Mathematics for Data Science

Lecture 2

Eva FEILLET[1]

[1]LISN
Paris-Saclay University

M1[AI], Fall 2025

Previously covered topics

- Vector space, example of $\mathbb{R}^n$, geometric interpretation
- Subspace (examples and proposition about the dimensions)
- Linear combination, span, linearly independent vectors, spanning list
- Basis, dimension, canonical basis of $\mathbb{R}^n$
- Surjectivity, injectivity, bijectivity, case of linear transformations
- Linear transformations, rank, image/range, kernel/nullspace, rank-nullity theorem

Preliminary remark

In course 2, we consider matrices with real-valued coefficients.

Table of Contents

Table of Contents

Matrices

To represent and manipulate vectors and linear maps on a computer, we use rectangular arrays of numbers known as **matrices**.

Definition

A matrix is a rectangular array of numbers, called **coefficients**.

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

where A has m rows and n columns. We say A is an $m \times n$ *matrix*.

Vocabulary : If $m = n$ then A is called a *square matrix*.

Matrices

Reminder: $(\mathcal{M}_{m,n}(\mathbb{R}), +, \cdot)$ is a $\mathbb{R}$-vector space.

Addition of matrices: Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices. Their sum is:

$$A + B = (a_{ij} + b_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}.$$

Multiplication by a scalar: If $\lambda \in \mathbb{R}$ and $A = (a_{ij})$ is an $m \times n$ matrix, then:

$$\lambda A = (\lambda a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}.$$

Matrices

Reminder: $(\mathcal{M}_{m,n}(\mathbb{R}), +, \cdot)$ is a $\mathbb{R}$-vector space.

Addition of matrices: Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices. Their sum is:

$$A + B = (a_{ij} + b_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}.$$

Multiplication by a scalar: If $\lambda \in \mathbb{R}$ and $A = (a_{ij})$ is an $m \times n$ matrix, then:

$$\lambda A = (\lambda a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}.$$

Warning

Matrix addition is only defined when the two matrices have the same size.

Matrix-Vector multiplication

Let $A \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$. Then:

$$Ax = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix}$$

Interpretation ?

Matrix-Vector multiplication

Let $A \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$. Then:

$$Ax = \begin{pmatrix} \sum_{j=1}^n a_{1j} x_j \\ \sum_{j=1}^n a_{2j} x_j \\ \vdots \\ \sum_{j=1}^n a_{mj} x_j \end{pmatrix}$$

Interpretation ?

Remark

- *This can be interpreted as a linear combination of the columns of A with weights given by the coordinates of x.*

Matrix-Matrix Multiplication

Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$. Then:

Matrix-Matrix Multiplication

Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$. Then:

$$AB = (c_{ij}) \in \mathbb{R}^{m \times p}, \quad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Matrix-Matrix Multiplication

Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$. Then:

$$AB = (c_{ij}) \in \mathbb{R}^{m \times p}, \quad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Be careful about sizes

- Matrix multiplication is only defined if the number of columns of A equals the number of rows of B.

Note: you can come back to matrix-vector multiplication by thinking of matrix B as stacked **column vectors**

Matrix-Matrix multiplication

Proposition

1. *Associativity: Let* $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$, $C \in \mathbb{R}^{p\times q}$. *Then*
$$(AB)C = A(BC).$$
2. *Distributivity (left): If* $A \in \mathbb{R}^{m\times n}$ *and* $B, C \in \mathbb{R}^{n\times p}$, *then* $A(B+C) = AB + AC$.
3. *Distributivity (right):If* $A, B \in \mathbb{R}^{m\times n}$ *and* $C \in \mathbb{R}^{n\times p}$, *then* $(A+B)C = AC + BC$.
4. *For any* $A \in \mathbb{R}^{m\times n}$, $I_m A = A = A I_n$.

Be careful about sizes

- Matrix multiplication is *not* commutative in general, i.e. $AB \neq BA$ in most cases.

