Hertie School

Data Science Summer School

Introduction to Linear Algebra (4-hour workshop)

https://ds3.ai/2022/linear-algebra.html

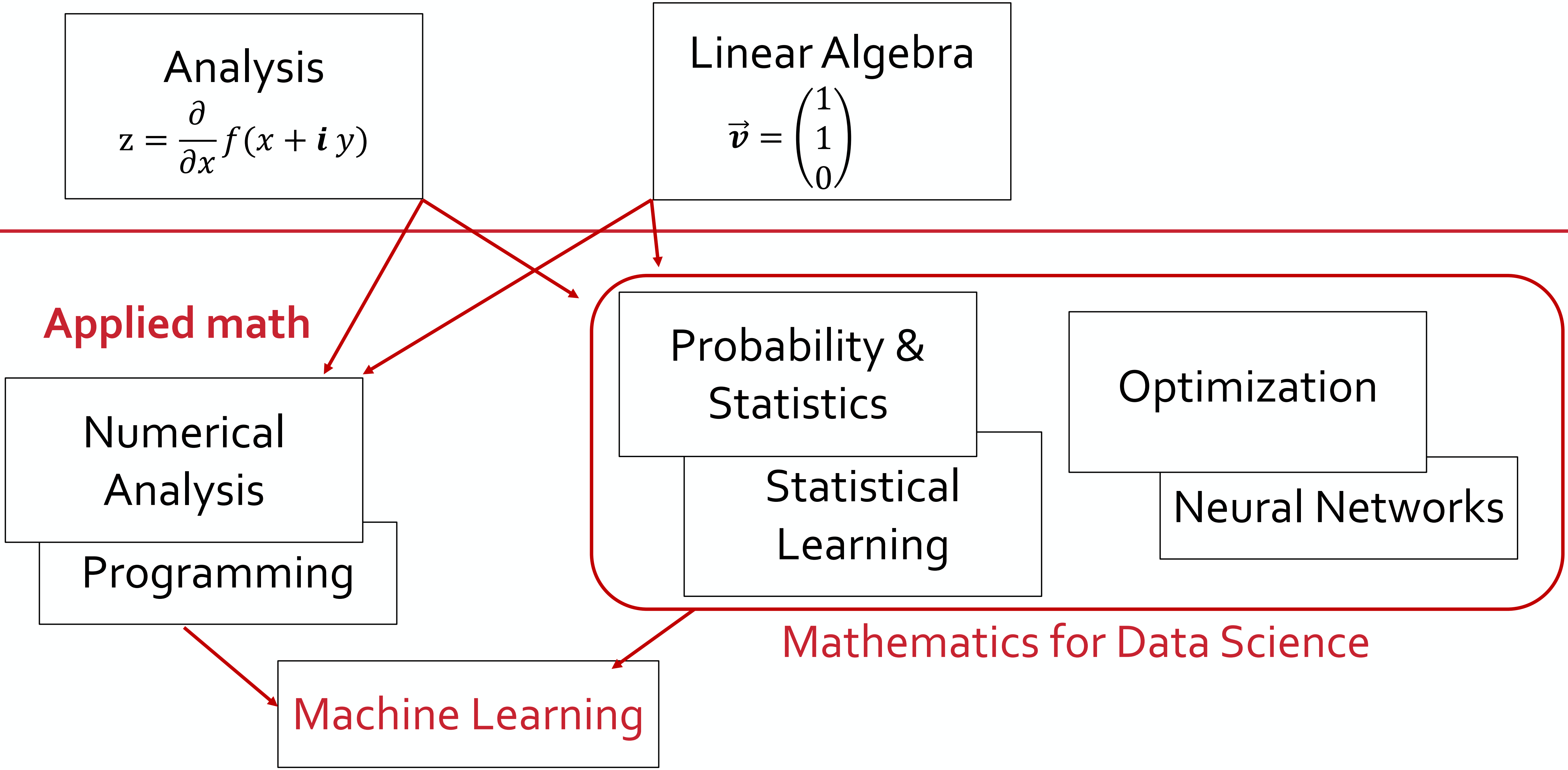
July 29, 2022

Lewin Stein

These slides are adopted from **GRAD-C23** Session3: Linear Algebra (1) by Lynn Kaack.

Math foundations (selection)

Hertie School



Learning objectives this workshop

- Objects: scalars, vectors, matrices, tensors
- Relations: vector space, linear independence, span, basis, rank
- Operations: transpose, determinant, matrix multiplication, trace, norm, inverse
- Preview to eigenvectors and eigenvalues (if there is still time)

There is a lot of material that one could cover for linear algebra.
I recommend you to work through the readings again at your own pace.

References & Sources

(e)-Books

- Gilbert Strang - Linear Algebra and Learning from Data – 2019
- Zico Kolter - Linear Algebra Review and Reference - 2008 www.cs.cmu.edu/~zkolter/course/15-884/linalg-review.pdf
- Kevin P. Murphy - Probabilistic Machine Learning – 2022

Free* online calculator

- www.wolframalpha.com

Videos

- Linear transformations and matrices | Chapter 3 https://www.youtube.com/watch?v=kYB8IZa5AuE
- The determinant | Chapter 6 https://www.youtube.com/watch?v=Ip3X9LOh2dk
- Inverse matrices, column space and null space | Chapter 7 https://www.youtube.com/watch?v=uQhTuRlWMxw
- Eigenvectors and eigenvalues | Chapter 14 https://www.youtube.com/watch?v=PFDu9oVAE-g

Hertie School

Basic concepts and notation

Application 1: Rotation in 2D

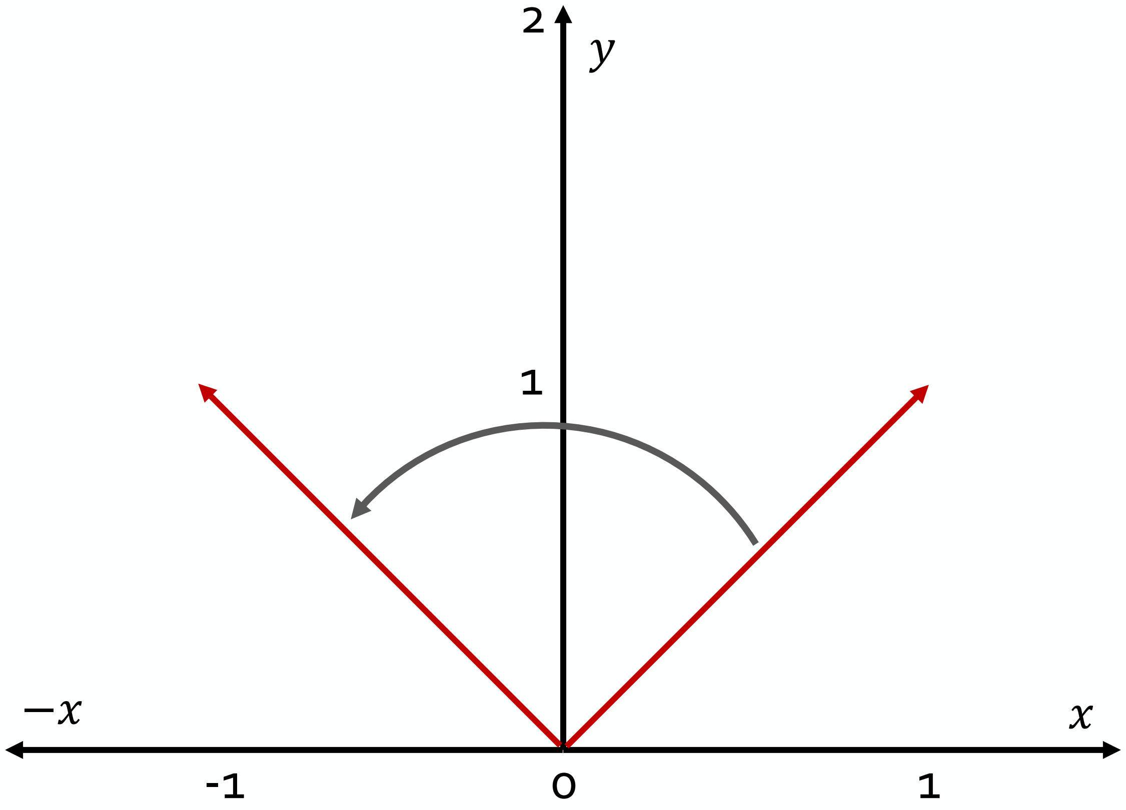
Vector
(1D line
of numbers)

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \overset{\theta = {}^{\pi}/_{2}}{\widehat{=}}$$

Matrix
(2D rectangle
of numbers)

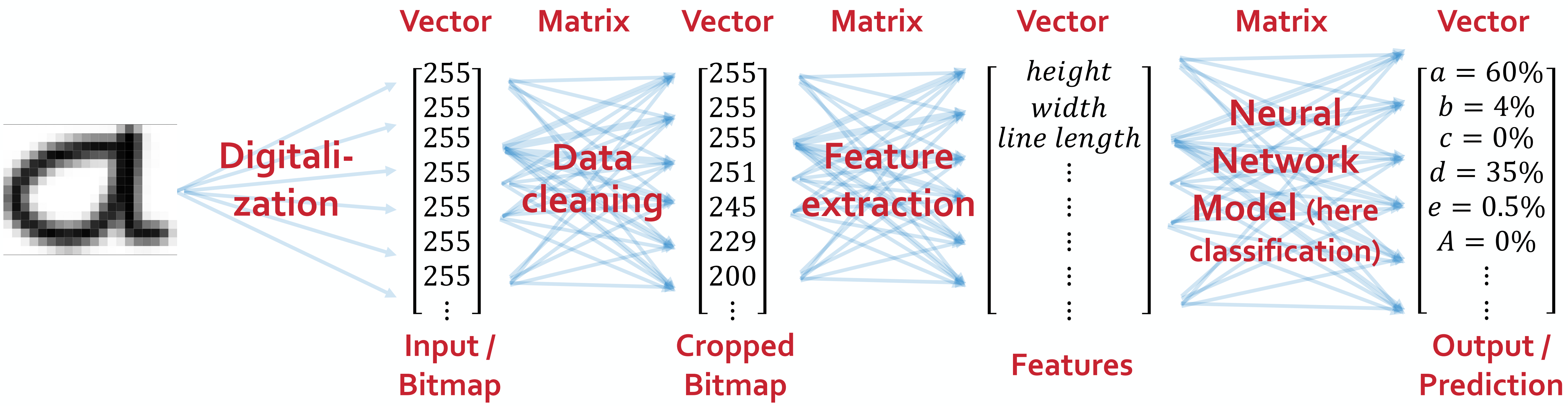
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Vector
(after -90° Rotation)



A matrix is both an object but also a way to manipulate vectors.
We call this manipulation **linear transformation**.

