma_2\boldsymbol{u}_2$ in $\mathbb{R}^3$ Change of basis vectors in $\mathbb{R}^3$

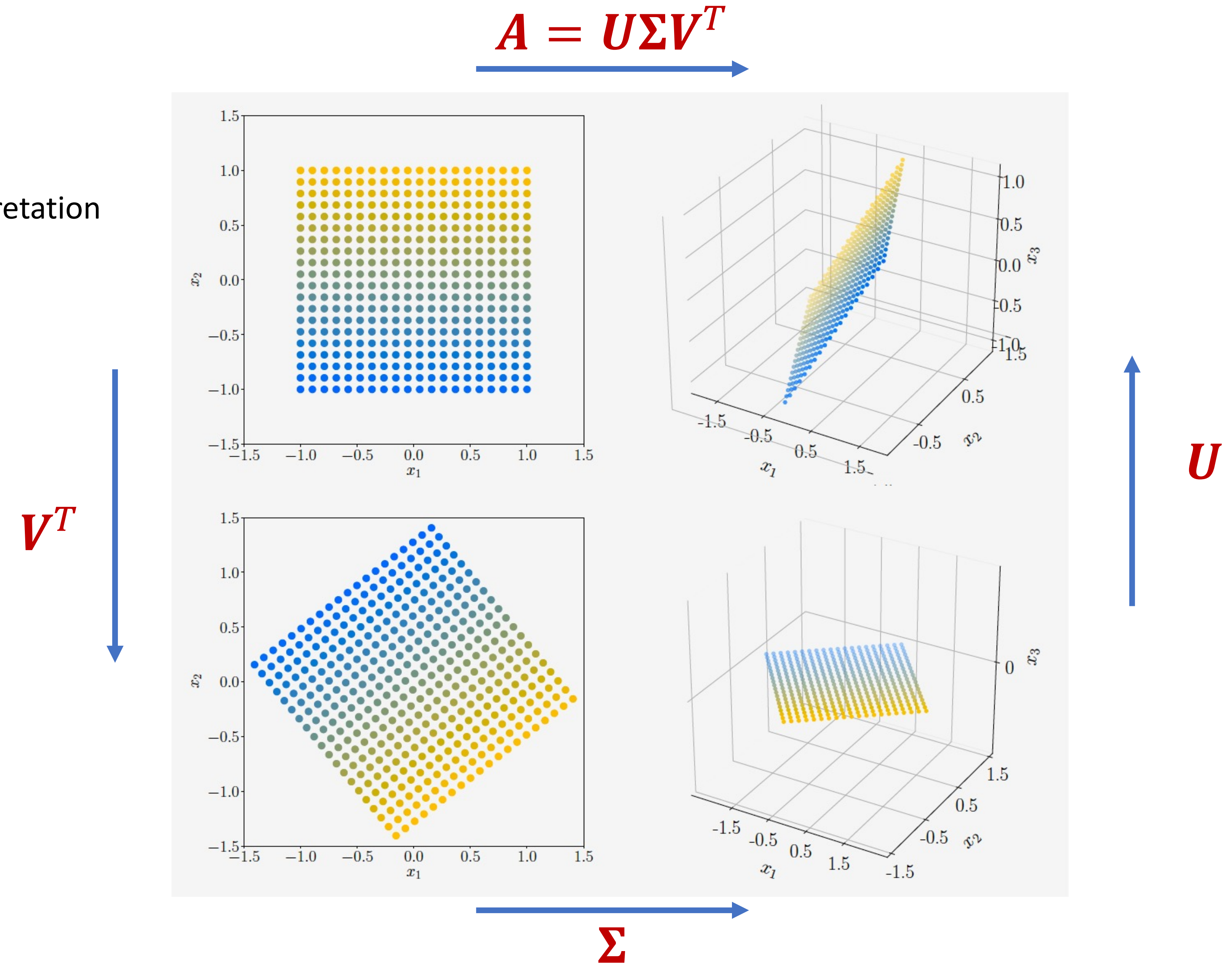


$\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are standard basis vectors in $\mathbb{R}^2$

Scaled $\sigma_1\boldsymbol{e}_1$ and $\sigma_2\boldsymbol{e}_2$ in $\mathbb{R}^3$

