Linear Algebra for Machine Learning

Paul MINCHELLA, Stéphane CHRÉTIEN
paul.minchella@lyon.unicancer.fr

Motivations for Linear Algebra Review

- Computational engine of mathematics:
  → **Numerical Analysis** *(Finite Elements)*; **Algebraic Geometry** *(Hodge Decomposition)*; **Statistics** *(Covariance Matrix, Data Shape)*
- Data science practitioners: diverse backgrounds
- Refresh key concepts often forgotten (e.g. eigenvalues)

Goal: develop dexterity with

- ✓ Linear Equations, Gaussian Elimination, Matrices
- ✓ Vector Spaces, Transformations, Basis Changes
- ✓ Diagonalization, Webpage Ranking, Covariance
- ✓ Orthogonality, Least Squares, SVD

Motivations for Linear Algebra Review

- Computational engine of mathematics:
  - → **Numerical Analysis** *(Finite Elements)*; **Algebraic Geometry** *(Hodge Decomposition)*; **Statistics** *(Covariance Matrix, Data Shape)*
- Data science practitioners: diverse backgrounds
- Refresh key concepts often forgotten (e.g. eigenvalues)

Goal: develop dexterity with

- ✓ Linear Equations, Gaussian Elimination, Matrices
- ✓ Vector Spaces, Transformations, Basis Changes
- ✓ Diagonalization, Webpage Ranking, Covariance
- ✓ Orthogonality, Least Squares, SVD

Example: Smoking in Smallville

Each year: 30% of nonsmokers start smoking, 20% of smokers quit. Initial population: 8000 smokers, 2000 nonsmokers. Questions:

- Numbers after 100 years?
- Numbers after n years?
- Is there a stable equilibrium?

Core points (why and how)

- **Goal**: reduce to fewer equations/variables via elimination.

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$

- **Basic elimination step (column 1):**
  1. Pivot: swap so $a_{11} \neq 0$.
  2. Normalize: $L_1 \leftarrow \frac{1}{a_{11}} L_1$.
  3. Zero below: for $i \geq 2$, $L_i \leftarrow L_i - a_{i1} L_1$.
- Iterate on remaining submatrix; back-substitute or continue to RREF.

Operations (preserve solution set of $Ax = b$):

Reason: Each operation equals left-multiplication by an invertible *elementary matrix* E, hence

$$Ax = b \iff (EA)x = Eb.$$

- Row swap: $L_i \leftrightarrow L_j$
- Scale: $L_i \leftarrow \lambda L_i$, $\lambda \neq 0$
- Row add: $L_i \leftarrow L_i + \lambda L_j$

Exercise

Solve the system

$$\begin{cases} x + y + z = 3 \\ 2x + y = 7 \\ 3x + 2z = 5 \end{cases}$$

Definition

A *matrix* is an $m \times n$ array of elements, where m is the number of rows and n is the number of columns.

$$A \in \mathcal{M}_{m\times n}(\mathbb{K}) \quad \text{(matrix with entries in a field } \mathbb{K}\text{, e.g. } \mathbb{R}, \mathbb{C}\text{)}.$$

We also write $A \in \mathbb{R}^{m\times n}$ for $\mathbb{K} = \mathbb{R}$.

Key properties of $\mathcal{M}_{m\times n}(\mathbb{K})$

- Vector space over $\mathbb{K}$: addition and scalar multiplication are defined entrywise.
- Dimension: $\dim \mathcal{M}_{m\times n}(\mathbb{K}) = mn$.
- Matrix multiplication defined if inner dimensions match $(A \in \mathcal{M}_{m\times n}, B \in \mathcal{M}_{n\times p})$.
- Multiplication is associative but **not commutative** in general.
- Identity matrix $I_n \in \mathcal{M}_{n\times n}(\mathbb{K})$: $AI_n = I_m A = A$.
- Invertibility only for square matrices $A \in \mathcal{M}_{n\times n}(\mathbb{K})$, with $\det(A) \neq 0$.

Idea

Matrix multiplication corresponds to composing two linear systems:

$$C = AB \quad \Longleftrightarrow \quad \text{Apply } B \text{ first, then } A.$$

Example (two 2×2 systems)

Let

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

First system (apply B to $\begin{pmatrix} x \\ y \end{pmatrix}$):

$$\begin{cases} z = 2x + 0y \\ w = x + 3y \end{cases}$$

Second system (apply A to $\begin{pmatrix} z \\ w \end{pmatrix}$):

$$\begin{cases} u = z + 2w = 4x + 6y \\ v = w = 1x + 3y \end{cases}$$

Overall transformation $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} u \\ v \end{pmatrix}$ is given by

$$C = AB = \begin{bmatrix} 4 & 6 \\ 1 & 3 \end{bmatrix}.$$

Definition: Dot Product

For vectors $\mathbf{v} = [a_1, \ldots, a_n]$ and $\mathbf{w} = [b_1, \ldots, b_n]$, the dot product is

$$\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} a_i b_i, \quad \text{and} \quad |\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

By the law of cosines, vectors $\mathbf{v}, \mathbf{w}$ are orthogonal iff $\mathbf{v} \cdot \mathbf{w} = 0$.

Matrix Multiplication

Let $A \in \mathcal{M}_{m \times n}(\mathbb{K})$ and $B \in \mathcal{M}_{p \times q}(\mathbb{K})$. Matrix multiplication AB is defined when $n = p$. If $(AB)_{ij}$ denotes the (i, j) entry, then

$$(AB)_{ij} = \mathrm{row}_i(A) \cdot \mathrm{col}_j(B).$$

Interpretation: Each matrix represents a linear map in a chosen basis. Therefore, multiplication of matrices (composition of linear maps) and the dot product (row · column) only make sense within the same basis. We will formalize this with the notions of *linear maps* and *basis*, introduced next.

Exercise

$$\begin{bmatrix} 2 & 7 \\ 3 & 3 \\ 1 & 5 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 37 & 46 & 55 & 64 \\ * & * & * & * \\ * & * & * & * \end{bmatrix}$$

Fill in the missing entries.

Definitions

- Transpose: $(A^\top)_{ij} = A_{ji}$.
  - Linearity: $(\alpha A + \beta B)^\top = \alpha A^\top + \beta B^\top$.
  - Involution: $(A^\top)^\top = A$.
  - Product rule: $(AB)^\top = B^\top A^\top$.
- A is **symmetric** if $A^\top = A$.
- A is **diagonal** if $A_{ij} \neq 0 \Rightarrow i = j$.
  - If A, B diagonal $\in \mathcal{M}_{n\times n}(\mathbb{K})$: $AB = BA, \quad (AB)_{ii} = a_{ii}\, b_{ii}$.
- Identity: $I_n = \mathrm{diag}(1, \ldots, 1),\ I_n A = A = A I_m$.
- Inverse: $A \in \mathcal{M}_{n\times n}$ is invertible $\iff \exists\, B$ s.t. $BA = AB = I_n$. Denote $A^{-1} = B$.

Definitions

- Transpose: $(A^\top)_{ij} = A_{ji}$.
  - Linearity: $(\alpha A + \beta B)^\top = \alpha A^\top + \beta B^\top$.
  - Involution: $(A^\top)^\top = A$.
  - Product rule: $(AB)^\top = B^\top A^\top$.
- A is **symmetric** if $A^\top = A$.
- A is **diagonal** if $A_{ij} \neq 0 \Rightarrow i = j$.
  - If A, B diagonal $\in \mathcal{M}_{n\times n}(\mathbb{K})$: $AB = BA, \quad (AB)_{ii} = a_{ii}\, b_{ii}$.
- Identity: $I_n = \mathrm{diag}(1, \ldots, 1)$, $I_n A = A = A I_m$.
- Inverse: $A \in \mathcal{M}_{n\times n}$ is invertible $\iff \exists\, B$ s.t. $BA = AB = I_n$. Denote $A^{-1} = B$.

Exercise

- Find 2×2 matrices (A, B) with $AB \neq BA$.
- Show $(AB)^\top = B^\top A^\top$. Deduce that $A^\top A$ is symmetric.

Definitions

- Transpose: $(A^\top)_{ij} = A_{ji}$.
    - Linearity: $(\alpha A + \beta B)^\top = \alpha A^\top + \beta B^\top$.
    - Involution: $(A^\top)^\top = A$.
    - Product rule: $(AB)^\top = B^\top A^\top$.
- A is **symmetric** if $A^\top = A$.
- A is **diagonal** if $A_{ij} \neq 0 \Rightarrow i = j$.
    - If A, B diagonal $\in \mathcal{M}_{n\times n}(\mathbb{K})$: $AB = BA, \quad (AB)_{ii} = a_{ii}\, b_{ii}$.
- Identity: $I_n = \mathrm{diag}(1, \ldots, 1)$, $I_n A = A = A I_m$.
- Inverse: $A \in \mathcal{M}_{n\times n}$ is invertible $\iff \exists\, B$ s.t. $BA = AB = I_n$. Denote $A^{-1} = B$.

Exercise

- Find 2×2 matrices (A, B) with $AB \neq BA$.
- Show $(AB)^\top = B^\top A^\top$. Deduce that $A^\top A$ is symmetric.