Application 2: Machine Learning



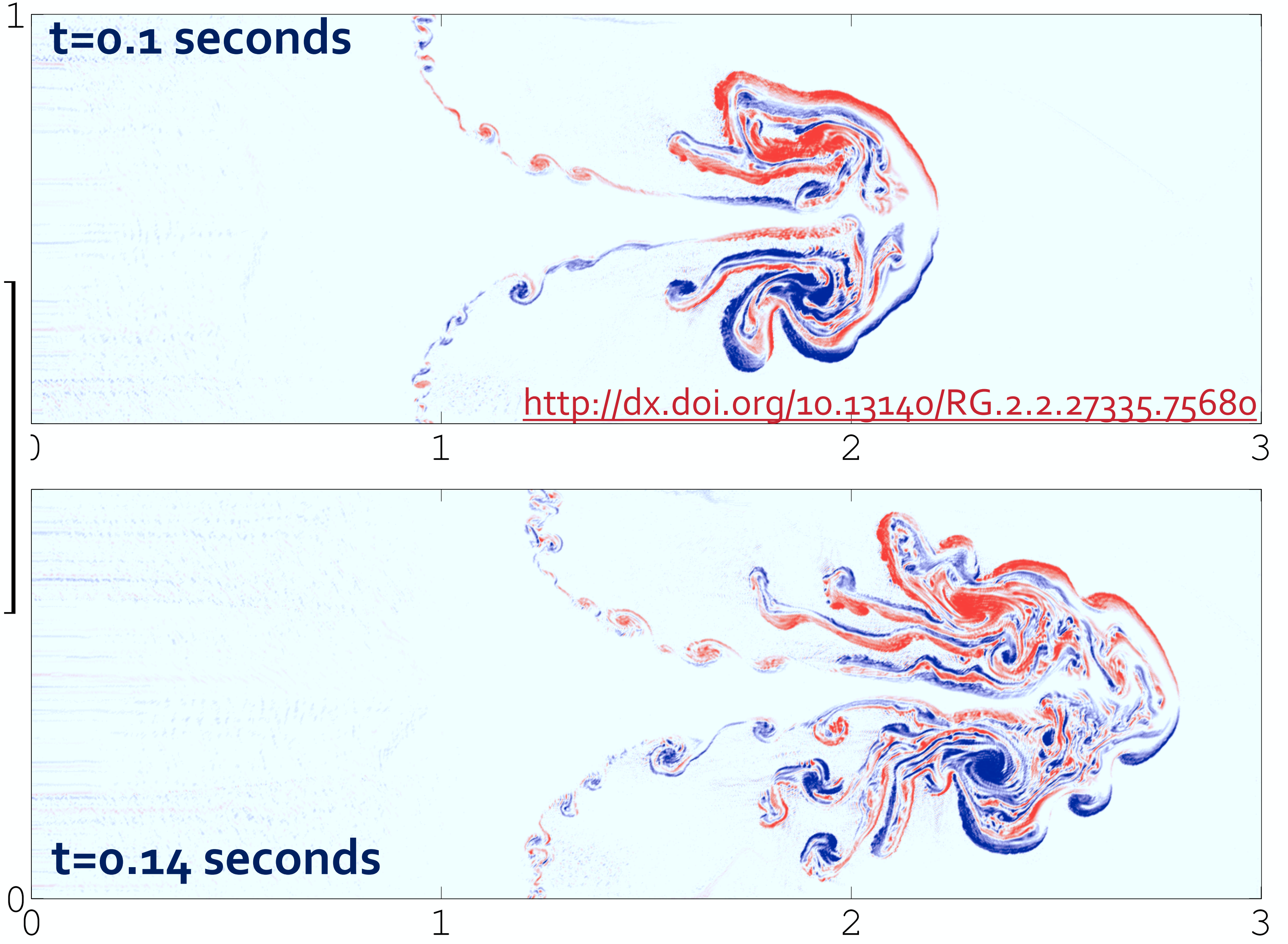
Application 3: Computational Fluid Dynamics

$$|\vec{\omega}| = |\nabla \times \vec{v}|$$

Matrix | **Vector** | **Vector**

$$\begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} density_{11} \\ density_{21} \\ density_{31} \\ \vdots \\ velocity_{11} \\ \vdots \\ pressure_{11} \\ \vdots \end{bmatrix} = \begin{bmatrix} density_{11} \\ density_{21} \\ density_{31} \\ \vdots \\ velocity_{11} \\ \vdots \\ pressure_{11} \\ \vdots \end{bmatrix}$$

discretized & **linearized** fluid equations | state @ old time | state @ new time



Vectors

- A vector $\mathbf{x} \in \mathbb{R}^n$ is a list of n numbers. As a column vector, we write

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Machine Learning slang:
"One-hot vector"
All zeros except for the i^th component, which is one

$$\boldsymbol{e_i} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Possible notation variants (no strict rule ☹):

- lower case (especially x, y, z, v, w)
- bold font $\boldsymbol{v}$
- vector on top $\vec{v}$
- underbar $\underline{v}$

Matrices

A matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ with m rows and n columns is arranged as such:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

We denote the entry of A in the ith row and jth column as A_{ij}.

Possible notation variants (no strict rule ☹):

- Upper case (especially A, B, C, I, Λ, M)
- bold font $\boldsymbol{M}$
- double underbar $\underline{\underline{M}}$

Special cases:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

Square matrix
Same number of rows and columns

$$\begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{bmatrix}$$

Diagonal matrix
Off-diagonal entries are zero

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Identity matrix
Ones on the main diagonal and zero elsewhere

Hertie School

Matrix ordering in (linear) computer storage

Row-major order: NumPy of Python, C/C++

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

```
for i=0:m-1
   for j=0:n-1
      A[i,j]=0
   end
end
```

The 1st storage address is a_{11}, the 2nd is a_{12}…
Hence $\boldsymbol{j}$ of $A[i,\boldsymbol{j}]$ is the fastest index and should be the innermost loop.

Column-major order: R, Julia, MATLAB, Fortran

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

```
for j=1:n
   for i=1:m
      A[i,j]=0
   end
end
```

The 1st storage address is a_{11}, the 2nd is a_{21}…
Hence $\boldsymbol{i}$ of $A[\boldsymbol{i},j]$ is the fastest index and should be the innermost loop.

The Transpose

Hertie School

Given $A \in \mathbb{R}^{m \times n}$, its transpose is written $A^T \in \mathbb{R}^{n \times m}$ and it is the $n \times m$ matrix whose entries are given by

$$(A^T)_{ij} = A_{ji}\,.$$

Transpose = flipping rows and columns

Properties:

- $(A^T)^T = A$
- $(A+B)^T = A^T + B^T$
- $(AB)^T = B^T A^T$
- A square matrix $A \in \mathbb{R}^{n \times n}$ is **symmetric** if $A = A^T$, and anti-symmetric if $A = -A^T$.

Tranpose of column vector $\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$

is row vector $\boldsymbol{v^T} = [v_1, v_2, v_3]$

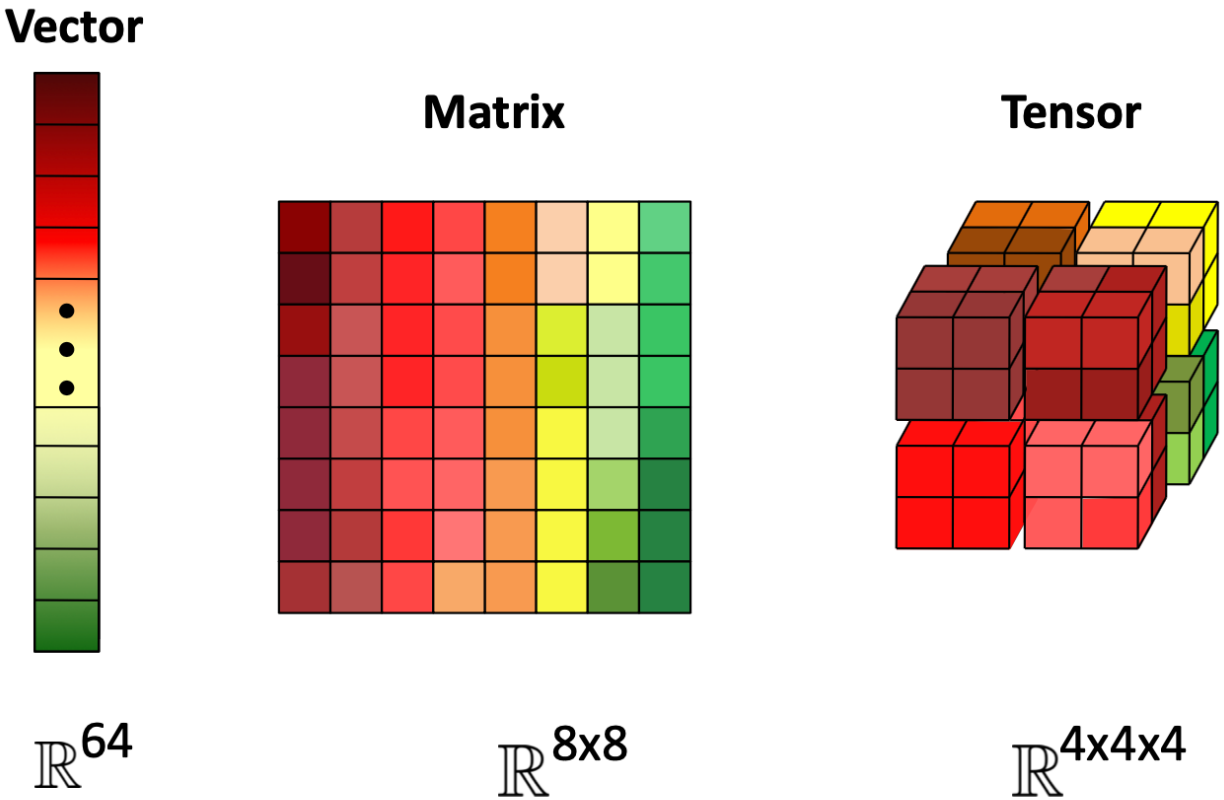
Tranpose of $A = \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$