Table of Contents

Matrix of a linear map

Proposition

If $A \in \mathbb{R}^{m \times n}$, the following mapping is a linear transformation.

$$\mathbb{R}^n \to \mathbb{R}^m, \quad x \mapsto Ax$$

Matrix of a linear map

Proposition

If $A \in \mathbb{R}^{m \times n}$, the following mapping is a linear transformation.

$$\mathbb{R}^n \to \mathbb{R}^m, \quad x \mapsto Ax$$

Matrix of a linear map

Suppose V and W are finite-dimensional vector spaces with bases $\mathcal{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{C} = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$, respectively, and $L : V \to W$ is a linear map. Then the matrix $A = (a_{ij})$ of L is defined by

$$L(\mathbf{v}_j) = a_{1j}\,\mathbf{w}_1 + \cdots + a_{mj}\,\mathbf{w}_m.$$

Note: the j-th column of $\mathbf{A}$ consists of the coordinates of $L(\mathbf{v}_j)$ in the chosen basis for W.

$$[L(x)]_{\mathcal{C}} = [L]_{\mathcal{B},\mathcal{C}}[x]_{\mathcal{B}}$$

Matrix of a linear map

Proposition (Another formulation)

Let $L : V \to W$ be a linear transformation between finite-dimensional vector spaces V and W ($\dim(V) = n$, $\dim(W) = m$). If $\{v_1, \ldots, v_n\}$ is a basis of V and $\{w_1, \ldots, w_m\}$ is a basis of W, then for each $1 \leq j \leq n$, there exists unique scalars a_{ij} such that

$$L(v_j) = \sum_{i=1}^{m} a_{ij} w_i$$

Note: Changing the basis in V or W changes the matrix representation.

Remarks

In other words, every matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ induces a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ given by

$$L\mathbf{x} = \mathbf{A}\mathbf{x}, \text{i.e. } [L(x)]_{\mathcal{C}} = [L]_{\mathcal{B},\mathcal{C}}[x]_{\mathcal{B}}$$

and the matrix of this map with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ is $\mathbf{A}$.

Conversely, every linear transformation $L : V \to W$ with $dim(V) = n, dim(W) = m$, can be described by a matrix $A \in \mathbb{R}^{m \times n}$.

Note: operator notation used above

Warning

Watch out for the dimensions: $A \in \mathbb{R}^{m \times n}$.

Reminder: Range, Rank, Kernel of a linear map

Let $L : V \mapsto W$ be a linear map, V and W vector spaces.

Range of a linear map

The range of L is the set of vectors $y \in W$ such that there exist a vector $x \in V$ that is mapped to y by L.

$$L(V) = \{y \in W \mid y = L(x),\ x \in V\}$$

Rank of a linear map

The **rank** of L is defined as $\mathrm{rank}(L) = \dim(L(V))$.

Kernel, Nullspace

We define the **kernel** of L as $\mathrm{null}(L) = \{\mathbf{x} \in V \mid L(\mathbf{x}) = \mathbf{0}\}$ (also denoted $ker(L)$).

Range of a linear map

With matrices, we now have a new way of writing the range (image) of a linear map:

Definition

Let $L : V \to W$ be a linear map. We define the range of L as

$$range(L) = \{w \in W \mid \exists v \in V, Lv = w\}$$

Note: operator notation / or replace with the matrix notation

Range and Rank of a matrix

- The **columnspace** (resp. **rowspace**) of a matrix $A \in \mathbb{R}^{m \times n}$ is the span of its columns, considered as vectors of $\mathbb{R}^m$ (resp. rows, considered as vectors of $\mathbb{R}^n$).
- The columnspace of A is also the range of the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ which is induced by A.
- The rowspace of A is the range of the linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ which is induced by $A^\top$.

Range and Rank of a matrix

- The **columnspace** (resp. **rowspace**) of a matrix $A \in \mathbb{R}^{m \times n}$ is the span of its columns, considered as vectors of $\mathbb{R}^m$ (resp. rows, considered as vectors of $\mathbb{R}^n$).
- The columnspace of A is also the range of the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ which is induced by A.
- The rowspace of A is the range of the linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ which is induced by $A^\top$.

Proposition

The dimension of the columnspace of A is the same as the dimension of the rowspace of A and it is called the **rank** *of A.*

$$rank(A) = dim(range(A)) = dim(range(A^\top))$$

Null space (kernel) and range (image)

We also have a new way of writing the null space (kernel) of a linear map.

Definition

Let $L : V \to W$ be a linear map. We define the nullspace of L as

$$null(L) = \{v \in V \mid Lv = 0\}$$

Link to the properties of injectivity/surjectivity

Proposition (Kernel and injectivity)

Let $A \in \mathbb{R}^{m\times n}$. The mapping $x \mapsto Ax$ is injective if and only if $\ker(A) = \{0\}$.

