Math Tutorial Booklet

to supplement the master's courses at the

Department of Statistics, LMU

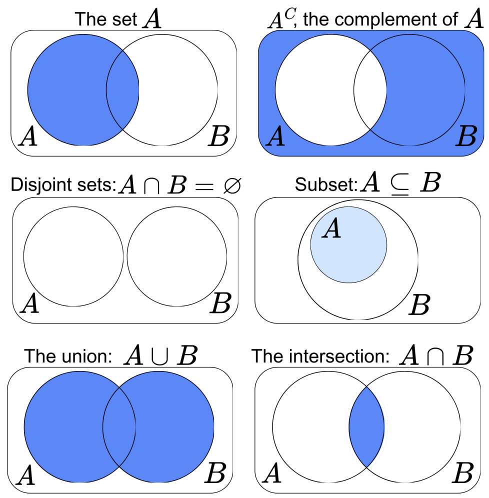
by Hannah Schulz-Künpel

Winter Semester 2024/25

Contents

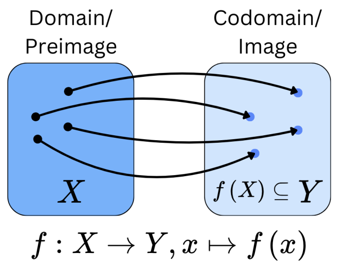
1 Some miscellaneous Basics

Illustration 1.1 (Set operations)



Quick reminder of the fundamental set operations. For a single element x in a set, we write $x \in A$.

Illustration 1.2



Visual representation of a function.

1.1 Normed and Metric Spaces

Definition 1.1 (Norm(ed Space)). Let $\mathcal{S}$ be a set (usually a vector space) over the real field $\mathbb{R}$. A norm on $\mathcal{S}$ is a function with domain $\mathcal{S}$ and codomain $[0, \infty[$, its value at an element (vector) x usually indicated by $\|x\|$ (or something similar), that satisfies the following properties:

1. $\|x\| = 0$ if and only if $x = 0$.
2. $\|\lambda x\| = |\lambda|\|x\|$ for all $x \in \mathcal{S}$ and $\lambda \in \mathbb{R}$.
3. (Triangle inequality) $\|x + y\| \leq \|x\| + \|y\|$ for all x and y in $\mathcal{S}$ [i.e. $\forall x, y \in \mathcal{S}$].

Definition 1.2 (Metric). A metric on the set $\mathcal{S}$ is a function $d : \mathcal{S} \times \mathcal{S} \to [0, \infty[$ that satisfies the following conditions (or axioms for metrics):

1. $\forall x, y \in \mathcal{S}$ we have $d(x, y) = 0$ if and only if $x = y$.
2. (Symmetry) $\forall x, y \in \mathcal{S}$ we have $d(x, y) = d(y, x)$.
3. (Triangle inequality) $\forall x, y, z \in \mathcal{S}$ we have

$$d(x, z) \leq d(x, y) + d(y, z) \,.$$

Definition 1.3. A metric space is a set $\mathcal{S}$ equipped with a metric; more precisely it is an ordered pair $(\mathcal{S}, d)$, where $\mathcal{S}$ is a set and d a metric on $\mathcal{S}$.

Remark 1.1. A norm on a vector space will always give rise to a metric on the same vector space by taking the norm of the difference between two vectors. Specifically, if $(\mathcal{S}, \|\cdot\|)$ is a normed vector space, then

$$d : \mathcal{S} \times \mathcal{S} \to \mathbb{R}, \quad (x, y) \mapsto \|x - y\|$$

is a metric on $\mathcal{S}$.

1.2 Open and Closed Sets

For the following definitions, let $\mathcal{S}$ be a set equipped with a metric d.

Definition 1.4 (ε-Balls). We define open balls and closed balls in $\mathcal{S}$. Consider $a \in \mathcal{S}$ and $\varepsilon > 0$:

a) The open ball with centre a and radius ε is the set

$$B_\varepsilon(a) = \{x \in \mathcal{S} : d(a, x) < \varepsilon\}$$

b) The closed ball with centre a and radius ε is the set

$$B_\varepsilon^-(a) = \{x \in \mathcal{S} : d(a, x) \leq \varepsilon\} \,.$$

Definition 1.5 (Open Set). A set $A \subset \mathcal{S}$ is open if and only if, for each $a \in A$, there exists $r > 0$, such that $B_r(a) \subset A$.