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

Transpose = mirror on diagonal axis

Tensors

A tensor is a generalization of a 2d array to more than 2 dimensions



Source: Murphy (2022)

Hertie School

Vector spaces

Vector addition and scaling

Vector space

The elements of an **n-dimensional vector space** $\mathcal{V}$ are the vectors $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$ (all with **n** entries).
Vectors can be modified by the basic operations:

- Adding: $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$
- Scaling: $\cdot : \mathbb{R} \times \mathcal{V} \to \mathcal{V}$

For $\mathbb{R}^2$ it looks like:

Operations are defined element-wise.

Examples:

$$\boldsymbol{v} + \boldsymbol{w} = (v_1 + w_1, \dots, v_n + w_n)$$

$$\boldsymbol{v} + \ 2\,\boldsymbol{w} = (v_1 + 2w_1, \dots, v_n + 2w_n)$$

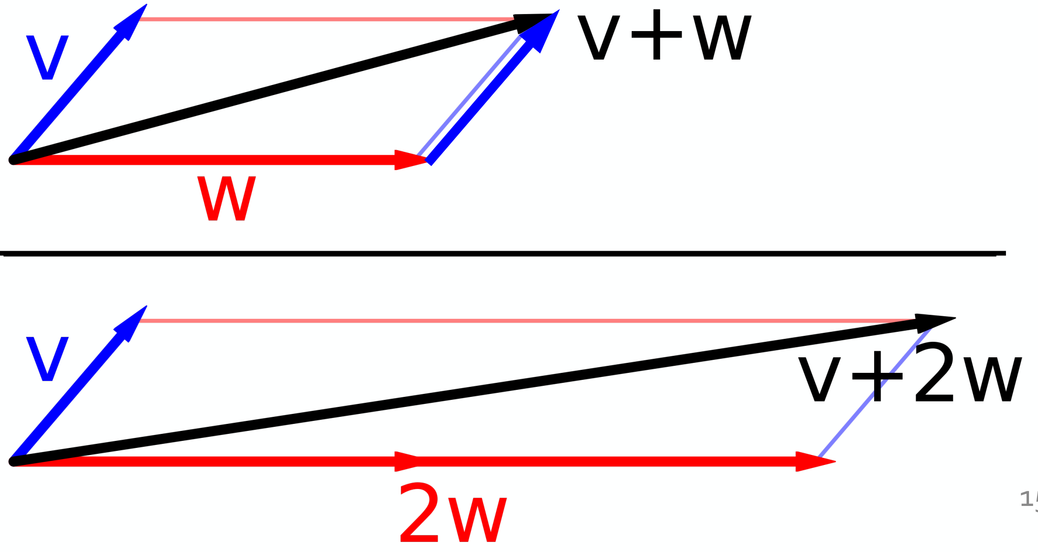


Image: Wikipedia

Sum symbol and Einstein notation

Sum symbol

$$\sum_{i=1}^{4} i = 1 + 2 + 3 + 4 = 10$$

$$\sum_{i=1}^{40} i = \frac{1}{2}\big((1 + 40) + (2 + 39) + \cdots + (38 + 3) + (39 + 2) + (40 + 1)\big) = \frac{41 \times 40}{2} = 820$$

Einstein notation:

If an index occurs twice in an expression, a sum sign is placed in front of it.

Therein, the index of the sum always runs over all entries (the full vector dimension).

$$x_i y_i = \sum_{i=1}^{n} x_i\ y_i \overbrace{=}^{\text{for } n=3} x_1\, y_1 + x_2\, y_2 + x_3\, y_3$$

Linear independence

Definition

A set of vectors $\{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^m$ is called **(linearly) independent** if no vector can be represented as a linear combination of the remaining vectors. Conversely, a vector which can be represented as a linear combination of the remaining vectors is said to be **(linearly) dependent**.

That is if

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

for some scalar values $\alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}$.

Linear independence

If

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

for some scalar values $\alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}$, then we say that $x_1, x_2, \dots, x_n$ are linearly dependent.

Question

Are the following vectors linearly independent?

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

Linear independence

If

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

for some scalar values $\alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}$, then we say that $x_1, x_2, \dots, x_n$ are linearly dependent.

Answer

Are the following vectors linearly dependent?

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

Can we find α_1 and α_2 such that $x_3 = \alpha_1 x_1 + \alpha_2 x_2$?

$$\alpha_1 1 + \alpha_2 1 = -3 \quad \Leftrightarrow \alpha_1 = -\alpha_2 - 3$$
$$\alpha_1 3 + \alpha_2 5 = 5$$

$$(-\alpha_2 - 3)3 + \alpha_2\, 5 = 5 \qquad /+9$$
$$\Leftrightarrow \quad 2\alpha_2 = 14 \qquad \Leftrightarrow \quad \alpha_2 = 7$$

Plugging in $\alpha_2 = 7$ yields $\alpha_1 = -10$.

x_1, x_2, x_3 are linearly dependent. In fact, any vector in $\mathbb{R}^2$ can be expressed by two linearly independent vectors, so three vectors must be linearly dependent.

Linear independence

Question

Are the following vectors linearly dependent?

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

Yes, they are linearly dependent because $x_3 = -2x_1 + x_2$.

If

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

for some scalar values $\alpha_1, \dots, \alpha_{n-1} \in \mathbb{R}$, then we say that $x_1, x_2, \dots, x_n$ are linearly dependent.

Hertie School

Linear independence

Span

The span of a set of vectors $\{x_1, x_2, \dots, x_n\}$ is the set of all vectors that can be expressed as a linear combination of $\{x_1, x_2, \dots, x_n\}$. That is,

$$\operatorname{span}(\{x_1, x_2, \dots, x_n\}) \equiv \left\{v\colon\ v = \sum_{i=1}^{n} \alpha_i x_i\,, \alpha_i \in \mathbb{R}\right\}.$$

If $\{x_1, x_2, \dots, x_n\}$ is a set of n linearly independent vectors where each $x_i \in \mathbb{R}^n$, then $\operatorname{span}(\{x_1, x_2, \dots, x_n\}) = \mathbb{R}^n$.

In other words, any vector $v \in \mathbb{R}^n$ can be written as a linear combination of $x_1, x_2, \dots, x_n$.

Linear independence

Question

Are the following vectors linearly independent?

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Linear independence

Answer is yes:

These are the **natural or standard** coordinate **basis vectors** of $\mathbb{R}^3$:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- **All** (3 in $\mathbb{R}^3$) are **linearly independent** (= full rank, rank definition follows).
- You can represent any point x ∈ $\mathbb{R}^3$ with those, by $x = x_1 e_1 + x_2 e_2 + x_3 e_3$.
- They are **normalized**: $||e_i|| = 1$ (|| || definition follows).
- They are **orthogonal**: $e_i^T e_j = 0 \text{ if } i \neq j$ (details follow).

right-handed coordinate system

Image: https://www.snb.ch

Linear independence

Basis of a space

The basis $\mathcal{B}$ is a set of linearly independent vectors that spans the whole space, meaning $\text{span}(\mathcal{B}) = \mathbb{R}^n$.

There are often multiple bases to choose from.