Matrix form of a linear system

A system of n linear equations in m unknowns is written as

Smoking in Smallville

Let $(n_t, s_t)^\top$ = (# nonsmokers , # smokers) at year t. Transition rule:

$$\begin{bmatrix} n_{t+1} \\ s_{t+1} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix} \begin{bmatrix} n_t \\ s_t \end{bmatrix}.$$

By iteration:

$$\begin{bmatrix} n_t \\ s_t \end{bmatrix} = \left(\begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix} \right)^t \begin{bmatrix} n_0 \\ s_0 \end{bmatrix}.$$

To study $t \gg 0$, we need to compute A^t with

$$A = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}.$$

Key Trick: Diagonalization

There exists a change-of-basis matrix B such that

$$B\,A\,B^{-1} = D \quad \text{(diagonal)}.$$

Then, for any integer m, we get

$$A^m = B^{-1}\,D^m\,B,$$

and computing D^m is easy (just raise diagonal entries).

$\Rightarrow$ Expensive repeated multiplications become trivial if A is diagonalizable.

Number Sets

- **Natural numbers:** $\mathbb{N} = \{0, 1, 2, 3, \dots\}$, non-negative integers.
- **Integers:** $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$, all negative and positive whole numbers, including 0.
- **Rational numbers:** $\mathbb{Q} = \left\{ \frac{p}{q} \,:\, p \in \mathbb{Z},\ q \in \mathbb{Z}^*,\ q \neq 0 \right\}$, ratios of integers.
- **Real numbers:** $\mathbb{R}$ is the *completion* of $\mathbb{Q}$; a totally ordered complete field. π, e, $\sqrt{2}$, $\varphi = \frac{1+\sqrt{5}}{2}$ are real numbers that are irrational.
- **Complex numbers:** $\mathbb{C} = \{x + \mathrm{i}y : x, y \in \mathbb{R},\ \mathrm{i}^2 = -1\}$, an algebraic extension of $\mathbb{R}$.

Hierarchy

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

Binary Operation (Internal Law)

Let E be a set. An *internal binary operation* on E is a map

$$\star : E \times E \to E, \qquad (x, y) \mapsto x \star y.$$

Examples: $+$ and $\times$ on $\mathbb{Z}$.

Group

A *group* is a pair $(G, \star)$ where $\star$ is an internal binary operation satisfying:

- **Associativity:** $(x \star y) \star z = x \star (y \star z)$.
- **Identity element:** there exists $e \in G$ such that $x \star e = e \star x = x$.
- **Inverse:** every $x \in G$ has an inverse x^{-1} with $x \star x^{-1} = e$.

If $x \star y = y \star x$ for all x, y, the group is *abelian*.

Name some familiar examples!

Example

$(\mathbb{Z}, +)$ is an abelian group. $(\mathbb{Z}, \times)$ is *not* a group: not every integer has a multiplicative inverse in $\mathbb{Z}$.

Ring

A *ring* $(A, +, \times)$ is a set with two operations:

- $(A, +)$ is an abelian group.
- $\times$ is associative and has a multiplicative identity 1.
- $\times$ distributes over $+$.

Field

A *field* is a ring $(K, +, \times)$ in which every nonzero element has a multiplicative inverse.

Examples

$\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are fields. $\mathbb{Z}$ is a ring but not a field.

Problem

Define, for $a, b \in \mathbb{Z}$, the operation

$$a \star b = a + b + 1.$$

1. Show that $\star$ is an *internal* binary operation on $\mathbb{Z}$.
2. Check associativity and commutativity of $\star$.
3. Determine the identity element e for $\star$.
4. For a given $a \in \mathbb{Z}$, find the inverse of a with respect to $\star$.
5. Conclude: is $(\mathbb{Z}, \star)$ a group? Is it abelian?

Your solution?

To fully grasp and master the notion of matrices, we need to introduce the formal framework that governs them: ***vector spaces*** *and* ***linear maps***.

Definition (Vector Space)

Let $\mathbb{K}$ be a field (e.g. $\mathbb{R}$ or $\mathbb{C}$). A *vector space over* $\mathbb{K}$ is a set V equipped with:

- an internal operation $+$ (vector addition)
$$(x, y) \mapsto x + y,$$
- an external operation (scalar multiplication)
$$\mathbb{K} \times V \to V, \qquad (\lambda, x) \mapsto \lambda x,$$

such that the following axioms hold for all $x, y, z \in V$ and $\lambda, \mu \in \mathbb{K}$:

1. $(V, +)$ is an abelian group (associativity, commutativity, neutral element 0, inverses).
2. Scalar compatibility: $(\lambda\mu)x = \lambda(\mu x)$.
3. Neutral element of scalars: $1_{\mathbb{K}} x = x$.
4. Distributivity: $(\lambda + \mu)x = \lambda x + \mu x$, and $\lambda(x + y) = \lambda x + \lambda y$.

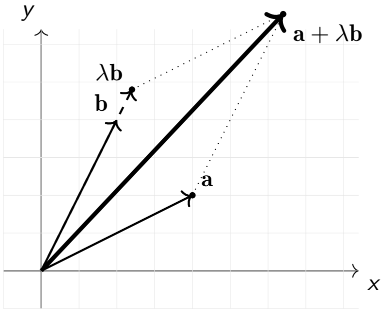
Examples of Vector Spaces

- $\mathbb{K}^n$ (the n-tuples of scalars from $\mathbb{K}$) is a vector space. Indeed:

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \lambda \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + \lambda b_1 \\ \vdots \\ a_n + \lambda b_n \end{pmatrix} \in \mathbb{K}^n$$

- The set $\mathbb{R}[X]$ of real-coefficient polynomials is a vector space over $\mathbb{R}$.
- The set $\mathcal{C}^0([a, b], \mathbb{R})$ of continuous functions from $[a, b]$ to $\mathbb{R}$ is a vector space.

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \lambda \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + \lambda b_1 \\ a_2 + \lambda b_2 \end{pmatrix} \in \mathbb{R}^2$$



Definition (Subspace)

Let V be a vector space over $\mathbb{K}$. A subset $F \subset V$ is a *subspace* if:

- $0_V \in F$,
- $\forall x, y \in F$, then $x + y \in F$ (closed under addition),
- $\forall \lambda \in \mathbb{K},\ x \in F$, then $\lambda x \in F$ (closed under scalar multiplication).

Definition (Subspace)

Let V be a vector space over $\mathbb{K}$. A subset $F \subset V$ is a *subspace* if:

- $0_V \in F$,
- $\forall x, y \in F$, then $x + y \in F$ (closed under addition),
- $\forall \lambda \in \mathbb{K},\ x \in F$, then $\lambda x \in F$ (closed under scalar multiplication).

Examples

- In $\mathbb{R}^3$, the set of vectors $(x, y, 0)$ is a subspace.
- $\{0\}$ and V are always subspaces (trivial subspaces).

Definition (Subspace)

Let V be a vector space over $\mathbb{K}$. A subset $F \subset V$ is a *subspace* if:

- $0_V \in F$,
- $\forall x, y \in F$, then $x + y \in F$ (closed under addition),
- $\forall \lambda \in \mathbb{K},\ x \in F$, then $\lambda x \in F$ (closed under scalar multiplication).

Examples

- In $\mathbb{R}^3$, the set of vectors $(x, y, 0)$ is a subspace.
- $\{0\}$ and V are always subspaces (trivial subspaces).

Properties

- If F, G are subspaces of V, then $F \cap G$ is also a subspace.
- The **sum** of subspaces is

$$F + G = \{x + y : x \in F,\ y \in G\},$$

which is again a subspace.

Span of a Set

Given a subset $A \subset V$, the set of all finite linear combinations of elements of A forms a subspace of V, denoted

$$\mathrm{Span}(A) = \{\lambda_1 v_1 + \cdots + \lambda_k v_k \mid k \in \mathbb{N},\ v_i \in A,\ \lambda_i \in \mathbb{K}\}.$$

It is called the *subspace spanned by A*.

Span of a Set

Given a subset $A \subset V$, the set of all finite linear combinations of elements of A forms a subspace of V, denoted

$$\mathrm{Span}(A) = \{\lambda_1 v_1 + \cdots + \lambda_k v_k \mid k \in \mathbb{N},\ v_i \in A,\ \lambda_i \in \mathbb{K}\}.$$

It is called the *subspace spanned by A*.

Linear Independence and Generating Set

- A family $(v_1, \ldots, v_p)$ is *linearly independent* if the only relation

$$\lambda_1 v_1 + \cdots + \lambda_p v_p = 0 \Rightarrow \forall\, i,\ \lambda_i = 0.$$

- It is a *generating set* of V if

$$\mathrm{Span}(v_1, \ldots, v_p) = V.$$

Definition

Let V be a vector space over a field $\mathbb{K}$.

- A **basis** of V is a family of vectors $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ in V that is
  1. linearly independent,
  2. and generates V (*i.e.*, $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = V$).
- The **dimension** of V, denoted $\dim(V)$, is the number of vectors in any basis of V.

Remark. The definition is well-posed: every two bases of a finite-dimensional vector space V have the same number of elements.

Theorem (Dimension of a Subspace)

Let F be a subspace of a finite-dimensional vector space V. Then

$$\dim(F) \leq \dim(V).$$

Moreover, if $F \neq V$, the inequality is strict.

Grassmann Formula

If F, G are finite-dimensional subspaces of V, then

$$\dim(F + G) = \dim(F) + \dim(G) - \dim(F \cap G).$$

Exercise

Consider the set

$$\mathcal{F} = \{[1,0],\ [0,1],\ [2,3]\} \subset \mathbb{R}^2.$$

- Is $\mathcal{F}$ linearly independent?
- Is $\mathcal{F}$ a generating family of $\mathbb{R}^2$?
- What is the cardinality of a maximal linearly independent subfamily of $\mathcal{F}$, *i.e.*, the dimension of $\mathrm{Span}(\mathcal{F})$?