Proposition (Equivalent statements about kernel and range of square matrix)

If $A \in \mathbb{R}^{n\times n}$, the following are equivalent:

1. *The transformation $x \mapsto Ax$ is bijective.*
2. $\mathrm{Im}(A) = \mathbb{R}^n$ *(i.e. A is surjective).*
3. $\ker(A) = \{0\}$ *(i.e. A is injective).*
4. $rank(A) = n$

Rank-nullity theorem for matrices

Theorem (Rank-nullity theorem for matrices)

Let $A \in \mathbb{R}^{m \times n}$. We have the following equality:

$$n = rank(A) + dim(ker(A))$$

Table of Contents

Diagonal matrices

Definition

A **diagonal matrix** in $\mathbb{R}^{n\times n}$ is a square matrix of the form

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix},$$

where $d_1, \ldots, d_n \in \mathbb{R}$.

In other words, all coefficients outside the main diagonal are zero.
Note: it is also denoted for short, $\mathrm{diag}(d_1, d_2, ...d_n)$

Properties: $D = D^\top$.

Diagonal matrices

Definition

A **diagonal matrix** in $\mathbb{R}^{n\times n}$ is a square matrix of the form

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix},$$

where $d_1, \ldots, d_n \in \mathbb{R}$.

In other words, all coefficients outside the main diagonal are zero.
Note: it is also denoted for short, $\mathrm{diag}(d_1, d_2, ...d_n)$

Properties: $D = D^\top$. We will also see that a diagonal matrix $D = \mathrm{diag}(d_1, \ldots, d_n)$ is invertible if and only if $d_i \neq 0$ for all i (see lecture on determinant).

Identity matrix

Definition

The **identity matrix** in $\mathbb{R}^n$, denoted I_n, is the diagonal matrix with all diagonal entries equal to 1:

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Proposition (Properties of the Identity Matrix)

Let A be an $n \times n$ matrix. Then:

1. $I_n\mathbf{x} = \mathbf{x}$ *for all* $\mathbf{x} \in \mathbb{R}^n$.
2. $AI_n = I_nA = A$
3. *In particular,* $I_nI_n = I_n$ *and* $(I_n)^{-1} = I_n$

Transpose of a matrix

Definition

If $\mathbf{A} \in \mathbb{R}^{m \times n}$, its *transpose* $\mathbf{A}^\top \in \mathbb{R}^{n \times m}$ is given by $(\mathbf{A}^\top)_{ij} = A_{ji}$ for each (i, j).

In other words, the columns of $\mathbf{A}$ become the rows of $\mathbf{A}^\top$, and the rows of $\mathbf{A}$ become the columns of $\mathbf{A}^\top$.

Transpose of a matrix

Definition

If $\mathbf{A} \in \mathbb{R}^{m \times n}$, its *transpose* $\mathbf{A}^\top \in \mathbb{R}^{n \times m}$ is given by $(\mathbf{A}^\top)_{ij} = A_{ji}$ for each (i, j).

In other words, the columns of $\mathbf{A}$ become the rows of $\mathbf{A}^\top$, and the rows of $\mathbf{A}$ become the columns of $\mathbf{A}^\top$.

Proposition (Properties of the Transpose)

Let A, B be matrices of compatible sizes and $\alpha \in \mathbb{R}$:

1. $(\mathbf{A}^\top)^\top = \mathbf{A}$
2. $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$
3. $(\alpha \mathbf{A})^\top = \alpha \mathbf{A}^\top$
4. $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$

Symmetric matrices

Definition

A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be **symmetric** if it is equal to its own transpose ($\mathbf{A} = \mathbf{A}^\top$), meaning that $A_{ij} = A_{ji}$ for all (i, j).

Note: antisymmetric matrix $\mathbf{A} = -\mathbf{A}^\top$

Note: In Lecture 4 we will remind the spectral theorem, e.g. for real-valued square matrices "If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric, then there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$."

The practical application of this theorem is a particular factorization of symmetric matrices, referred to as the **eigendecomposition** or **spectral decomposition**.

Other particular matrices

- Matrix of zeros, matrix of ones
- Triangular matrix
- Band matrix
- Block-diagonal matrix
- Shift matrix, circulant matrix
- ...

More examples here.

