Introduction to Convexity

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 500

1. The Euclidean Space $\mathbb{R}^n$
   - The Euclidean Space $\mathbb{R}^n$
   - Flats
   - Dimension
   - Hyperplanes
   - Affine Transformations
   - Length, Distance and Angle
   - Open Sets and Closed Sets
   - Convergence and Compactness
   - Continuity

Subsection 1

The Euclidean Space $\mathbb{R}^n$

Vector Space Operations in $\mathbb{R}^3$

- In three-dimensional coordinate geometry a **point** or **vector** is determined by its **coordinates** x, y, z relative to some rectangular coordinate system.
- We identify the point or vector with the ordered triple (x, y, z).
- Vectors are **added** together according to a **parallelogram law**, which is equivalent to the addition of corresponding coordinates.
- The word **scalar** is used as a synonym for real number.
- The **product** of a scalar and a vector is equivalent to the multiplication of each coordinate of the vector by the scalar.
- Thus, if (x, y, z) and (u, v, w) are vectors, and λ is a scalar, then

$$\begin{array}{rcl} (x,y,z)+(u,v,w) & = & (x+u,y+v,z+w); \\ \lambda(x,y,z) & = & (\lambda x, \lambda y, \lambda z). \end{array}$$

- These equations can be extended in the natural way to define vector addition and scalar multiplication of real n-tuples.

Euclidean Space $\mathbb{R}^n$

- For each positive integer n, denote by $\mathbb{R}^n$ the set of all n-tuples $(x_1, \ldots, x_n)$ of real numbers.
- Then $\mathbb{R}^n$ is called the n-**dimensional Euclidean space**.
- Each element $\boldsymbol{x} = (x_1, \ldots, x_n)$ of $\mathbb{R}^n$ is called a **point** or **vector** of $\mathbb{R}^n$ and the real numbers $x_1, \ldots, x_n$ are called the **coordinates** of $\boldsymbol{x}$.
- For $n = 1$, we identify the 1-tuple $\boldsymbol{x} = (x_1)$ with the real number x_1 itself, so that $\mathbb{R}^1$ becomes simply $\mathbb{R}$, the set of real numbers.
- For $n = 1, 2, 3$, we often write x, (x, y), (x, y, z) instead of (x_1), (x_1, x_2), (x_1, x_2, x_3).
- Geometrically, $\mathbb{R}^1$ can be thought of as a line, $\mathbb{R}^2$ as a plane, and $\mathbb{R}^3$ as the set of points in space.
- Lower case Roman letters such as $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ will denote points of $\mathbb{R}^n$, lower case Roman and Greek letters such as $x, y, z, \lambda, \mu, \nu$ will denote scalars, and capital Roman letters such as A, B, C will denote subsets of $\mathbb{R}^n$.

Addition and Scalar Multiplication

- **Addition** and **scalar multiplication** in $\mathbb{R}^n$ are defined coordinatewise.
- Thus, if $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$, and λ is a scalar, then

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \ldots, x_n + y_n) \quad \text{and} \quad \lambda \boldsymbol{x} = (\lambda x_1, \ldots, \lambda x_n).$$

- The vector $(0, \ldots, 0)$ of $\mathbb{R}^n$, all of whose coordinates are 0, is denoted by $\mathbf{0}$ and is called the **zero vector** or **origin** of $\mathbb{R}^n$.
- The vector in $\mathbb{R}^n$ whose only non-zero coordinate is a 1 in the ith position is denoted by $\boldsymbol{e}_i$ and is called the ***i*th elementary vector**.
- A point of $\mathbb{R}^n$ all of whose coordinates are integers is called a **lattice point**.
- The vector $(-1)\boldsymbol{x}$ is written simply as $-\boldsymbol{x}$.
- **Vector subtraction** is defined by the rule $\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{x} + (-1)\boldsymbol{y}$.
- It is sometimes convenient to write $\frac{\boldsymbol{x}}{\lambda}$ for $\frac{1}{\lambda}\boldsymbol{x}$.

$\mathbb{R}^n$ as a Real Vector Space

- The set $\mathbb{R}^n$, equipped with the above operations of vector addition and scalar multiplication, is a **real vector space**.
- This means that, if $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$, then the following relations hold:
  (i) $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$;
  (ii) $\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}$;
  (iii) $\boldsymbol{x} + \mathbf{0} = \boldsymbol{x}$;
  (iv) $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$;
  (v) $1\boldsymbol{x} = \boldsymbol{x}$;
  (vi) $\lambda(\mu\boldsymbol{x}) = (\lambda\mu)\boldsymbol{x}$;
  (vii) $\lambda(\boldsymbol{x} + \boldsymbol{y}) = \lambda\boldsymbol{x} + \lambda\boldsymbol{y}$;
  (viii) $(\lambda + \mu)\boldsymbol{x} = \lambda\boldsymbol{x} + \mu\boldsymbol{x}$.

Extending Operations on Sets

- We extend the operations of vector addition and scalar multiplication to subsets of $\mathbb{R}^n$ by defining:

$$A+B=\{\boldsymbol{a}+\boldsymbol{b}:\boldsymbol{a}\in A,\boldsymbol{b}\in B\} \quad \text{and} \quad \lambda A=\{\lambda\boldsymbol{a}:\boldsymbol{a}\in A\},$$

  where $A,B\subseteq\mathbb{R}^n$ and $\lambda\in\mathbb{R}$.
- The set $A+B$ is called the **vector sum** of A and B.
- It follows from the above definitions that both sets $A+B$ and λA are empty when A is empty.
- We write $-A$ for the set $(-1)A$, and $A-B$ for the set $A+(-B)$.
- It is sometimes convenient to write $\frac{A}{\lambda}$ for $\frac{1}{\lambda}A$.

Symmetric Sets

- The set A in $\mathbb{R}^n$ is said to be **0-symmetric**, or simply **symmetric**, if $-A = A$.
- Geometrically, A is symmetric if it is its own reflection in the origin.
- Examples of symmetric sets in $\mathbb{R}^2$ are:
  - ellipses centered at the origin;
  - parallelograms with centers at the origin;
  - lines through the origin;
  - $\mathbb{R}^2$ itself.

Translates

- The set $\{\boldsymbol{a}\}+B$, where $\boldsymbol{a}\in\mathbb{R}^n$, is often written as $\boldsymbol{a}+B$ and is called a **translate** of B or, more precisely, the **translate of** B **by** $\boldsymbol{a}$.
- It is an easy exercise in set theory to show that

$$A+B=\bigcup(\boldsymbol{a}+B:\boldsymbol{a}\in A),$$

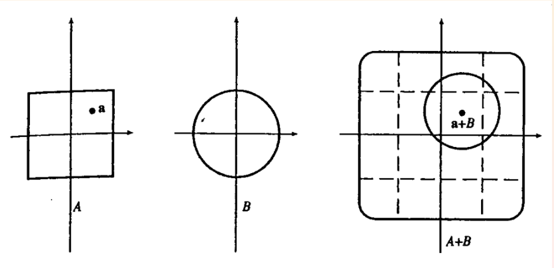
  i.e., $A+B$ is the union of all translates of B by vectors $\boldsymbol{a}$ in A.
- This result can help us to visualize $A+B$ in simple cases.

Example

- Suppose that A and B are the square and the circular disc in $\mathbb{R}^2$ defined by the equations

$$A = \{(x, y) : |x|, |y| \leq 1\}, \quad B = \{(x, y) : x^2 + y^2 \leq 1\}.$$

- Then $\boldsymbol{a} + B$ is the circular disc with center $\boldsymbol{a}$ and radius 1;
- $A + B$ is the union of all such discs for $\boldsymbol{a} \in A$.



Caution with Set Operations

- Vector addition and scalar multiplication, when applied to sets in $\mathbb{R}^n$, do not have all the properties one might expect, and the reader is warned to be cautious.
- For example, it is not always true that $A+A=2A$.

  To see this, let A consist of distinct points $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{R}^n$.

  Then $A+A=\{2\boldsymbol{a},2\boldsymbol{b},\boldsymbol{a}+\boldsymbol{b}\}$, whereas $2A=\{2\boldsymbol{a},2\boldsymbol{b}\}$.

Properties of Set Operations

- Properties (i)-(viii) above do, however, partially generalize to give the following easily verified results:
  (i)* $A+B=B+A$;
  (ii)* $A+(B+C)=(A+B)+C$;
  (iii)* $A+\mathbf{0}=A$;
  (iv)* $\mathbf{0}\in A+(-A)$ when $A\neq\emptyset$;
  (v)* $1A=A$;
  (vi)* $\lambda(\mu A)=(\lambda\mu)A$;
  (vii)* $\lambda(A+B)=\lambda A+\lambda B$;
  (viii)* $(\lambda+\mu)A\subseteq\lambda A+\mu A$.

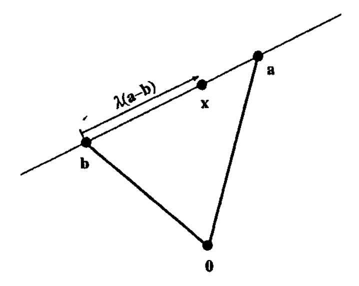
Subsection 2

Flats

Equation of a Line in $\mathbb{R}^3$

- For each point $\boldsymbol{x}$ on the line through distinct points $\boldsymbol{a}$ and $\boldsymbol{b}$ of $\mathbb{R}^3$, there exists a unique scalar λ such that

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{b} + \lambda(\boldsymbol{a} - \boldsymbol{b}) \\ &= \lambda \boldsymbol{a} + (1-\lambda)\boldsymbol{b}. \end{aligned}$$



- Conversely, each point $\boldsymbol{x}$ of this form lies on the line through $\boldsymbol{a}$ and $\boldsymbol{b}$.
- Thus the line through $\boldsymbol{a}$ and $\boldsymbol{b}$ is the set $\{\lambda \boldsymbol{a} + (1-\lambda)\boldsymbol{b} : \lambda \in \mathbb{R}\}$, which can also be written in the symmetrical form $\{\lambda \boldsymbol{a} + \mu \boldsymbol{b} : \lambda + \mu = 1\}$.
- We note that the subset

$$\{\lambda \boldsymbol{a} + (1-\lambda)\boldsymbol{b} : 0 \le \lambda \le 1\} = \{\lambda \boldsymbol{a} + \mu \boldsymbol{b} : \lambda, \mu \ge 0, \lambda + \mu = 1\}$$

of the line through $\boldsymbol{a}$ and $\boldsymbol{b}$ is the line segment joining $\boldsymbol{a}$ and $\boldsymbol{b}$.