Exercise

Consider the set

$$\mathcal{F} = \{[1,0],\ [0,1],\ [2,3]\} \subset \mathbb{R}^2.$$

- Is $\mathcal{F}$ linearly independent?
- Is $\mathcal{F}$ a generating family of $\mathbb{R}^2$?
- What is the cardinality of a maximal linearly independent subfamily of $\mathcal{F}$, *i.e.*, the dimension of $\mathrm{Span}(\mathcal{F})$?

Canonical Example

In $\mathbb{R}^n$, the canonical family

$$e_1 = (1,0,\ldots,0), \quad e_2 = (0,1,0,\ldots,0),\ \ldots,\ e_n = (0,\ldots,0,1)$$

is a basis of $\mathbb{R}^n$, and its dimension is n.

Theorem (Incomplete Basis Theorem)

Let V be a finite-dimensional vector space with $\dim(V) = n$. Suppose

$$(v_1, \ldots, v_p), \quad p < n,$$

is a linearly independent family of vectors in V.
Then there exist additional vectors $v_{p+1}, \ldots, v_n \in V$ such that

$$(v_1, \ldots, v_p, v_{p+1}, \ldots, v_n)$$

is a basis of V.
In other words, ***any linearly independent family can be extended to a basis****.*

Theorem (Incomplete Basis Theorem)

Let V be a finite-dimensional vector space with $\dim(V) = n$. Suppose

$$(v_1, \dots, v_p), \quad p < n,$$

is a linearly independent family of vectors in V.
Then there exist additional vectors $v_{p+1}, \dots, v_n \in V$ such that

$$(v_1, \dots, v_p, v_{p+1}, \dots, v_n)$$

is a basis of V.
In other words, ***any linearly independent family can be extended to a basis****.*

Example

In $\mathbb{R}^3$, consider $\{\,(1,0,0),\,(0,1,0)\,\}$. This family is linearly independent but not a basis (only $p = 2 < n = 3$).
Adding $(0,0,1)$ yields

$$\{(1,0,0),\,(0,1,0),\,(0,0,1)\},$$

which forms the canonical basis of $\mathbb{R}^3$.

Theorem (Extracted Basis Theorem)

Let V be a finite-dimensional vector space with $\dim(V) = n$. Suppose

$$(v_1, \dots, v_p), \quad p \geq n,$$

is a generating family of V.
Then there exists a subfamily

$$(v_{i_1}, \dots, v_{i_n})$$

that forms a basis of V.
In other words, ***any generating family contains a basis****.*

Example

In $\mathbb{R}^3$, consider the generating family

$$\{(1,0,0),\ (0,1,0),\ (0,0,1),\ (1,1,1)\}.$$

This set spans $\mathbb{R}^3$. By removing the redundant vector $(1,1,1)$, we obtain the canonical basis

$$\{(1,0,0),\ (0,1,0),\ (0,0,1)\},$$

Conceptual Motivation

A **linear transformation** is a mapping between vector spaces that preserves their structure. It takes a vector as input and produces another vector, in such a way that:

- vector addition is preserved,
- scalar multiplication is preserved.

These maps are fundamental because they capture the essence of "structure-preserving" operations in linear algebra.

Conceptual Motivation

A **linear transformation** is a mapping between vector spaces that preserves their structure. It takes a vector as input and produces another vector, in such a way that:

- vector addition is preserved,
- scalar multiplication is preserved.

These maps are fundamental because they capture the essence of "structure-preserving" operations in linear algebra.

Definition

Let V and W be vector spaces. A map $T : V \to W$ is a *linear transformation* if

$$T(c\mathbf{v}_1 + \mathbf{v}_2) = c\,T(\mathbf{v}_1) + T(\mathbf{v}_2), \quad \forall\, \mathbf{v}_1, \mathbf{v}_2 \in V,\ c \in \mathbb{K}.$$

Conceptual Motivation

A **linear transformation** is a mapping between vector spaces that preserves their structure. It takes a vector as input and produces another vector, in such a way that:

- vector addition is preserved,
- scalar multiplication is preserved.

These maps are fundamental because they capture the essence of "structure-preserving" operations in linear algebra.

Definition

Let V and W be vector spaces. A map $T : V \to W$ is a *linear transformation* if

$$T(c\mathbf{v}_1 + \mathbf{v}_2) = c\,T(\mathbf{v}_1) + T(\mathbf{v}_2), \quad \forall\, \mathbf{v}_1, \mathbf{v}_2 \in V,\; c \in \mathbb{K}.$$

Exercise

Check whether the map T is a linear transformation, where $T : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by

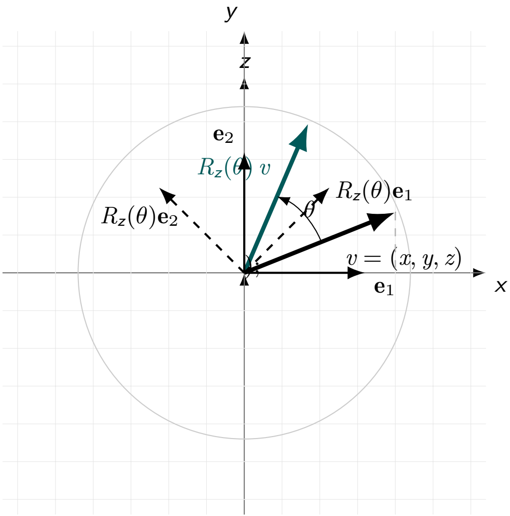
$$T(x, y) = (2x + y,\; -x + 3y)$$

Canonical Examples

Let $V = \mathbb{R}^3$.

- **Identity** $Id : \mathbb{R}^3 \to \mathbb{R}^3$, $Id(\mathbf{x}) = \mathbf{x}$. Matrix: I_3.
- **Homothety (Scaling)** $H_\alpha(\mathbf{x}) = \alpha\,\mathbf{x}$ for $\alpha \in \mathbb{R}$. Matrix: αI_3.
- **Rotation about the *z*-axis by angle θ**

$$R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad R_z(\theta)\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \\ z \end{bmatrix}.$$



$$R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R_z(\theta) \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \\ z \end{bmatrix}$$

Definitions

Let V, W be vector spaces.

- $\mathcal{L}(V, W) :=$ set of linear maps $T : V \to W$.
- **Endomorphism**: $T \in \mathcal{L}(V, V)$.
- **Isomorphism**: bijective linear map $T \in \mathcal{L}(V, W)$.
- **Automorphism**: bijective endomorphism $T \in \mathcal{L}(V, V)$.

Algebraic Structure

On $\mathcal{L}(V) := \mathcal{L}(V, V)$,

- With pointwise addition $+$ and composition $\circ$, $(\mathcal{L}(V), +, \circ)$ is a (not-necessarily commutative) **ring** with identity Id.
- Distributivity: $T \circ (S_1 + S_2) = T \circ S_1 + T \circ S_2$ and $(S_1 + S_2) \circ T = S_1 \circ T + S_2 \circ T$.

Definitions

For $T \in \mathcal{L}(V, W)$:

$$\ker(T) := \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}_W\}, \qquad \operatorname{Im}(T) := \{T(\mathbf{v}) \mid \mathbf{v} \in V\} \subseteq W.$$

Key Properties

- T is injective $\iff \ker(T) = \{\mathbf{0}_V\}$.
- If $(\mathbf{v}_i)_{i \in I}$ generates V, then $\operatorname{Im}(T) = \operatorname{Span}\big(T(\mathbf{v}_i) \ : \ i \in I\big)$.

Rank–nullity theorem

Let $T \in \mathcal{L}(V, W)$ with $\dim(V) < \infty$. Then

$$\underbrace{\dim \ker(T)}_{\text{nullity}} + \underbrace{\dim \operatorname{Im}(T)}_{\text{rank}} = \dim(V).$$

Consequences

- T injective $\iff \dim \ker(T) = 0 \iff \operatorname{rank}(T) = \dim(V)$.
- If $\dim(W) < \infty$, then T surjective $\iff \operatorname{rank}(T) = \dim(W)$.
- If $\dim(V) = \dim(W)$, then: injective $\iff$ surjective $\iff$ T is an isomorphism.

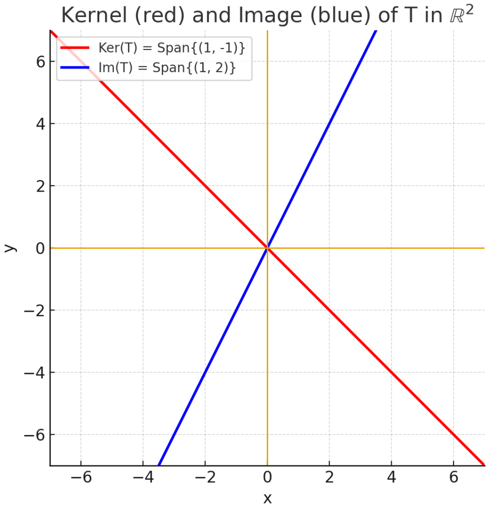
In $\mathbb{R}^2$

Consider the following linear map:

$$T : \mathbb{R}^2 \to \mathbb{R}^2$$
$$(x, y) \mapsto (x + y,\ 2x + 2y)$$

1. What type of morphism is T (e.g., injective, surjective, isomorphism) ? Prove the linear aspect.
2. Determine $\ker(T)$ and $\mathrm{Im}(T)$. Is the rank-null theorem verified ?
3. Plot $\ker(T)$ in red and $\mathrm{Im}(T)$ in blue in the Cartesian plane, and provide their equations. Interpret your result.

