LINEAR ALGEBRA FOR DATA SCIENCE

ISEP 1ère année

2023-2024

SESSION 12th - December 22nd 2023

Cours partially based on … see the References

Outline

- Linear Algebra Basics
- Eigenvalue Decomposition
- Singular Value Decomposition (SVD)*

Outline

- Linear Algebra Basics
- Eigenvalue Decomposition
- Singular Value Decomposition (SVD)*

Basic concepts

- Vector in R^n is an ordered set of n real numbers.
  - e.g. v = (1,6,3,4) is in R^4
  - A column vector: $\begin{pmatrix} 1 \\ 6 \\ 3 \\ 4 \end{pmatrix}$
  - A row vector: $\begin{pmatrix} 1 & 6 & 3 & 4 \end{pmatrix}$
- m-by-n matrix is an object in $R^{m \times n}$ with m rows and n columns, each entry is filled with a (typically) real number:

$$\begin{pmatrix} 1 & 2 & 8 \\ 4 & 78 & 6 \\ 9 & 3 & 2 \end{pmatrix}$$

Basic concepts

Vector norms: A norm of a vector $\|x\|$ is informally a measure of the "length" of the vector.

- The p-norm $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$

- Common norms: L_1, L_2 (Euclidean)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

- L_{infinity}

$$\|x\|_\infty = \max_i |x_i|$$

Basic concepts

We will use lower case letters for vectors. For instance, we denote the vector $\boldsymbol{x}$ and its elements x_i.

- Vector dot (inner) product:

$$x^T y \in \mathbb{R} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ x_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i .$$

If u•v=0, $||u||_2 \neq 0$, $||v||_2 \neq 0$ → *u and v are orthogonal*

If u•v=0, $||u||_2 = 1$, $||v||_2 = 1$ → *u and v are orthonormal*

- Vector outer product:

$$xy^T \in \mathbb{R}^{m\times n} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

Basic concepts

We will use upper case letters for matrices. The elements are referred to as a_{ij} .

- Matrix product:

$$A \in \mathbb{R}^{m\times n} \qquad B \in \mathbb{R}^{n\times p}$$

$$C = AB \in \mathbb{R}^{m\times p}$$

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

e.g.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

$$C = AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

Basic concepts: Special matrices

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$ diagonal   also denoted $diag(a,b,c)$

$$\begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}$$ upper-triangular

$$\begin{pmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{pmatrix}$$ lower-triangular

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$ I (identity matrix)

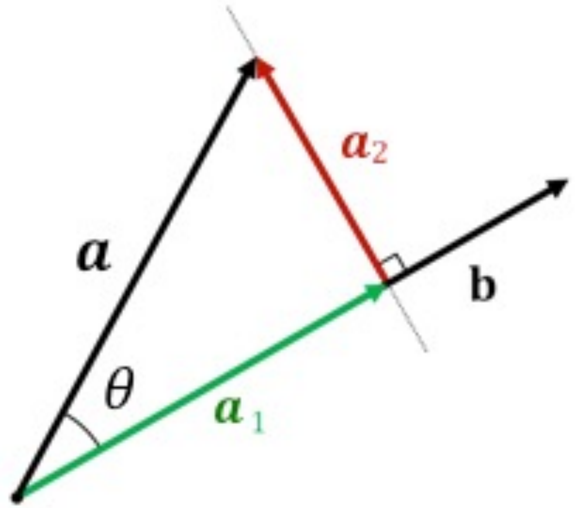
0= $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ the null vector in dimension 3

Basic concepts

Orthogonal projection of a vector $\boldsymbol{a}$ on (or onto) another vector **b** (different from the null vector)

Let $\boldsymbol{a}$ and $\boldsymbol{b}$ be two vectors such that $\boldsymbol{b} \neq \mathbf{0}$, then, the orthogonal projection of $\boldsymbol{a}$ onto $\boldsymbol{b}$ is a vector $\boldsymbol{a_1}$ given by:

$$\boldsymbol{a_1} = \frac{\boldsymbol{a} \bullet \boldsymbol{b}}{\boldsymbol{b} \bullet \boldsymbol{b}}\, \boldsymbol{b} = \| \boldsymbol{a} \| \, (\cos\theta) \frac{\boldsymbol{b}}{\| \boldsymbol{b} \|}$$



θ is the angle between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$

Basic concepts

Definition: Given a *n x m* matrix **A,** the transpose of **A** is the *m x n* matrix, denoted $\mathbf{A}^T$ whose element *(i,j)* corresponds to the element *(j,i)* of **A**