Flats

- The **line** through distinct points $\boldsymbol{a}$ and $\boldsymbol{b}$ of $\mathbb{R}^n$ is the set $\{\lambda\boldsymbol{a}+\mu\boldsymbol{b} : \lambda+\mu=1\}$.
- Clearly this set contains both $\boldsymbol{a}$ and $\boldsymbol{b}$, and its points can be placed into a bijective correspondence with the points of the real line $\mathbb{R}$ itself.
- The set A in $\mathbb{R}^n$ is called a **flat** if whenever it contains two points, it also contains the entire line through them.
- Expressed algebraically, A is a flat if $\lambda\boldsymbol{a}+\mu\boldsymbol{b}\in A$ whenever $\boldsymbol{a},\boldsymbol{b}\in A$ and $\lambda+\mu=1$.
- Equivalently, A is a flat if $\lambda A+\mu A\subseteq A$ whenever $\lambda+\mu=1$.
- Synonyms for flat used by other authors are: **affine set**, **affine variety**, **affine manifold**, **linear variety**, and **linear manifold**.
- The empty set, singletons, lines, and $\mathbb{R}^n$ itself are examples of flats in $\mathbb{R}^n$. Planes are flats in $\mathbb{R}^3$.

Flats Containing the Origin

- Let A be a flat in $\mathbb{R}^n$ which contains the origin.
- Suppose that $\boldsymbol{a}, \boldsymbol{b} \in A$ and $\lambda \in \mathbb{R}$.
- Since A is a flat and $\boldsymbol{a}, \mathbf{0} \in A$, $\lambda \boldsymbol{a} + (1-\lambda)\mathbf{0} \in A$, i.e., $\lambda \boldsymbol{a} \in A$. Thus A is closed under scalar multiplication.
- Since A is a flat and $\boldsymbol{a}, \boldsymbol{b} \in A$, $\frac{1}{2}\boldsymbol{a} + \frac{1}{2}\boldsymbol{b} \in A$. But A is closed under scalar multiplication. So $2(\frac{1}{2}\boldsymbol{a} + \frac{1}{2}\boldsymbol{b}) \in A$, i.e., $\boldsymbol{a} + \boldsymbol{b} \in A$. Thus A is closed under addition.
- Hence A is a non-empty subset of $\mathbb{R}^n$ which is closed under addition and scalar multiplication, i.e., A is a subspace of the real vector space $\mathbb{R}^n$.
- Trivially, a subspace of $\mathbb{R}^n$ is a flat containing the origin.
- We have shown that flats through the origin in $\mathbb{R}^n$ are precisely the subspaces of $\mathbb{R}^n$.

Relation Between Flats and Subspaces

Theorem

The non-empty flats in $\mathbb{R}^n$ are precisely the translates of subspaces of $\mathbb{R}^n$.

- Suppose first that A is a non-empty flat in $\mathbb{R}^n$. Let $\boldsymbol{a} \in A$.

  We show that $A - \boldsymbol{a}$ is a flat. Let $\boldsymbol{x}, \boldsymbol{y} \in A - \boldsymbol{a}$ and $\lambda + \mu = 1$.

  Then $\boldsymbol{x} + \boldsymbol{a}, \boldsymbol{y} + \boldsymbol{a} \in A$. So

$$\lambda(\boldsymbol{x} + \boldsymbol{a}) + \mu(\boldsymbol{y} + \boldsymbol{a}) = \lambda\boldsymbol{x} + \mu\boldsymbol{y} + \boldsymbol{a} \in A.$$

  Thus, $\lambda\boldsymbol{x} + \mu\boldsymbol{y} \in A - \boldsymbol{a}$, and $A - \boldsymbol{a}$ is a flat.

  Since $A - \boldsymbol{a}$ contains the origin, it must be a subspace of $\mathbb{R}^n$.

  Hence the non-empty flat A is the translate of the subspace $A - \boldsymbol{a}$ of $\mathbb{R}^n$ by the vector $\boldsymbol{a}$.

Relation Between Flats and Subspaces (Cont'd)

- Suppose next that A is a subspace of $\mathbb{R}^n$ and that $\boldsymbol{u} \in \mathbb{R}^n$. We show that $A + \boldsymbol{u}$ is a flat. Let $\boldsymbol{x}, \boldsymbol{y} \in A + \boldsymbol{u}$ and $\lambda + \mu = 1$. Then there exist $\boldsymbol{a}, \boldsymbol{b} \in A$ such that $\boldsymbol{x} = \boldsymbol{a} + \boldsymbol{u}$, $\boldsymbol{y} = \boldsymbol{b} + \boldsymbol{u}$. So

$$\lambda \boldsymbol{x} + \mu \boldsymbol{y} = \lambda \boldsymbol{a} + \mu \boldsymbol{b} + \boldsymbol{u} \in A + \boldsymbol{u},$$

since $\lambda \boldsymbol{a} + \mu \boldsymbol{b} \in A$, as A is a subspace of $\mathbb{R}^n$.

This shows that $A + \boldsymbol{u}$ is a flat.

Uniqueness of Subspace

Corollary

Each non-empty flat in $\mathbb{R}^n$ is the translate of precisely one subspace of $\mathbb{R}^n$.

- Let A be a non-empty flat in $\mathbb{R}^n$. Suppose that A is a translate of both the subspaces B and C of $\mathbb{R}^n$. Then C must be a translate of B. So there exists $\boldsymbol{b} \in \mathbb{R}^n$ such that $C = B + \boldsymbol{b}$. Since 0 lies in C, it follows that $-\boldsymbol{b}$, and hence $\boldsymbol{b}$, lies in B. Thus $C = B + \boldsymbol{b} \subseteq B$. By symmetry, $B \subseteq C$. Hence $B = C$, and A is the translate of precisely one subspace of $\mathbb{R}^n$.

Parallel Flats

- The observation that two (distinct) lines in $\mathbb{R}^2$ are parallel if and only if one is a translate of the other prompts the following definition.
- In $\mathbb{R}^n$ a flat A is said to be **parallel** to a flat B if each is a translate of the other.
- The relation of parallelism is an equivalence relation on the family of all flats in $\mathbb{R}^n$.
- This notion of parallelism does not quite accord with that used in elementary geometry on two counts:
  - Firstly, a flat is considered to be parallel to itself.
  - Secondly, it only allows parallelism between flats of the same dimension. For example, we cannot speak of a line being parallel to a plane.
- The preceding corollary shows that each non-empty flat in $\mathbb{R}^n$ is parallel to precisely one subspace of $\mathbb{R}^n$.

Closure Under Intersections

Theorem

The intersection of an arbitrary family of flats in $\mathbb{R}^n$ is a flat.

- Let $(A_i : i \in I)$ be a family of flats in $\mathbb{R}^n$.
  Let $\boldsymbol{a}, \boldsymbol{b} \in \bigcap(A_i : i \in I)$ and $\lambda + \mu = 1$.
  Then $\boldsymbol{a}, \boldsymbol{b} \in A_i$. As A_i is a flat, $\lambda\boldsymbol{a} + \mu\boldsymbol{b} \in A_i$, for each $i \in I$.
  Thus, $\lambda\boldsymbol{a} + \mu\boldsymbol{b} \in \bigcap(A_i : i \in I)$.
  This shows that the intersection is a flat.

Affine Hull

- The **affine hull** $\text{aff}A$ of a set A in $\mathbb{R}^n$ is the intersection of all flats in $\mathbb{R}^n$ containing A.
- Such flats exist, since $\mathbb{R}^n$ is a flat containing A.
- In view of the preceding theorem, $\text{aff}A$ is a flat which contains A.
- Moreover, if B is any flat in $\mathbb{R}^n$ containing A, then $\text{aff}A \subseteq B$.
- Thus, we may refer to $\text{aff}A$ as the smallest flat in $\mathbb{R}^n$ containing A.
- Clearly, A is a flat if and only if $A = \text{aff}A$.
- Moreover, $\text{aff}(\text{aff}A) = \text{aff}A$.
- Another easy result is that, if $A \subseteq B$, then $\text{aff}A \subseteq \text{aff}B$.

Affine Hull in $\mathbb{R}^3$

- In the space $\mathbb{R}^3$:
  - The affine hull of two distinct points is the line through them;
  - The affine hull of three non-collinear points is the plane which they determine;
  - The affine hull of four non-coplanar points is the whole space $\mathbb{R}^3$ itself.

Generalized Flat Relation

- By definition, a set A in $\mathbb{R}^n$ is a flat if $\lambda \boldsymbol{a} + \mu \boldsymbol{b} \in A$ whenever $\boldsymbol{a}, \boldsymbol{b} \in A$ and $\lambda + \mu = 1$.
- This defining relation of a flat implies a more general one, as we now establish in the following fundamental theorem.

Theorem

Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ be points of a flat A in $\mathbb{R}^n$. Let $\lambda_1 + \cdots + \lambda_m = 1$. Then $\lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_m \boldsymbol{a}_m \in A$.

- Let $\boldsymbol{a} \in A$. Then the points $\boldsymbol{a}_1 - \boldsymbol{a}, \ldots, \boldsymbol{a}_m - \boldsymbol{a}$ lie in the subspace $A - \boldsymbol{a}$ of $\mathbb{R}^n$, whence so too does the point

$$\lambda_1(\boldsymbol{a}_1 - \boldsymbol{a}) + \cdots + \lambda_m(\boldsymbol{a}_m - \boldsymbol{a}) = \lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_m \boldsymbol{a}_m - \boldsymbol{a}.$$

Hence $\lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_m \boldsymbol{a}_m \in A$.

Affine Combinations and the Affine Hull

- A point $\boldsymbol{x}$ is said to be an **affine combination** of points $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_m$ in $\mathbb{R}^n$ if there exist scalars $\lambda_1,\ldots,\lambda_m$ with $\lambda_1+\cdots+\lambda_m=1$ such that

$$\boldsymbol{x}=\lambda_1\boldsymbol{a}_1+\cdots+\lambda_m\boldsymbol{a}_m.$$

- The preceding theorem can now be expressed as: Every affine combination of points of a flat in $\mathbb{R}^n$ belongs to that flat.
- The affine hull of a set was defined by means of flats containing that set.
- The following theorem expresses the affine hull of a set in terms of points of the set itself.

Theorem

Let A be a set in $\mathbb{R}^n$. Then $\operatorname{aff} A$ is the set of all affine combinations of points of A.

Proof

- Denote by B the set of all affine combinations of points of A.

  That $B \subseteq \operatorname{aff} A$ follows from the preceding theorem and the inclusion $A \subseteq \operatorname{aff} A$.

  We next show that B is a flat. If $\boldsymbol{x}, \boldsymbol{y} \in B$, then $\boldsymbol{x} = \lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_m \boldsymbol{a}_m$, $\boldsymbol{y} = \mu_1 \boldsymbol{b}_1 + \cdots + \mu_p \boldsymbol{b}_p$, for some $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_p \in A$, and scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_p$ with $\lambda_1 + \cdots + \lambda_m = 1$, $\mu_1 + \cdots + \mu_p = 1$. Let $\lambda + \mu = 1$. Then

$$\lambda \boldsymbol{x} + \mu \boldsymbol{y} = \lambda \lambda_1 \boldsymbol{a}_1 + \cdots + \lambda \lambda_m \boldsymbol{a}_m + \mu \mu_1 \boldsymbol{b}_1 + \cdots + \mu \mu_p \boldsymbol{b}_p$$

and

$$\begin{aligned} & \lambda\lambda_1 + \cdots + \lambda\lambda_m + \mu\mu_1 + \cdots + \mu\mu_p \\ & = \lambda(\lambda_1 + \cdots + \lambda_m) + \mu(\mu_1 + \cdots + \mu_p) \\ & = \lambda + \mu = 1. \end{aligned}$$

  Thus $\lambda \boldsymbol{x} + \mu \boldsymbol{y} \in B$. So B is a flat. Since B is a flat and $B \supseteq A$, it follows that $B \supseteq \operatorname{aff} A$. Hence $B = \operatorname{aff} A$.