SVD

Geometric Interpretation



SVD Algorithm

Consider the Eigendecomposition of a Symmetric Positive Definite (SPD) matrix:

$$\boldsymbol{S} = \boldsymbol{S}^T = \boldsymbol{PDP}^T$$

This is similar in form to:

$$\boldsymbol{S} = \boldsymbol{U\Sigma V}^T$$

In other words, the SVD of an SPD is an EigenDecomposition of $\boldsymbol{S} = \boldsymbol{S}^T$

SVD Algorithm

Given $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, then $\boldsymbol{S} = \boldsymbol{A}^T\boldsymbol{A}$ is a symmetric positive semi-definite matrix by *SPSD Theorem*

By *Spectral Theorem*, there exists an orthonormal basis and $\boldsymbol{S}$ can be diagonalized.

$$\boldsymbol{S} = \boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{PDP}^T = \boldsymbol{P}\begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{bmatrix}\boldsymbol{P}^T$$

$\boldsymbol{P}$ is an orthogonal matrix made of orthonormal eigen basis

SVD Algorithm

Assume that the SVD of $\boldsymbol{A} = \boldsymbol{U\Sigma V}^T$ exists, then:

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

Since $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$

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

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

SVD Algorithm

To obtain $\boldsymbol{U}$, we use the same procedure except we compute for:

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

Since $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$

$$\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{U}\begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^2 \end{bmatrix}\boldsymbol{U}^T$$

$\therefore \boldsymbol{P} = \boldsymbol{U}\ and\ \lambda_i = \sigma_i^2$

SVD Algorithm – Connecting $\boldsymbol{U}$ and $\boldsymbol{V}$

Since the columns of $\boldsymbol{V}$ are orthonormal, for $i \neq j$:

$$(\boldsymbol{A}v_i)^T \boldsymbol{A}v_j = 0$$

$$(\boldsymbol{A}v_i)^T \boldsymbol{A}v_j = {v_i}^T \boldsymbol{A}^T \boldsymbol{A}v_j = {v_i}^T \lambda_j v_j = \lambda_j {v_i}^T v_j = 0$$

SVD Algorithm – Connecting $\boldsymbol{U}$ and $\boldsymbol{V}$

For i:

$$(\boldsymbol{A}\boldsymbol{v}_i)^T\boldsymbol{A}\boldsymbol{v}_i = \sigma_i^2{\boldsymbol{v}_i}^T\boldsymbol{v}_i \Rightarrow \|\boldsymbol{A}\boldsymbol{v}_i\| = \sigma_i\|\boldsymbol{v}_i\| = \sigma_i\sqrt{{\boldsymbol{v}_i}^T\boldsymbol{v}_i} = \sigma_i$$

Furthermore, $\boldsymbol{AV} = \boldsymbol{U\Sigma} \rightarrow \boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{u}_i\sigma_i \; for \; i = 1 \ldots r$

$\boldsymbol{u}_i$ is simply the normalized image of $\boldsymbol{A}\boldsymbol{v}_i$:

$$\boldsymbol{u}_i = \frac{\boldsymbol{A}\boldsymbol{v}_i}{\|\boldsymbol{A}\boldsymbol{v}_i\|} = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i$$

SVD Algorithm – Connecting $\boldsymbol{U}$ and $\boldsymbol{V}$

Therefore,

$$\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i = \boldsymbol{u}_i \sigma_i$$

The above equation holds for $i = 1, \ldots, r$ where $r = \min(m, n)$

If $m > n = r$, we know that for $i > r$, $\boldsymbol{u}_i$ vectors are orthonormal

If $r = m < n$, we know that for $i > r$, $\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{0}$

Therefore,

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{\Sigma} \Rightarrow \boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

Since for orthogonal matrix: $\boldsymbol{V}^{-1} = \boldsymbol{V}^T$

Example: Perform SVD on $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \end{bmatrix}$

We need to solve for $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$: $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\boldsymbol{V}^T$

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

The diagonal elements of $\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}$ are the Eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$

$$det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \begin{bmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{bmatrix} = 0, \lambda_1 = \sigma_1^2 = 3, \lambda_2 = \sigma_2^2 = 1$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Example: Perform SVD on $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \end{bmatrix}$

Solving for $\boldsymbol{V}$:

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_1 = \lambda_1\boldsymbol{v}_1 \Rightarrow \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\boldsymbol{v}_1 = 3\boldsymbol{I}\boldsymbol{v}_1 \Rightarrow \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_2 = \lambda_2\boldsymbol{v}_2 \Rightarrow \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\boldsymbol{v}_2 = \boldsymbol{I}\boldsymbol{v}_2 \Rightarrow \boldsymbol{v}_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$\boldsymbol{V} = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2] = \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} \Rightarrow \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Example: Perform SVD on $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \end{bmatrix}$