Transpose: You can think of it as

- "flipping" the rows and columns

OR

- "reflecting" vector/matrix on line

e.g.
$$\begin{pmatrix} a \\ b \end{pmatrix}^T = \begin{pmatrix} a & b \end{pmatrix}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

Properties:

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- $(A+B)^T = A^T + B^T$

Determinant (1/2)

The determinant of a *n x n* matrix **A**, denoted *det*(**A**) or |**A**| is a real number, some examples:

- For a 1x1 matrix $\mathbf{A} = [a]$, *det*(**A**) = a
- For a 2x2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $det(\mathbf{A}) = (ad\text{-}cb)$
- For a 3x3 matrix $\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ is:

$$det(\mathbf{A}) = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

Determinant (2/2)

General definition:

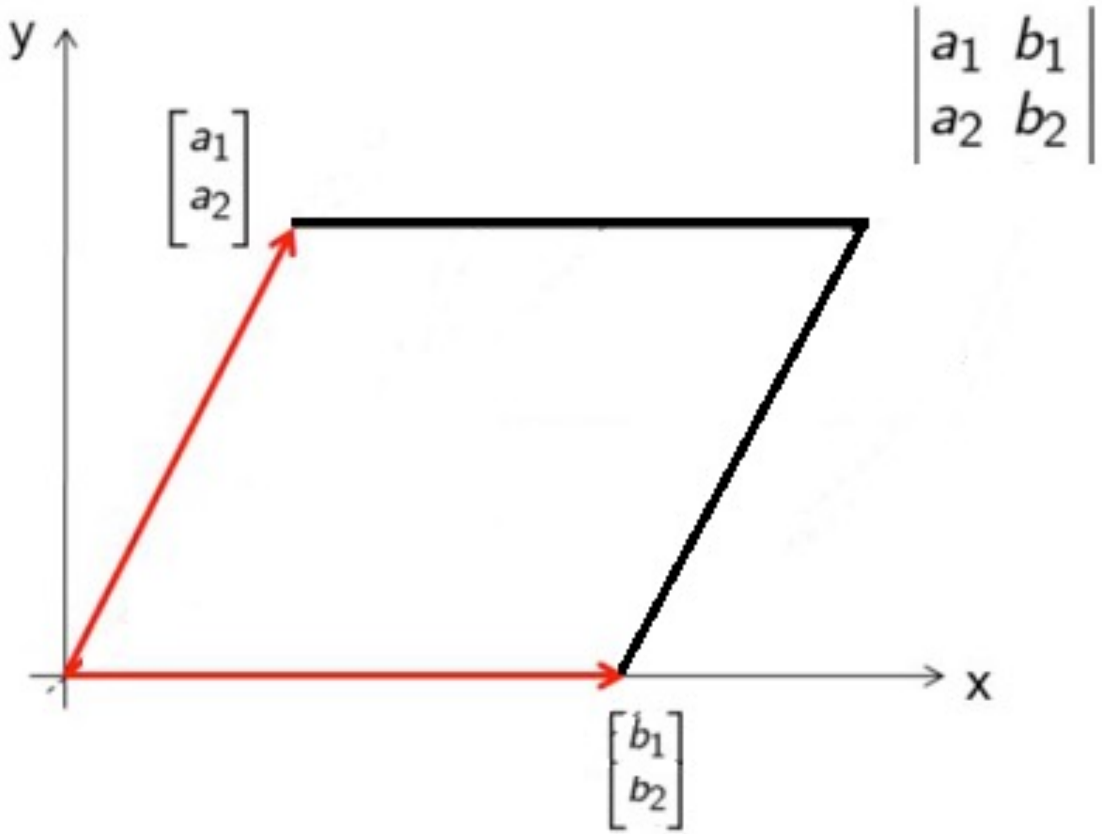
For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{ij}]$ is the sum of n terms of the form $\pm a_{1j} \det(A_{1j})$, with plus and minus signs alternating, where the entries $a_{11}, a_{12}, \ldots, a_{1n}$ are from the first row of A. In symbols,

$$\begin{aligned}\det(A) &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + \cdots + (-1)^{1+n} a_{1n}\det(A_{1n}) \\ &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j})\end{aligned}$$

Where the submatrix $\mathbf{A}_{ij}$ is obtained by crossing out row *i* and column *j*