Definition 1.6 (Closed Set). A set $M \subset \mathcal{S}$ is said to be closed if its complement $M^C := X \setminus M$ is open.

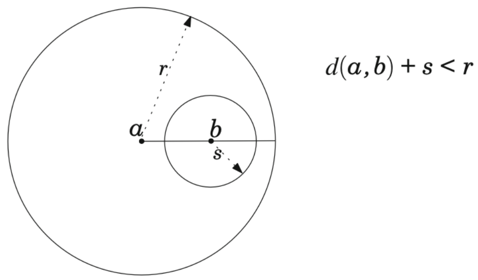
Proposition 1.1. *An open ball $B_r(a)$ is an open set.*

Proof. This is another example showing the importance of the triangle inequality. Let $b \in B_r(a)$. Choose s, such that $0 < s < r - d(a, b)$. Then $B_s(b) \subset B_r(a)$, because if $x \in B_s(b)$ we have

$$d(x, a) \leq d(x, b) + d(b, a) < s + d(b, a) < r.$$

□

Illustration 1.3 $B_r(a)$ and $B_s(b)$ for Euclidean distance.



This would enable us to guess that the condition $0 < s < r - d(a, b)$ leads to $B_s(b) \subset B_r(a)$. Source: [4, p.38].

Proposition 1.2. *A closed ball $B_r^-(a)$ is a closed set.*

Proposition 1.3. *(i) The union of open subsets in $\mathcal{S}$ is always open.*

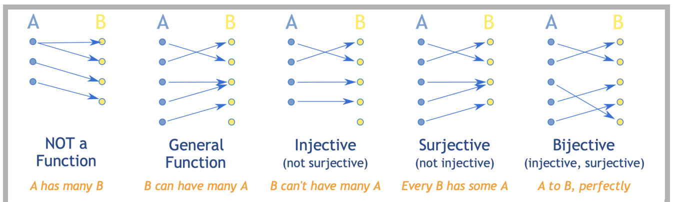
(ii) The intersection of closed subsets in $\mathcal{S}$ is always closed.

> **Note 1.1.** *The proofs of Propositions 1.2 and 1.3 may be found in [4], but are also very intuitive - feel free to try yourself!*

Definition 1.7 (Bounded set)**.** A set $A \subset \mathcal{S}$ is said to be bounded if there exist $a \in \mathcal{S}$ and $r > 0$, such that $A \subset B_r(a)$.

1.3 Minimal Function Properties

Illustration 1.4 Injective, Surjective, Bijective Functions



Nice representation of basic function characteristics. Source: MathIsFun.

Definition 1.8 (Continuous function)**.** Consider two metric spaces (X, d_X) and (Y, d_Y). Let $f : X \to Y$ and let $a \in X$. We say that f is continuous at the point a if the following condition is satisfied:

$$\forall \varepsilon > 0 \; \exists \delta > 0, \text{ such that } d_Y(f(x), f(a)) < \varepsilon \; \forall x \in X \text{ that satisfy } d_X(x, a) < \delta.$$

We say that f is continuous if it is continuous at every point $a \in X$.

> **Remark 1.2.** The condition of Definition 1.8 is called the ***epsilon-delta criterion***. Equivalently, we could say that a function is continuous at $a \in X$ if
>
> $$\forall \varepsilon > 0 \; \exists \delta > 0, \textit{ such that } f\left(B_\delta(a)\right) \subset B_\varepsilon(f(a)),$$
>
> where it should be understood by the context whether $B_r(\cdot)$ refers to X (left) or Y (right).