Example

Corollary

Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^n$. Then

$$\mathrm{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\} = \{\lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_m \boldsymbol{a}_m : \lambda_1 + \cdots + \lambda_m = 1\}.$$

Example: Each point $\boldsymbol{x} = (x_1, \ldots, x_n)$ of $\mathbb{R}^n$ can be expressed as an affine combination of the zero vector $\mathbf{0}$ and the elementary vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ as follows:

$$\boldsymbol{x} = (1 - x_1 - \cdots - x_n)\mathbf{0} + x_1 \boldsymbol{e}_1 + \cdots + x_n \boldsymbol{e}_n.$$

The corollary now shows that $\mathrm{aff}\{\mathbf{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\} = \mathbb{R}^n$.

Linear Hull

- Let A be a non-empty set in $\mathbb{R}^n$.
- We recall that a point of the form $\lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_m \boldsymbol{a}_m$, where $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in A$ and $\lambda_1, \ldots, \lambda_m$ are scalars, is said to be a **linear combination** of points of A.
- The set of all such linear combinations is the smallest subspace of $\mathbb{R}^n$ which contains A, and is called here the **linear hull** of A and we denote it by $\text{lin} A$.
- Since $\text{lin} A$ is a flat containing $A \cup \{\mathbf{0}\}$, it follows that $\text{aff}(A \cup \{\mathbf{0}\}) \subseteq \text{lin} A$.
- On the other hand, $\text{aff}(A \cup \{\mathbf{0}\})$ is a subspace of $\mathbb{R}^n$ containing A, so $\text{lin} A \subseteq \text{aff}(A \cup \{\mathbf{0}\})$.
- We conclude that $\text{lin} A = \text{aff}(A \cup \{\mathbf{0}\})$.
- We define $\text{lin} \varnothing = \{\mathbf{0}\}$.
- This ensures that $\text{lin} \varnothing$ is the smallest subspace of $\mathbb{R}^n$ which contains $\varnothing$, and that $\text{lin} \varnothing = \text{aff}(\varnothing \cup \{\mathbf{0}\})$.

Addition and Scalar Multiplication

- We conclude the section by examining how flats behave with respect to the operations of addition and scalar multiplication.

Theorem

Let A, B be flats in $\mathbb{R}^n$ and let α be a scalar. Then $A+B$ and αA are flats.

- Let $\lambda+\mu=1$. Since A and B are flats, $\lambda A+\mu A \subseteq A$ and $\lambda B+\mu B \subseteq B$. Thus,

$$\lambda(A+B)+\mu(A+B)=(\lambda A+\mu A)+(\lambda B+\mu B) \subseteq A+B;$$
$$\lambda(\alpha A)+\mu(\alpha A)=\alpha(\lambda A+\mu A) \subseteq \alpha A.$$

  This shows that $A+B$ and αA are flats.

Corollary

Let $A_1,\ldots,A_m$ be flats in $\mathbb{R}^n$ and let $\lambda_1,\ldots,\lambda_m$ be scalars. Then $\lambda_1 A_1+\cdots+\lambda_m A_m$ is a flat.

Scalar Distributivity

- We saw in the last section that it is not in general true that $A+A=2A$.
- It is true, however, when A is a flat.

Theorem

Let A be a flat in $\mathbb{R}^n$ and let $\lambda_1,\ldots,\lambda_m$ be scalars with $\lambda_1+\cdots+\lambda_m\neq 0$. Then

$$(\lambda_1+\cdots+\lambda_m)A=\lambda_1 A+\cdots+\lambda_m A.$$

- Write $\lambda=\lambda_1+\cdots+\lambda_m$. Then, using a previous theorem, we deduce that

$$\begin{array}{rcl}(\lambda_1+\cdots+\lambda_m)A & \subseteq & \lambda_1 A+\cdots+\lambda_m A \\ & = & \lambda(\frac{\lambda_1}{\lambda}A+\cdots+\frac{\lambda_m}{\lambda}A) \\ & \subseteq & \lambda A \\ & = & (\lambda_1+\cdots+\lambda_m)A.\end{array}$$

Thus $(\lambda_1+\cdots+\lambda_m)A=\lambda_1 A+\cdots+\lambda_m A$.

Subsection 3

Dimension

Affine Dependence

- The set A in $\mathbb{R}^n$ is said to be **affinely dependent** if there exists $\boldsymbol{a} \in A$ such that $\boldsymbol{a} \in \text{aff}(A \backslash \{\boldsymbol{a}\})$.
- Thus in $\mathbb{R}^3$:
  - A set of three points is affinely dependent if and only if it is collinear;
  - A set of four points is affinely dependent if and only if it is coplanar;
  - Any set having more than four points is affinely dependent.

Affine Independence

- A set in $\mathbb{R}^n$ which is not affinely dependent is said to be **affinely independent**.
- In $\mathbb{R}^3$:
  - A set of three points is affinely independent precisely when it is the vertex set of a non-degenerate triangle;
  - A set of four points is affinely independent precisely when it is the vertex set of a non-degenerate tetrahedron.
- In $\mathbb{R}^n$, the empty set, every singleton, and every set consisting of two points are affinely independent.
- Since any set in $\mathbb{R}^n$ which contains an affinely dependent set is itself affinely dependent, it follows that every subset of an affinely independent set is affinely independent.

Criterion for Affine Dependence

Theorem

Let A be a set in $\mathbb{R}^n$. Then A is affinely dependent if and only if there exist distinct points $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_m$ of A and scalars $\lambda_1,\ldots,\lambda_m$, not all zero, such that

$$\lambda_1\boldsymbol{a}_1+\cdots+\lambda_m\boldsymbol{a}_m=\mathbf{0} \quad \text{and} \quad \lambda_1+\cdots+\lambda_m=0.$$

- Suppose that A is affinely dependent. Then there exists $\boldsymbol{a}_1\in A$ such that $\boldsymbol{a}_1\in \mathrm{aff}(A\backslash\{\boldsymbol{a}_1\})$. By a previous theorem, there exist (distinct) points $\boldsymbol{a}_2,\ldots,\boldsymbol{a}_m$ of $A\backslash\{\boldsymbol{a}_1\}$ and scalars $\mu_2,\ldots,\mu_m$, such that $\boldsymbol{a}_1=\mu_2\boldsymbol{a}_2+\cdots+\mu_m\boldsymbol{a}_m$ and $\mu_2+\cdots+\mu_m=1$. Write $\lambda_1=-1$, $\lambda_2=\mu_2$, ..., $\lambda_m=\mu_m$. Then $\lambda_1,\ldots,\lambda_m$ are not all zero and satisfy the conclusion.

Criterion for Affine Dependence (Cont'd)

- Suppose next that there exist distinct points $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_m$ of A, and scalars $\lambda_1,\ldots,\lambda_m$, not all zero, which satisfy the hypothesis.

  Suppose that $\lambda_1 \neq 0$. Then

$$\boldsymbol{a}_1 = -\frac{1}{\lambda_1}(\lambda_2\boldsymbol{a}_2+\cdots+\lambda_m\boldsymbol{a}_m) \quad \text{and} \quad -\frac{1}{\lambda_1}(\lambda_2+\cdots+\lambda_m)=1,$$

  which shows that $\boldsymbol{a}_1$ is an affine combination of $\boldsymbol{a}_2,\ldots,\boldsymbol{a}_m$. Hence $\boldsymbol{a}_1 \in \operatorname{aff}\{\boldsymbol{a}_2,\ldots,\boldsymbol{a}_m\} \subseteq \operatorname{aff}(A\backslash\{\boldsymbol{a}_1\})$. So A is affinely dependent.

Corollary

A subset $\{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_m\}$ of $\mathbb{R}^n$ is affinely dependent if and only if there exist scalars $\lambda_1,\ldots,\lambda_m$, not all zero, such that

$$\lambda_1\boldsymbol{a}_1+\cdots+\lambda_m\boldsymbol{a}_m=\boldsymbol{0} \quad \text{and} \quad \lambda_1+\cdots+\lambda_m=0.$$

Uniqueness

Corollary

Let $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$ be an affinely independent set in $\mathbb{R}^n$. Then each point of $\mathrm{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$ can be expressed uniquely in the form

$$\lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_m \boldsymbol{a}_m, \quad \text{where} \quad \lambda_1 + \cdots + \lambda_m = 1.$$

- A previous corollary shows that each point of $\mathrm{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$ can be expressed in the desired form.

  To establish the uniqueness, suppose that

$$\begin{gathered} \lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_m \boldsymbol{a}_m = \mu_1 \boldsymbol{a}_1 + \cdots + \mu_m \boldsymbol{a}_m, \\ \lambda_1 + \cdots + \lambda_m = \mu_1 + \cdots + \mu_m = 1. \end{gathered}$$

Uniqueness (Cont'd)

- Then
$$(\lambda_1 - \mu_1)\boldsymbol{a}_1 + \cdots + (\lambda_m - \mu_m)\boldsymbol{a}_m = \mathbf{0}$$
with $(\lambda_1 - \mu_1) + \cdots + (\lambda_m - \mu_m) = 0$.
Since $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}_{\neq}$ is affinely independent, the preceding corollary shows that the scalars $\lambda_1 - \mu_1, \ldots, \lambda_m - \mu_m$ must be zero.
Thus $\lambda_1 = \mu_1$, ..., $\lambda_m = \mu_m$, and the uniqueness is established.

Cardinality of Affinely Independent Sets

- We mentioned that any set of more than four points in $\mathbb{R}^3$ is affinely dependent.

Corollary

An affinely independent set in $\mathbb{R}^n$ cannot contain more than $n+1$ points.

- It suffices to show that every set of the form $\{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_m\}_{\neq}$ in $\mathbb{R}^n$, where $m>n+1$, is affinely dependent. Let $\{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_m\}_{\neq}$ be a set in $\mathbb{R}^n$, where $m>n+1$. Then the system of the $n+1$ linear simultaneous equations

$$\lambda_1\boldsymbol{a}_1+\cdots+\lambda_m\boldsymbol{a}_m=\mathbf{0}, \quad \lambda_1+\cdots+\lambda_m=0,$$

in the m unknowns $\lambda_1,\ldots,\lambda_m$ is homogeneous. Since $m>n+1$, it has a non-trivial solution. Hence, $\{\boldsymbol{a}_1,\ldots,\boldsymbol{a}_m\}_{\neq}$, is affinely dependent by a previous corollary.