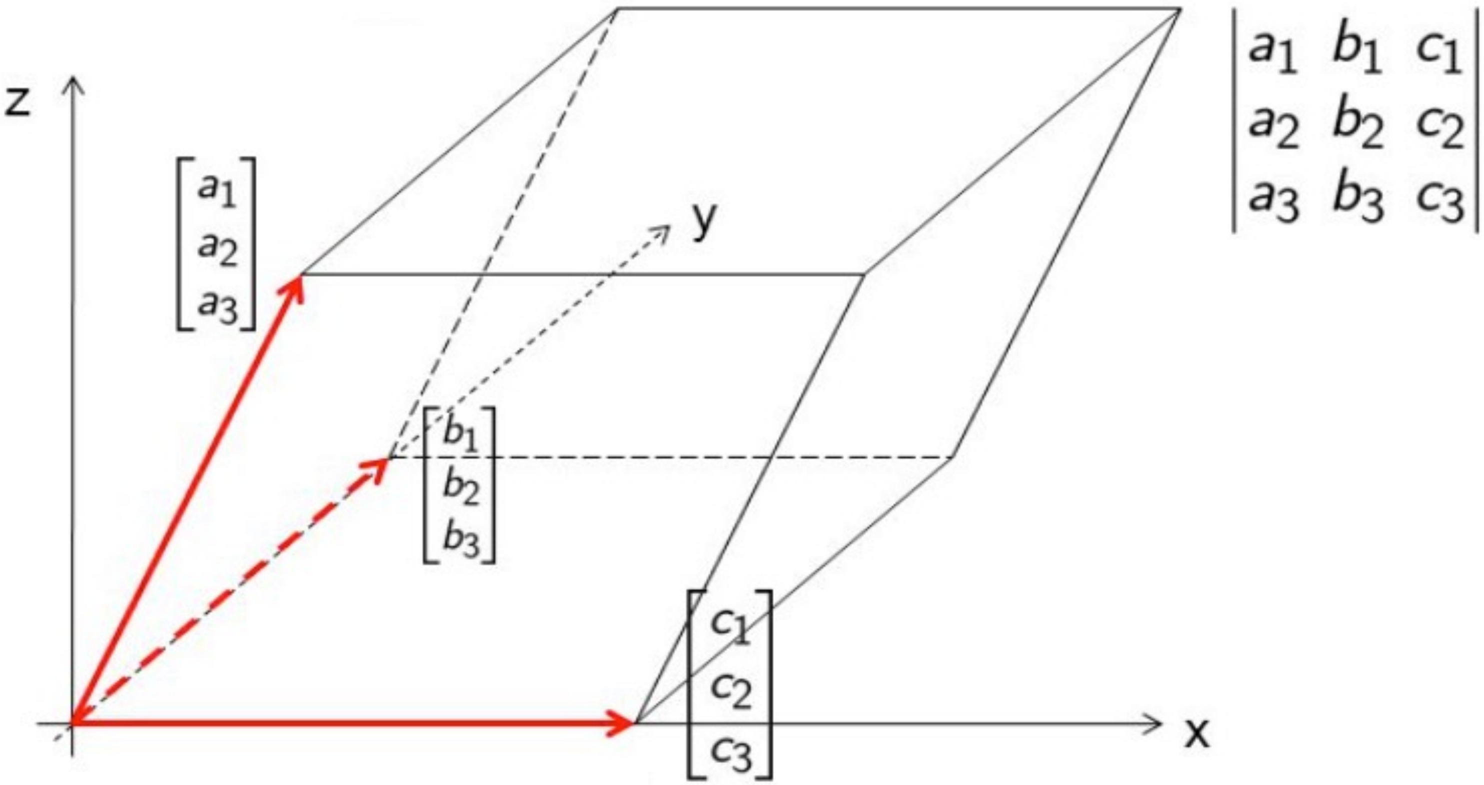
Determinant - geometrical interpretation

- For a 2x2 matrix $\mathbf{A}$, $|\det(\mathbf{A})|$={area of the parallelogram determined by the columns of $\mathbf{A}$}



Determinant - geometrical interpretation

- For a 3x3 matrix $\mathbf{A}$, $|\det(\mathbf{A})|$={volume of the parallelepiped determined by the columns of $\mathbf{A}$}



Determinant - properties

Properties of Determinants

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two $n \times n$ matrices.