Geometric interpretation



- Every vector $v \in \mathbb{R}^2$ is sent *onto the blue line*: $T(v) \in \operatorname{im}(T) = \operatorname{Span}\{(1,2)\}$...
- ... except vectors on the red line that are mapped to the origin: $T(v) = 0 \iff v \in \ker(T) = \operatorname{Span}\{(1,-1)\}$.

Hence T is a rank-1 linear map that *collapses the plane* onto $\operatorname{Span}\{(1,2)\}$ along $\operatorname{Span}\{(1,-1)\}$.

Definitions

Let U, V be subspaces of a vector space E.

$$U + V = \{u + v : \; u \in U, \; v \in V\}.$$

We say that E **is the direct sum of** U **and** V, written $E = U \oplus V$, if

$$E = U + V \quad \text{and} \quad U \cap V = \{0\}.$$

In this case, every decomposition $x = u + v$ is *unique*.

Criteria and Dimensions

For finite-dimensional spaces:

$$E = U \oplus V \iff E = U + V \text{ and } U \cap V = \{0\}. \qquad \dim(U \oplus V) = \dim U + \dim V.$$

Rank-nullity theorem

Let $T : E \to E$ be linear with $\dim E < \infty$.

$$\dim \ker T + \underbrace{\dim \operatorname{im} T}_{\mathrm{rg}(T)} = \dim E.$$

If moreover $\ker T \cap \operatorname{im} T = \{0\}$, then

$$E = \ker T \oplus \operatorname{im} T.$$

Useful cases:

- If T is a **projection** ($T^2 = T$), then $E = \ker T \oplus \operatorname{im} T$.
- If T is **symmetric** (real matrix A with $A^\top = A$), then $(\operatorname{im} T)^\perp = \ker T$ and thus $E = \ker T \overset{\perp}{\oplus} \operatorname{im} T$.

Check the previous example to see an application of this theorem with $\mathbb{R}^2 = \ker T \oplus \operatorname{im} T$!

Definition

Let $T : V \to W$ be a linear transformation, where V and W are vector spaces with bases $\mathcal{B}_V = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ and $\mathcal{B}_W = \{\mathbf{f}_1, \dots, \mathbf{f}_m\}$, respectively.
The **matrix of T with respect to these bases** is the $m \times n$ matrix

$$M_{ij} \text{ such that } T(\mathbf{e}_j) = \sum_{i=1}^{m} M_{ij}\,\mathbf{f}_i.$$

Equivalently, the j-th column of the matrix is the coordinate vector of $T(\mathbf{e}_j)$ in the basis $\mathcal{B}_W$.

Examples

What are the matrix of the following linear maps?

- $\mathbb{R}^2 \to \mathbb{R}^2$: $T(x, y) = (2x + y,\ -x + 3y)$
- $\mathbb{R}^3 \to \mathbb{R}^3$: $T(x, y, z) = (x + z,\ y,\ 2z)$
- $\mathbb{R}^2 \to \mathbb{R}^3$: $T(x, y) = (x,\ y,\ x + y)$

Motivation

Although a linear transformation is an abstract object, once a basis is chosen it can be represented by a matrix. Different choices of basis yield different representations. This motivates the study of **change of basis**.

Motivation

Although a linear transformation is an abstract object, once a basis is chosen it can be represented by a matrix. Different choices of basis yield different representations. This motivates the study of **change of basis**.

Example

The sets

$$B_1 = \{[1,0],[1,1]\}, \quad B_2 = \{[1,1],[1,-1]\}$$

are both bases of $\mathbb{K}^2$. Let $\mathbf{v}_{B_i}$ denote the coordinates of $\mathbf{v}$ in basis B_i.
For example:

$$[0,1]_{B_1} = 0\cdot[1,0] + 1\cdot[1,1] = 1\cdot[1,1] + 0\cdot[1,-1] = [1,0]_{B_2}.$$

Motivation

Although a linear transformation is an abstract object, once a basis is chosen it can be represented by a matrix. Different choices of basis yield different representations. This motivates the study of **change of basis**.

Example

The sets

$$B_1 = \{[1,0],[1,1]\}, \quad B_2 = \{[1,1],[1,-1]\}$$

are both bases of $\mathbb{K}^2$. Let $\mathbf{v}_{B_i}$ denote the coordinates of $\mathbf{v}$ in basis B_i.
For example:

$$[0,1]_{B_1} = 0 \cdot [1,0] + 1 \cdot [1,1] = 1 \cdot [1,1] + 0 \cdot [1,-1] = [1,0]_{B_2}.$$

Algorithm: Coordinates in a Basis

To write a vector $\mathbf{b}$ in terms of a basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$:

1. Form the matrix A with columns $\mathbf{v}_i$.
2. Solve $A\mathbf{x} = \mathbf{b}$ (e.g. by Gaussian elimination).

Exercise

Write the vector $[2, 1]$ in terms of the bases B_1 and B_2 above.

Exercise

Write the vector $[2,1]$ in terms of the bases B_1 and B_2 above.

General Principle

To represent a linear transformation $T: V \to W$, we need bases for both spaces:

$$B_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \quad \text{(basis of } V\text{)}, \quad B_2 = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\} \quad \text{(basis of } W\text{)}.$$

The matrix $M_{B_2 B_1}$ representing T (input in B_1, output in B_2) has as its i^{th} column the coordinates of $T(\mathbf{v}_i)$ expressed in basis B_2.

Example: Rotation by 90° in $\mathbb{R}^2$

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ rotate vectors counterclockwise by 90°.

- With respect to the standard basis $B_1 = \{[1,0],[0,1]\}$:

$$T([1,0]) = [0,1], \quad T([0,1]) = [-1,0],$$

hence

$$T_{B_1 B_1} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

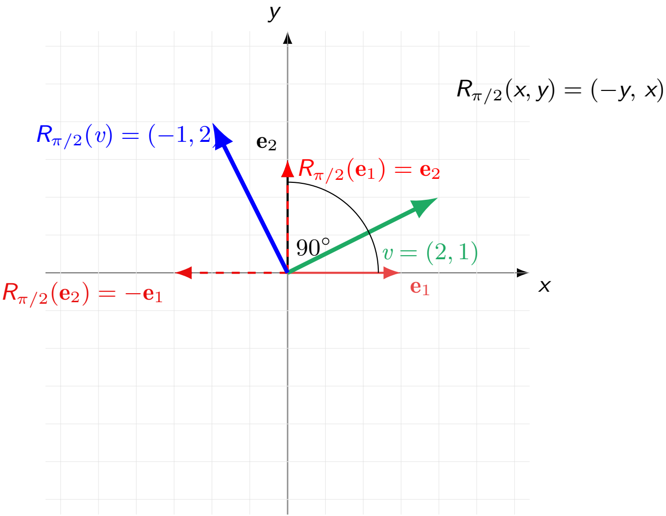
- Using B_1 (input) and $B_2 = \{[1,1],[1,-1]\}$ (output, from Example 3):

$$T([1,0]) = [0,1] = \tfrac{1}{2}[1,1] - \tfrac{1}{2}[1,-1],$$

$$T([1,1]) = [-1,1] = 0 \cdot [1,1] - 1 \cdot [1,-1],$$

so

$$M_{B_2 B_1} = \begin{bmatrix} 1/2 & 0 \\ 1/2 & -1 \end{bmatrix}.$$


y
$R_{\pi/2}(x, y) = (-y,\, x)$
$R_{\pi/2}(v) = (-1, 2)$
$\mathbf{e}_2$
$R_{\pi/2}(\mathbf{e}_1) = \mathbf{e}_2$
$90°$
$v = (2, 1)$
$R_{\pi/2}(\mathbf{e}_2) = -\mathbf{e}_1$
$\mathbf{e}_1$
x

Suppose we have the coordinates of a vector with respect to basis B_1, but we need them with respect to another basis B_2.

Definition

Let B_1 and B_2 be two bases of a vector space V. The *change of basis matrix* Δ_{21} converts coordinates from B_1 (old basis) to B_2 (new basis).

Suppose we have the coordinates of a vector with respect to basis B_1, but we need them with respect to another basis B_2.

Definition

Let B_1 and B_2 be two bases of a vector space V. The *change of basis matrix* Δ_{21} converts coordinates from B_1 (old basis) to B_2 (new basis).

Proposition

Given $B_1 = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $B_2 = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ of V:

1. Form the $n \times 2n$ matrix
$$\left[\, B_2 \mid B_1 \,\right],$$
i.e., first block $=$ columns of B_2, second block $=$ columns of B_1.
2. Row reduce to get identity on the left.
3. The right block is the change of basis matrix Δ_{21}.

Proof ?

Suppose we have the coordinates of a vector with respect to basis B_1, but we need them with respect to another basis B_2.

Definition

Let B_1 and B_2 be two bases of a vector space V. The *change of basis matrix* Δ_{21} converts coordinates from B_1 (old basis) to B_2 (new basis).

Proposition

Given $B_1 = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $B_2 = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ of V:

1. Form the $n \times 2n$ matrix
$$\left[\, B_2 \mid B_1 \,\right],$$
i.e., first block $=$ columns of B_2, second block $=$ columns of B_1.
2. Row reduce to get identity on the left.
3. The right block is the change of basis matrix Δ_{21}.

Proof ?

Example

Let $B_1 = \{[1,0],[0,1]\}, \quad B_2 = \{[1,1],[1,-1]\}$. Find Δ_{12}, the change of basis matrix from B_1 to B_2.