Solving for $\boldsymbol{U}$:

$$\boldsymbol{u}_1 = \frac{\boldsymbol{A}\boldsymbol{v}_1}{\sigma_1} = \frac{\begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}}{\sqrt{3}} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{bmatrix}$$

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

Example: Perform SVD on $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & 1 \end{bmatrix}$

Solving for $\boldsymbol{U}$:

$$\boldsymbol{u}_3 = \boldsymbol{u}_1 \times \boldsymbol{u}_2 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \end{bmatrix} \times \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \end{bmatrix}$$

$$\boldsymbol{U} = [\boldsymbol{u}_1 \quad \boldsymbol{u}_2 \quad \boldsymbol{u}_3] = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{bmatrix}$$

```
>>> a = np.array([[1, 0],[-1, 1],[0, 1]])
>>> a.shape
(3, 2)
>>> np.linalg.svd(a)
(array([[-4.08248290e-01, -7.07106781e-01, -5.77350269e-01],
       [ 8.16496581e-01, -5.55111512e-17, -5.77350269e-01],
       [ 4.08248290e-01, -7.07106781e-01,  5.77350269e-01]]),
array([1.73205081, 1.        ]),
array([[-0.70710678,  0.70710678],
       [-0.70710678, -0.70710678]]))
>>> np.sqrt(3)
1.7320508075688772
>>> 1/np.sqrt(3)
0.5773502691896258
>>> 1/np.sqrt(6)
0.4082482904638631
>>> 1/np.sqrt(2)
0.7071067811865475
```

Eigendecomposition vs SVD

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

Exists for square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ with basis eigenvectors of $\mathbb{R}^n$

$\boldsymbol{P}$ vectors are not necessarily orthogonal. Hence, may not represent rotations

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

Exists for any matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$

$\boldsymbol{U}$ and $\boldsymbol{V}^T$ vectors are orthonormal. Hence, they represent rotations

Eigendecomposition vs SVD

$$\boldsymbol{A} = \boldsymbol{PDP}^{-1}$$

Linear Mapping

- Change of basis in the domain
- Independent scaling of new basis.
- Mapping from domain to codomain.
- Change of basis in the codomain

Domain and codomain must have the same dimension

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

Linear Mapping

- Change of basis in the domain
- Independent scaling of new basis.
- Mapping from domain to codomain.
- Change of basis in the codomain

Domain and codomain may have different dimensions

Eigendecomposition vs SVD

$$\boldsymbol{A} = \boldsymbol{PDP}^{-1}$$

$\boldsymbol{P}$ and $\boldsymbol{P}^{-1}$ are inverses of each other

$\boldsymbol{D}$: real or complex eigenvalues

If $\boldsymbol{A}$ is symmetric, the Eigendecomposition is equal to SVD

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

$\boldsymbol{U}$ and $\boldsymbol{V}$ are not necessarily inverses of each other

$\boldsymbol{\Sigma}$: the non-zero entries are real and positive

$\boldsymbol{\Sigma}$: the non-zero eigenvalues are square root of non-zero eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$ which are equal to non-zero eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$

If $\boldsymbol{A}$ is symmetric, the SVD is equal to EigenDecomposition

Matrix Approximation

Low-Rank Approximation of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$

Assume SVD: $\boldsymbol{A} = \boldsymbol{U\Sigma V}^T$

Assume the singular values in $\boldsymbol{\Sigma}$ are sorted in descending order

Rank 1 approximation of $\boldsymbol{A}$:

$$\boldsymbol{A} \approx \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T$$

Rank 2 approximation of $\boldsymbol{A}$:

$$\boldsymbol{A} \approx \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T$$