- $\boldsymbol{A}$ is invertible if and only if $det(\boldsymbol{A}) \neq 0$.
  - $\boldsymbol{A}$ is invertible if there exists another matrix denoted $\boldsymbol{A}^{-1}$ such that $\boldsymbol{A}\,\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\,\boldsymbol{A} = \boldsymbol{I}$
- $det(\boldsymbol{AB}) = (det(\boldsymbol{A}))(det(\boldsymbol{B}))$ (if they exist).
- $det(\boldsymbol{A}^T) = det(\boldsymbol{A})$.
- If $\boldsymbol{A}$ is triangular, then $det(\boldsymbol{A})$ is the product of the entries on the main diagonal of $\boldsymbol{A}$.

Linear combination

Given vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ in $\mathbb{R}^n$ and given scalars $c_1, c_2, \ldots, c_p$, the vector $\mathbf{y}$ defined by :

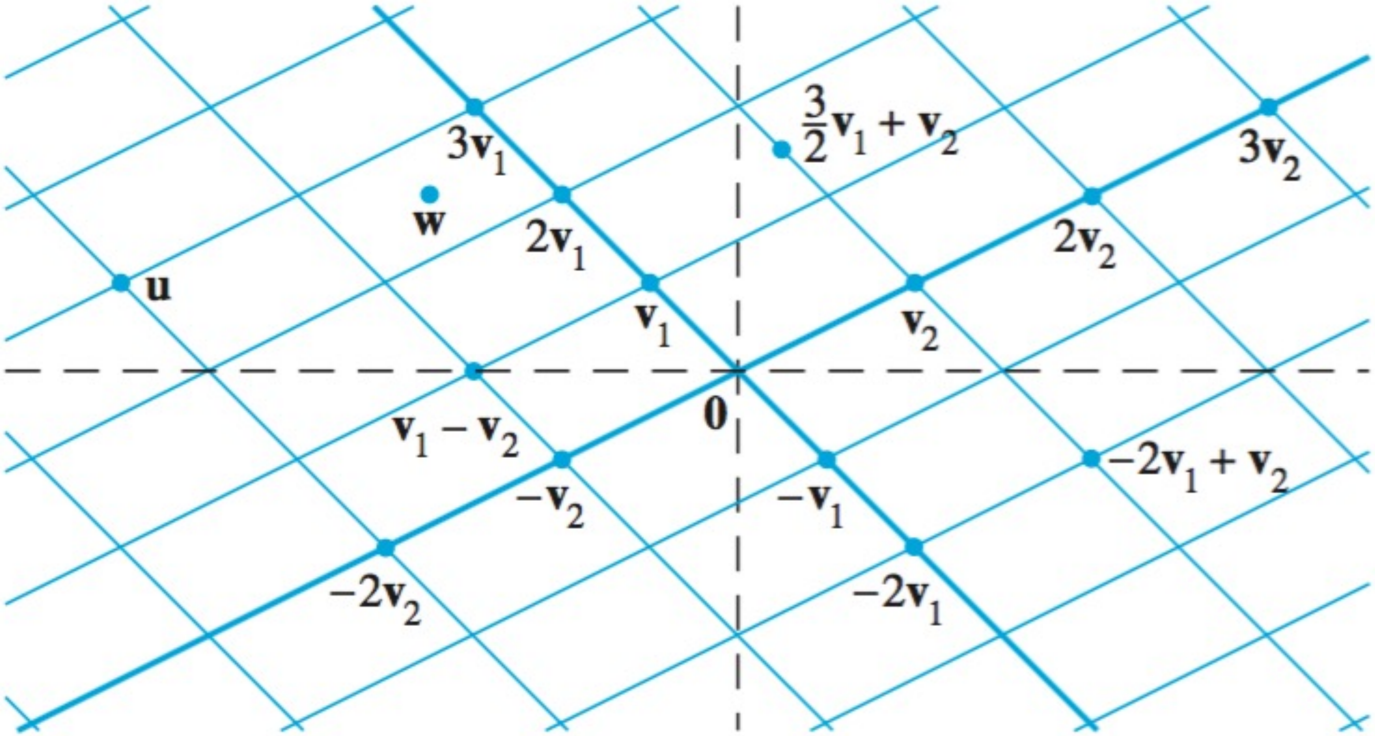
$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_p\mathbf{v}_p$$

is called a **linear combination** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ with weights $c_1, c_2, \ldots, c_p$.

- The weights can be any real number, including zero.