> **Note 1.2.** *There are several other ways to define continuity of functions, some of which we might explore in the future. The above version is just what is possible with the previous definitions in this booklet.*

Proposition 1.4. *Consider two metric spaces (X, d_X) and (Y, d_Y) and a function $f : X \to Y$. The following **three** conditions are equivalent:*

1. *f is continuous.*
2. *$f^{-1}(U)$ is open in X whenever U is open in Y.*
3. *$f^{-1}(U)$ is closed in X whenever U is closed Y.*

Proof. See [4, p. 54] □

2 Systems of Linear Equations & Matrix Algebra

Definition 2.1. *(i)* A **linear equation** in the variables $x_1, \ldots, x_n$, $n \in \mathbb{N}_{>0}$, is an equation that can be written in the form

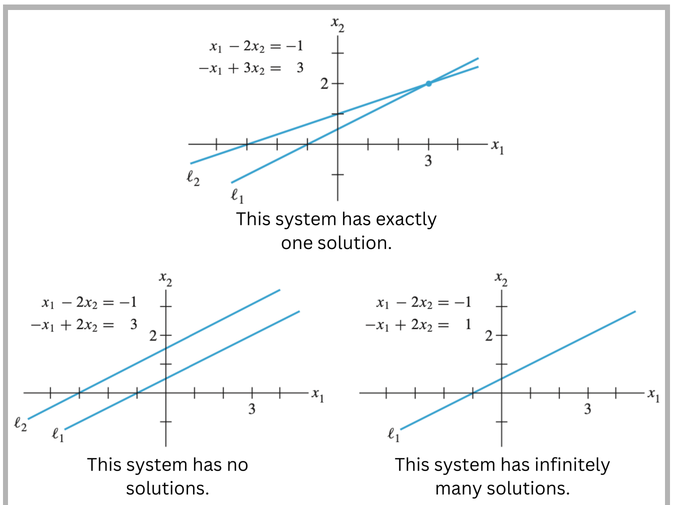
$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$$

where b and the **coefficients** $a_1, \ldots, a_n$ are real or complex numbers, usually known in advance.

(ii) A **system of linear equations** (or a *linear system*) is a collection of one or more linear equations involving the same variables - say, $x_1, \ldots, x_n$.

(iii) A **solution** of a given system of linear equations is a tuple $(s_1, s_2, \ldots, s_n)$ of numbers that makes each equation a true statement when the values $s_1, \ldots, s_n$ are substituted for $x_1, \ldots, x_n$, respectively.

Illustration 2.1 Visual example of linear systems



Source: [3, p.3]

Definition 2.2 (Matrix)**.** With $m, n \in \mathbb{N}_{>0}$, a $m \times n$ matrix $\boldsymbol{A}$ is an $m \cdot n$-tuple of elements $a_{ij}, i = 1, \ldots, m, j = 1, \ldots, n$, which is ordered according to a rectangular scheme consisting of m rows and n columns:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{m2} & \cdots & a_{nm} \end{bmatrix}.$$

In the most common case of $a_{ij} \in \mathbb{R}$, $\boldsymbol{A}$ is a *real-valued matrix.*
For an arbitrary $n \in \mathbb{N}_{>0}$, we refer to a $n \times 1$ matrix $\boldsymbol{v}$ as a (column) **vector**. The (row) vector containing the same values, but in a $1 \times n$ matrix, is denoted by $\boldsymbol{v}^\top$, see Definition 2.3 for the $^\top$-notation.

Matrix Addition/Multiplication

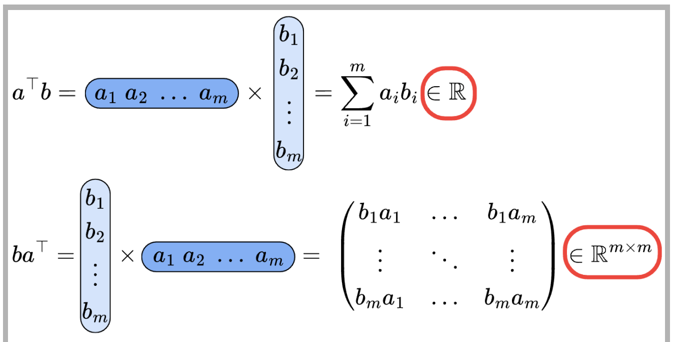
The sum of two real-valued matrices $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$ is defined as the elementwise sum, i.e.,

$$\boldsymbol{A}+\boldsymbol{B} := \begin{bmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & & \vdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{bmatrix} \in \mathbb{R}^{m\times n}.$$