$\diamond$ Practice with numpy

Exercise : Let A, B be real-valued symmetric matrices of size $n \times n, n > 1$. Suppose that their product is symmetric. (a) Show that A and B commute. (b) Deduce that every diagonal matrix commutes with all other diagonal matrices.

Table of Contents

Invertible matrix

Definition

Let A be an $n \times n$ matrix. If there exists a matrix $B \in \mathbb{R}^{n\times n}$ such that

$$AB = BA = I_n,$$

then A is said to be **invertible** (or **nonsingular**), and B is called the **inverse** of A. The inverse of A is denoted by A^{-1}.

Invertible matrix

Definition

Let A be an $n \times n$ matrix. If there exists a matrix $B \in \mathbb{R}^{n \times n}$ such that

$$AB = BA = I_n,$$

then A is said to be **invertible** (or **nonsingular**), and B is called the **inverse** of A. The inverse of A is denoted by A^{-1}.

Caution

Not every square matrix is invertible.

Invertible matrix

Alternative definition through the lens of linear maps

Definition

Let A be an $n \times n$ matrix. If the linear transformation $x \mapsto Ax$ is bijective, we say that A is *invertible* and denote its inverse by A^{-1}. It satisfies:

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

where I_n is the $n \times n$ identity matrix.

Properties of invertible matrices

Proposition

If A and B are invertible $n \times n$ matrices, and $\alpha \in \mathbb{R} \setminus \{0\}$, then:

1. $(A^{-1})^{-1} = A$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $(A^\top)^{-1} = (A^{-1})^\top$
4. $(\alpha A)^{-1} = \frac{1}{\alpha}A^{-1}$
5. $I_n^{-1} = I_n$

Theorem of Equivalent Statements for an Invertible Matrix

Theorem

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

1. *Invertibility: A is invertible.*
2. *Trivial kernel:* $\ker(A) = \{\mathbf{0}\}$.
3. *Full rank:* $\mathrm{rank}(A) = n$.
4. *The columns (or rows) of A are linearly independent.*
5. *Span: The columns of A span $\mathbb{R}^n$, i.e.* $\mathrm{range}(A) = \mathbb{R}^n$.
6. *Linear map: A is bijective as a linear transformation $\mathbb{R}^n \to \mathbb{R}^n$.*
7. *A is surjective as a linear transformation $\mathbb{R}^n \to \mathbb{R}^n$.*
8. *Lin. stm.: For any $\mathbf{b} \in \mathbb{R}^n$, equation $A\mathbf{x} = \mathbf{b}$ has a unique solution.*
9. *Equivalence to identity: A is row-equivalent[a] to the identity matrix I_n.*
10. *Determinant:* $\det(A) \neq 0$ *(see next lecture on determinant).*

[a] Two matrices are said to be **row equivalent** if one can be changed to the other by a sequence of elementary row operations.

Theorem of basis change

Theorem

Let $\mathcal{B} = (u_1, u_2, ...u_n)$ and $\mathcal{B}' = (v_1, v_2, ...v_n)$ be two bases of a vector space E, $L : E \mapsto E$ a linear map, $A = [L]_{\mathcal{B}}$ the matrix of L in $\mathcal{B}$ and $B = [L]_{\mathcal{B}'}$ the matrix of L in $\mathcal{B}'$.
Let P be the matrix such that the j^{th} column is $[v_j]_{\mathcal{B}}$, the coordinates of basis vector v_j of $\mathcal{B}'$ in the basis $\mathcal{B}$.

$$P = [[v_1]_{\mathcal{B}} \dots [v_n]_{\mathcal{B}}], \textit{ so that } [x]_{\mathcal{B}} = P[x]_{\mathcal{B}'}$$

Then P is invertible and we have

$$B = P^{-1}AP.$$

Definition

With the above notations, A and B are called **similar** matrices. (FR: *matrices semblables*.)