Linear combination example

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$



Linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$.

$$\mathbf{u} = 3\mathbf{v}_1 - 2\mathbf{v}_2 \qquad \mathbf{w} = \tfrac{5}{2}\mathbf{v}_1 - \tfrac{1}{2}\mathbf{v}_2$$

Linear independence

- A set of vectors is linearly independent if none of them can be written as a linear combination of the others.
- Vectors $v_1,\ldots,v_k$ are linearly independent if $c_1v_1+\ldots+c_kv_k = 0$ **implies** $c_1=\ldots=c_k=0$

$$\begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

e.g. $\begin{pmatrix} 1 & 0 \\ 2 & 3 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

The equality holds <u>if and only if</u> (u,v)=(0,0), i.e. the columns are linearly independent.

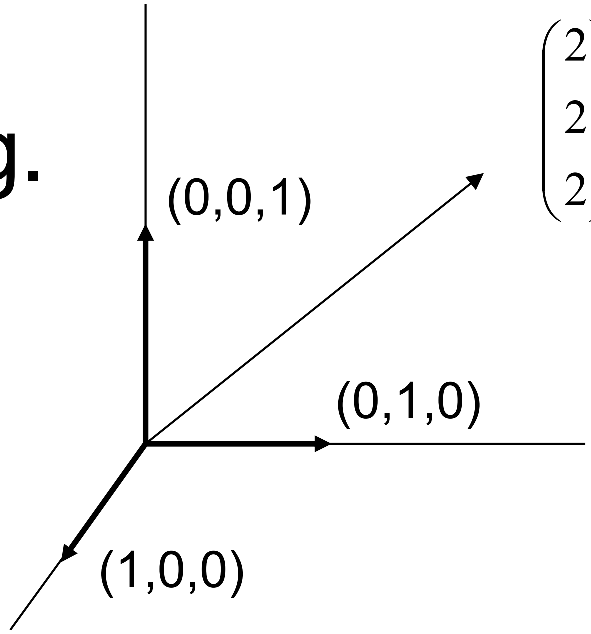
$$x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad x_2 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix} \quad x_3 = \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$$

They are <u>not</u> linearly independent since $x_3 = -2x_1 + x_2$

Span of a vector space

- If all vectors in a vector space may be expressed as linear combinations of a set of vectors $v_1,\ldots,v_k$, then $v_1,\ldots,v_k$ spans this space.
- The cardinality of this set is the dimension of the vector space.

e.g.



$$\begin{pmatrix}2\\2\\2\end{pmatrix} = 2\begin{pmatrix}1\\0\\0\end{pmatrix} + 2\begin{pmatrix}0\\1\\0\end{pmatrix} + 2\begin{pmatrix}0\\0\\1\end{pmatrix}$$

- A basis of a vector space V is a maximal set of linearly independent vectors that span V. A basis is also called a linearly independent spanning set.

Rank of a Matrix

- rank(A) (the rank of a m-by-n matrix A) is

  The maximal number of linearly independent columns

  =The maximal number of linearly independent rows

  =The dimension of col(A)

  =The dimension of row(A)

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}$$

- If A is n by m, then
  - rank(A)<= min(m,n)
  - If n=rank(A), then A has full row rank
  - If m=rank(A), then A has full column rank

Inverse of a matrix

- Inverse of a square matrix $\mathbf{A}$, denoted by $\mathbf{A}^{-1}$ is the *unique* matrix such that.
  - $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ (identity matrix)
- $\mathbf{A}^{-1}$ exists if and only if $\det(\mathbf{A}) \neq 0$
- If $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ exist, then
  - $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$,
- $(\mathbf{A}^{\mathsf{T}})^{-1} = (\mathbf{A}^{-1})^{\mathsf{T}}$
- For orthonormal matrices $\mathbf{A}^{-1} = \mathbf{A}^{\mathsf{T}}$
- For diagonal matrices $\mathbf{D}^{-1} = \mathrm{diag}\{d_1^{-1}, \ldots, d_n^{-1}\}$
  - If $d_i^{-1} \neq 0 \; \forall \, i$.