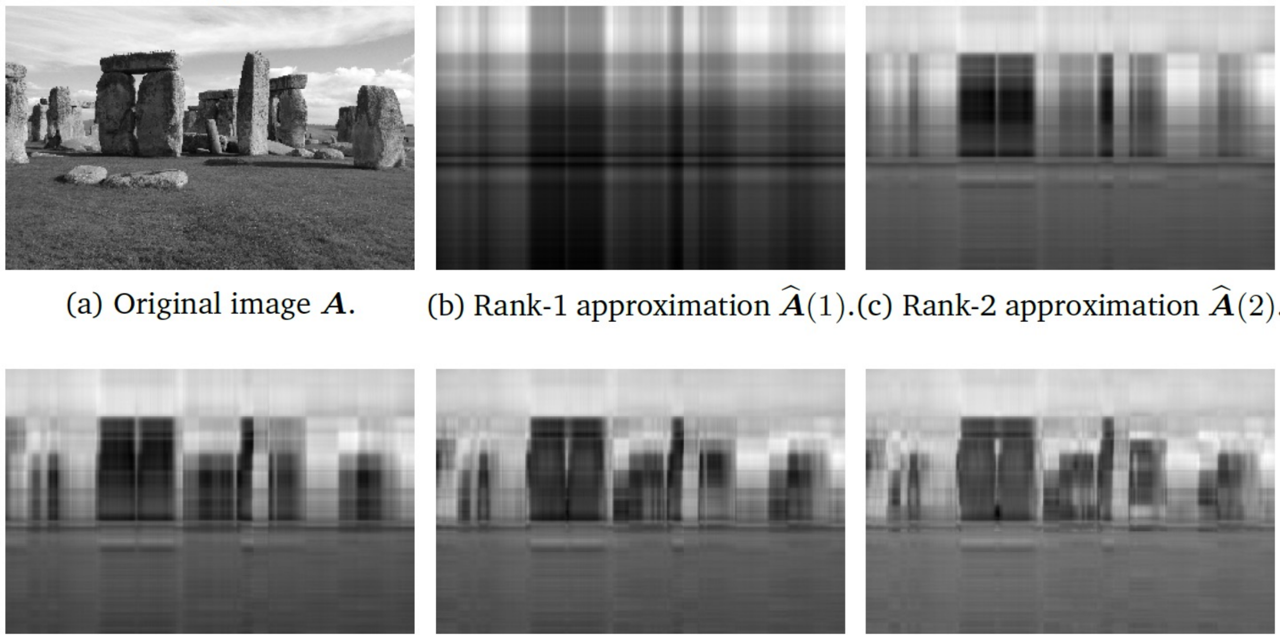
Rank k $(where\ k \leq r = number\ of\ non-zero\ singular\ values)$ approximation of $\boldsymbol{A}$:

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

Application of Low-Rank Approximation

For example, given a 640×480 grayscale image. The total number to represent the image is $307{,}200$

A rank 3 approximation is only $3 \times (640 + 480) + 3 = 3{,}363$ which is just 1% of the original size



(a) Original image $\boldsymbol{A}$. (b) Rank-1 approximation $\widehat{\boldsymbol{A}}(1)$. (c) Rank-2 approximation $\widehat{\boldsymbol{A}}(2)$.

(d) Rank-3 approximation $\widehat{\boldsymbol{A}}(3)$. (e) Rank-4 approximation $\widehat{\boldsymbol{A}}(4)$. (f) Rank-5 approximation $\widehat{\boldsymbol{A}}(5)$.

Distance/Norm of Low-Rank Approximation

Given $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, distance or norm measures how far is the low-rank approximation $\widehat{\boldsymbol{A}}(k)$ from $\boldsymbol{A}$

Definition (Spectral Norm of a Matrix): For $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$, the spectral norm of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is:

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

Theorem (Spectral Norm): The spectral norm of $\boldsymbol{A}$ is its largest singular value σ_1.

Distance/Norm of Low-Rank Approximation

Theorem (Eckart-Young): Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ of rank r and any matrix $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ of rank k. For any $k \leq r$ with $\widehat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$, it holds that:

$$\widehat{\boldsymbol{A}}(k) = \underset{rank(\boldsymbol{B})=k}{\operatorname{argmin}} \|\boldsymbol{A} - \boldsymbol{B}\|_2,$$

$$\left\|\boldsymbol{A} - \widehat{\boldsymbol{A}}(k)\right\|_2 = \sigma_{k+1}$$

Eckart-Young Theorem

We can justify $\left\| \boldsymbol{A} - \widehat{\boldsymbol{A}}(k) \right\|_2 = \sigma_{k+1}$ since:

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

Eckart-Young Theorem shows that $\widehat{\boldsymbol{A}}(k)$ is an optimal low-rank approximation of $\boldsymbol{A}$

In Summary

Determinants are signed volume of matrices

Matrix decomposition helps in the interpretability of matrices

Matrix approximation is useful in signal compression/approximation

SPECTRAL NORMALIZATION FOR GENERATIVE ADVERSARIAL NETWORKS, Miyato et al ICLR 2018