Intuition

To invert a square matrix A, we want A^{-1} with $A\,A^{-1} = I$. Augment and row-reduce:

$$(A \,|\, I) \;\sim\; (I \,|\, A^{-1}).$$

Each row operation equals left-multiplication by an invertible *elementary matrix*, so the same product that turns A into I turns I into A^{-1}.

Elementary Row Operations (preserve solutions of $Ax = b$)

1. Swap rows: $L_i \leftrightarrow L_j$
2. Scale a row: $L_i \leftarrow \lambda L_i$ $(\lambda \neq 0)$
3. Row add: $L_i \leftarrow L_i + \lambda L_j$

If $(A \,|\, b) \mapsto (EA \,|\, Eb)$ with E elementary, then $Ax = b \iff (EA)x = Eb$. Thus A is invertible iff it can be reduced to I; applying the same steps to I yields A^{-1}.

Augmented Notation: Worked Example

$$\left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right)$$

Augmented Notation: Worked Example

$$\left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & -5 & -2 & 3 \end{array}\right) \quad \text{(since } L_2 \leftarrow 3L_2 - 2L_1\text{)}$$

Augmented Notation: Worked Example

$$\left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & -5 & -2 & 3 \end{array}\right) \quad \text{(since } L_2 \leftarrow 3L_2 - 2L_1\text{)}$$

$$\sim \left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1 & \frac{2}{5} & -\frac{3}{5} \end{array}\right) \quad \text{(since } L_2 \leftarrow -\frac{1}{5}L_2\text{)}$$

Augmented Notation: Worked Example

$$\left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & -5 & -2 & 3 \end{array}\right) \quad (\text{since } L_2 \leftarrow 3L_2 - 2L_1)$$

$$\sim \left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1 & \frac{2}{5} & -\frac{3}{5} \end{array}\right) \quad (\text{since } L_2 \leftarrow -\tfrac{1}{5}L_2)$$

$$\sim \left(\begin{array}{cc|cc} 3 & 0 & \frac{3}{5} & \frac{3}{5} \\ 0 & 1 & \frac{2}{5} & -\frac{3}{5} \end{array}\right) \quad (\text{since } L_1 \leftarrow L_1 - L_2)$$

Augmented Notation: Worked Example

$$\left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & -5 & -2 & 3 \end{array}\right) \quad \text{(since } L_2 \leftarrow 3L_2 - 2L_1\text{)}$$

$$\sim \left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1 & \frac{2}{5} & -\frac{3}{5} \end{array}\right) \quad \text{(since } L_2 \leftarrow -\tfrac{1}{5}L_2\text{)}$$

$$\sim \left(\begin{array}{cc|cc} 3 & 0 & \frac{3}{5} & \frac{3}{5} \\ 0 & 1 & \frac{2}{5} & -\frac{3}{5} \end{array}\right) \quad \text{(since } L_1 \leftarrow L_1 - L_2\text{)}$$

$$\sim \left(\begin{array}{cc|cc} 1 & 0 & \frac{1}{5} & \frac{1}{5} \\ 0 & 1 & \frac{2}{5} & -\frac{3}{5} \end{array}\right) \quad \text{(since } L_1 \leftarrow \tfrac{1}{3}L_1\text{).}$$

Augmented Notation: Worked Example

$$\left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right)$$

$$\sim \left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & -5 & -2 & 3 \end{array}\right) \quad \text{(since } L_2 \leftarrow 3L_2 - 2L_1\text{)}$$

$$\sim \left(\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 0 & 1 & \frac{2}{5} & -\frac{3}{5} \end{array}\right) \quad \text{(since } L_2 \leftarrow -\tfrac{1}{5} L_2\text{)}$$

$$\sim \left(\begin{array}{cc|cc} 3 & 0 & \frac{3}{5} & \frac{3}{5} \\ 0 & 1 & \frac{2}{5} & -\frac{3}{5} \end{array}\right) \quad \text{(since } L_1 \leftarrow L_1 - L_2\text{)}$$

$$\sim \left(\begin{array}{cc|cc} 1 & 0 & \frac{1}{5} & \frac{1}{5} \\ 0 & 1 & \frac{2}{5} & -\frac{3}{5} \end{array}\right) \quad \text{(since } L_1 \leftarrow \tfrac{1}{3} L_1\text{).}$$

Hence

$$A^{-1} = \begin{pmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{2}{5} & -\frac{3}{5} \end{pmatrix}, \quad \text{and} \quad A\,A^{-1} = I.$$

Smallville Recap

We want to analyze the long-term behavior of the system:

- What happens after n years?
- Does the system converge to an equilibrium state?

Key idea: a matrix represents a linear transformation in some basis. Choosing a "better" basis can simplify the representation. Goal: compute

$$\lim_{t\to\infty} A^t.$$

Definition: Eigenvalues and Eigenvectors

Let $T : V \to V$, with $V = \mathbb{R}^n$. If there exists a basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ such that

$$T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i, \quad \lambda_i \in \mathbb{R},$$

then the matrix of T in basis B is diagonal:

$$M_{BB} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Next Step

Given a matrix A representing T, we must find vectors $\mathbf{v}$ and scalars λ such that

$$A\mathbf{v} = \lambda\mathbf{v} \iff (A - \lambda I_n) \cdot \mathbf{v} = 0.$$

Key Idea

For a matrix A, an eigenvalue λ must satisfy

$$(A - \lambda I)\mathbf{v} = \mathbf{0} \quad \text{for some nonzero vector } \mathbf{v}.$$

This means $\ker(A - \lambda I) \neq \{\mathbf{0}\}$. Equivalently, since a square matrix has a nontrivial kernel iff its determinant vanishes:

$$\underbrace{\det(A - \lambda I)}_{\text{Characteristic polynomial } \chi_A(\lambda)} = 0.$$

- The roots of $\chi_A(\lambda)$ are the **eigenvalues** of A.
- The corresponding nonzero solutions $\mathbf{v}$ are the **eigenvectors**.
- The set of eigenvalues of A is called the **spectrum** of A, denoted by $\mathrm{Sp}(A)$.

Eigenspace

Given an eigenvalue λ_i, the **eigenspace** associated with λ_i is

$$E_{\lambda_i} \;=\; \ker(A-\lambda_i I) = \{\, \mathbf{v} \in \mathbb{R}^n \;\mid\; A\mathbf{v} = \lambda_i \mathbf{v} \,\}.$$

It is a subspace of $\mathbb{R}^n$, spanned by all eigenvectors corresponding to λ_i.

Remark

For each eigenvalue λ_i of a matrix A, there are two notions of multiplicity:

- The **algebraic multiplicity** of λ_i is its multiplicity as a root of the characteristic polynomial $\chi_A(\lambda)$.
- The **geometric multiplicity** of λ_i is the dimension of its eigenspace $\dim E_{\lambda_i}$.

These satisfy

$$1 \;\leq\; \dim E_{\lambda_i} \;\leq\; \text{algebraic multiplicity of } \lambda_i.$$

*A matrix is **diagonalizable** precisely when, for **each eigenvalue**, the geometric and algebraic **multiplicities coincide**.*

Example 1: Eigenvalues

Let

$$A = \begin{bmatrix} 7 & 2 \\ 3 & 8 \end{bmatrix}.$$

Compute $\det(A - \lambda I)$ and find the eigenvalues of A.

Example 2: Eigenvectors

For each eigenvalue λ of A, solve

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

Find a basis for the eigenspace corresponding to λ.

Example 1: Eigenvalues

Let

$$A = \begin{bmatrix} 7 & 2 \\ 3 & 8 \end{bmatrix}.$$

Compute $\det(A - \lambda I)$ and find the eigenvalues of A.

Example 2: Eigenvectors

For each eigenvalue λ of A, solve

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

Find a basis for the eigenspace corresponding to λ.

Observation

In Example 3, the eigenvalues appeared on the diagonal of the diagonalized matrix D. Next, we will see how the eigenvectors determine this change of basis.

Setup

Let $T : V \to V$ be a linear automorphism. Suppose B_1 and B_2 are two bases of V. We denote by $M_{B_1 B_1}$ and $M_{B_2 B_2}$ the matrices of T expressed in these bases.

Setup

Let $T : V \to V$ be a linear automorphism. Suppose B_1 and B_2 are two bases of V. We denote by $M_{B_1 B_1}$ and $M_{B_2 B_2}$ the matrices of T expressed in these bases.

Change of Basis Relation

Let Δ_{12} be the change-of-basis matrix from B_2 to B_1:

$$\mathbf{v}_{B_1} = \Delta_{12}\, \mathbf{v}_{B_2}.$$

Then the matrices of T in the two bases are related by

$$M_{B_2 B_2} = \Delta_{21}\, M_{B_1 B_1}\, \Delta_{12}, \quad \text{where } \Delta_{21} = \Delta_{12}^{-1}.$$

Setup

Let $T : V \to V$ be a linear automorphism. Suppose B_1 and B_2 are two bases of V. We denote by $M_{B_1 B_1}$ and $M_{B_2 B_2}$ the matrices of T expressed in these bases.

Change of Basis Relation

Let Δ_{12} be the change-of-basis matrix from B_2 to B_1:

$$\mathbf{v}_{B_1} = \Delta_{12}\, \mathbf{v}_{B_2}.$$

Then the matrices of T in the two bases are related by

$$M_{B_2 B_2} = \Delta_{21}\, M_{B_1 B_1}\, \Delta_{12}, \quad \text{where} \;\; \Delta_{21} = \Delta_{12}^{-1}.$$

Key Point

Two matrices $M_{B_1 B_1}$ and $M_{B_2 B_2}$ representing the same linear transformation in different bases are said to be **similar**. Similarity preserves eigenvalues and many structural properties, but the form of the matrix depends on the chosen basis.