Outline

- Linear Algebra Basics
- Eigenvalue Decomposition
- Singular Value Decomposition (SVD)*

Eigenvalues & Eigenvectors

- **Eigenvectors** (for a square $m \times m$ matrix $\mathbf{S}$)

$$\mathbf{S}\mathbf{v} = \lambda \mathbf{v}$$

(right) eigenvector

$\mathbf{v} \in \mathbb{R}^m \neq \mathbf{0}$

eigenvalue

λ scalar

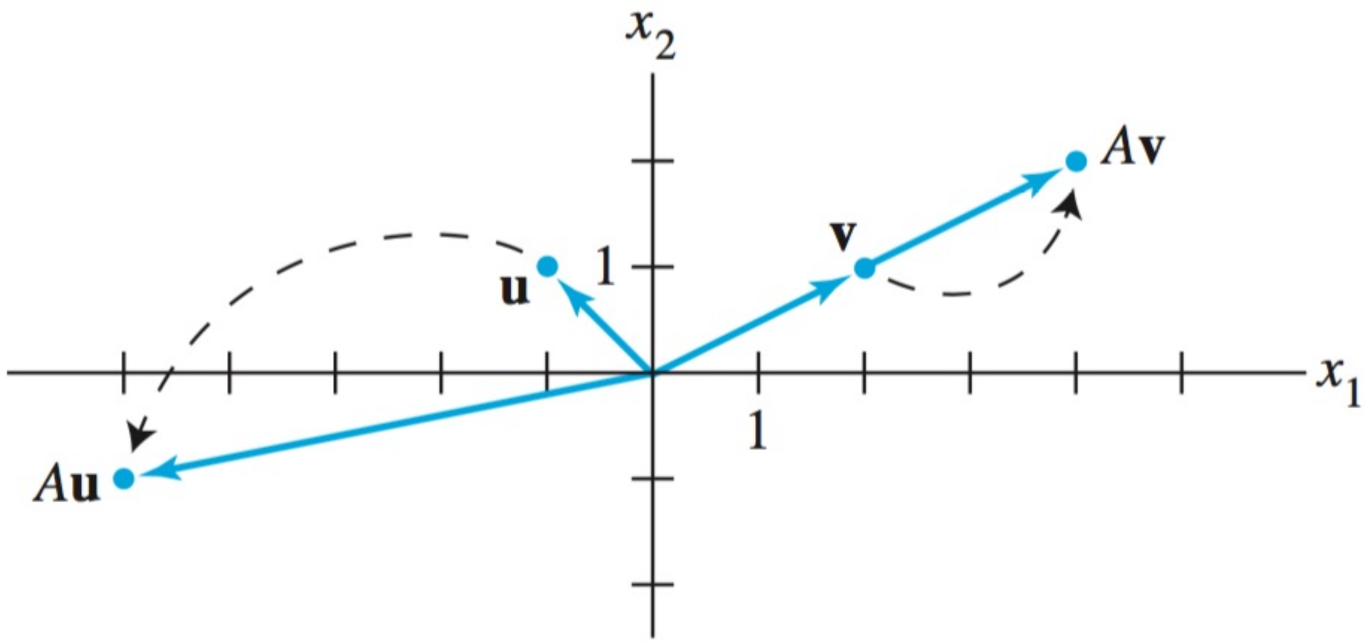
$\mathbf{v}$ is an *eigenvector* of $\mathbf{S}$ associated to the *eigenvalue* λ

Example

$$\begin{pmatrix} 6 & -2 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Eigenvalues & Eigenvectors interpretation

Let $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$



A only *stretches* or *dilates* **v.**
In contrast, **u** is not an eigenvector of **A**.

How to calculate Eigenvalues & Eigenvectors?

If λ is an eigenvalue of an mxm matrix **S,** then, we can find the associated eigenvectors by solving the equation:

$$(\mathbf{S} - \lambda \mathbf{I})\, \mathbf{v} = \mathbf{0}.$$

This equation has a non-zero solution if $|\mathbf{S} - \lambda \mathbf{I}| = 0$

The equation $|\mathbf{S} - \lambda \mathbf{I}| = 0$ is called the **characteristic equation** of $\mathbf{S}$. It is a m-th order equation in λ and can have **at most *m* distinct solutions**.

Remark:

- Eigenvectors are non-zero vectors
- Eigenvalues can be zero, and even complex.

Exercice 1

What are the eigenvalues and eigenvectors of the following matrix?

$$S = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Solution exercice 1

- Let $$S = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$
- Then $$S - \lambda I = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \Rightarrow (2-\lambda)^2 - 1 = 0.$$
- The eigenvalues are 1 and 3.
- The eigenvectors:

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Plug in these values and solve for eigenvectors.