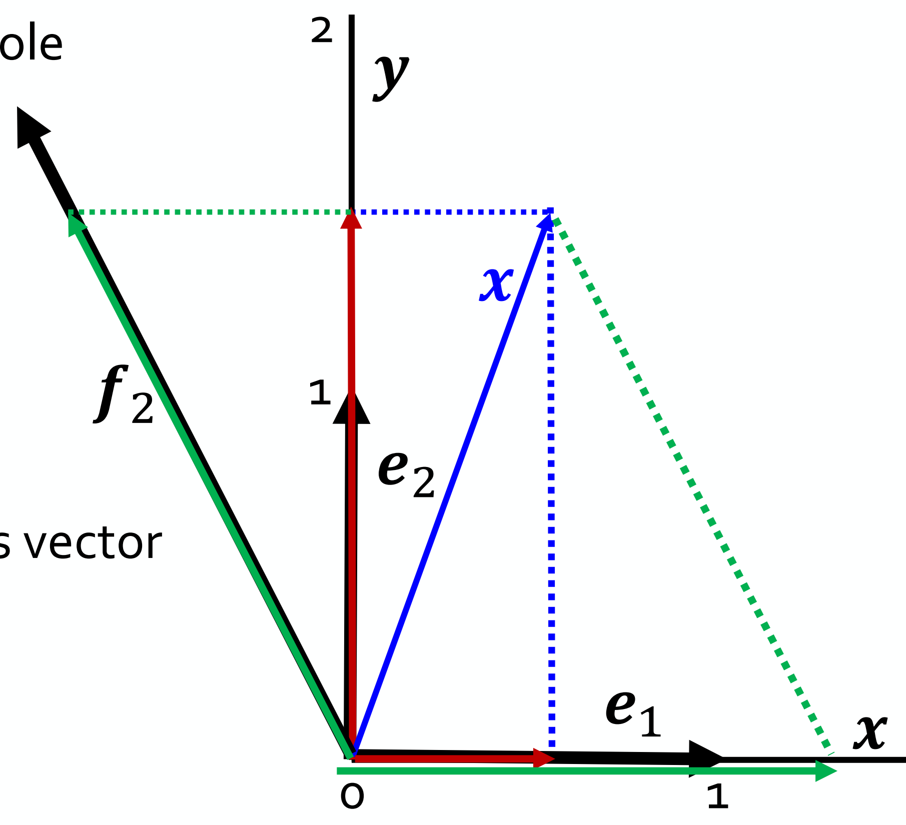
Natural/standard basis uses the coordinate vectors $\boldsymbol{e}_i$.

A vector $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ can be viewed as a list of coefficients for each basis vector

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n,$$

$$\boldsymbol{x} = \xi_1\boldsymbol{e}_1 + \xi_2\boldsymbol{f}_2 + \cdots + \xi_n\boldsymbol{e}_n.$$

In $\mathbb{R}^2$:



Rank

Definition

The **column rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of columns of A that constitute a linearly independent set.

The **row rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of columns of A that constitute a linearly independent set.

The **rank** of a matrix is equal to its column rank is equal to its row rank. It is the dimension of its column space.

Rank

Examples

The **column rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of columns of A that constitute a linearly independent set.

Examples:

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

Rank

Examples

The **column rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of columns of A that constitute a linearly independent set.

Examples:

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix}$$

$n = 3$ columns in A
$r = 2$ columns in A

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$$

$n = 3$ columns in B
$r = 3$ columns in B
(full rank)

Rank

Properties

- For $A \in \mathbb{R}^{m \times n}$, we have $\text{rank}(A) \leq \min(m, n)$. If $\text{rank}(A) = \min(m, n)$ then A is said to be **full rank**.

- For $A \in \mathbb{R}^{m \times n}$, we have $\text{rank}(A) = \text{rank}(A^T)$

- For $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$ we have $\text{rank}(AB) \leq \min\big(\text{rank}(A), \text{rank}(B)\big)$

- For $A, B \in \mathbb{R}^{m \times n}$, we have $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$

Determinant

Definition

The determinant of a square matrix $A \in \mathbb{R}^{n\times n}$ is a function det: $\mathbb{R}^{n\times n} \to \mathbb{R}$, and it is denoted $|A|$ or det A.

Notation ambiguity ☹ (context helps out):

- $|a|$ or $|\boldsymbol{v}|$ = absolute value (if scalar/vector)
- $|A|$ or det $\boldsymbol{A}$ = determinant (if matrix)

green includes "+"
red includes "-"

$$\left|\begin{bmatrix} a_{11} \end{bmatrix}\right| = a_{11}$$

$$\left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right| = a_{11}a_{22} - a_{12}a_{21}$$

$$\left|\begin{bmatrix} \color{green}{a_{11}} & \color{red}{a_{12}} & \color{green}{a_{13}} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right| =$$

Determinant

Definition

The determinant of a square matrix $A \in \mathbb{R}^{n\times n}$
is a function det: $\mathbb{R}^{n\times n} \to \mathbb{R}$,
and it is denoted $|A|$ or det A.

Notation ambiguity ☹ (context helps out):

- $|a|$ or $|\boldsymbol{v}|$ = absolute value (if scalar/vector)
- $|A|$ or det $\boldsymbol{A}$ = determinant (if matrix)

green includes "+"
red includes "-"

$$\left|\begin{bmatrix} a_{11} \end{bmatrix}\right| = a_{11}$$

$$\left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right| = a_{11}a_{22} - a_{12}a_{21}$$

$$\left|\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right| = +a_{11}\left|\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}\right|$$

Determinant

Definition

The determinant of a square matrix $A \in \mathbb{R}^{n\times n}$ is a function det: $\mathbb{R}^{n\times n} \to \mathbb{R}$, and it is denoted $|A|$ or det A.

Notation ambiguity ☹ (context helps out):

- $|a|$ or $|\boldsymbol{v}|$ = absolute value (if scalar/vector)
- $|A|$ or det $\boldsymbol{A}$ = determinant (if matrix)

green includes "+"
red includes "-"

$$\left|\begin{bmatrix} a_{11} \end{bmatrix}\right| = a_{11}$$

$$\left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right| = a_{11}a_{22} - a_{12}a_{21}$$

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

Determinant

Definition

The determinant of a square matrix $A \in \mathbb{R}^{n\times n}$
is a function det: $\mathbb{R}^{n\times n} \to \mathbb{R}$,
and it is denoted $|A|$ or det A.

Notation ambiguity ☹ (context helps out):

- $|a|$ or $|\boldsymbol{v}|$ = absolute value (if scalar/vector)
- $|A|$ or det $\boldsymbol{A}$ = determinant (if matrix)

green includes "+"
red includes "-"

$$\left|\begin{bmatrix} a_{11} \end{bmatrix}\right| = a_{11}$$

$$\left|\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right| = a_{11}a_{22} - a_{12}a_{21}$$

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

Determinant interpreted geometrically

(in standard/natural basis)

The determinant | Chapter 6, Essence of linear algebra | Minute 0:00-5:30

www.youtube.com/watch?v=Ip3X9LOh2dk

Determinant

Visualization example

Consider the 2x2 matrix $A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}$

with row vectors $a_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $a_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$.

The absolute value of the determinant measures the area of the parallelogram (gray area) given by the row vectors:

$$|\det A| = |2 - 9| = 7.$$

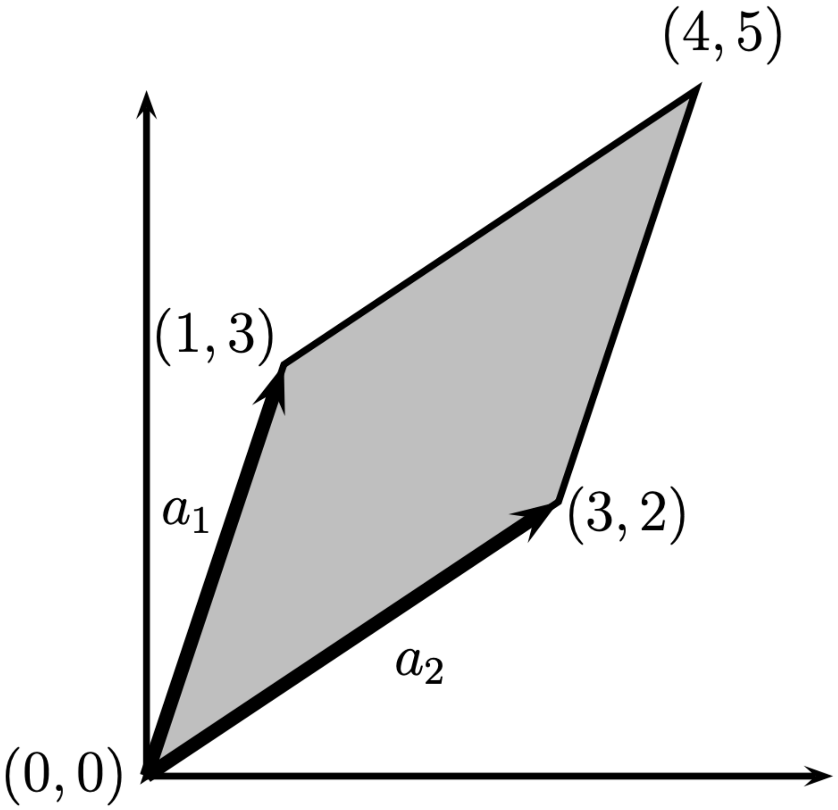


Image: Kolter

Determinant

Question

How big is the (gray) area between the vectors $\begin{bmatrix}1\\1\end{bmatrix}$ and $\begin{bmatrix}1\\2\end{bmatrix}$?