For real-valued matrices $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times k}$, the elements c_{ij} of the product $\boldsymbol{C} = \boldsymbol{AB} \in \mathbb{R}^{m\times k}$ are computed as

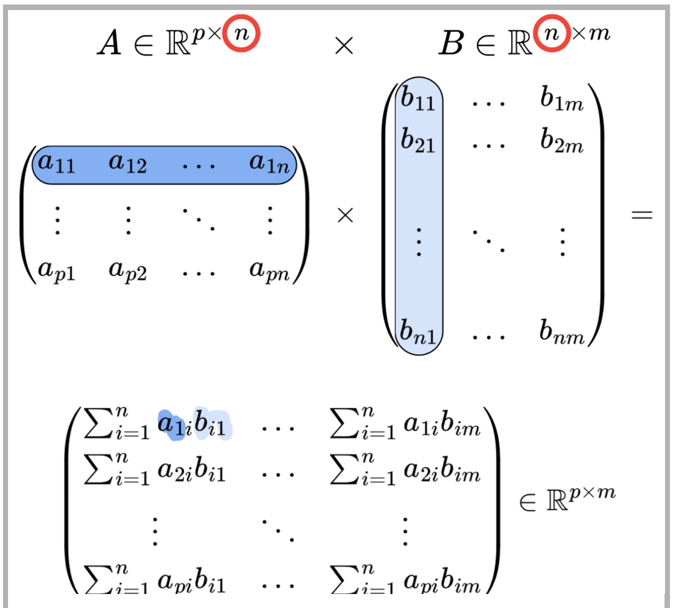
$$c_{ij} = \sum_{l=1}^{n} a_{il}b_{lj}, \quad i = 1,\dots,m, \quad j = 1,\dots,k.$$

Illustration 2.2 Vector multiplication



An illustration of the fact that multiplying a row with a column vector will yield a result of different dimension that multiplying a column with a row vector - a **scalar** vs a **matrix**!

Illustration 2.3 Matrix multiplication



Not all matrices may be multiplied together! The inner dimension (red circles in above illustration) always has to match. The dimension of the result will then be determined by the outer dimensions of the multiplied matrices.

Notation 2.1. *We denote both the matrix an vector of zero-elements as* $\mathbf{0}$*, with the dimension usually inferred from context. Otherwise, the dimension may be specified via the index, such as* $\mathbf{0}_n$ *or* $\mathbf{0}_{n\times m}$*. Furthermore, for* $n \in \mathbb{N}_{>0}$*, we define the real-valued identity matrix as*

$$\boldsymbol{I}_n := \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

Matrix Operation Properties

- Associativity:
  $\forall \boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{n\times p}, \boldsymbol{C} \in \mathbb{R}^{p\times q}:$
  $$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$$
- Distributivity:
  $\forall \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}, \boldsymbol{C}, \boldsymbol{D} \in \mathbb{R}^{n\times p}:$
  $$(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC}+\boldsymbol{BC}$$
  $$\boldsymbol{A}(\boldsymbol{C}+\boldsymbol{D}) = \boldsymbol{AC}+\boldsymbol{AD}$$
- Multiplication with the identity matrix:
  $$\forall \boldsymbol{A} \in \mathbb{R}^{m\times n}: \boldsymbol{I}_m\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{A}$$
  Note that $\boldsymbol{I}_m \neq \boldsymbol{I}_n$ for $m \neq n$.
- Multiplication with a scalar:
  $\forall \boldsymbol{A} \in \mathbb{R}^{m\times n}, \lambda \in \mathbb{R}:$
  $$\lambda\boldsymbol{A} = \boldsymbol{B}, \quad \text{with } b_{ij} = \lambda a_{ij}.$$

Definition 2.3 (Transpose)**.** For a $m \times n$ matrix $\boldsymbol{A}$, the $n \times m$ matrix $\boldsymbol{B}$ with $b_{ij} = a_{ji}$, $\forall i \in \{1,\dots,m\}, j \in \{1,\dots,n\}$, $m,n \in \mathbb{N}_{>0}$ is called the transpose of $\boldsymbol{A}$. We write $\boldsymbol{B} = \boldsymbol{A}^\top$.