Affine Hull and Affine Independence

Corollary

Let A be an affinely independent subset of $\mathbb{R}^n$. Suppose that $\boldsymbol{a}$ is a point of $\mathbb{R}^n$ not lying in $\operatorname{aff} A$. Then the set $A \cup \{\boldsymbol{a}\}$ is affinely independent.

- We argue by contradiction. Suppose that $A \cup \{\boldsymbol{a}\}$ is affinely dependent. Then there exist distinct points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ of A and scalars $\lambda, \lambda_1, \ldots, \lambda_m$, not all zero, such that $\lambda \boldsymbol{a} + \lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_m \boldsymbol{a}_m = \mathbf{0}$ and $\lambda + \lambda_1 + \cdots + \lambda_m = 0$. The scalar λ cannot be zero, for then A is affinely dependent. Thus the equation can be used to express $\boldsymbol{a}$ as an affine combination of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$. So $\boldsymbol{a} \in \operatorname{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$. This, however, contradicts the hypothesis that $\boldsymbol{a} \notin \operatorname{aff} A$. Hence $A \cup \{\boldsymbol{a}\}$ is affinely independent.

Example

- In $\mathbb{R}^n$ the set $\{\mathbf{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ is affinely independent.
- To see this, suppose that the scalars $\lambda, \lambda_1, \ldots, \lambda_n$ satisfy

$$\lambda \mathbf{0} + \lambda_1 \boldsymbol{e}_1 + \cdots + \lambda_n \boldsymbol{e}_n = \mathbf{0} \quad \text{and} \quad \lambda + \lambda_1 + \cdots + \lambda_n = 0.$$

  The first of these equations shows that $\lambda_1, \ldots, \lambda_n$ are all zero.

  Hence λ must also be zero from the second equation.

  The corollary now shows that the set $\{\mathbf{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ is affinely independent.

Independent Generators

- In $\mathbb{R}^3$ as a simple case-by-case consideration shows, each r-dimensional flat ($r = 0,1,2,3$) is the affine hull of some affinely independent set of $r+1$ points.
- For example, a plane is the affine hull of any three of its points which are not collinear.
- Previous examples show that $\mathbb{R}^3$ is the affine hull of the affinely independent set $\{\mathbf{0}, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$.
- This suggests that we might assign a *dimension* r to a flat in $\mathbb{R}^n$ if it is the affine hull of some affinely independent set of $r+1$ points.
- Before we can formalize this idea, however, two results need to be established:
  (i) Every flat in $\mathbb{R}^n$ is the affine hull of some finite affinely independent set;
  (ii) If two affinely independent sets in $\mathbb{R}^n$ have the same affine hull, then they have the same number of elements.

Dimension Theorem

Theorem

Every flat in $\mathbb{R}^n$ is the affine hull of some finite affinely independent subset of $\mathbb{R}^n$. Moreover, the number of elements in such a subset is determined uniquely by the flat itself.

- Consider the non-trivial case of a flat A in $\mathbb{R}^n$ which is neither empty nor a singleton. Let m be the largest positive integer such that A contains an affinely independent subset of $m+1$ elements. Such an m exists by a previous corollary, and $m \geq 1$, since A contains at least two points. Let $\{\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$ be an affinely independent subset of A. Since A is a flat, $\text{aff}\{\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\} \subseteq A$. Now $A \subseteq \text{aff}\{\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$, for otherwise there would exist some point $\boldsymbol{a}$ of A not lying in $\text{aff}\{\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$ and, by a previous corollary, $\{\boldsymbol{a}, \boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$ would be an affinely independent subset of A having $m+2$ elements, so contradicting the definition of m. Hence $A = \text{aff}\{\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$.

Dimension Theorem (Cont'd)

- We now complete the proof by showing that m is the dimension of the unique subspace B of $\mathbb{R}^n$ that is parallel to A.

  This we do by proving that the subset $\{\boldsymbol{a}_1 - \boldsymbol{a}_0, \ldots, \boldsymbol{a}_m - \boldsymbol{a}_0\}$ of B is a basis for B. Let $\boldsymbol{b} \in B$. Then $\boldsymbol{b} = \boldsymbol{x} - \boldsymbol{a}_0$ for some $\boldsymbol{x} \in A$. Thus, there exist scalars $\lambda_0, \lambda_1, \ldots, \lambda_m$ such that

$$\boldsymbol{x} = \lambda_0 \boldsymbol{a}_0 + \lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_m \boldsymbol{a}_m$$

  and $\lambda_0 + \lambda_1 + \cdots + \lambda_m = 1$. Hence,

$$\boldsymbol{b} = \boldsymbol{x} - \boldsymbol{a}_0 = \lambda_1 (\boldsymbol{a}_1 - \boldsymbol{a}_0) + \cdots + \lambda_m (\boldsymbol{a}_m - \boldsymbol{a}_0).$$

  This shows that $\{\boldsymbol{a}_1 - \boldsymbol{a}_0, \ldots, \boldsymbol{a}_m - \boldsymbol{a}_0\}$ spans B.

Dimension Theorem (Cont'd)

- Finally, suppose that $\mu_1, \ldots, \mu_m$ satisfy

$$\mu_1(\boldsymbol{a}_1 - \boldsymbol{a}_0) + \cdots + \mu_m(\boldsymbol{a}_m - \boldsymbol{a}_0) = \mathbf{0}.$$

Then

$$\begin{array}{c} -(\mu_1 + \cdots + \mu_m)\boldsymbol{a}_0 + \mu_1 \boldsymbol{a}_1 + \cdots + \mu_m \boldsymbol{a}_m = \mathbf{0}, \\ -(\mu_1 + \cdots + \mu_m) + \mu_1 + \cdots + \mu_m = 0. \end{array}$$

But $\{\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}_{\neq}$ is affinely independent. So all of $\mu_1, \ldots, \mu_m$ are zero. Thus $\{\boldsymbol{a}_1 - \boldsymbol{a}_0, \ldots, \boldsymbol{a}_m - \boldsymbol{a}_0\}$ is linearly independent. We conclude that $\{\boldsymbol{a}_1 - \boldsymbol{a}_0, \ldots, \boldsymbol{a}_m - \boldsymbol{a}_0\}$ is a basis for B.

Hence, m is the dimension of B, and so is uniquely determined by A.

Dimension of Flats

- A flat in $\mathbb{R}^n$ which is the affine hull of some affinely independent set of $r+1$ points is said to have **dimension** r and is called an r-**flat**.
- It follows from the theorem that each flat in $\mathbb{R}^n$ has a unique dimension r attached to it, and from a previous corollary that $r \leq n$.
- The empty flat is the affine hull of the (affinely independent) empty set, and so has dimension -1.
- Clearly every singleton (point) has dimension 0 and every line has dimension 1.
- We have already seen that $\mathbb{R}^n$ is the affine hull of the affinely independent set $\{\mathbf{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$, whence $\mathbb{R}^n$ has dimension n.

Dimension of Subsets

- The concept of dimension is extended to arbitrary subsets of $\mathbb{R}^n$ by defining the **dimension** $\dim A$ of a set A in $\mathbb{R}^n$ to be the dimension of the flat $\text{aff} A$.
- We note that when a flat in $\mathbb{R}^n$ is also a subspace of $\mathbb{R}^n$ its dimension as defined above coincides with its dimension as a subspace of the real vector space $\mathbb{R}^n$.
- Hence we may apply the term dimension unambiguously both to flats and subspaces of $\mathbb{R}^n$.

Dimension Equation

Theorem

Let A and B be flats in $\mathbb{R}^n$ which have a non-empty intersection. Then

$$\dim(A+B)+\dim(A\cap B)=\dim A+\dim B.$$

- Let $\boldsymbol{c}\in A\cap B$. Then $A-\boldsymbol{c}$ and $B-\boldsymbol{c}$ are subspaces of $\mathbb{R}^n$. So, by the dimension theorem of elementary linear algebra,
$\dim((A-\boldsymbol{c})+(B-\boldsymbol{c}))+\dim((A-\boldsymbol{c})\cap(B-\boldsymbol{c}))=\dim(A-\boldsymbol{c})+\dim(B-\boldsymbol{c})$,
that is, $\dim(A+B-2\boldsymbol{c})+\dim((A\cap B)-\boldsymbol{c})=\dim(A-\boldsymbol{c})+\dim(B-\boldsymbol{c})$.
The proof of the preceding theorem shows that the dimension of a non-empty flat in $\mathbb{R}^n$ coincides with the dimension of the unique subspace of $\mathbb{R}^n$ which is parallel to it. It follows from this last result that the dimension of any translate of a flat is the same as the dimension of the flat itself. Thus, the last equation above simplifies to $\dim(A+B)+\dim(A\cap B)=\dim A+\dim B$.

Affine Bases

- An **affine basis** for a flat in $\mathbb{R}^n$ is any affinely independent set in $\mathbb{R}^n$ whose affine hull is that flat.
- A previous theorem shows that every flat has an affine basis.
- By definition, every affine basis for an r-flat has precisely $r+1$ elements.

  Example: $\{\mathbf{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ is an affine basis for $\mathbb{R}^n$.
- The next result shows that every affinely independent subset of a set in $\mathbb{R}^n$ can be extended to an affine basis for the affine hull of the set.

Extension to an Affine Basis

Theorem

Let B be an affinely independent subset of a set A in $\mathbb{R}^n$. Then there exists an affine basis for $\operatorname{aff} A$ that lies in A and contains B.

- Consider the non-empty family $\mathscr{F}$ of all affinely independent subsets of A which contain B. Since no affinely independent set in $\mathbb{R}^n$ contains more than $n+1$ points, there must exist some member C of $\mathscr{F}$ that is not properly contained in any other member of $\mathscr{F}$. Since C is a subset of A, we have $\operatorname{aff} C \subseteq \operatorname{aff} A$. We claim that $\operatorname{aff} C = \operatorname{aff} A$.

  Suppose that $\operatorname{aff} C \subset \operatorname{aff} A$. Since $\operatorname{aff} A$ is the smallest flat containing A, we cannot have $A \subseteq \operatorname{aff} C$, whence there exists some point $\boldsymbol{a}$ of A not lying in $\operatorname{aff} C$. We can now use a previous corollary to deduce that $C \cup \{\boldsymbol{a}\}$ is a member of $\mathscr{F}$ which properly contains C. This contradicts the choice of C. Thus $\operatorname{aff} C = \operatorname{aff} A$ and C is an affine basis of $\operatorname{aff} A$.

Barycentric Coordinates

Corollary

Let A be a subset of $\mathbb{R}^n$. Then A contains an affine basis for aff A.

- Let $\{\boldsymbol{a}_0, \ldots, \boldsymbol{a}_r\}$ be an affine basis for a non-empty r-flat A in $\mathbb{R}^n$.
- Then, by a previous corollary, each point $\boldsymbol{x}$ of A can be expressed uniquely in the form

$$\boldsymbol{x} = \lambda_0 \boldsymbol{a}_0 + \cdots + \lambda_r \boldsymbol{a}_r, \quad \text{where} \quad \lambda_0 + \cdots + \lambda_r = 1.$$

- The scalars $\lambda_0, \ldots, \lambda_r$ are called the **barycentric coordinates** of $\boldsymbol{x}$ relative to (the ordered affine basis) $\boldsymbol{a}_0, \ldots, \boldsymbol{a}_r$.
- A previous example shows that the barycentric coordinates of a point $\boldsymbol{x} = (x_1, \ldots, x_n)$ of $\mathbb{R}^n$ relative to $\boldsymbol{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ are $1 - x_1 - \cdots - x_n$, $x_1, \ldots, x_n$.