Definition

A matrix $A \in \mathbb{R}^{n\times n}$ is **diagonalizable** if there exists a basis B_2 of eigenvectors of A.
In that basis, the matrix of A is diagonal:

$$D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

Definition

A matrix $A \in \mathbb{R}^{n\times n}$ is **diagonalizable** if there exists a basis B_2 of eigenvectors of A.
In that basis, the matrix of A is diagonal:

$$D = \operatorname{diag}(\lambda_1, \dots, \lambda_n).$$

Change of Basis Relation

Let $P := M_{B_1 B_2}$ **invertible** be the change-of-basis matrix from the eigenbasis B_2 to the reference basis B_1.
Then

$$A = P\,D\,P^{-1}.$$

Definition

A matrix $A \in \mathbb{R}^{n\times n}$ is **diagonalizable** if there exists a basis B_2 of eigenvectors of A.
In that basis, the matrix of A is diagonal:

$$D = \mathrm{diag}(\lambda_1, \dots, \lambda_n).$$

Change of Basis Relation

Let $P := M_{B_1 B_2}$ **invertible** be the change-of-basis matrix from the eigenbasis B_2 to the reference basis B_1.
Then

$$A = P\,D\,P^{-1}.$$

Theorem

For every integer $p \geq 0$,

$$A^p = P\,D^p\,P^{-1}.$$

Key Point

Diagonalization transforms a difficult computation (A^p) into a simple one (because simply $D^p = \mathrm{diag}(\lambda_1^p, \dots, \lambda_n^p)$, just raising eigenvalues to powers).

Characterizations

Let $A \in \mathbb{R}^{n\times n}$. The following are equivalent:

- A is diagonalizable $\iff$ there exists a basis of $\mathbb{R}^n$ formed by eigenvectors of A.
- A is similar to a diagonal matrix:
$$\exists\, P \text{ invertible}, \quad P^{-1}AP = D.$$
- The sum of the dimensions of all eigenspaces equals n.

Characterizations

Let $A \in \mathbb{R}^{n\times n}$. The following are equivalent:

- A is diagonalizable $\iff$ there exists a basis of $\mathbb{R}^n$ formed by eigenvectors of A.
- A is similar to a diagonal matrix:
$$\exists\, P \text{ invertible}, \quad P^{-1}AP = D.$$
- The sum of the dimensions of all eigenspaces equals n.

Practical Conditions

- If A has n distinct eigenvalues $\Rightarrow$ A is diagonalizable.
- If A is symmetric (real entries), then A is diagonalizable (Spectral Theorem).
- Otherwise: compare **algebraic multiplicity** (from characteristic polynomial) and **geometric multiplicity** (dimension of eigenspace). Diagonalizability requires equality for every eigenvalue.

Trace

The **trace** of a square matrix $A \in \mathcal{M}_n(\mathbb{R})$ is

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

If A is diagonalizable, $A = PDP^{-1}$ with $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, then

$$\mathrm{tr}(A) = \mathrm{tr}(D) = \lambda_1 + \cdots + \lambda_n.$$

Determinant

The **determinant** of a square matrix $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ is a scalar denoted $\det(A)$.

Formal recursive definition:

- For $n = 1$, $\det([a_{11}]) = a_{11}$.
- For $n \geq 2$,

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(M_{1j}),$$

where M_{1j} is the $(n-1) \times (n-1)$ submatrix obtained by deleting row 1 and column j.

Geometric meaning: $\det(A)$ represents the scaling factor of volumes induced by the linear transformation associated with A, with its sign indicating whether orientation is preserved or reversed.
If A is diagonalizable as above, then

$$\det(A) = \det(D) = \lambda_1 \cdot \lambda_2 \cdots \lambda_n.$$

Spectrum and Transpose

Since $\det(A - \lambda I) = \det((A - \lambda I)^\top) = \det(A^\top - \lambda I)$, we have

$$\mathrm{Sp}(A) = \mathrm{Sp}(A^\top).$$

Let's once again consider

$$\begin{aligned} T : \mathbb{R}^2 &\to \mathbb{R}^2 \\ (x, y) &\mapsto (x + y,\ 2x + 2y) \end{aligned}$$

We showed in 36 that T is an endomorphism and not invertible.

1. What is the matrix representation of T in the canonical basis of $\mathbb{R}^2$? Let's denote it A.
2. Is A diagonalizable? If so, what are its eigenvalues?
3. Determine a diagonal matrix D and an invertible matrix P (and P^{-1}) such that

$$A = P\,D\,P^{-1}.$$

Diagonalization Method

Let

$$A = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}, \quad \begin{bmatrix} n_0 \\ s_0 \end{bmatrix} = \begin{bmatrix} 2000 \\ 8000 \end{bmatrix}.$$

Show that

$$\lim_{t \to \infty} A^t = \begin{bmatrix} 2/5 & 2/5 \\ 3/5 & 3/5 \end{bmatrix}.$$

Hence, what is the equilibrium ?

Solution ?

Diagonalization Method

Let

$$A = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}, \quad \begin{bmatrix} n_0 \\ s_0 \end{bmatrix} = \begin{bmatrix} 2000 \\ 8000 \end{bmatrix}.$$

Show that

$$\lim_{t \to \infty} A^t = \begin{bmatrix} 2/5 & 2/5 \\ 3/5 & 3/5 \end{bmatrix}.$$

Hence, what is the equilibrium ?

Solution ?

Observation

Since $\begin{bmatrix} 2/5 & 2/5 \\ 3/5 & 3/5 \end{bmatrix} \begin{bmatrix} 2000 \\ 8000 \end{bmatrix} = \begin{bmatrix} 4000 \\ 6000 \end{bmatrix}$, even though the initial population had far more smokers, the equilibrium state shifts toward more nonsmokers than at the start.

Definition

A square matrix $Q \in \mathbb{R}^{n \times n}$ is called **orthogonal** if

$$Q^\top Q = QQ^\top = I_n.$$

Equivalently, $Q^{-1} = Q^\top$.

Intuition

An orthogonal matrix represents a linear transformation that:

- preserves inner products and lengths,
- preserves orthogonality,
- is a composition of rotations and reflections.

Remark: Isometry Property

Orthogonal matrices preserve lengths:

$$\|Q\mathbf{x}\|^2 = (Q\mathbf{x})^\top (Q\mathbf{x}) = \mathbf{x}^\top Q^\top Q\mathbf{x} = \mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|^2.$$

Thus, $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$.

Why "orthogonal"?

If $Q = [\mathbf{q}_1\ \mathbf{q}_2\ \ldots\ \mathbf{q}_n]$, then

$$\mathbf{q}_i \cdot \mathbf{q}_j = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases},$$

so the column vectors form an **orthonormal basis** of $\mathbb{R}^n$.

Example

The rotation matrix in $\mathbb{R}^2$,

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

is orthogonal: $R(\theta)^\top R(\theta) = I_2$.

Remark

The set of all $n \times n$ orthogonal matrices forms a group under multiplication, called the **orthogonal group** $O(n)$. The subgroup of matrices with determinant 1 is the **special orthogonal group** $SO(n)$, corresponding to pure rotations.

Theorem (Spectral Theorem)

Every real symmetric matrix is diagonalizable. That is, if $A \in \mathbb{R}^{n \times n}$ and $A^\top = A$, then there exists an orthogonal matrix Q such that

$$Q^\top A Q = D,$$

where D is diagonal.

Example

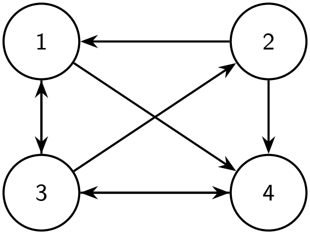
Consider

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

- Show that A is symmetric.
- Compute its eigenvalues and eigenvectors.
- Verify that A is diagonalizable via an orthogonal change of basis.

Motivation

Model the web as a weighted, directed graph: vertices = websites, edges = links. If site j has ℓ_j outgoing links, each outgoing edge carries weight $1/\ell_j$. This yields a column-stochastic transition matrix T; to allow random jumps, add a uniform "teleportation" matrix R.



From Graph to Matrices

With ℓ_j the out-degree of vertex j,

$$T_{ij} = \begin{cases} \frac{1}{\ell_j} & \text{if } j \to i \text{ is an edge} \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, we construct this matrix:

$$T = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ \frac{1}{2} & 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{3} & 0 \end{bmatrix}.$$

Teleportation and the Google Matrix

The matrix $\mathbf{1}$ denotes the column vector of size n with all entries equal to 1. Hence $\mathbf{1}\mathbf{1}^\top$ is the $n \times n$ matrix of all 1's.

$$R = \frac{1}{n}\,\mathbf{1}\mathbf{1}^\top$$

is therefore the matrix where every entry is $\frac{1}{n}$. This models the fact that with probability p, a user may *randomly jump* to any website, independently of links.

The Google matrix combines both behaviors:

$$G = (1-p)\,T + p\,R \;\in\; \mathbb{R}^{n\times n}, \quad \text{with typical choice } p \approx 0.15.$$

This construction ensures that G is stochastic, irreducible, and aperiodic, so it admits a unique stationary distribution. This stationary distribution reflects the long-term importance (rank) of each page, which is the core of Google's PageRank algorithm.