Note 2.1. *In general,* $\boldsymbol{A}^\top$ *can be obtained by writing the columns of* $\boldsymbol{A}$ *as the rows of* $\boldsymbol{A}^\top$*.*

Definition 2.4 (Square and symmetric matrices)**.** When a matrix has the same number of rows and columns, it is called *square*. A square $n \times n$ matrix $\boldsymbol{A}$ is symmetric if $\boldsymbol{A} = \boldsymbol{A}^\top$.

Definition 2.5 (Inverse)**.** Consider a real-valued square matrix $A \in \mathbb{R}^{n\times n}$. If a matrix $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ with the property that $\boldsymbol{AB} = \boldsymbol{I}_n = \boldsymbol{BA}$ exists, $\boldsymbol{A}$ is called **invertible** and $\boldsymbol{B}$ is called the inverse of $\boldsymbol{A}$, denoted by $\boldsymbol{A}^{-1}$.

Calculation Rule 2.1. Inverse of Matrices

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} = \boldsymbol{A}^{-1}\boldsymbol{A}$$
$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$
$$(\boldsymbol{A}+\boldsymbol{B})^{-1} \neq \boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}$$

***Calculation Rule 2.2.* Transpose of Matrices**

$$\left(\boldsymbol{A}^\top\right)^\top = \boldsymbol{A}$$
$$(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$$
$$(\boldsymbol{A}+\boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$$

Illustration 2.4 Special Matrices

For $n \in \mathbb{N}_{>0}$ non-zero elements $a_1, a_2, \dots, a_n$, the following is called a **diagonal matrix:**

$$\text{diag}(a_1, a_2, \dots, a_n) = \begin{pmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & a_n \end{pmatrix}$$

Also, for $*$ denoting some arbitrary non-zero element, we often look at

$$\begin{pmatrix} * & * & \dots & * & * \\ 0 & * & \dots & * & * \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & * & * \\ 0 & 0 & \dots & 0 & * \end{pmatrix} \quad \& \quad \begin{pmatrix} * & 0 & \dots & 0 & 0 \\ * & * & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ * & \dots & * & * & 0 \\ * & * & \dots & * & * \end{pmatrix}$$

upper triangular matrices **lower triangular matrices**

Notation 2.2 (Compact representation and augmented matrix)**.** *Using matrix algebra, we can write a* ***system of linear equations*** *(see Definition 2.1) as*

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

The corresponding ***augmented matrix*** *is given by*

$$\left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right]$$

(where the vertical line is optional).

Definition 2.6. A rectangular matrix is in **echelon form** (or row echelon form) if it has the following three properties:

1. All nonzero rows are above any rows of all zeros.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zeros.

If a matrix in echelon form satisfies the following two additional conditions, then it is in **reduced echelon form**:

4. The leading entry in each nonzero row is equal to 1.
5. Each leading 1 is the only nonzero entry in its column.

Illustration 2.5 Row Echelon Form

EXAMPLE The following matrices are in echelon form. The leading entries ($\blacksquare$) may have any nonzero value; the starred entries ($*$) may have any value (including zero).

$$\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & \blacksquare & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * \end{bmatrix}$$

The following matrices are in reduced echelon form because the leading entries are 1's, and there are 0's below *and above* each leading 1.

$$\begin{bmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & * & 0 & 0 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 1 & 0 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 1 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * \end{bmatrix}$$

Source: Example 1 of [3]

***Calculation Rule 2.3.* Gaussian Elimination**

Gaussian Elimination refers to the repeated application of the following three operations in suitable order:

- **(Replacement)** Replace one row by the sum of itself and a multiple of another row. (I.e. add to one row a multiple of another row.)
- **(Interchange)** Interchange two rows.
- **(Scaling)** Multiply all entries in a row by a nonzero constant.

Among other applications, we can solve a system of linear equations by applying these operations to its augmented matrix to obtain the row echelon form and performing back substitution.