Scalars of Point Relative to Affine Basis

Theorem

Let $\{\boldsymbol{a}_0,\ldots,\boldsymbol{a}_r\}$ be an affine basis for a non-empty r-flat A in $\mathbb{R}^n$. Let $\lambda_0,\ldots,\lambda_r$ be the barycentric coordinates of a point $\boldsymbol{x}=(x_1,\ldots,x_n)$ of A relative to $\boldsymbol{a}_0,\ldots,\boldsymbol{a}_r$. Then there exist scalars a_{ij} $(i=0,\ldots,r,\ j=0,\ldots,n)$ such that, for $i=0,\ldots,r$,

$$\lambda_i = a_{i0}+a_{i1}x_1+\cdots+a_{in}x_n.$$

- Extend, if necessary, $\{\boldsymbol{a}_0,\ldots,\boldsymbol{a}_r\}$ to an affine basis $\{\boldsymbol{a}_0,\ldots,\boldsymbol{a}_n\}$ for $\mathbb{R}^n$. Each point $\boldsymbol{x}=(x_1,\ldots,x_n)$ of $\mathbb{R}^n$ can be written uniquely in the form

$$\boldsymbol{x}=\lambda_0\boldsymbol{a}_0+\cdots+\lambda_n\boldsymbol{a}_n, \quad \text{where} \quad \lambda_0+\cdots+\lambda_n=1.$$

  In particular, each of the points $\mathbf{0},\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$ can be so expressed. Write $\boldsymbol{e}_0=\mathbf{0}$.

Scalars of Point (Cont'd)

- Then there are scalars b_{ij} $(i=0,\ldots,n;\ j=0,\ldots,n)$ such that, for $i=0,\ldots,n$,

$$\boldsymbol{e}_i = b_{0i}\boldsymbol{a}_0 + \cdots + b_{ni}\boldsymbol{a}_n \quad \text{and} \quad b_{0i} + \cdots + b_{ni} = 1.$$

Write $\boldsymbol{x} = (1 - x_1 - \cdots - x_n)\boldsymbol{e}_0 + x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n$.

Then $\boldsymbol{x} = \mu_0\boldsymbol{a}_0 + \cdots + \mu_n\boldsymbol{a}_n$, where, for $i=0,\ldots,n$,

$$\mu_i = b_{i0}(1 - x_1 - \cdots - x_n) + b_{i1}x_1 + \cdots + b_{in}x_n.$$

A routine verification shows that $\mu_0 + \cdots + \mu_n = 1$.

Since the representation of $\boldsymbol{x}$ in this form is unique, we can deduce that $\lambda_i = \mu_i$, for $i=0,\ldots,n$.

We complete the proof by putting $a_{i0} = b_{i0}$ for $i=0,\ldots,n$, and $a_{ij} = b_{ij} - b_{i0}$ $(i=0,\ldots,n,\ j=1,\ldots,n)$, and noting that $\boldsymbol{x} \in A$ if and only if $\lambda_{r+1} = 0,\ldots,\lambda_n = 0$.

Non-Meetings 1-Flats

Theorem

Let L and M be two lines that lie in a 2-flat A of $\mathbb{R}^n$ and which do not meet. Then L and M are parallel.

- Let $\boldsymbol{a}, \boldsymbol{b}$ be distinct points of L, and let $\boldsymbol{c}, \boldsymbol{d}$ be distinct points of M. Since $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\}$ is affinely independent, it will form an affine basis for A. Thus $\boldsymbol{d} = \alpha\boldsymbol{a} + \beta\boldsymbol{b} + \gamma\boldsymbol{c}$ for some $\alpha, \beta, \gamma \in \mathbb{R}$ with $\alpha + \beta + \gamma = 1$. A typical point on M, the line joining $\boldsymbol{c}$ and $\boldsymbol{d}$, has the form

$$(1-\theta)\boldsymbol{c} + \theta\boldsymbol{d} = \theta\alpha\boldsymbol{a} + \theta\beta\boldsymbol{b} + (\theta(\gamma-1)+1)\boldsymbol{c},$$

  for some $\theta \in \mathbb{R}$. Since the latter point does not lie on L for any θ, we must have $\gamma = 1$ and $\boldsymbol{d} = \alpha(\boldsymbol{a}-\boldsymbol{b}) + \boldsymbol{c}$. Hence $\boldsymbol{d} - (\boldsymbol{c}-\boldsymbol{a}) = \alpha(\boldsymbol{a}-\boldsymbol{b}) + \boldsymbol{c} - \boldsymbol{c} + \boldsymbol{a} = (\alpha+1)\boldsymbol{a} - \alpha\boldsymbol{b} \in L$. Thus, $M - (\boldsymbol{c}-\boldsymbol{a}) \subseteq L$. Since $M - (\boldsymbol{c}-\boldsymbol{a})$ is a line, we must have $M - (\boldsymbol{c}-\boldsymbol{a}) = L$. Thus L and M are parallel.

Subsection 4

Hyperplanes

Linear Equations

- Consider the following system of m linear equations in n real variables $x_1, \ldots, x_n$:

$$\left\{\begin{array}{rcl} a_{11}x_1 + \cdots + a_{1n}x_n & = & a_{10} \\ & \vdots & \\ a_{m1}x_1 + \cdots + a_{mn}x_n & = & a_{m0} \end{array}\right.$$

  where a_{ij} are given scalars.
- By the **solution set** of this system is meant the set of all n-tuples $(x_1, \ldots, x_n)$ of $\mathbb{R}^n$ that satisfy it.
- The solution set of the system is clearly the intersection of the solution sets of the m linear equations which comprise it.

Solution Sets and Hyperplanes

- An easy verification shows that the solution set of any one of the individual linear equations is a flat.
- So the solution set of the whole system is a flat.
- Later in the section, we shall show that every flat is the solution set of some system of linear equations.
- In general, the solution set of a single linear equation $a_1x_1 + \cdots + a_nx_n = a_0$ is an $(n-1)$-dimensional flat in $\mathbb{R}^n$.
- In the study of convexity in $\mathbb{R}^n$, flats of dimension $n-1$ play a key role, and are given their own name, **hyperplanes**.

Hyperplanes

- To be precise, we should refer not to a *hyperplane*, but to a *hyperplane in* $\mathbb{R}^n$.
- When no ambiguity is likely to arise, however, we do speak simply of a hyperplane.
- A hyperplane:
  - in $\mathbb{R}^1$ is a point;
  - in $\mathbb{R}^2$ is a line;
  - in $\mathbb{R}^3$ is a plane.
- Thus:
  - A hyperplane in $\mathbb{R}^2$ has an equation of the form $ax + by + c = 0$, where not both of a and b are zero;
  - A hyperplane in $\mathbb{R}^3$ has an equation of the form $ax + by + cz + d = 0$, where not all of a, b and c are zero.

Characterization of Hyperplanes

Theorem

A set H in $\mathbb{R}^n$ is a hyperplane if and only if there exist scalars $c_0, c_1, \ldots, c_n$, where not all $c_1, \ldots, c_n$ are zero, such that

$$H = \{(x_1, \ldots, x_n) : c_0 + c_1 x_1 + \cdots + c_n x_n = 0\}.$$

- Let $H = \{(x_1, \ldots, x_n) : c_0 + c_1 x_1 + \cdots + c_n x_n = 0\}$, where $c_0, c_1, \ldots, c_n$ are scalars and not all $c_1, \ldots, c_n$ are zero, say $c_1 \neq 0$. Let $\boldsymbol{u} = (u_1, \ldots, u_n)$, $\boldsymbol{v} = (v_1, \ldots, v_n)$ lie in H and let $\lambda + \mu = 1$. Then

$$\begin{aligned} & c_0 + c_1(\lambda u_1 + \mu v_1) + \cdots + c_n(\lambda u_n + \mu v_n) \\ &= \lambda(c_0 + c_1 u_1 + \cdots + c_n u_n) + \mu(c_0 + c_1 v_1 + \cdots + c_n v_n) \\ &= \lambda 0 + \mu 0 = 0. \end{aligned}$$

  Thus $\lambda \boldsymbol{u} + \mu \boldsymbol{v} \in H$ and H is a flat.

Characterization of Hyperplanes (Cont'd)

- Define points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ of H by the equations $\boldsymbol{a}_1 = (-\frac{c_0}{c_1}, 0, 0, \ldots, 0)$ and $\boldsymbol{a}_2 = (-\frac{c_0+c_2}{c_1}, 1, 0, \ldots, 0), \ldots, \boldsymbol{a}_n = (-\frac{c_0+c_n}{c_1}, 0, 0, \ldots, 1)$. Since H is a flat, $\text{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} \subseteq H$.

  We now establish the opposite inclusion. Let $\boldsymbol{x} \in H$. Then the equations

$$\boldsymbol{x} = (x_1, \ldots, x_n) = (1 - x_2 - \cdots - x_n)\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n$$

  express $\boldsymbol{x}$ as an affine combination of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. So $\boldsymbol{x} \in \text{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ Hence, $H \subseteq \text{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ and, therefore, $H = \text{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$.

Characterization of Hyperplanes (Cont'd)

- To show that the set $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ is affinely independent, suppose that $\lambda_1, \ldots, \lambda_n$ satisfy

$$\lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_n \boldsymbol{a}_n = \boldsymbol{0} \quad \text{and} \quad \lambda_1 + \cdots + \lambda_n = 0.$$

  Comparing the ith coordinates $(i = 2, \ldots, n)$ on both sides of the first of these equations, we find that $\lambda_2, \ldots, \lambda_n$ are all zero. Thus, so too is λ_1, from the second equation.

  So, $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ is affinely independent.

  But $H = \text{aff}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$, and so H is an $(n-1)$-dimensional flat, i.e., H is a hyperplane.

Characterization of Hyperplanes (Converse)

- Conversely, suppose that H is a hyperplane in $\mathbb{R}^n$. Let $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ be an affine basis for H. Extend this to an affine basis $\{\boldsymbol{b}_0, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ for $\mathbb{R}^n$. Then each $\boldsymbol{x} = (x_1, \ldots, x_n)$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\boldsymbol{x} = \lambda_0 \boldsymbol{b}_0 + \lambda_1 \boldsymbol{b}_1 + \cdots + \lambda_n \boldsymbol{b}_n, \quad \text{where} \quad \lambda_0 + \lambda_1 + \cdots + \lambda_n = 1.$$

Thus $\lambda_0, \lambda_1, \ldots, \lambda_n$ are the barycentric coordinates of $\boldsymbol{x}$ relative to the (ordered) affine basis $\boldsymbol{b}_0, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$. By a previous theorem, there exist scalars $c_0, c_1, \ldots, c_n$ such that

$$\lambda_0 = c_0 + c_1 x_1 + \cdots + c_n x_n.$$

Since $\boldsymbol{x} \in H$ iff $\lambda_0 = 0$, $H = \{(x_1, \ldots, x_n) : c_0 + c_1 x_1 + \cdots + c_n x_n = 0\}$. Not all of $c_1, \ldots, c_n$ are zero, for this would imply that either H is empty (if $c_0 \neq 0$) or $\mathbb{R}^n$ (if $c_0 = 0$), both of which contradict the assumption that H is an $(n-1)$-dimensional flat.