Answer

Half the absolute value of the determinant measures the area of the triangel (gray area) between the row vectors:

$$0.5 \left|\det\begin{bmatrix}1 & 1\\1 & 2\end{bmatrix}\right| = 0.5\ |2-1| = 0.5.$$

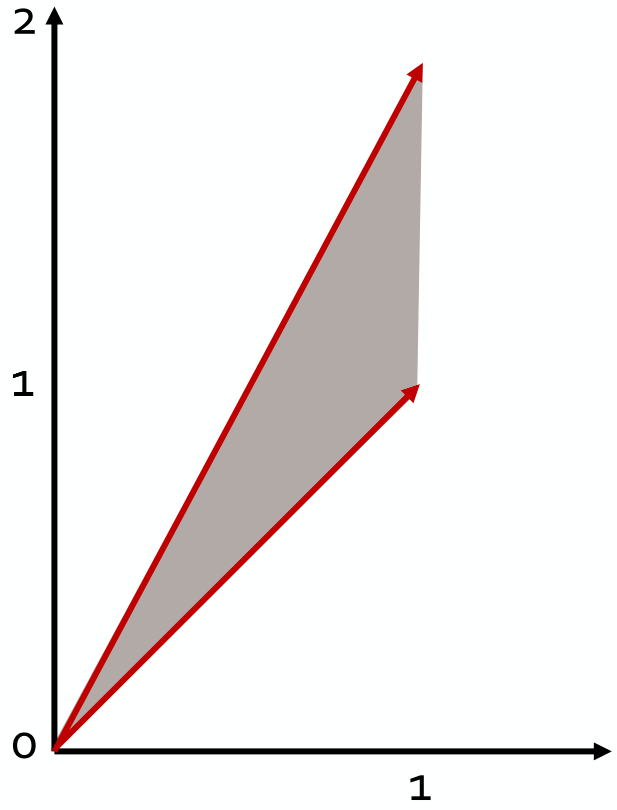


Image: Kolter

Determinant

Properties

1. The determinant of identity is $|I| = 1$ (the volume of a unit hypercube is 1)
2. Given $A \in \mathbb{R}^{n \times n}$, if we multiply **a single row** in A by a scalar $t \in \mathbb{R}$, then the determinant of the new matrix is $t|A|$.
3. If we exchange two rows a_i^T and a_j^T of A then the determinant of the new matrix is $-|A|$.
4. A matrix $A \in \mathbb{R}^{n \times n}$ has full rank if $|A| \neq 0$.

More Properties:

- For $A \in \mathbb{R}^{n \times n}$, $|A| = |A^T|$.
- For $A, B \in \mathbb{R}^{n \times n}$, $|AB| = |A|\,|\mathrm{B}|$.
- For $A \in \mathbb{R}^{n \times n}$, if $|A| = 0$ we call A "singular" (more details follow).
- For $A \in \mathbb{R}^{n \times n}$and A non-singular, $|A^{-1}| = 1/|A|$ (A^{-1} definition of inverse follows).

Hertie School

Matrix multiplication

Matrix multiplication

Matrix-matrix

The product $\cdot$ of two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ is

$$A \cdot B = A\,B = C \in \mathbb{R}^{m \times p},$$

where $C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj} = \boxed{A_{ik} B_{kj}.}$

Einstein notation!

The number of columns n in A must be equal to the number of rows n in B.

The example on the left:

$$c_{12} = a_{11} b_{12} + a_{12} b_{22}$$
$$c_{33} = a_{31} b_{13} + a_{32} b_{23}$$

```
Common programming notation:
matmul(A,B)
A.B
A*B
```

dangerous: may be pointwise multiplication

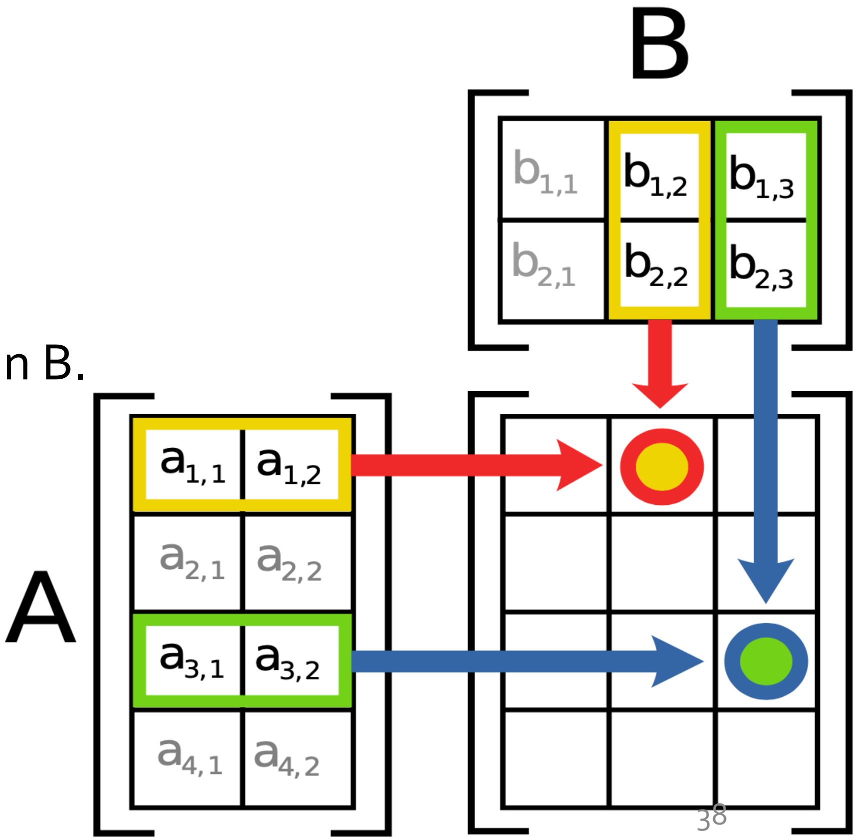


Image: Wikipedia

Matrix multiplication

Example

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 0 \end{bmatrix} = \begin{bmatrix} 1+5 & ? \\ ? & ? \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ ? & ? \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ -3+5 & ? \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 2 & -4 \end{bmatrix}$$

Let $v_1 = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$, $v_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

Questions

$$\boldsymbol{v_1\ v_2} = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 - 3 = -2$$

$$\boldsymbol{v_2\ v_1} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 3 & 0 & -3 \end{bmatrix}$$

Matrix multiplication

Properties

- Matrix multiplication is associative: $(AB)C = A(BC)$

Verify for (i, j)th entry with matrices $A \in \mathbb{R}^{m\times n}$, B $\in \mathbb{R}^{n\times p}$ and $C \in \mathbb{R}^{p\times q}$:

$$\big((AB)C\big)_{ij} = \sum_{k=1}^{p} (AB)_{ik} C_{kj} = \sum_{k=1}^{p} \left(\sum_{l=1}^{n} A_{il} B_{lk} \right) C_{kj}$$

$$= \sum_{k=1}^{p} \left(\sum_{l=1}^{n} A_{il} B_{lk} C_{kj} \right) = \sum_{l=1}^{n} \left(\sum_{k=1}^{p} A_{il} B_{lk} C_{kj} \right)$$

$$= \sum_{l=1}^{n} A_{il} \left(\sum_{k=1}^{p} B_{lk} C_{kj} \right) = \sum_{l=1}^{n} A_{il} (BC)_{lj} = \big(A(BC)\big)_{ij}$$

Matrix multiplication

Properties

- Matrix multiplication is associative: $(AB)C = A(BC)$
- Distributive: $A(B + C) = AB + AC$
- In general (!) not communitative: $AB \neq BA$

Identity Matrix

Identiy matrix

$$\mathrm{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

It has the property that for all $A \in \mathbb{R}^{m \times n}$,

$$AI = A = IA$$

Hertie School

Vector-vector multiplication

Dot product / scalar product (inner product)

Given two vectors $x, y \in \mathbb{R}^n$, the inner product of x and y is written as $\langle x, y \rangle$, which is also known as the dot product or scalar product (in Euclidean space)

$$x^T y = [x_1 \; x_2 \; \ldots x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \cdots + x_n y_n$$

Properties:

- Symmetric: $x^T y = y^T x$
- Dot product of two orthogonal vectors is zero.

Possible notation variants:

- $\langle x, y \rangle$
- $x^T y$
- $\boldsymbol{x}^T \boldsymbol{y}$
- $\vec{x} \cdot \vec{y}$

Hertie School

Vector-vector multiplication

Outer product

Given two vectors $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$, the outer product $xy^T \in \mathbb{R}^{m \times n}$ is a matrix, whose entries are given by $(xy^T)_{ij} = x_i y_i$

$$xy^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} [y_1 \; y_2 \; \dots y_n] = \begin{bmatrix} x_1 y_1 & \cdots & x_1 y_n \\ x_2 y_1 & & x_2 y_n \\ \vdots & \ddots & \vdots \\ x_m y_1 & \cdots & x_m y_n \end{bmatrix}$$

Hertie School

Vector-matrix multiplication

Given a $A \in \mathbb{R}^{m \times n}$ and a vector $x \in \mathbb{R}^n$, their product is a vector $y = Ax \in \mathbb{R}^m$. There are different ways to understand this

By rows: $Ax = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \\ a_{31}x_1 + a_{32}x_2 \end{bmatrix}$

By columns: $Ax = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix}$

Ax is a linear combination of the columns of A

$$Ax = x_1 a_1 + x_2 a_2$$

Vector-matrix multiplication

Columnspace of A

Given a $A \in \mathbb{R}^{m \times n}$, the linear combinations of the columns fill the column space of A.

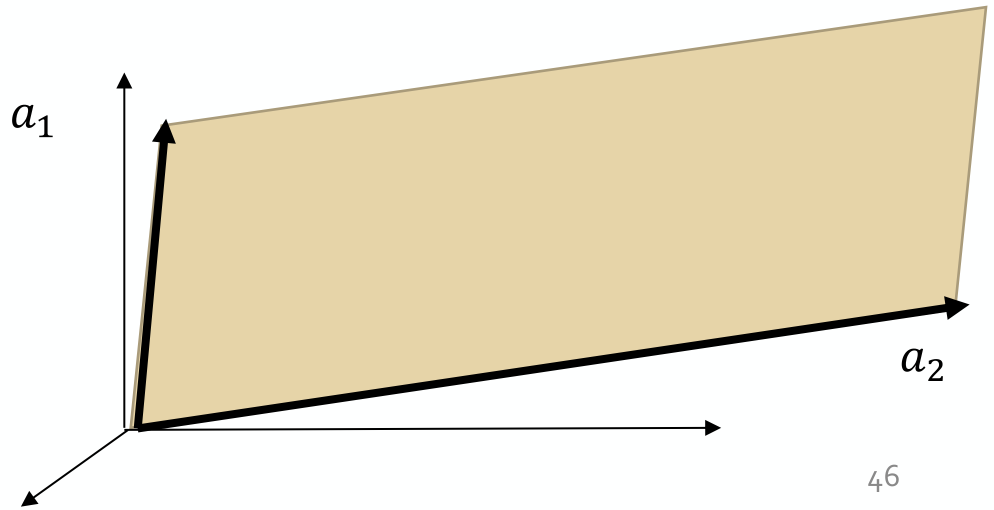
Example:
If $A \in \mathbb{R}^{3 \times 2}$has two linearly independent columns a_1 and a_2, then the linear combination of the columns of A

$$Ax = x_1 a_1 + x_2 a_2$$

corresponds to any point on a plane given by a_1 and a_2.