3 Vector Spaces

Definition 3.1. A **vector space** is a nonempty set V of objects, called vectors, on which are defined two operations, called addition and multiplication by scalars, subject to the **ten axioms** (or rules) listed below. The axioms must hold for all vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and scalars $c, d \in \mathbb{R}$:

(i) $\boldsymbol{u} + \boldsymbol{v} \in V$.

(ii) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.

(iii) $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$.

(iv) There exists a **zero** vector $\boldsymbol{0}$, so that $\boldsymbol{u} + \boldsymbol{0} = \boldsymbol{u}$.

(v) For each $\boldsymbol{u} \in V$, $\exists -\boldsymbol{u} \in V$, so that $\boldsymbol{u} + (-\boldsymbol{u}) = \boldsymbol{0}$.

(vi) $c\boldsymbol{u} \in V$.

(vii) $c(\boldsymbol{u} + \boldsymbol{v}) = c\boldsymbol{u} + c\boldsymbol{v}$.

(viii) $(c + d)\boldsymbol{u} = c\boldsymbol{u} + d\boldsymbol{u}$.

(ix) $c(d\boldsymbol{u}) = (cd)\boldsymbol{u}$.

(x) $1\boldsymbol{u} = \boldsymbol{u}$.

Definition 3.2 (Subspace)**.** A subspace of a vector space V is a subset H of V that has three properties:

a. The zero vector of V is in H.

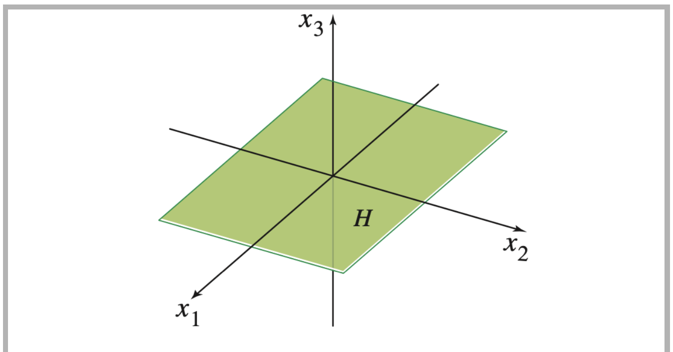
b. H is closed under vector addition. That is, for each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u}+\mathbf{v}$ is in H.

c. H is closed under multiplication by scalars. That is, for each $\mathbf{u}$ in H and each scalar c, the vector $c\mathbf{u}$ is in H.

Example 3.1 (Some vector spaces)

1. The set $V=\{\mathbf{0}\}$ with the operations $\mathbf{0}+\mathbf{0}=\mathbf{0}$ und $c\mathbf{0}=\mathbf{0}$, $\forall c\in\mathbb{R}$. is a vector space.
2. The set of all $m\times n$ - matrices with the usual operations for addition and scalar multiplication is a vector space. Here, the zero vector is the zero matrix $\mathbf{0}\in\mathbb{R}^{m\times n}$.
3. It is easily verified that for any $n\in\mathbb{N}_{>0}$, $\mathbb{R}^n$ is a vector space.

Illustration 3.1



The (x_1,x_2) plane as a subspace of $\mathbb{R}^3$. Source: Figure 7 of [3]

Definition 3.3 (Span)**.** The set of all linear combinations of given vectors $\boldsymbol{a}_1,\dots,\boldsymbol{a}_k\in\mathbb{R}^n$ is called the span or the set spanned by the vectors:

$$\operatorname{span}\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_k\}=\{c_1\boldsymbol{a}_1+\dots+c_k\boldsymbol{a}_k : c_1,\dots,c_k\in\mathbb{R}\}$$

Proposition 3.1. *For any subset* $S=\{s_1,\dots,s_k\}$ *with* $s_1,\dots,s_k\in V$, $\operatorname{span}\{s_1,\dots,s_k\}$ *is a subspace of* V.