Characterization of r-Flats

Corollary

In $\mathbb{R}^n$ each r-flat ($r = -1, \ldots, n$) can be expressed as the intersection of $n-r$ hyperplanes, and so is the solution set of some system of $n-r$ linear equations.

- The only (-1)-flat in $\mathbb{R}^n$ is the empty set, which is the intersection of the $n+1$ hyperplanes $x_1 = 0$, ..., $x_n = 0$, $x_1 + \cdots + x_n = 1$.
  The only n-flat in $\mathbb{R}^n$ is $\mathbb{R}^n$ itself, which is the intersection of no hyperplanes.
  Consider now the case of an r-flat A in $\mathbb{R}^n$, where $r = 0, \ldots, n-1$. Let $\{\boldsymbol{a}_0, \ldots, \boldsymbol{a}_r\}$ be an affine basis for A. Extend this to an affine basis $\{\boldsymbol{a}_0, \ldots, \boldsymbol{a}_n\}$ for $\mathbb{R}^n$. Then each $\boldsymbol{x}$ in $\mathbb{R}^n$ can be expressed uniquely in the form

$$\boldsymbol{x} = \lambda_0 \boldsymbol{a}_0 + \cdots + \lambda_n \boldsymbol{a}_n, \quad \text{where} \quad \lambda_0 + \cdots + \lambda_n = 1.$$

Characterization of r-Flats (Cont'd)

- Now A is the set in $\mathbb{R}^n$ consisting precisely of those $\boldsymbol{x}$'s whose barycentric coordinates $\lambda_{r+1}, \ldots, \lambda_n$ are all zero.

  But each of the sets $\{\boldsymbol{x} : \lambda_i = 0\}$ is the hyperplane

  $$\text{aff}\{\boldsymbol{a}_0, \ldots, \boldsymbol{a}_{i-1}, \boldsymbol{a}_{i+1}, \ldots, \boldsymbol{a}_n\}.$$

  It now follows that A is the intersection of the $n-r$ hyperplanes with equations $\lambda_{r+1} = 0, \ldots, \lambda_n = 0$.

Uniqueness of Constants

- Given a hyperplane H in $\mathbb{R}^n$, there exist scalars $c_0, c_1, \ldots, c_n$, with not all $c_1, \ldots, c_n$ zero, such that

$$H = \{(x_1, \ldots, x_n) : c_0 + c_1x_1 + \cdots + c_nx_n = 0\}.$$

- We now consider to what extent H determines the scalars $c_0, c_1, \ldots, c_n$.
- It certainly does not determine them uniquely, for the scalars $\theta c_0, \theta c_1, \ldots, \theta c_n$, where $\theta \neq 0$, serve equally well in the equation for H.
- Suppose that $d_0, d_1, \ldots, d_n$ are also scalars such that

$$H = \{(x_1, \ldots, x_n) : d_0 + d_1x_1 + \cdots + d_nx_n = 0\}.$$

- Assume that $c_1 \neq 0$, and let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ be the points of H specified as in the first part of the proof of the preceding theorem.

Uniqueness of Constants (Cont'd)

- Substituting the coordinates of the $\boldsymbol{a}_i$ into the above equation for H in terms of the d's, we deduce that $d_i = \frac{d_1}{c_1} c_i$ for $i = 0, \ldots, n$.
- Since not all of $d_0, d_1, \ldots, d_n$ can be zero, we deduce that d_1, and hence $\frac{d_1}{c_1}$, is not zero.
- Writing $\theta = \frac{d_1}{c_1}$, we find that $d_0 = \theta c_0, d_1 = \theta c_1$, ..., $d_n = \theta c_n$.
- Thus the hyperplane H determines the scalars $c_0, c_1, \ldots, c_n$ to within a common non-zero scalar multiple.

Halfspaces

- The importance of hyperplanes in $\mathbb{R}^n$ is that they divide the whole space into two halfspaces in a natural way.
  Example: A line in $\mathbb{R}^2$ with equation $ax + by + c = 0$ divides $\mathbb{R}^2$ into the two halfplanes determined by the inequalities $ax + by + c \leq 0$ and $ax + by + c \geq 0$.
- A hyperplane in $\mathbb{R}^n$ with equation $c_0 + c_1x_1 + \cdots + c_nx_n = 0$ divides $\mathbb{R}^n$ into the two halfspaces determined by the inequalities

$$c_0 + c_1x_1 + \cdots + c_nx_n \leq 0 \quad \text{and} \quad c_0 + c_1x_1 + \cdots + c_nx_n \geq 0.$$

- Let $c_0, c_1, \ldots, c_n$ be scalars, where not all $c_1, \ldots, c_n$ are zero.
  Then a set of either of the forms

$$\{(x_1, \ldots, x_n) : c_0 + c_1x_1 + \cdots + c_nx_n \leq 0\} \quad \text{or}$$
$$\{(x_1, \ldots, x_n) : c_0 + c_1x_1 + \cdots + c_nx_n \geq 0\}$$

  is called a **closed halfspace** in $\mathbb{R}^n$.

Halfspaces Determined by a Hyperplane

- A set of either of the forms
$$\{(x_1,\ldots,x_n) : c_0 + c_1x_1 + \cdots + c_nx_n < 0\} \quad \text{or}$$
$$\{(x_1,\ldots,x_n) : c_0 + c_1x_1 + \cdots + c_nx_n > 0\}$$
is called a **open halfspace** in $\mathbb{R}^n$.
- If the scalars $c_0, c_1, \ldots, c_n$ are replaced, respectively, by $\theta c_0, \theta c_1, \ldots, \theta c_n$, for some $\theta \neq 0$, then we obtain the same pair of closed halfspaces and the same pair of open halfspaces, although the order of the halfspaces is reversed when $\theta < 0$.
- Thus, if H is a hyperplane in $\mathbb{R}^n$ with equation $c_0 + c_1x_1 + \cdots + c_nx_n = 0$, then the above pair of closed halfspaces and the above pair of open halfspaces are determined by H (independent of equation).
- Hence we may refer unambiguously to the closed halfspaces and the open halfspaces **determined by** H.
- We say that the closed (open) halfspaces determined by H are **opposite** to one another.

Example

- Any line through two points lying in opposite halfspaces determined by a hyperplane in $\mathbb{R}^n$ meets the hyperplane.
- Suppose that the hyperplane H has equation $c_0 + c_1x_1 + \cdots + c_nx_n = 0$, and that the points $\boldsymbol{a} = (a_1, \ldots, a_n)$ and $\boldsymbol{b} = (b_1, \ldots, b_n)$ lie in opposite halfspaces determined by H.
- Omitting the trivial case when either of $\boldsymbol{a}$ or $\boldsymbol{b}$ lies on H,

$$c_0 + c_1a_1 + \cdots + c_na_n = \alpha < 0 \quad \text{and} \quad c_0 + c_1b_1 + \cdots + c_nb_n = \beta > 0.$$

- The points on the line L through $\boldsymbol{a}$ and $\boldsymbol{b}$ are precisely those points of the form $(\lambda a_1 + (1-\lambda)b_1, \ldots, \lambda a_n + (1-\lambda)b_n)$, where the scalar λ assumes all real values.
- We find, by substituting these coordinates into the equation of H, that $\lambda = \frac{\beta}{\beta - \alpha}$ corresponds to the unique point of intersection of L and H.
- This value of λ satisfies $0 \leq \lambda \leq 1$. So the portion of L lying, between $\boldsymbol{a}$ and $\boldsymbol{b}$, the so-called **line segment** joining $\boldsymbol{a}$ and $\boldsymbol{b}$, meets H.

Characterization of Parallel Hyperplanes

Theorem

Let H and H' be hyperplanes in $\mathbb{R}^n$ with respective equations $c_0 + c_1 x_1 + \cdots + c_n x_n = 0$ and $c'_0 + c'_1 x_1 + \cdots + c'_n x_n = 0$. Then H and H' are parallel if and only if there exists a scalar θ such that $c'_1 = \theta c_1$, ..., $c'_n = \theta c_n$.

- Suppose first that H and H' are parallel, say $H' = H + \boldsymbol{a}$, where $\boldsymbol{a} = (a_1, \ldots, a_n)$. Then $(x_1, \ldots, x_n) \in H$ if and only if

$$\begin{aligned} & c'_0 + c'_1(x_1 + a_1) + \cdots + c'_n(x_n + a_n) \\ = & c'_0 + c'_1 a_1 + \cdots + c'_n a_n + c'_1 x_1 + \cdots + c'_n x_n = 0. \end{aligned}$$

  Thus, by the above remarks on the representation of hyperplanes by linear equations, there exists a θ, such that $c'_1 = \theta c_1, \ldots, c'_n = \theta c_n$.

Characterization of Parallel Hyperplanes

- Suppose next that $c_1' = \theta c_1, \ldots, c_n' = \theta c_n$, where θ is a (non-zero) scalar. Then, for $d_0 = \frac{c_0'}{\theta}$, H' is represented by the equation
$$d_0 + c_1 x_1 + \cdots + c_n x_n = 0.$$
Let $\boldsymbol{b} = (b_1, \ldots, b_n)$ satisfy
$$c_1 b_1 + \cdots + c_n b_n = c_0 - d_0.$$
Then H' also has the equation
$$c_0 + c_1(x_1 - b_1) + \cdots + c_n(x_n - b_n) = 0.$$
Thus $\boldsymbol{x} = (x_1, \ldots, x_n) \in H'$ if and only if $\boldsymbol{x} - \boldsymbol{b} \in H$.
Hence $H' = H + \boldsymbol{b}$. This shows that H and H' are parallel.

Relative Position of Hyperplanes

Corollary

Two parallel hyperplanes in $\mathbb{R}^n$ are either identical or disjoint. Two non-parallel hyperplanes in $\mathbb{R}^n$ must meet.

- Let H and H' be parallel hyperplanes in $\mathbb{R}^n$. Then they have respective equations

$$c_0 + c_1 x_1 + \cdots + c_n x_n = 0 \quad \text{and} \quad c_0' + \theta c_1 x_1 + \cdots + \theta c_n x_n = 0,$$

say, where θ is a non-zero scalar. If $c_0' = \theta c_0$, then H and H' are identical. Otherwise they are disjoint.