The columnspace is also called the **range** and it is defined as the span of the columns of A:

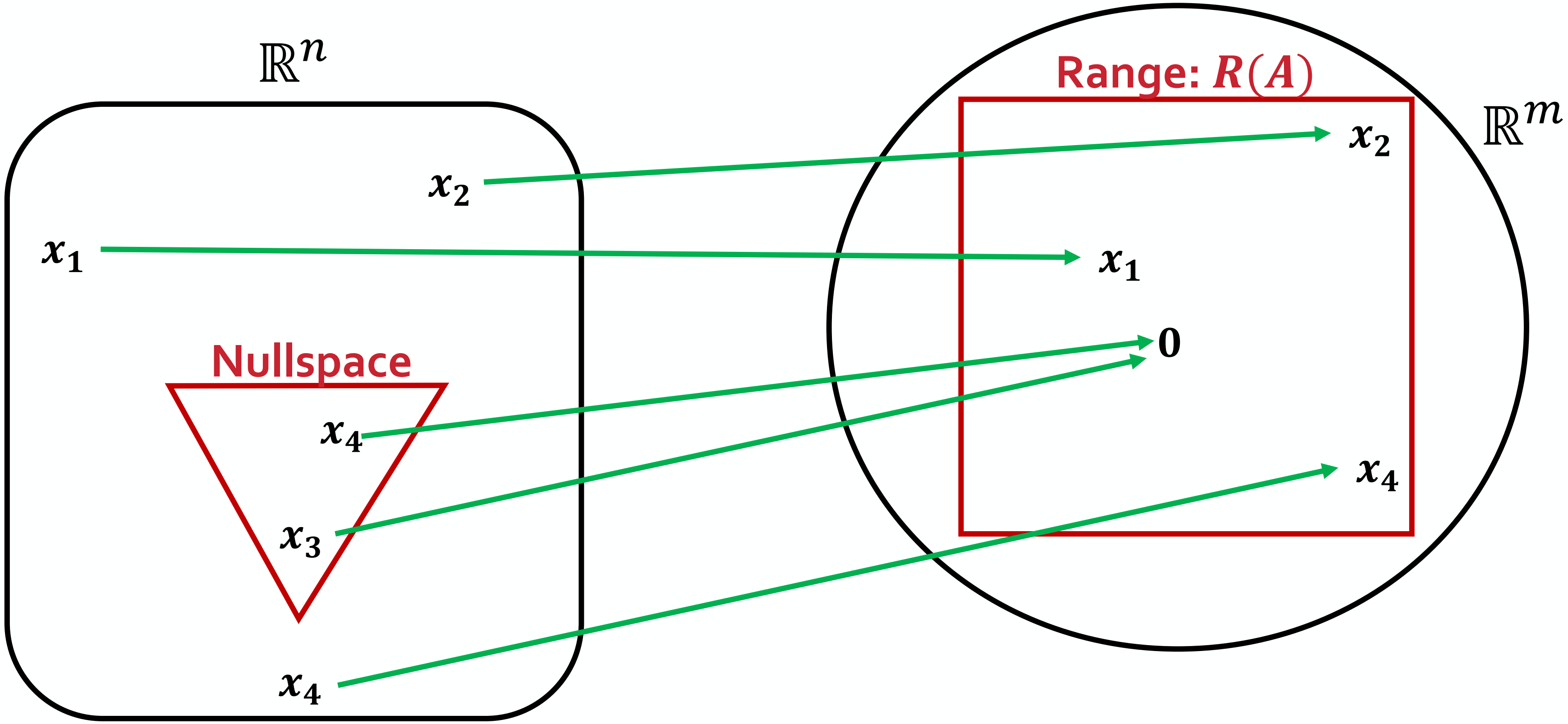
$$\mathcal{R}(A) = \{v \in \mathbb{R}^m : v = Ax, x \in \mathbb{R}^n\}.$$



Hertie School

Vector-matrix multiplication

Let vector $\boldsymbol{x_i} \in \mathbb{R}^n$ be transformed by matrix $A \in \mathbb{R}^{m\times n}$ to $\boldsymbol{y_i} \in \mathbb{R}^m$: $\quad \mathrm{A}\,\boldsymbol{x_i} = \boldsymbol{y_i}$.



Hertie School

More Operations and (their) Properties

Hertie School

Trace

Trace

The trace of a square matrix $A \in \mathbb{R}^{n \times n}$, denoted or is the sum of diagonal elements in the matrix:

$$\mathrm{tr}A = \sum_{i=1}^{n} A_{ii}\,.$$

Properties:

- For $A \in \mathbb{R}^{n \times n}$, $\mathrm{tr}A = \mathrm{tr}A^T$
- For $A, B \in \mathbb{R}^{n \times n}$, $\mathrm{tr}(A + B) = \mathrm{tr}A + \mathrm{tr}B$
- For $A \in \mathbb{R}^{n \times n}$, and $t \in \mathbb{R}$, $\mathrm{tr}(tA) = t\,\mathrm{tr}A$
- For A, B such that AB is square, $\mathrm{tr}AB = \mathrm{tr}BA$
- For A, B, C such that ABC is square, $\mathrm{tr}ABC = \mathrm{tr}BCA = \mathrm{tr}CAB$ and so on (cyclic permutation property)
- Trace trick: $x^T A x = tr(x^T A x) = tr(x x^T A)$

Norms

Definition of a norm

A norm is any function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ that satisfies four properties:

1. For all $x \in \mathbb{R}^n$, $f(x) \geq 0$ (non-negativity)
2. $f(x) = 0$ if and only if $x = 0$ (definiteness)
3. For all $x \in \mathbb{R}^n$, t $\in \mathbb{R}$, $f(tx) = |t| f(x)$ (homogeneity)
4. For all $x, y \in \mathbb{R}^n$, $f(x+y) \leq f(x) + f(y)$ (triangle inequality)

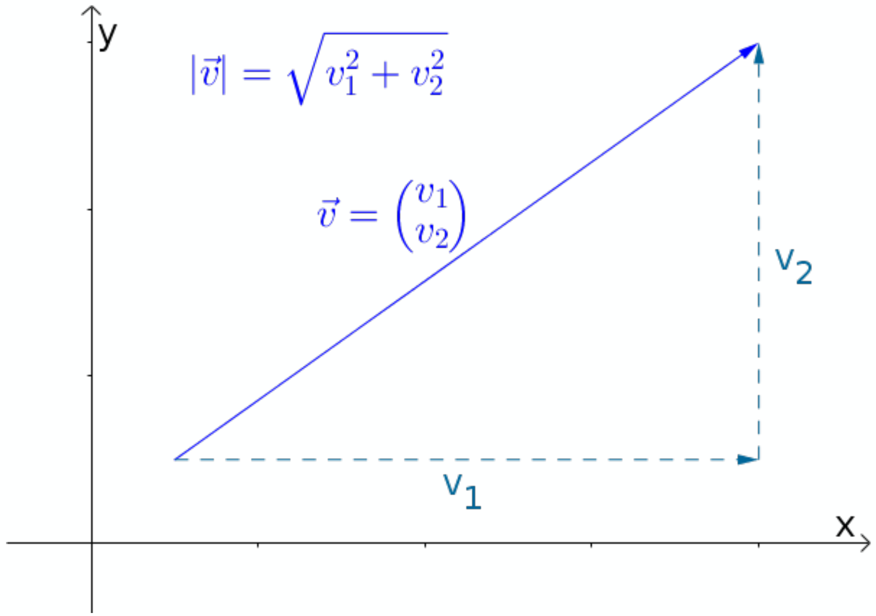
ℓ_2-Norm (Euclidian norm)

The ℓ_2-Norm on $\mathbb{R}^n$ is defined for $x \in \mathbb{R}^n$ as

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

$$\|x\|_2^2 = x^T x$$

This is just the "length" of a vector!