Proof. Properties a and c from Definition 3.2 clearly hold by setting $c_1=\dots=c_k=0$ and $c_i=c$, $c_j=0$ $\forall\{1,\dots,k\}\ni j\neq i$, respectively. So, it suffices to show that $\operatorname{span}\{s_1,\dots,s_k\}$ is closed under linear combinations. Let $u,v\in\operatorname{span}\{s_1,\dots,s_k\}$ and λ,μ be constants. By the definition of $\operatorname{span}\{s_1,\dots,s_k\}$, there are constants c_i and d_i such that:

$$\begin{aligned} u &= c_1s_1+c_2s_2+\dots \\ v &= d_1s_1+d_2s_2+\dots \\ \Rightarrow \lambda u+\mu v &= \lambda\sum_{i=1}^k c_is_i+\mu\sum_{i=1}^k d_is_i \\ &= (\lambda c_1+\mu d_1)\,s_1+(\lambda c_2+\mu d_2)\,s_2+\dots \end{aligned}$$

This last sum is a linear combination of elements of S, and is thus in $\operatorname{span}\{s_1,\dots,s_k\}$. Then $\operatorname{span}\{s_1,\dots,s_k\}$ is closed under linear combinations, and is thus a subspace of V. □

Definition 3.4. An indexed set of vectors $\{\mathbf{v}_1,\dots,\mathbf{v}_p\}$ is said to be **linearly independent** if the vector equation

$$x_1\mathbf{v}_1+x_2\mathbf{v}_2+\dots+x_p\mathbf{v}_p=\mathbf{0}$$

has only the trivial solution. The set $\{\mathbf{v}_1,\dots,\mathbf{v}_p\}$ is said to be **linearly dependent** if there exist weights $c_1,\dots,c_p$, not all zero, such that

$$c_1\mathbf{v}_1+c_2\mathbf{v}_2+\dots+c_p\mathbf{v}_p=\mathbf{0}\,.$$

Definition 3.5 (Basis)**.** Let H be a subspace of a vector space V. An indexed set of vectors $\mathcal{B}=\{\mathbf{b}_1,\dots,\mathbf{b}_p\}$ in V is a basis for H if

(i) $\mathcal{B}$ is a linearly independent set, and

(ii) the subspace spanned by $\mathcal{B}$ coincides with H; that is,
$$H=\operatorname{Span}\{\mathbf{b}_1,\dots,\mathbf{b}_p\}\,.$$

Example 3.2 (Basis of $\mathbb{R}^3$)

The standard basis of $\mathbb{R}^3$ is $\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$,

but so is $\left\{\begin{pmatrix}a\\0\\0\end{pmatrix},\begin{pmatrix}0\\b\\0\end{pmatrix},\begin{pmatrix}0\\0\\c\end{pmatrix}\right\}$ for any $a,b,c\in\mathbb{R}$.

Can you think of more bases for $\mathbb{R}^3$?

Definition 3.6. If a vector space V is spanned by a finite set, then V is said to be **finite-dimensional**, and the dimension of V, written as $\dim(V)$, ***is the number of vectors in a basis for*** V. The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero. If V is not spanned by a finite set, then V is said to be **infinite-dimensional**.

Proposition 3.2 (The Basis Theorem)**.** *Let V be a p-dimensional vector space, $p\geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Any set of exactly p elements that spans V is automatically a basis for V.*

Proof. See Theorem 12 on page 229 of [3]. □

Proposition 3.3. *If a vector space V has a basis $\mathcal{B}=\{\mathbf{b}_1,\dots,\mathbf{b}_n\}$, then any set in V containing more than n vectors must be linearly dependent.*

Proof. See Theorem 9 on page 227 of [3]. □

Definition 3.7 (Column space)**.** The column space of an $m\times n$ matrix A, written as $\operatorname{Col}(A)$, is the set of all linear combinations of the columns of A. If $A=\begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}$, then

$$\operatorname{Col}(A)=\operatorname{Span}\{\mathbf{a}_1,\dots,\mathbf{a}_n\}\,.$$

Definition 3.8. (Kernel) The **kernel** or **null space** of an $m \times n$ matrix A, written as Nul A, is the set of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. In set notation,

$$\text{Ker}(A) = \text{Nul}(A) = \{\mathbf{x} : \mathbf{x} \text{ is in } \mathbb{R}^n \text{ and } A\mathbf{x} = \mathbf{0}\} .$$

