# Advanced Data Analysis with Python

Cecilia Graiff

September 4, 2025

ALMAnaCH, Inria Paris - École d'Affaires Publiques, SciencesPo

# Recap of Core Concepts: Linear Algebra

# Vectors

- A vector is an ordered list of numbers:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

- **Notation**: By $x \in \mathbb{R}^n$, we denote a vector with $n$ entries.
- Components: $x_i$ is the $i$-th element of $\mathbf{x}$.
- Transpose: $\mathbf{x}^\top = [x_1, x_2, \ldots, x_n]$

# What is a Matrix?

- A matrix is a bidimensional arrangements of elements of a set $S$:

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

- **Notation**: By $A \in \mathbb{R}^{m \times n}$, we denote a matrix with $m$ rows and $n$ columns, where the entries of A are real numbers.
- Each element: $a_{ij}$ = row $i$, column $j$
- Special matrices: identity $I$, diagonal $D$, zero 0

# Matrix Operations

- Addition/Subtraction: element-wise

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

# Matrix Operations

- Addition/Subtraction: element-wise

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

- Scalar multiplication: $\alpha A = [\alpha a_{ij}]$

# Matrix Operations

- Addition/Subtraction: element-wise

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

- Scalar multiplication: $\alpha A = [\alpha a_{ij}]$
- Matrix multiplication: $C = AB^{m \times p}$, where

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \quad \text{for } A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$$

# Matrix Operations

- Addition/Subtraction: element-wise

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

- Scalar multiplication: $\alpha A = [\alpha a_{ij}]$
- Matrix multiplication: $C = AB^{m \times p}$, where

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \quad \text{for } A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$$

- Transpose: $(A^\top)_{ij} = a_{ji}$

# Matrix Addition

Given two matrices $A$ and $B$ of the same size, the sum is:

$$C = A + B, \quad c_{ij} = a_{ij} + b_{ij}$$

**Example:**

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

$$A + B = \begin{bmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$$

# Matrix Subtraction

$$C = A - B, \quad c_{ij} = a_{ij} - b_{ij}$$

**Example:**

$$A = \begin{bmatrix} 4 & 5 \\ 6 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

$$A - B = \begin{bmatrix} 4-1 & 5-2 \\ 6-3 & 7-4 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}$$

# Scalar Multiplication

Multiply every element of a matrix by a scalar $k$:

$$kA = [ka_{ij}]$$

**Example:**

$$k = 2, \quad A = \begin{bmatrix} 1 & -1 \\ 0 & 3 \end{bmatrix}$$

$$2A = \begin{bmatrix} 2 & -2 \\ 0 & 6 \end{bmatrix}$$

# Matrix Multiplication

For $A$ of size $m \times n$ and $B$ of size $n \times p$:

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

**Example:**

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$AB = \begin{bmatrix} 1 \cdot 0 + 2 \cdot 1 & 1 \cdot 1 + 2 \cdot 0 \\ 3 \cdot 0 + 4 \cdot 1 & 3 \cdot 1 + 4 \cdot 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}$$

# Transpose of a Matrix

The transpose of $A$ is denoted $A^T$ and flips rows and columns:

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

**Example:**

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

# Inverse of a Matrix

For a square matrix $A$, $A^{-1}$ satisfies:

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

**Example:**

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}, \quad A^{-1} = \frac{1}{2 \cdot 2 - 3 \cdot 1} \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}$$

# Matrix-Vector Multiplication

- Multiply $A \in \mathbb{R}^{m \times n}$ by $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{y} = A\mathbf{x}, \quad y_i = \sum_{j=1}^{n} a_{ij} x_j$$

- Example: regression model in matrix form

$$\mathbf{y} = X\boldsymbol{\beta} + \varepsilon$$

- Key idea: linear transformations

# Important Matrix Properties

- $(A + B)^\top = A^\top + B^\top$
- $(AB)^\top = B^\top A^\top$
- Associativity: $A(BC) = (AB)C$
- Identity: $AI = IA = A$
- Inverse: $AA^{-1} = A^{-1}A = I$ (if $A$ is square and invertible)