Theorem of basis change

Same statement, other notations :

1. $[L]_{\mathcal{B}'} = P^{-1}[L]_{\mathcal{B}}P$
2. We can also write that $\mathrm{P}_{\mathcal{B}'}^{\mathcal{B}}$ is the matrix to change from $\mathcal{B}$ to $\mathcal{B}'$, i.e. $[x]_{\mathcal{B}'} = \mathrm{P}_{\mathcal{B}'}^{\mathcal{B}}[x]_{\mathcal{B}}$ and $[x]_{\mathcal{B}} = \mathrm{P}_{\mathcal{B}}^{\mathcal{B}'}[x]_{\mathcal{B}'}$.
   Then
   $$\mathrm{P}_{\mathcal{B}}^{\mathcal{B}'}\,\mathrm{P}_{\mathcal{B}'}^{\mathcal{B}} = I_n.$$
   and
   $$[L]_{\mathcal{B}'} = \mathrm{P}_{\mathcal{B}'}^{\mathcal{B}}\,[L]_{\mathcal{B}}\,\mathrm{P}_{\mathcal{B}}^{\mathcal{B}'}$$
3. Alternatively
   $$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{Id}_E)\,\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{Id}_E) = \mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(\mathrm{Id}_E) = I_n.$$
   $$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(L) = \mathrm{P}_{\mathcal{B}'}^{\mathcal{B}}\,\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(L)\,\mathrm{P}_{\mathcal{B}}^{\mathcal{B}'}$$
   $$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(L) = \mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(\mathrm{Id}_E)\,\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(L)\,\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{Id}_E)$$

Similar matrices VS equivalent matrices[1]

Definition (Equivalent matrices)

We consider two rectangular $m \times n$ matrices A and B. They are called **equivalent** if there exist an invertible $n \times n$ matrix P and an invertible $m \times m$ matrix Q such that

$$B = Q^{-1}AP$$

Equivalent matrices represent the same linear transformation $V \mapsto W$ under two different choices of a pair of bases of V and W, with P and Q being the change of basis matrices in V and W respectively.

[1] FR: *matrices semblables vs équivalentes*

Similar matrices VS equivalent matrices[1]

Definition (Equivalent matrices)

We consider two rectangular $m \times n$ matrices A and B. They are called **equivalent** if there exist an invertible $n \times n$ matrix P and an invertible $m \times m$ matrix Q such that

$$B = Q^{-1}AP$$

Equivalent matrices represent the same linear transformation $V \mapsto W$ under two different choices of a pair of bases of V and W, with P and Q being the change of basis matrices in V and W respectively.

By contrast, the notion of similarity is only defined for square matrices. Two $n \times n$ matrices A and B are similar if they represent the same endomorphism $V \mapsto V$ under different choices of basis for V. Similar matrices are equivalent (taking $Q = P$), but not reciproquely.

[1] FR: *matrices semblables vs équivalentes*

Orthogonal matrix

Definition

A square matrix U is called *orthogonal* if:

$$U^\top U = UU^\top = I.$$

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

Interpretation: the columns (and rows) of an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ form an orthonormal basis of $\mathbb{R}^n$

Proposition

The determinant of a real-valued orthogonal matrix is either 1 *or* -1*.*

Table of Contents

Trace of a matrix

Definition (Trace)

If $A = (a_{ij})$ is an $n \times n$ matrix, its *trace* is:

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

The trace of a matrix is obtained by summing its diagonal coefficients.

Trace of a matrix

Proposition (Properties of the Trace)

Let A, B *be* $n \times n$ *matrices, and* $\lambda \in \mathbb{R}$*:*

1. $\mathrm{Tr}(I_n) = n$
2. $\mathrm{Tr}(A + B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$
3. $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$
4. $\mathrm{Tr}(\lambda A) = \lambda \mathrm{Tr}(A)$

Trace of a matrix

Proposition (Properties of the Trace)

Let A, B be $n \times n$ matrices, and $\lambda \in \mathbb{R}$:

1. $\mathrm{Tr}(I_n) = n$
2. $\mathrm{Tr}(A + B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$
3. $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$
4. $\mathrm{Tr}(\lambda A) = \lambda \mathrm{Tr}(A)$

Proposition (Invariance under similarity)

Let $A \in \mathbb{R}^{n \times n}$ a matrix and $P \in \mathbb{R}^{n \times n}$ an invertible matrix.

$$\mathrm{Tr}(P^{-1}AP) = \mathrm{Tr}(A)$$

Table of Contents

Linear systems

Definition

The general form of a **system of linear equations** in the unknowns $x_1, \ldots, x_n$ is:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \quad\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

where $a_{ij} \in \mathbb{R}$ and $b_i \in \mathbb{R}$ are called coefficients.

Voc: Every n-tuple $(x_1, \ldots, x_n) \in \mathbb{R}^n$ that satisfies all equations is called a *solution* of the system.

Example of linear systems

- A system of equations **without a solution**.
- A system with a **unique solution**.
- A system with **redundancy** (infinitely many solutions).