The case of symmetric matrices

For symmetric matrices, eigenvectors for distinct eigenvalues are **orthogonal**

$$Sv_{\{1,2\}} = \lambda_{\{1,2\}} v_{\{1,2\}}, \text{ and } \lambda_1 \neq \lambda_2 \Rightarrow v_1 \bullet v_2 = 0$$

- All eigenvalues of a real symmetric matrix are **real**.

- *Reminder:* The **covariance matrix** is symmetric!

The case of symmetric matrices

A symmetric $n \times n$ real matrix $\mathbf{A}$ is said to be **positive semi definite** if

$$\forall \mathbf{x} \in \mathbb{R}^n \quad \mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} \geq 0$$

<u>Consequence</u>: All eigenvalues of a positive semidefinite matrix are **non-negative**

$$\forall \mathbf{x} \in \mathbb{R}^n \quad \mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} \geq 0 \text{ then if } \mathbf{S}\mathbf{v} = \lambda \mathbf{v} \Rightarrow \lambda \geq 0$$

- *<u>Clue:</u>* The **variance matrix** is **positive semidefinite**

Consider our example

$$S = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Real, symmetric matrix

- The eigenvalues are 1 and 3 are real nonnegative, then $\mathbf{S}$ is positive semidefinite*.
- The eigenvectors are orthogonal:

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

*Plus specifically, positive definite car all eigenvalues are positive.

Diagonalization

Matrix diagonalization theorem

- Let $\mathbf{S} \in \mathbb{R}^{m \times m}$ be a **square** matrix with ***m*** **linearly independent eigenvectors.**
- **Theorem**: it exists an **eigen decomposition**

diagonal

$$\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$$

Unique for distinct eigen-values

- Columns of ***U*** are **eigenvectors** of ***S***
- Diagonal elements of $\mathbf{\Lambda}$ are **eigenvalues** of $\mathbf{S}$

$$\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_m), \quad \lambda_i \geq \lambda_{i+1}$$

Diagonal decomposition: why/how

Let ***U*** have the eigenvectors as columns: $U = \begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix}$

Then, ***SU*** can be written

$$SU = S\begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \ldots & \lambda_n v_n \end{bmatrix} = \begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ldots & \\ & & \lambda_n \end{bmatrix}$$

Thus $\boldsymbol{SU=U\Lambda}$, or $\boldsymbol{U^{-1}SU=\Lambda}$

And $\boldsymbol{S=U\Lambda U^{-1}}$.

Diagonal decomposition – example (1/2)

Recall $S = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}; \lambda_1 = 1, \lambda_2 = 3.$

The eigenvectors $\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ form $U = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

Inverting, we have $U^{-1} = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix}$

Recall $UU^{-1} = I$.

Then, $\boldsymbol{S=U\Lambda U^{-1}} =$ $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix}$

Diagonal decomposition – example (2/2)

Let's divide ***U*** (and multiply ***U***$^{-1}$) by $\sqrt{2}$ so they are orthonormal

Then, ***S***=

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

$\boldsymbol{Q}$ $\qquad\qquad$ Λ $\qquad\qquad$ $(\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T)$

A square matrix **Q** is said **orthogonal** if its column (row) vectors are orthonormal.
<u>Consequence</u> If **Q** orthogonal, then $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$

Symmetric Eigen Decomposition

- If $\mathbf{S} \in \mathbb{R}^{m \times m}$ is a **symmetric** matrix:
- **Theorem**: Exists a (unique) **eigen decomposition** $S = Q \Lambda Q^T$
- where ***Q*** is **orthogonal:**
  - $\boldsymbol{Q^{-1} = Q^T}$
  - Columns of ***Q*** are normalized eigenvectors
  - Columns are orthogonal.
  - The eigenvalues are real.

Outline

- Linear Algebra Basics
- Eigenvalue Decomposition
- Singular Value Decomposition (SVD)*

Singular Value Decomposition

For an $m \times n$ matrix $\mathbf{A}$ of rank r there exists a factorization (Singular Value Decomposition = **SVD**) as follows:

$$A = U \Sigma V^T$$

m×m | *m×n* | *V* is *n×n*

The columns of ***U*** are orthogonal eigenvectors of $\boldsymbol{AA^T}$.

The columns of ***V*** are orthogonal eigenvectors of $\boldsymbol{A^TA}$.

Eigenvalues $\lambda_1 \ldots \lambda_r$ of $\boldsymbol{AA^T}$ are the eigenvalues of $\boldsymbol{A^TA}$.