Definition 3.9 (Linear transformation)**.** A **linear transformation** T from a vector space V into a vector space W is a rule that assigns to each vector $\mathbf{x}$ in V a unique vector $T(\mathbf{x})$ in W, such that

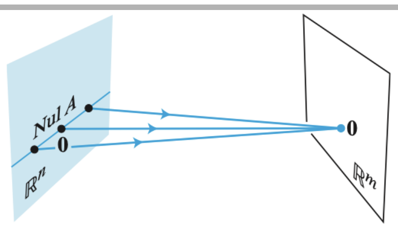
(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}$ in V, and

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u}$ in V and all scalars c.

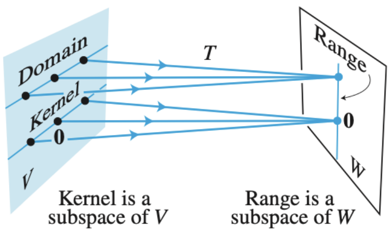
Definition 3.10. Let $T : V \to W$ be a linear mapping between two vector spaces V and W.

- The image set is defined as
$$\text{Im}(T) = \{\boldsymbol{w} \in W : T(\boldsymbol{v}) = \boldsymbol{w} \text{ for a } \boldsymbol{v} \in V\}.$$
and the rank of T is defined as the dimension of the image, i.e. $\text{rank}(T) = \dim(\text{Im}(T))$.
- The kernel is defined as
$$\text{Ker}(T) = \{\boldsymbol{v} \in V : T(\boldsymbol{v}) = \mathbf{0}\} .$$

Illustration 3.2



A more dynamic description of $\text{Ker}(A)$ is the set of all x in $\mathbb{R}^n$ that are mapped into the zero vector of $\mathbb{R}^m$ via the linear transformation $x \longmapsto Ax$. Source: Figure 1 on page 201 of [3].



Subspaces associated with a linear transformation. Source: Figure 2 on page 206 of [3].

Definition 3.11 (Rank)**.** The rank of a matrix A is defined as the dimension or its column space, i.e.

$$\text{rank}(A) = \dim(\text{Col}(A)) .$$

A matrix $A \in \mathbb{R}^{m \times n}$ (where m can be equal to n) is said to have *full rank*, if $\text{rank}(A) = \max\{m, n\}$.

Note 3.1. *Sometimes, equivalently, one refers to* ***column rank*** *and* ***row rank*** *as the number of linearly independent columns or rows, respectively, of a matrix.*
A fundamental fact is that the column rank always equals the row rank. This is equivalent to the following fact:

$$\text{rank}(A) = \text{rank}(A^\top) .$$

Calculation Rule 3.1. **Determining the rank**

Generally, the rank of a matrix is calculated by bringing it into a simpler form (like row Echelon form via Gaussian elimination). The rank then equals the number of the number of non-zero rows and also the number of pivots (or basic columns).

Theorem 3.1 (The Rank (nullity) Theorem)**.** *If a matrix A has n columns, then* $\text{rank}(A) + \dim(\text{Ker}(A)) = n$.

Note 3.2. *In linear algebra, subspaces often arise in two variants:*

- *As the solution set of an linear system* $A\mathbf{x} = \mathbf{0}$.
- *As a span of given vectors* $\text{span}\{v_1, \ldots, v_n\}$.

Furthermore, we refer to those variables in a system of linear equations that are not constrained by the equations, allowing them to take any value, as ***free variables****. They correspond to columns in the matrix that do not have a leading 1 (pivot) in reduced row echelon form. The number of free variables in* $A\mathbf{x} = \mathbf{0}$ *will always correspond to the dimension of the kernel of* A, $\dim(\text{Ker}(A))$.

References

[1] Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. *Mathematics for Machine Learning*. Cambridge University Press, 2020.

[2] A. Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer International Publishing, 2020.

[3] D.C. Lay, S.R. Lay, and J. McDonald. *Linear Algebra and Its Applications*. Pearson, 2016.

[4] R. Magnus. *Metric Spaces: A Companion to Analysis*. Springer Undergraduate Mathematics Series. Springer International Publishing, 2022.

[5] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer New York, 2013.