Ranking Vector

We start with the uniform distribution:

$$\mathbf{v}(0) = \left[\frac{1}{n}, \quad \frac{1}{n}, \quad \dots, \quad \frac{1}{n}\right]^\top,$$

which represents an equal probability of being at any page initially.
Iterating the process,

$$\mathbf{v}(t+1) = G\,\mathbf{v}(t),$$

converges to the unique fixed point

$$\mathbf{v}_\infty = \left(\lim_{t\to\infty} G^t\right) \cdot \mathbf{v}(0).$$

The vector $\mathbf{v}_\infty$ is the **PageRank vector**: its i-th entry gives the long-term probability of a user visiting page i.
Pages with larger entries are ranked higher in search results.

Theorem (Perron-Frobenius)

If M is a column-stochastic matrix with all entries positive, then:

- 1 *is an eigenvalue of* M*,*
- *the associated eigenvector* $\mathbf{v}_\infty$ *has strictly positive entries,*
- $\mathbf{v}_\infty$ *can be normalized so that its entries sum to* 1*,*
- *the iteration* $M^t\mathbf{v}(0)$ *converges to* $\mathbf{v}_\infty$*.*

Theorem (Perron-Frobenius)

If M is a column-stochastic matrix with all entries positive, then:

- 1 *is an eigenvalue of* M*,*
- *the associated eigenvector* $\mathbf{v}_\infty$ *has strictly positive entries,*
- $\mathbf{v}_\infty$ *can be normalized so that its entries sum to* 1*,*
- *the iteration* $M^t\mathbf{v}(0)$ *converges to* $\mathbf{v}_\infty$*.*

Application to PageRank

The PageRank vector is defined by solving

$$G\mathbf{v} = \mathbf{v},$$

that is, finding the eigenvector of G associated with eigenvalue 1.
Challenge: for the web, n is in the billions. Direct eigenvector computation is infeasible.
Practical solution: approximate $\mathbf{v}_\infty$ by iterating

$$\mathbf{v}(m) = G^m\mathbf{v}(0),$$

for moderate m, until convergence is reached.

Setup

We have k observations of m variables:

$$X = \{p_1, \dots, p_k\}, \quad p_i = (p_{i1}, \dots, p_{im}) \in \mathbb{R}^m.$$

For each coordinate j, let $\mu_j(X)$ be the mean. Define the centered data matrix:

$$N_{ij} = p_{ij} - \mu_j(X).$$

Setup

We have k observations of m variables:

$$X = \{p_1, \dots, p_k\}, \quad p_i = (p_{i1}, \dots, p_{im}) \in \mathbb{R}^m.$$

For each coordinate j, let $\mu_j(X)$ be the mean. Define the centered data matrix:

$$N_{ij} = p_{ij} - \mu_j(X).$$

Questions of Interest

- How is the data spread across directions in $\mathbb{R}^m$?
- Is the variance larger in some directions than others?
- Do subsets of the data cluster in certain patterns?

Setup

We have k observations of m variables:

$$X = \{p_1, \ldots, p_k\}, \quad p_i = (p_{i1}, \ldots, p_{im}) \in \mathbb{R}^m.$$

For each coordinate j, let $\mu_j(X)$ be the mean. Define the centered data matrix:

$$N_{ij} = p_{ij} - \mu_j(X).$$

Questions of Interest

- How is the data spread across directions in $\mathbb{R}^m$?
- Is the variance larger in some directions than others?
- Do subsets of the data cluster in certain patterns?

Definition

The **covariance matrix** of X is

$$\mathrm{cov}(X) = N^\top N.$$

Example : Centering, Covariance, and Eigenanalysis

Consider the dataset

$$X = \{(1,1), (2,2), (2,3), (3,2), (3,3), (4,4)\}.$$

Tasks:

1. Compute the coordinate-wise mean $\boldsymbol{\mu} = (\mu_1, \mu_2)$.
2. Form the centered data matrix N with entries
$$N_{ij} = p_{ij} - \mu_j.$$
3. Compute the covariance matrix
$$\mathrm{cov}(X) = N^\top N \quad \text{(optionally normalized by } \tfrac{1}{k} \text{ or } \tfrac{1}{k-1}\text{)}.$$
4. Find the eigenvalues and associated eigenvectors of $\mathrm{cov}(X)$.

Theorem (Variance along Eigenvectors)

Order the eigenvalues of $\mathrm{cov}(X)$ *as*

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m.$$

Then the data variance along each direction is proportional to the corresponding eigenvalue, in the direction of the associated eigenvector.

Theorem (Variance along Eigenvectors)

Order the eigenvalues of $\mathrm{cov}(X)$ *as*

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m.$$

Then the data variance along each direction is proportional to the corresponding eigenvalue, in the direction of the associated eigenvector.

Interpretation

- The largest eigenvalue $\lambda_{\max}$ indicates the direction of greatest data spread.
- Smaller eigenvalues correspond to directions with less variation.
- For 2D data: eigenvectors give the principal axes of the ellipse approximating the data cloud, and eigenvalues determine their lengths.

Theorem (Variance along Eigenvectors)

Order the eigenvalues of $\mathrm{cov}(X)$ *as*

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m.$$

Then the data variance along each direction is proportional to the corresponding eigenvalue, in the direction of the associated eigenvector.

Interpretation

- The largest eigenvalue $\lambda_{\max}$ indicates the direction of greatest data spread.
- Smaller eigenvalues correspond to directions with less variation.
- For 2D data: eigenvectors give the principal axes of the ellipse approximating the data cloud, and eigenvalues determine their lengths.

Key Point: PCA

Theorem 14 is the foundation of **Principal Component Analysis (PCA)**, a fundamental tool in applied mathematics, statistics, and machine learning.

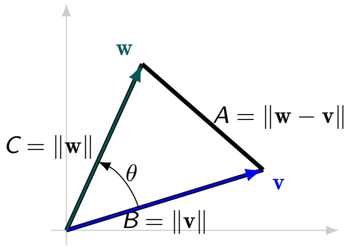
Why orthogonality matters

In data science we approximate: we **minimize distance between model and data**. Squared distances are quadratic, so minimization leads to linear systems. Vector calculus links minimization with **orthogonal projections** onto subspaces.

Law of Cosines (Al-Kashi)

For a triangle with side lengths A, B, C and opposite angles a, b, c,

$$A^2 = B^2 + C^2 - 2BC\cos(c).$$



$$A^2 = B^2 + C^2 - 2BC\cos(\theta)$$

Exercise (warm-up)

Let $\mathbf{v}, \mathbf{w}$ start at the origin and c be the angle between them. Apply the law of cosines to the triangle with sides $A = \|\mathbf{w} - \mathbf{v}\|, \quad B = \|\mathbf{v}\|, \quad C = \|\mathbf{w}\|$ to show

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \, \|\mathbf{w}\| \cos(c).$$

Motivation

Fitting data requires restricting model complexity: *a good fit minimizes the error between data and model, without overfitting.*

Key Point

Least squares = **projection onto a subspace**. Understanding this requires the geometry of orthogonality.

Example: Line of Best Fit in $\mathbb{R}^2$

Data set: $X = \{(1,6),(2,5),(3,7),(4,10)\}$.
We want the line $y = ax + b$ that minimizes the total squared error:

$$\text{error} = \sqrt{\sum_{(x_i,y_i)\in X} (y_i - (ax_i + b))^2}.$$

Minimizing the error is the same as minimizing the content of the square root.
Equivalently, solve the least squares system:

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \\ 7 \\ 10 \end{bmatrix}.$$

Interpretation: we seek the projection of $\begin{bmatrix} 6 & 5 & 7 & 10 \end{bmatrix}^\top$ onto the column space of the matrix.

Ordinary Least Squares in Simple Linear Regression

We consider the model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \qquad i = 1, \ldots, N,$$

where ε_i are random errors with zero mean. The aim is to estimate (β_0, β_1) by minimizing the total squared error.

Derivation of the Optimal Coefficients

We minimize the quadratic error

$$f(\beta_0, \beta_1) = \sum_{i=1}^{N} \big(Y_i - (\beta_0 + \beta_1 X_i)\big)^2.$$

First-order conditions:

$$\begin{cases} \dfrac{\partial f}{\partial \beta_0} = -2\sum_{i=1}^{N} \big(Y_i - \beta_0 - \beta_1 X_i\big) = 0, \\ \dfrac{\partial f}{\partial \beta_1} = -2\sum_{i=1}^{N} X_i\big(Y_i - \beta_0 - \beta_1 X_i\big) = 0. \end{cases}$$

Dividing by N and introducing

$$\bar{X} = \tfrac{1}{N}\sum_{i=1}^{N} X_i, \quad \bar{Y} = \tfrac{1}{N}\sum_{i=1}^{N} Y_i,$$

we obtain the system

$$\begin{cases} \bar{Y} = \beta_0 + \beta_1 \bar{X}, \\ \dfrac{1}{N}\sum_{i=1}^{N} X_i Y_i = \beta_0 \bar{X} + \beta_1 \tfrac{1}{N}\sum_{i=1}^{N} X_i^2. \end{cases}$$

Covariance and Variance Forms

Define

$$\mathrm{Cov}(X, Y) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}), \qquad \mathrm{Var}(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2.$$

Useful expansions:

$$\frac{1}{N}\sum(X_i - \bar{X})(Y_i - \bar{Y}) = \frac{1}{N}\sum X_i Y_i - \bar{X}\bar{Y},$$
$$\frac{1}{N}\sum(X_i - \bar{X})^2 = \frac{1}{N}\sum X_i^2 - \bar{X}^2.$$

Therefore

$$\hat{\beta}_1 = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}.$$

Final

Substituting the first into the second and rearranging yields

$$\hat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N} X_i Y_i - \bar{X}\bar{Y}}{\frac{1}{N}\sum_{i=1}^{N} X_i^2 - \bar{X}^2} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}.$$

Finally,

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Remark: Interpretation of $\hat{\beta}_1$

- Numerator $\mathrm{Cov}(X, Y)$: co-variation, the linear effect of X on Y.
- Denominator $\mathrm{Var}(X)$: variability of X itself.
- Hence $\hat{\beta}_1$ measures the **average change in Y per unit change in X**, *i.e.*, the "linear effect" of X normalized by its own variability.