# Inner Product (Dot Product)

**Definition:** For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

**Properties:**

- Commutative: $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
- Linear: $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$
- Relation to norm: $\|\mathbf{x}\|^2 = \mathbf{x} \cdot \mathbf{x}$

**Example:**

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \ \mathbf{x} \cdot \mathbf{y} = 1 * 4 + 2 * 5 + 3 * 6 = 32$$

# Outer Product

**Definition:** For $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{y} \in \mathbb{R}^n$,

$$\mathbf{x} \otimes \mathbf{y} = \mathbf{x}\mathbf{y}^\top \in \mathbb{R}^{m \times n}$$

**Example:**

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \; \mathbf{y} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}, \; \mathbf{x}\mathbf{y}^\top = \begin{bmatrix} 1*4 & 1*5 \\ 2*4 & 2*5 \\ 3*4 & 3*5 \end{bmatrix} = \begin{bmatrix} 4 & 5 \\ 8 & 10 \\ 12 & 15 \end{bmatrix}$$

**Notes:**

- Outer product creates a matrix from two vectors.
- Used in covariance matrices: $Cov(\mathbf{x}) = \frac{1}{n}\sum_i (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$

# Applied Example: Inner Product

**Context: Political Science — Voting Similarity**

Suppose we have two legislators' voting records on $n$ bills:

$$\mathbf{x}, \mathbf{y} \in \{0, 1\}^n \quad (1 = \text{yes}, 0 = \text{no})$$

**Inner product:**

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

counts the number of agreements between legislators.

**Example:**

$$\mathbf{x} = [1, 0, 1, 1], \quad \mathbf{y} = [1, 1, 0, 1] \quad \Rightarrow \mathbf{x} \cdot \mathbf{y} = 2$$

**Interpretation:** The two legislators voted the same way on 2 out of 4 bills.

# Applied Example: Outer Product

**Context: Health Data — Covariance Matrix**

Given a centered vector of biomarker measurements for a patient:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

**Outer product:**

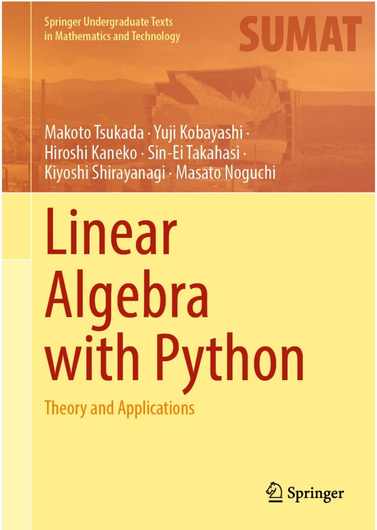
$$\mathbf{x}\mathbf{x}^\top = \begin{bmatrix} x_1^2 & x_1x_2 & x_1x_3 \\ x_2x_1 & x_2^2 & x_2x_3 \\ x_3x_1 & x_3x_2 & x_3^2 \end{bmatrix}$$

**Interpretation:** This is a single-patient contribution to the covariance matrix.

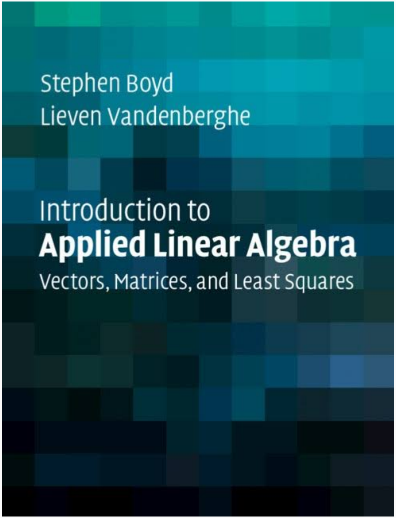
# Inner vs Outer Product

| **Inner Product** | **Outer Product** |
|---|---|
| Scalar result | Matrix result |
| Measures similarity | Constructs linear map |
| $\mathbf{x}^\top \mathbf{y}$ | $\mathbf{x}\mathbf{y}^\top$ |
| Used in projection, norm | Used in covariance, rank-1 matrices |

## Recommended Readings




Linear Algebra with Python




Introduction to Applied Linear Algebra