Relative Position of Hyperplanes (Cont'd)

- Let H and H' be non-parallel hyperplanes in $\mathbb{R}^n$ having respective equations

$$c_0 + c_1 x_1 + \cdots + c_n x_n = 0 \quad \text{and} \quad c_0' + c_1' x_1 + \cdots + c_n' x_n = 0.$$

Then there is no scalar θ such that $c_1' = \theta c_1, \ldots, c_n' = \theta c_n$. It follows that $n \geq 2$. Suppose that $c_1 \neq 0$. Then, for some $j \in \{2, \ldots, n\}$, $c_j' \neq \frac{c_1'}{c_1} c_j$, say $c_2' \neq \frac{c_1'}{c_1} c_2$. It is easily verified that the point

$$\left(\frac{c_0' c_2 - c_0 c_2'}{c_1 c_2' - c_1' c_2}, \frac{c_0 c_1' - c_0' c_1}{c_1 c_2' - c_1' c_2}, 0, \ldots, 0 \right)$$

lies in $H \cap H'$. So H and H' meet.

Subsection 5

Affine Transformations

Affine Transformations

- A mapping $T:\mathbb{R}^n \to \mathbb{R}^m$ is called an **affine transformation** if $T(\lambda\boldsymbol{x}+\mu\boldsymbol{y}) = \lambda T(\boldsymbol{x}) + \mu T(\boldsymbol{y})$ whenever $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$ and $\lambda+\mu=1$.
- A simple example of an affine transformation is the mapping $T:\mathbb{R}^3 \to \mathbb{R}^3$ defined by the equation $T(x,y,z) = (x,y,1)$.

  Geometrically, T is the orthogonal projection of $\mathbb{R}^3$ onto the plane with equation $z=1$.
- For each vector $\boldsymbol{q} \in \mathbb{R}^n$, the mapping $T:\mathbb{R}^n \to \mathbb{R}^n$ defined by the equation $T(x) = \boldsymbol{x}+\boldsymbol{q}$ is an affine transformation called the **translation** of $\mathbb{R}^n$ through $\boldsymbol{q}$.

Affine versus Linear Transformations

- Clearly every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is also an affine one.
- That not every affine transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is linear, follows from the observation that it need not map the zero vector of $\mathbb{R}^n$ to the zero vector of $\mathbb{R}^m$.
- See the two examples of affine transformations given above.
- The exact relationship between linear and affine transformations is given in the following result.

Relation Between Affine and Linear Transformations

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be an affine transformation. Then T is linear if and only if $T(\mathbf{0}) = \mathbf{0}$.

- In view of the remarks above, it will suffice to show that T is linear when $T(\mathbf{0}) = \mathbf{0}$.

  Suppose, then, that $T(\mathbf{0}) = \mathbf{0}$. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. Then

$$T(\lambda \mathbf{x}) = T(\lambda \mathbf{x} + (1-\lambda)\mathbf{0}) = \lambda T(\mathbf{x}) + (1-\lambda) T(\mathbf{0}) = \lambda T(\mathbf{x}).$$

  Using this last result, we deduce that

$$\begin{aligned} T(\mathbf{x}+\mathbf{y}) &= T\left(2\left(\tfrac{1}{2}\mathbf{x} + \tfrac{1}{2}\mathbf{y}\right)\right) = 2T\left(\tfrac{1}{2}\mathbf{x} + \tfrac{1}{2}\mathbf{y}\right) \\ &= 2\left(\tfrac{1}{2}T(\mathbf{x}) + \tfrac{1}{2}T(\mathbf{y})\right) = T(\mathbf{x}) + T(\mathbf{y}). \end{aligned}$$

  Thus T is linear.

Matrix Form of an Affine Transformation

- In the following discussion, all vectors considered will be identified with column vectors in the natural way.

Theorem

The affine transformations $T:\mathbb{R}^n \to \mathbb{R}^m$ are precisely those mappings $T:\mathbb{R}^n \to \mathbb{R}^m$ which can be expressed in the form $T(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{q}$, for some real $m \times n$ matrix $\boldsymbol{Q}$ and some real $m \times 1$ matrix $\boldsymbol{q}$.

- It is easily verified that a mapping of the type under consideration is an affine transformation.

  Assume, then, that $T:\mathbb{R}^n \to \mathbb{R}^m$ is an affine transformation. Let $T(\boldsymbol{0}) = \boldsymbol{q}$. Then the mapping $T':\mathbb{R}^n \to \mathbb{R}^m$ defined by the equation $T'(\boldsymbol{x}) = T(\boldsymbol{x}) - \boldsymbol{q}$ is readily shown to be an affine transformation with $T'(\boldsymbol{0}) = \boldsymbol{0}$. The theorem shows that T' is linear, whence there is a real $m \times n$ matrix $\boldsymbol{Q}$ such that $T'(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x}$. Thus $T(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{q}$.

Remarks

- The affine transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ determines the matrices $\boldsymbol{Q}$ and $\boldsymbol{q}$ uniquely:
  - The jth column of $\boldsymbol{Q}$ must be $T(\boldsymbol{e}_j) - T(\mathbf{0})$;
  - $\boldsymbol{q}$ must be $T(\mathbf{0})$.
- The above representation of an affine transformation in terms of matrices shows easily that, if $T : \mathbb{R}^n \to \mathbb{R}^m$ is an affine transformation, $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r \in \mathbb{R}^n$ and $\lambda_1 + \cdots + \lambda_r = 1$, then

$$T(\lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_r \boldsymbol{a}_r) = \lambda_1 T(\boldsymbol{a}_1) + \cdots + \lambda_r T(\boldsymbol{a}_r).$$

Affine Transformations and Flats

Corollary

Let $T:\mathbb{R}^n \to \mathbb{R}^m$ be an affine transformation and let A be a set in $\mathbb{R}^n$. Then $T(\text{aff}\, A) = \text{aff}\, T(A)$. If A is a flat, then so too is $T(A)$.

- A point $\boldsymbol{x}$ lies in $T(\text{aff}\, A)$ if and only if there exist $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r \in A$ and $\lambda_1, \ldots, \lambda_r$ with $\lambda_1 + \cdots + \lambda_r = 1$ such that

$$\boldsymbol{x} = T(\lambda_1 \boldsymbol{a}_1 + \cdots + \lambda_r \boldsymbol{a}_r) = \lambda_1 T(\boldsymbol{a}_1) + \cdots + \lambda_r T(\boldsymbol{a}_r),$$

  that is, if and only if $\boldsymbol{x} \in \text{aff}\, T(A)$. Thus $T(\text{aff}\, A) = \text{aff}\, T(A)$.

  If A is a flat, then $\text{aff}\, T(A) = T(\text{aff}\, A) = T(A)$. This shows that $T(A)$ is a flat.

Non-Singular Affine Transformations

- Consider an affine transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ of $\mathbb{R}^n$ into itself.
- By the theorem, there exist a real $n \times n$ matrix $\boldsymbol{Q}$ and a real $n \times 1$ matrix $\boldsymbol{q}$ such that $T(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{q}$.
- The affine transformation T is said to be **non-singular** if the determinant $\det \boldsymbol{Q}$ of the matrix $\boldsymbol{Q}$ is non-zero, that is if $\boldsymbol{Q}$ has an inverse, i.e., is non-singular.

Invertible Affine Transformations

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an affine transformation. Then T has an inverse if and only if T is non-singular. When T is non-singular, its inverse $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is an affine transformation.

- Let $\boldsymbol{Q}$ be a real $n \times n$ matrix and $\boldsymbol{q}$ a real $n \times 1$ matrix such that $T(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{q}$ for all $\boldsymbol{x}$ in $\mathbb{R}^n$. Suppose first that $\boldsymbol{Q}$ is non-singular. Then $\det \boldsymbol{Q}$ is non-zero and $\boldsymbol{Q}$ has an inverse $\boldsymbol{Q}^{-1}$. For each $\boldsymbol{y}$ in $\mathbb{R}^n$, the equation $T(\boldsymbol{x}) = \boldsymbol{y}$ has the unique solution $\boldsymbol{x} = \boldsymbol{Q}^{-1}\boldsymbol{y} - \boldsymbol{Q}^{-1}\boldsymbol{q}$.

  It follows that T has an inverse, which is the affine transformation $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ defined by the equation $T^{-1}(\boldsymbol{y}) = \boldsymbol{Q}^{-1}\boldsymbol{y} - \boldsymbol{Q}^{-1}\boldsymbol{q}$ for $\boldsymbol{y}$ in $\mathbb{R}^n$.

  Suppose next that $\det \boldsymbol{Q}$ is zero. Then there exists a non-zero vector $\boldsymbol{z}$ in $\mathbb{R}^n$ such that $\boldsymbol{Q}(\boldsymbol{z}) = \boldsymbol{0}$. Hence $T(\boldsymbol{z}) = T(\boldsymbol{0})$ and T is not injective. Hence T has no inverse.

Affine Transformations and Affinely Independent Sets

Theorem

Let $\{\boldsymbol{a}_0,\ldots,\boldsymbol{a}_r\}_{\neq}$ and $\{\boldsymbol{b}_0,\ldots,\boldsymbol{b}_r\}_{\neq}$ be affinely independent sets in $\mathbb{R}^n$. Then there exists a non-singular affine transformation $T:\mathbb{R}^n\to\mathbb{R}^n$ such that $T(\boldsymbol{a}_i)=\boldsymbol{b}_i$, for $i=0,\ldots,r$.

- Extend the sets $\{\boldsymbol{a}_0,\ldots,\boldsymbol{a}_r\}_{\neq}$ and $\{\boldsymbol{b}_0,\ldots,\boldsymbol{b}_r\}_{\neq}$, respectively, to affine bases $\{\boldsymbol{a}_0,\ldots,\boldsymbol{a}_n\}$ and $\{\boldsymbol{b}_0,\ldots,\boldsymbol{b}_n\}$ for $\mathbb{R}^n$. Then each $\boldsymbol{x}$ in $\mathbb{R}^n$ can be written uniquely in the form $\boldsymbol{x}=\lambda_0\boldsymbol{a}_0+\cdots+\lambda_n\boldsymbol{a}_n$, $\lambda_0+\cdots+\lambda_n=1$.

  Define a mapping $T:\mathbb{R}^n\to\mathbb{R}^n$ by the equation

$$T(\boldsymbol{x})=\lambda_0\boldsymbol{b}_0+\cdots+\lambda_n\boldsymbol{b}_n.$$

  It is routine to verify that T is a bijective affine transformation. Hence T is a non-singular affine transformation such that $T(\boldsymbol{a}_i)=\boldsymbol{b}_i$, for $i=0,\ldots,r$.

Affine Transformations and Flats

Corollary

Let A and B be flats in $\mathbb{R}^n$ of the same dimension. Then there is a non-singular affine transformation $T: \mathbb{R}^n \to \mathbb{R}^n$ such that $T(A) = B$.

- If A and B are both empty, then T can be taken as the identity mapping of $\mathbb{R}^n$ onto itself.

  Suppose, then, that A and B are non-empty and have affine bases $\{\boldsymbol{a}_0, \ldots, \boldsymbol{a}_r\}_{\neq}$ and $\{\boldsymbol{b}_0, \ldots, \boldsymbol{b}_r\}_{\neq}$, respectively.