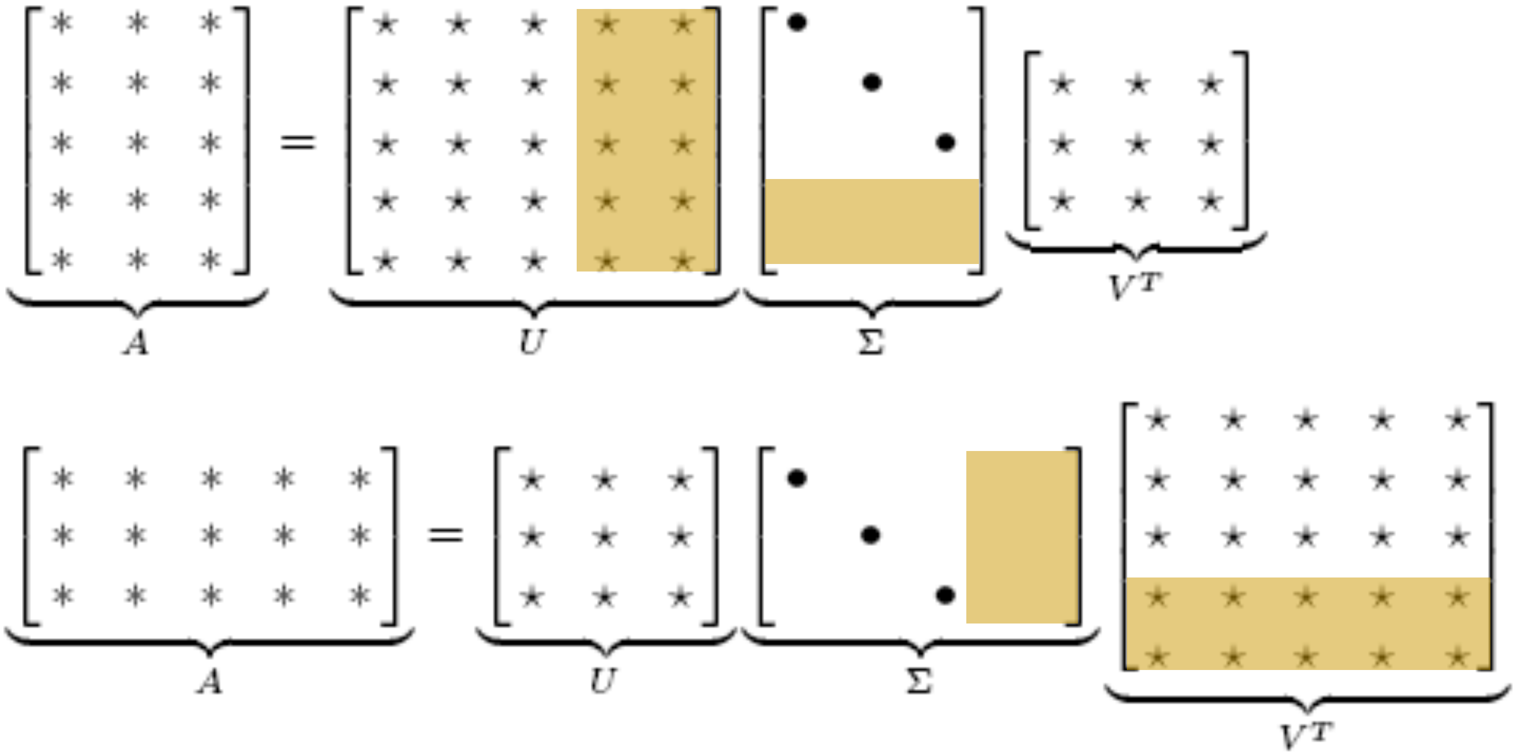
$$\sigma_i = \sqrt{\lambda_i}$$

$$\Sigma = diag(\sigma_1 \ldots \sigma_r)$$

Singular values.

Singular Value Decomposition

- Illustration of SVD dimensions and sparseness



SVD example

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

Thus m=3, n=2. Its SVD is

$$\begin{bmatrix} 0 & 2/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Typically, the singular are values arranged in decreasing order.

References

- David Lay « **Linear Algebra and Its Applications** », 5th Edition. Publisher: Pearson (2011).
- "**Review of Linear Algebra** » Recitation by Leman Akoglu (2010).
- Thomas Hoffman **"Linear algebra background".** Visited on September 2018 https://web.stanford.edu/class/cs276a/handouts/lecture15.pdf