Definition

Subspaces $W, W' \subset V$ are **orthogonal** if

$$\mathbf{w} \cdot \mathbf{w}' = 0 \quad \forall\, \mathbf{w} \in W,\ \mathbf{w}' \in W'.$$

A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is **orthonormal** if $\mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij}$.
A matrix A is **orthonormal** when its columns are orthonormal vectors.

Fundamental Subspaces of a Matrix

Let $A \in \mathbb{R}^{m \times n}$.

- **Column space (image):**
$$C(A) \equiv \mathrm{Im}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^m.$$
Dimension = **rank** of A.
- **Null space (kernel):**
$$N(A) \equiv \ker(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\} \subset \mathbb{R}^n.$$
- **Row space:**
$$R(A) = \mathrm{Im}(A^\top) = \{\mathbf{y}^\top A,\ \mathbf{y} \in \mathbb{R}^m\} = [\text{span of row vectors of } A] \subset \mathbb{R}^n.$$
- **Rank–nullity theorem:**
$$n = \dim N(A) + \dim R(A).$$
- **Orthogonal complement:** For a subspace $W \subset V$,
$$W^\perp = \{\mathbf{v} \in V \mid \mathbf{v} \cdot \mathbf{w} = 0,\ \forall\, \mathbf{w} \in W\}.$$

Example

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 3}.$$

- Compute $C(A) = \mathrm{Im}(A)$: span of the column vectors. Is it all of $\mathbb{R}^3$?
- Compute $C^\perp$.
- Compute $N(A) = \ker(A)$: solve $A\mathbf{x} = \mathbf{0}$ explicitly.
- Determine $R(A)$: span of row vectors. Compare $\dim R(A)$ with $\dim C(A)$.
- Verify the rank–nullity theorem:

$$n = 3 = \dim N(A) + \dim R(A).$$

Worked Example: Solution

- **Column Space**
$$C(A) = \mathrm{Span}\{(1,2,1)^\top, (2,4,1)^\top\}, \quad \dim = 2 < 3.$$
So $C(A) \neq \mathbb{R}^3$.
- **Orthogonal Complement**
$$C(A)^\perp = \ker(A^\top) = \mathrm{Span}\{(-2,1,0)^\top\}.$$
- **Null Space**
$$N(A) = \ker(A) = \mathrm{Span}\{(1,-2,1)^\top\}, \quad \dim = 1.$$
- **Row Space**
$$R(A) = \mathrm{Span}\{(1,2,3), (1,1,1)\}, \quad \dim = 2.$$
Note $\dim R(A) = \dim C(A) = 2$.
- **Rank–Nullity Theorem**
$$3 = \dim N(A) + \dim R(A) = 1 + 2.$$

Consistency: $\dim C(A)^\perp = 1$, and indeed $(-2,1,0)$ is orthogonal to both generators of $C(A)$.

Exercise

For A as above:

1. Prove $N(A) = R(A)^{\perp}$ and $N(A^{\top}) = C(A)^{\perp}$.
2. Prove any $\mathbf{v} \in V$ decomposes uniquely as $\mathbf{v} = \mathbf{w}' + \mathbf{w}''$ with $\mathbf{w}' \in W$, $\mathbf{w}'' \in W^{\perp}$.
3. Prove that the closest vector in W to $\mathbf{v}$ is exactly $\mathbf{w}'$.

Let A denote the matrix on the left in the last displayed equation in Example 78, and let $\mathbf{b} = [6, 5, 7, 10]^\top$. Then

$$A^\top A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}$$

so

$$\left(A^\top A\right)^{-1} = \begin{bmatrix} 3/2 & -1/2 \\ -1/2 & 1/5 \end{bmatrix}$$

Continuing with the computation, we have

$$A \cdot \left(A^\top A\right)^{-1} \cdot A^\top = \frac{1}{10} \begin{bmatrix} 7 & 4 & 1 & -2 \\ 4 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 \\ -2 & 1 & 4 & 7 \end{bmatrix}$$

Putting everything together, we see that indeed

$$A \cdot \left(A^\top A\right)^{-1} \cdot A^\top \cdot \mathbf{b} = \begin{bmatrix} 4.9 \\ 6.3 \\ 7.7 \\ 9.1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 3.5 \\ 1.4 \end{bmatrix}$$

where $[3.5, 1.4]$ is the solution we obtained using partials.

Quadratic fit for Example 4.2

Fit a degree–2 model $y = ax^2 + bx + c$ to the data of Example 4.2. Set up the least–squares system

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}}_{A} \begin{bmatrix} c \\ b \\ a \end{bmatrix} = \underbrace{\begin{bmatrix} 6 \\ 5 \\ 7 \\ 10 \end{bmatrix}}_{\mathbf{b}}$$

and carry out the same analysis as for the linear fit:

- Derive the normal equations $A^\top A\,\mathbf{y} = A^\top \mathbf{b}$ (with $\mathbf{y} = [c, b, a]^\top$).
- Show that the solution $\mathbf{y}$ gives $\mathbf{b}' = A\mathbf{y}$ equal to the projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$.
- Verify that this agrees with the solution found by minimizing via partial derivatives.

Towards SVD

Real-world problems often involve *non-square* matrices. **Singular Value Decomposition (SVD)** is “diagonalization for non-square matrices” and will generalize these ideas.

Singular Value Decomposition Theoreme

Let M be an $m \times n$ matrix of rank r. There exist matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ with orthonormal columns, and a diagonal matrix $\Sigma \in \mathbb{R}^{m\times n}$ with nonzero entries $\sigma_1, \ldots, \sigma_r$, such that

$$M = U\Sigma V^\top.$$

Singular Value Decomposition Theoreme

Let M be an $m \times n$ matrix of rank r. There exist matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ with orthonormal columns, and a diagonal matrix $\Sigma \in \mathbb{R}^{m\times n}$ with nonzero entries $\sigma_1, \ldots, \sigma_r$, such that

$$M = U\Sigma V^\top.$$

Key Ideas of the Proof

- $M^\top M$ is symmetric $\Rightarrow$ diagonalizable with orthonormal eigenvectors.
- If $M^\top M\mathbf{v}_i = \lambda_i \mathbf{v}_i$, define singular values $\sigma_i = \sqrt{\lambda_i}$.
- Construct $\mathbf{q}_i = \frac{1}{\sigma_i} M\mathbf{v}_i$; these vectors are orthonormal in $\mathbb{R}^m$.
- Collect $\{\mathbf{q}_i\}$ as columns of U, $\{\mathbf{v}_i\}$ as columns of V.
- Then $U^\top MV = \Sigma$, hence $M = U\Sigma V^\top$.

Singular Value Decomposition Theoreme

Let M be an $m \times n$ matrix of rank r. There exist matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ with orthonormal columns, and a diagonal matrix $\Sigma \in \mathbb{R}^{m\times n}$ with nonzero entries $\sigma_1, \dots, \sigma_r$, such that

$$M = U\Sigma V^\top.$$

Key Ideas of the Proof

- $M^\top M$ is symmetric $\Rightarrow$ diagonalizable with orthonormal eigenvectors.
- If $M^\top M\mathbf{v}_i = \lambda_i \mathbf{v}_i$, define singular values $\sigma_i = \sqrt{\lambda_i}$.
- Construct $\mathbf{q}_i = \frac{1}{\sigma_i} M\mathbf{v}_i$; these vectors are orthonormal in $\mathbb{R}^m$.
- Collect $\{\mathbf{q}_i\}$ as columns of U, $\{\mathbf{v}_i\}$ as columns of V.
- Then $U^\top MV = \Sigma$, hence $M = U\Sigma V^\top$.

Interpretation

SVD generalizes diagonalization to non-square matrices. It expresses any matrix as:

$$\text{(orthogonal change of basis)} \times \text{(scaling)} \times \text{(orthogonal change of basis)}.$$

Example: Computing an SVD

Compute the Singular Value Decomposition of

$$M = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 2 & 0 \end{bmatrix}.$$

(Hint: start with $M^\top M$ and find eigenvalues/eigenvectors.)

Exercise: Verification and Rank-One Approximation

Check that indeed $M = U\Sigma V^\top$. What is the best rank-one approximation of M?

Example: Decomposition into Rank-One Matrices

Write

$$M = \mathbf{u}_1\sigma_1\mathbf{v}_1^\top + \cdots + \mathbf{u}_r\sigma_r\mathbf{v}_r^\top,$$

and interpret this as a decomposition into rank-one matrices. Discuss its use in applications such as image compression.

Exercise: Least Squares via SVD

Show that least squares approximation is an instance of SVD: minimizing $\|M\mathbf{x} - \mathbf{b}\|$ reduces to

$$\mathbf{y} = V^\top\mathbf{x} = \frac{1}{\Sigma}U^\top\mathbf{b}.$$