Norms

Taxicab/Manhattan Norm (ℓ_1-Norm)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

Euclidian norm (ℓ_2-Norm)

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

Maximum norm (ℓ_∞-Norm)

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

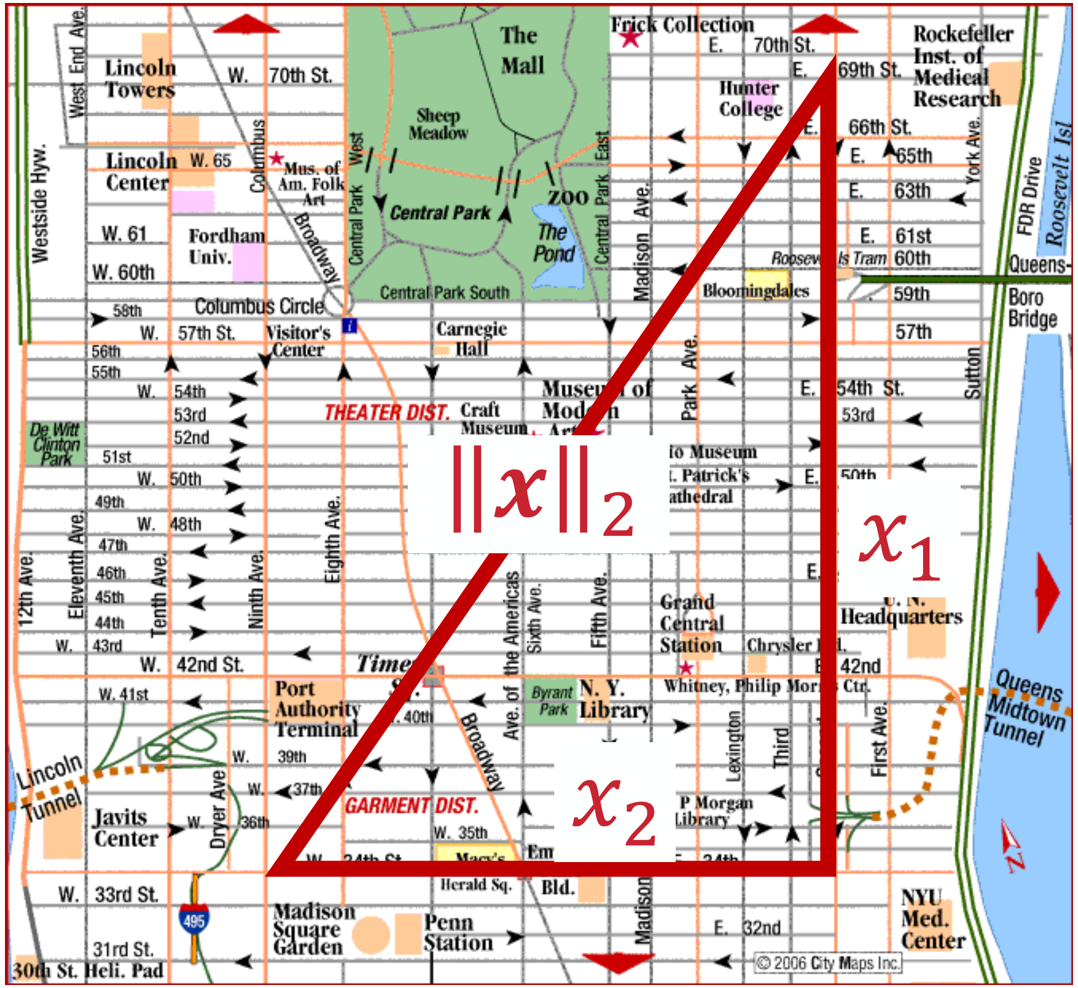


Image: Deisenroth

Norms

Common norms

Norms for vectors

ℓ_1-Norm

$$||x||_1 = \sum_{i=1}^{n} |x_i|$$

ℓ_2-Norm

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

ℓ_∞-Norm

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

ℓ_p-Norm defined by $p \geq 1$

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

Norms for matrices exist, e.g. the Frobenius norm $\|A\|_F = \sqrt{tr(A^T A)}$

With a norm, we can derive a unit vector $u = \frac{v}{\|v\|}$

Inverse

Definition

The inverse of a square matrix $A \in \mathbb{R}^{n \times n}$ is denoted A^{-1}, and it is the unique matrix such that

$$A^{-1}A = I = AA^{-1}.$$

Note: Not all matrices have inverses (e.g., non-square matrices by definition do not).

If A^{-1}exists, then A is **invertible** or non-singular.

Condition for invertibility:

- A square matrix A is invertible if it is full rank.
- A square matrix A is invertible if $\det A \neq 0$.

Inverse

Properties

For square matrices $A, B \in \mathbb{R}^{n\times n}$ that are invertible, we have that

- $(A^{-1})^{-1} = A$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^{-1})^{T} = (A^{T})^{-1}$ This is sometimes abbreviated as $(A^{-1})^{T} = A^{-T}$.
- $\det A^{-1} = (\det\ A)^{-1}$

For a 2x2 matrix we compute the inverse as:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad A^{-1} = \frac{1}{|A|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

For a block diagonal matrix, the inverse is obtained by inverting each block separately, e.g.

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix}$$

Orthogonal matrices

Orthogonal and normalized vectors

Two vectors $x, y \in \mathbb{R}^n$ are **orthogonal** if $x^T y = 0$.
A vector $x \in \mathbb{R}^n$ is **normalized** if $\|x\|_2 = 1$.

Orthogonal matrices

A square matrix $U \in \mathbb{R}^{n \times n}$ is **orthogonal** if all its columns are orthogonal to each other and are normalized. (The column vectors then are referred to as being orthonormal.) It follows that

$$U^T U = I = U U^T.$$

Derivation for $U \in \mathbb{R}^{3\times 3}$: $\begin{bmatrix} u_1^T \\ u_2^T \\ u_3^T \end{bmatrix} [\mathrm{u}_1\ u_2\ u_3] = \begin{bmatrix} u_1^T u_1 & \dots & \dots \\ u_2^T u_1 & u_2^T u_2 & \dots \\ u_3^T u_1 & \dots & u_3^T u_3 \end{bmatrix}$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

Orthogonal matrices

Orthogonal and normalized vectors

Two vectors $x, y \in \mathbb{R}^n$ are **orthogonal** if $x^T y = 0$.
A vector $x \in \mathbb{R}^n$ is **normalized** if $\|x\|_2 = 1$.

Orthogonal matrices

A square matrix $U \in \mathbb{R}^{n\times n}$ is **orthogonal** if all its columns are orthogonal to each other and are normalized. (The column vectors then are referred to as being orthonormal.) It follows that

$$U^T U = I = U U^T.$$

This means that $U^T = U^{-1}$ (the transpose is the inverse for an orthogonal matrix).

Property:
Operating on a vector with an orthogonal matrix will not change its Euclidean norm $\|Ux\|_2 = \|x\|_2$.

Hertie School

Eigenvectors and Eigenvalues

Eigenvectors and eigenvalues

Definition

Given a square matrix $A \in \mathbb{R}^{n\times n}$, we say that $\lambda \in \mathbb{C}$ is an **eigenvalue** of A and $x \in \mathbb{C}^n$ is the corresponding **eigenvector** if

$$A\,x = \lambda\,x, \qquad x \neq 0.$$

The **eigenvalues** can be **determined** as follows:

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

If $(A\ -\lambda\,I)$ had full rank, x can never land on nullspace 0. Hence, its rank is smaller and

$$\det(A\ -\lambda\,I) = 0.$$

This leads to a polynomial equation which roots λ are the **eigenvalues.** Pluging a certain λ_i back gives

$$(A\ -\lambda_i\,I)\,x_i = 0,$$

which can be solved for the corresponding **eigenvector(s)** x_i.

Hertie School

Eigenvectors and eigenvalues

Definition

Given a square matrix $A \in \mathbb{R}^{n\times n}$, we say that $\lambda \in \mathbb{C}$ is an **eigenvalue** of A and $x \in \mathbb{C}^n$ is the corresponding **eigenvector** if

$$Ax = \lambda x, \qquad x \neq 0.$$

Calculation

Characteristic equation of A to find n eigenvalues λ_i:

$$\det(A - \lambda I) = 0.$$

To find associated eigenvectors x_i, we solve the eigenvector equation $(A - \lambda_i I)x_i = 0.$

Remark

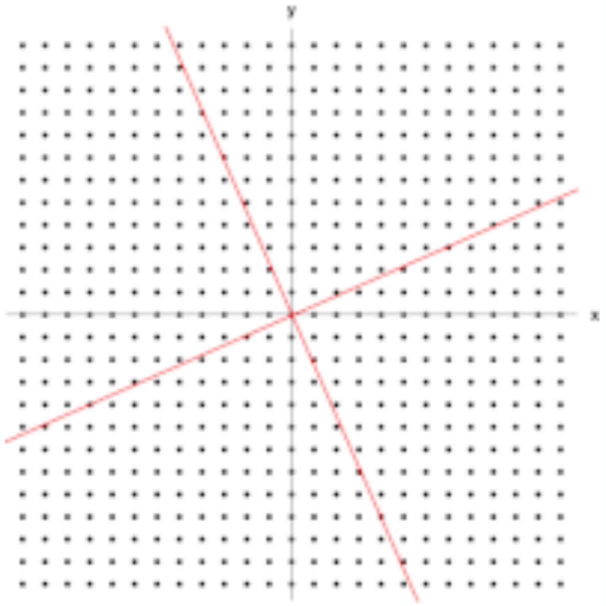
We can write all eigenvector equations simultaneously as

$$AX = X\Lambda,$$

where columns of $X \in \mathbb{R}^{n\times n}$ are the eigenvectors of A, and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$

Visualization

Image: Wikipedia



Hertie School

Eigenvectors and eigenvalues of $A = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$

coupled system

Eigenvalues

$$\det\left(\begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = (1-\lambda)(-1)(1+\lambda) = 0 \quad \rightarrow \quad \lambda_1 = 1,\ \lambda_2 = -1$$

Eigenvector 1

$$\begin{bmatrix} 1-1 & 2 \\ 0 & -1-1 \end{bmatrix} v_1 = \begin{bmatrix} 0 & 2 \\ 0 & -2 \end{bmatrix}\begin{bmatrix} \text{any} \\ 0 \end{bmatrix} = 0 \quad \rightarrow \quad v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

(Check: $A\, v_1 = \lambda_1\, v_1$)

Eigenvector 2

$$\begin{bmatrix} 1+1 & 2 \\ 0 & -1+1 \end{bmatrix} v_2 = \begin{bmatrix} 2 & 2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \alpha \\ -\alpha \end{bmatrix} = 0 \quad \rightarrow \quad v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