Geometric interpretation in 2D

In dimension 2, each equation corresponds to a line in the plane.

- If the lines intersect at a point $\Rightarrow$ unique solution.
- If the lines are parallel and disjoint $\Rightarrow$ no solution.
- If the lines coincide $\Rightarrow$ infinitely many solutions.

Matrix Formulation

The system can be written as a matrix-vector product:

$$Ax = b$$

where

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Solving linear systems

Example

Particular solution of system (also called **special solution**)

Method for finding a general solution :

1. Find a particular solution to $Ax = b$.
2. Find all solutions to $Ax = 0$.
3. Combine the solutions from steps 1. and 2. to the general solution.

Note: Neither the general nor the particular solution is unique.

Elementary transformations of linear systems

Elementary transformations : keep the solution set the same, but transform the equation system into a simpler form.

- Exchange of two equations
- Addition of one equation to another
- Multiplication of an equation by a scalar $\lambda \in \mathbb{R}^\star$

Elementary transformations of linear systems

Elementary transformations : keep the solution set the same, but transform the equation system into a simpler form.

- Exchange of two equations
- Addition of one equation to another
- Multiplication of an equation by a scalar $\lambda \in \mathbb{R}^\star$

Gaussian elimination is a constructive algorithmic way for transforming any system of linear equations into a particular, more simple form called the **row-echelon form**.

Gaussian elimination

Definition (Row-echelon form)

A matrix is in row-echelon form if

1. All rows that contain only zeros are at the bottom of the matrix; correspondingly, all rows that contain at least one nonzero element are on top of rows that contain only zeros.
2. Looking at nonzero rows only, the first nonzero number from the left (also called the **pivot** or the **leading coefficient**) is always strictly to the leading coefficient right of the pivot of the row above it.
3. All entries below a pivot are zero.

The row-echelon form makes it easier to determine a particular solution.

Proposition

Every $m \times n$ matrix is row-equivalent to a unique reduced row-echelon form.

Gaussian Elimination

Idea: Reduce the system $Ax = b$ to an equivalent triangular system using elementary row operations. *See the MML book for a detailed exercise*. In

particular, practice with the examples of Section 2.1.

Gaussian elimination

Definition (Reduced Row Echelon Form)

The matrix describing an equation system is in **reduced row-echelon form** (also: row-reduced echelon form or row canonical form) if

1. It is in row-echelon form.
2. Every pivot is 1.
3. The pivot is the only nonzero entry in its column.

Calculating an Inverse Matrix by Gaussian Elimination

How to calculate an inverse Matrix by Gaussian elimination ?

Calculating an Inverse Matrix by Gaussian Elimination

How to calculate an inverse Matrix by Gaussian elimination ?

We apply elimination to the augmented matrix $[A \mid I_n]$ to obtain $[I_n \mid A^{-1}]$. If we bring the augmented equation system into reduced row-echelon form, we can read out the inverse on the right-hand side of the equation system. Hence, determining the inverse of a matrix is equivalent to solving systems of linear equations.

Connection with the rank

Rank of a matrix A: The maximum number of linearly independent rows (or columns) of A.

Interpretation in Gaussian Elimination:

- The number of *pivots* obtained in row echelon form equals $\mathrm{rank}(A)$.
- The rank determines whether a system has zero, one, or infinitely many solutions.

Cases:

- If $\mathrm{rank}(A) < \mathrm{rank}([A|b])$: the system is **inconsistent** (no solution).
- If $\mathrm{rank}(A) = \mathrm{rank}([A|b]) = n$: the system has a **unique solution**.
- If $\mathrm{rank}(A) = \mathrm{rank}([A|b]) < n$: the system has **infinitely many solutions**.

Properties of the reduced row echelon form

Pivot columns of A (in the *original* A) form a basis of the column space.
Nonzero rows of the reduced row echelon form build a basis of the row space.
Note: The number of free variables in a linear system is also called **nullity**, cf rank nullity theorem. For a system of n equations with n unknowns, with matrix A, $n =$ number of free variables $+ \; rank(A)$.

Next class

Test (30 min) on Lecture 1 + Lecture 2
Recap and practice : https://prismia.chat/shared/linear-algebra
(except Dot products and orthogonality)
Lecture 3: Determinant, Diagonalization
Further reading : https://mml-book.github.io/book/mml-book.pdf