(Check: $A\, v_2 = \lambda_2\, v_2$)

Eigenbasis

$$X = [v_1, v_2] = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \quad \rightarrow \quad X^{-1} = \frac{1}{\det(X)}\begin{bmatrix} -1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$$

(Check: $X\, X^{-1} = I$)

$$X^{-1} A\, X = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix} X = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} X = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \mathrm{diag}(\lambda_1, \lambda_2) = \Lambda$$

Basis transformation

decoupled system

→ great to solve differential equation systems

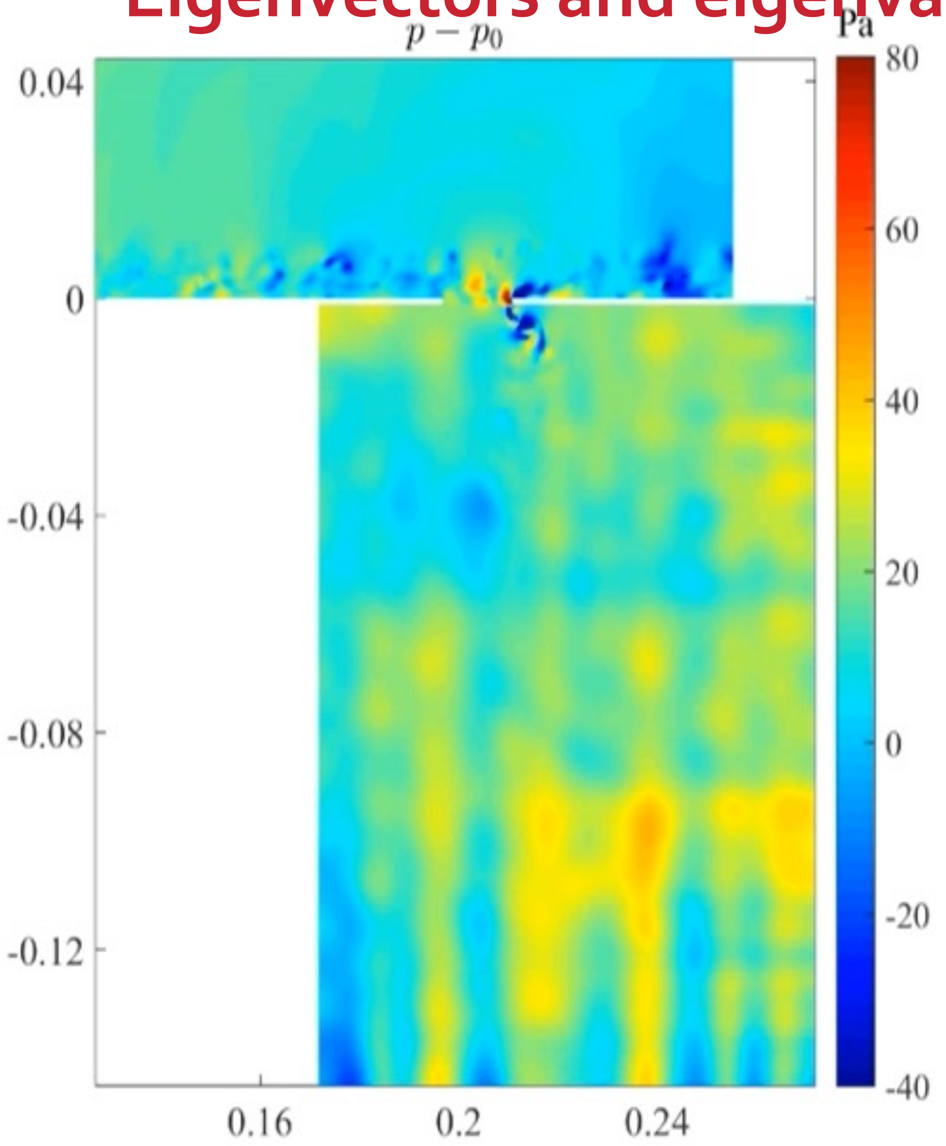
→ great to extract dominant features of a system

Eigenvectors and eigenvalues

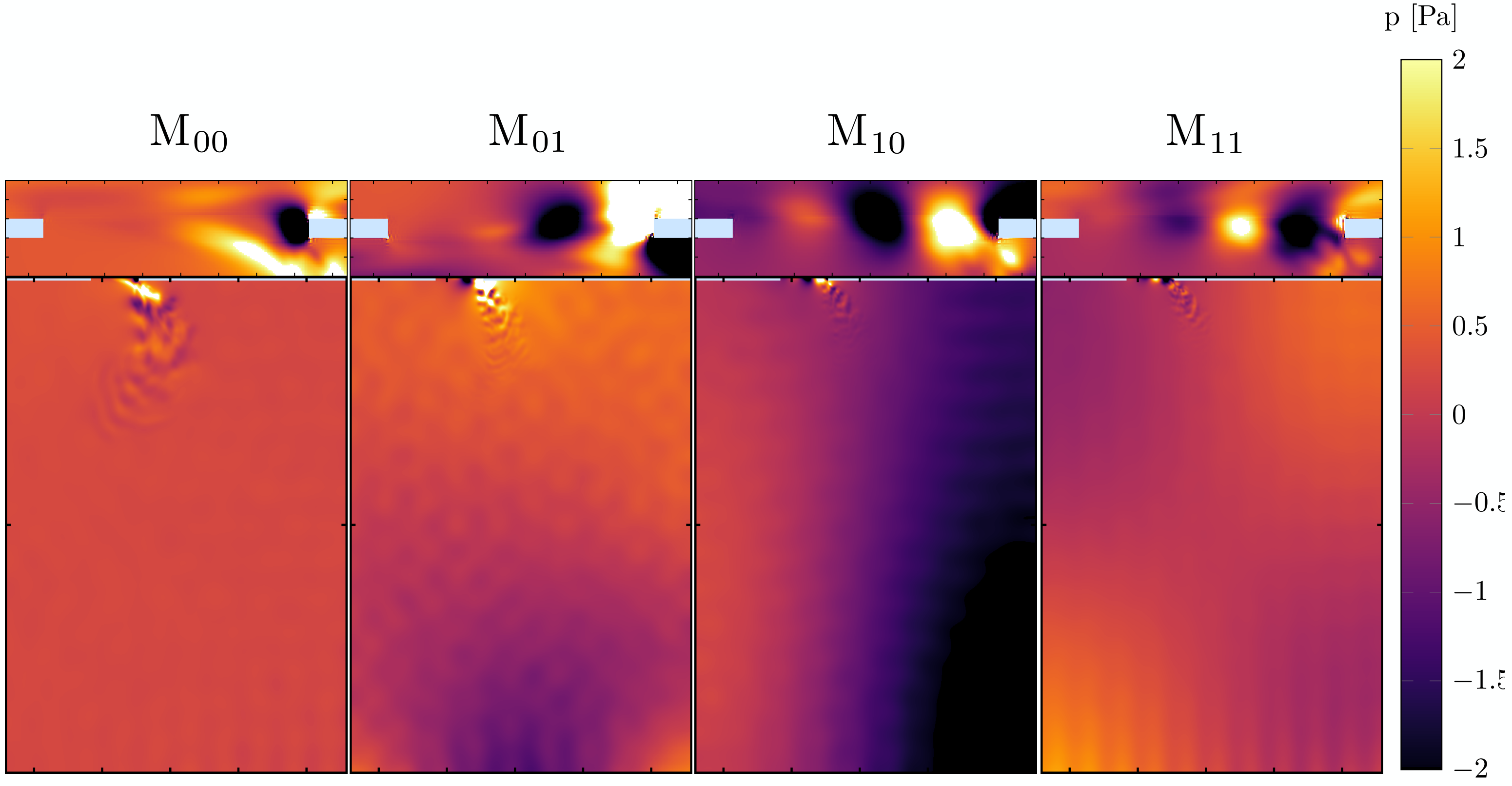
Properties

- The trace of a matrix A is equal to the sum of its **eigenvalues**: $trA = \sum_{i=1}^{n} \lambda_i$
- The determinant of A is equal to the product of its **eigenvalues**: $|A| = \prod_{i=1}^{n} \lambda_i$
- If A is non-singular, then $A^{-1}x_i = \left(\frac{1}{\lambda_i}\right) x_i$.
- The **eigenvalues** of a diagonal matrix $D = diag(d_1, \ldots, d_n)$ are just the diagonal entries $d_1, \ldots, d_n$.

Eigenvectors and eigenvalues



$Let \quad \boldsymbol{A}\, v_{\text{old state}} = v_{\text{ new state}}$ and λ_i be an **eigenvalue** of $\boldsymbol{A}$**.**
$Real(\lambda_i)$ can be interpreted as growth rate of an **eigenvector i.**
$Imag(\lambda_i)$ can be interpreted as frequency of an **eigenvector i.**



We found a coupling between whirls and acoustic modes of the same frequency.

https://doi.org/10.14279/depositonce-8183

Eigenvectors and eigenvalues interpreted geometrically

Eigenvectors and eigenvalues | Chapter 14, Essence of linear algebra | Minute 0:45-10:45

www.youtube.com/watch?v=PFDu9oVAE-g

Backup slides

Hertie School

From equation system (of scalars) to vector and matrix

$$4x_1 - 5x_2 = -13$$
$$-2x_1 + 3x_2 = 9$$

$$\longleftrightarrow$$

$$\boldsymbol{A}\,x = b$$

Scalar

In this example, the variables $x_1$and $x_2 \in \mathbb{R}$ are scalars.

Vector

$x_1$and x_2 can be written as a vector $x \in \mathbb{R}^n$:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

Matrix

We need to multiply x with a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ to get b:

$$\boldsymbol{A} = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}$$