  Let T be as in the theorem. Then, by a previous corollary,

$$T(A) = T(\text{aff}\{\boldsymbol{a}_0, \ldots, \boldsymbol{a}_r\}) = \text{aff}\{\boldsymbol{b}_0, \ldots, \boldsymbol{b}_r\} = B.$$

r-Dimensional Flats in $\mathbb{R}^n$

- Suppose that B is an r-dimensional flat ($r \geq 1$) in $\mathbb{R}^n$ and that $A = \text{aff}\{\mathbf{0}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_r\}$.
- Then A and B are flats of the same dimension.
- By the corollary, there exists a non-singular affine transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $T(A) = B$.
- The flat A consists precisely of those points $(x_1, \ldots, x_n)$ for which $x_{r+1} = 0$, ..., $x_n = 0$.
- Hence A can be identified with $\mathbb{R}^r$ by associating the point $(x_1, \ldots, x_n)$ of A with the point $(x_1, \ldots, x_r)$ of $\mathbb{R}^r$.
- Under this identification $T(\mathbb{R}^r) = B$.
- Thus every r-dimensional flat ($r \geq 1$) can be considered to be an affine copy of $\mathbb{R}^r$.
- This identification is often helpful when working with r-dimensional sets in $\mathbb{R}^n$, for we may consider them as subsets of $\mathbb{R}^r$ and make use of the resulting algebraic simplification.

Subsection 6

Length, Distance and Angle

The Inner Product

- The **inner product** $\boldsymbol{x} \cdot \boldsymbol{y}$ of vectors $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$ is the real number defined by the equation

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1 y_1 + \cdots + x_n y_n.$$

- The following properties of the inner product are immediate consequences of its definition.
- For $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$:
  (i) $\boldsymbol{x} \cdot \boldsymbol{x} \geq 0$, and $\boldsymbol{x} \cdot \boldsymbol{x} = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$;
  (ii) $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x}$;
  (iii) $(\lambda \boldsymbol{x} + \mu \boldsymbol{y}) \cdot \boldsymbol{z} = \lambda(\boldsymbol{x} \cdot \boldsymbol{z}) + \mu(\boldsymbol{y} \cdot \boldsymbol{z})$.

The Norm and the Distance

- The **norm** or **length** $\|\boldsymbol{x}\|$ of a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ in $\mathbb{R}^n$ is the non-negative real number defined by the equation

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}, \quad \text{whence} \quad \|\boldsymbol{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

- The **distance** between points $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$ of $\mathbb{R}^n$ is the non-negative real number

$$\|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2},$$

i.e., the length of the vector $\boldsymbol{x} - \boldsymbol{y}$, or $\boldsymbol{y} - \boldsymbol{x}$.

Properties of the Norm

- The norm of the zero vector is 0.
- The norm of each elementary vector $\boldsymbol{e}_i$ is 1.
- In general, any vector in $\mathbb{R}^n$ which has norm 1 is called a **unit vector**.
- The following properties of the norm are simple consequences of its definition.
- For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$:
  - (i) $\|\boldsymbol{x}\| \geq 0$, and $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$;
  - (ii) $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$;
  - (iii) $\|\lambda \boldsymbol{x} + \mu \boldsymbol{y}\|^2 = \lambda^2 \|\boldsymbol{x}\|^2 + 2\lambda\mu \boldsymbol{x} \cdot \boldsymbol{y} + \mu^2 \|\boldsymbol{y}\|^2$.

Inequalities Involving the Norm

Theorem

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Then:

(i) $|\boldsymbol{x} \cdot \boldsymbol{y}| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$ (**Cauchy-Schwarz Inequality**);

(ii) $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ (**Triangle Inequality**);

(iii) $|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|| \leq \|\boldsymbol{x} - \boldsymbol{y}\|$;

(iv) if, for some $\alpha > 0$, $\|\boldsymbol{x} + \lambda \boldsymbol{y}\| \geq \|\boldsymbol{x}\|$ whenever $0 < \lambda < \alpha$, then $\boldsymbol{x} \cdot \boldsymbol{y} \geq 0$.

- We only prove (iv), since (i), (ii), and (iii) are standard results. Let $\alpha > 0$ be such that $\|\boldsymbol{x} + \lambda \boldsymbol{y}\| \geq \|\boldsymbol{x}\|$ whenever $0 < \lambda < \alpha$. Then, whenever $0 < \lambda < \alpha$,

$$\|\boldsymbol{x}\|^2 \leq \|\boldsymbol{x} + \lambda \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + 2\lambda \boldsymbol{x} \cdot \boldsymbol{y} + \lambda^2 \|\boldsymbol{y}\|^2.$$

Hence $\boldsymbol{x} \cdot \boldsymbol{y} + \frac{1}{2}\lambda \|\boldsymbol{y}\|^2 \geq 0$. Letting $\lambda \to 0_+$ in the last inequality, we deduce that $\boldsymbol{x} \cdot \boldsymbol{y} \geq 0$.

Angle Between Vectors

- The Cauchy-Schwarz inequality allows us to introduce the concept of angle into $\mathbb{R}^n$.
- The angle between non-zero vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of $\mathbb{R}^n$ is the unique real number θ satisfying the conditions

$$\cos\theta = \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|} \quad \text{and} \quad 0 \le \theta \le \pi.$$

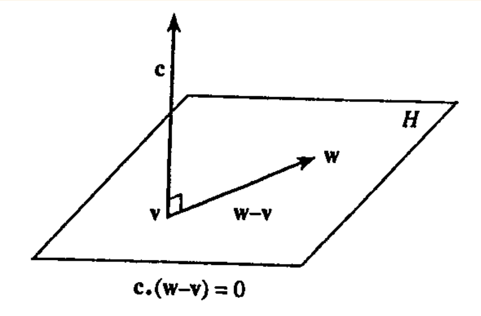
- This definition accords with the usual one of elementary geometry.
- The angle between $\boldsymbol{x}$ and $\boldsymbol{y}$ is called **acute** or **obtuse** according as $\boldsymbol{x} \cdot \boldsymbol{y}$ is positive or negative.
- Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, whether zero or not, are said to be **orthogonal** if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$.

Normal Vectors to a Hyperplane

- Consider a hyperplane H in $\mathbb{R}^n$ with equation $c_0 + c_1x_1 + \cdots + c_nx_n = 0$.
- This equation can be written in the form $c_0 + \boldsymbol{c} \cdot \boldsymbol{x} = 0$, where $\boldsymbol{c}$ is the non-zero vector $(c_1, \ldots, c_n)$ and $\boldsymbol{x}$ is $(x_1, \ldots, x_n)$.
- Such a vector $\boldsymbol{c}$ is said to be a **normal vector** to H.
- By the discussion on the representation of hyperplanes by means of linear equations, it follows that the normal vectors of H are precisely those vectors of the form $\lambda\boldsymbol{c}$ for some non-zero scalar λ.
- Thus H has exactly two unit normal vectors, namely $\pm\frac{\boldsymbol{c}}{\|\boldsymbol{c}\|}$.
- Hence, given any hyperplane H in $\mathbb{R}^n$, it may be assumed that it has an equation of the form $c_0 + \boldsymbol{c} \cdot \boldsymbol{x} = 0$, where $\boldsymbol{c}$ is a unit vector.

Normal Vectors to a Hyperplane (Cont'd)

- This concept of a normal vector generalizes the one familiar in elementary geometry.
- Suppose that $\boldsymbol{v}$ and $\boldsymbol{w}$ lie in a hyperplane H in $\mathbb{R}^n$ with equation $c_0 + \boldsymbol{c} \cdot \boldsymbol{x} = 0$. Then $c_0 + \boldsymbol{c} \cdot \boldsymbol{v} = 0$ and $c_0 + \boldsymbol{c} \cdot \boldsymbol{w} = 0$. So $\boldsymbol{c} \cdot (\boldsymbol{w} - \boldsymbol{v}) = 0$.
- This shows that $\boldsymbol{c}$ is orthogonal to every vector which is the difference of two vectors in H.



Orthogonal Complement

- Let A be a subspace of $\mathbb{R}^n$.
- Then the **orthogonal complement** $A^\perp$ of A is the set of all those vectors in $\mathbb{R}^n$ which are orthogonal to all the vectors in A, i.e.,

$$A^\perp = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \cdot \boldsymbol{a} = 0, \text{ for all } \boldsymbol{a} \in A\}.$$

- It follows easily from this definition that $A^\perp$ is a subspace of $\mathbb{R}^n$ which intersects A in the set $\{\mathbf{0}\}$.
- A standard result of linear algebra asserts that each vector of $\mathbb{R}^n$ can be expressed uniquely in the form $\boldsymbol{a} + \boldsymbol{b}$, where $\boldsymbol{a} \in A$ and $\boldsymbol{b} \in A^\perp$.
- Thus $A + A^\perp = \mathbb{R}^n$.

Orthonormal Sequences

- A sequence $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m$ of vectors in $\mathbb{R}^n$ is said to be an **orthonormal sequence** if $\boldsymbol{u}_i \cdot \boldsymbol{u}_j$ is 1 or 0 according as $i = j$ or $i \neq j$.
- The simplest example of such a sequence is the sequence $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$ of elementary vectors in $\mathbb{R}^n$.
- In an orthonormal sequence, each term is a unit vector, each two terms are orthogonal, and no two terms are the same.
- The terms of an orthonormal sequence $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m$ in $\mathbb{R}^n$ form a linearly independent set $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m\}$.

  To see this, suppose that scalars $\lambda_1,\ldots,\lambda_m$ are such that $\lambda_1\boldsymbol{u}_1 + \cdots + \lambda_m\boldsymbol{u}_m = \boldsymbol{0}$. Then, for $i = 1,\ldots,m$,

$$\lambda_i = (\lambda_1\boldsymbol{u}_1 + \cdots + \lambda_m\boldsymbol{u}_m) \cdot \boldsymbol{u}_i = \boldsymbol{0} \cdot \boldsymbol{u}_i = 0.$$

  This shows that $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m\}$ is linearly independent.
- Hence $\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m\}$ is an **orthonormal basis** for the subspace $\mathrm{lin}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m\}$ of $\mathbb{R}^n$.

Orthonormal Sequences (Cont'd)

- Thus each point $\boldsymbol{x}$ of $\text{lin}\{\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m\}$ can be written uniquely as a linear combination of $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_m$, say

$$\boldsymbol{x} = \mu_1\boldsymbol{u}_1 + \cdots + \mu_m\boldsymbol{u}_m.$$

Then, for $i = 1,\ldots,m$,

$$\boldsymbol{x}\cdot\boldsymbol{u}_i = (\mu_1\boldsymbol{u}_1 + \cdots + \mu_m\boldsymbol{u}_m)\cdot\boldsymbol{u}_i = \mu_i.$$

We conclude that

$$\boldsymbol{x} = (\boldsymbol{x}\cdot\boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (\boldsymbol{x}\cdot\boldsymbol{u}_m)\boldsymbol{u}_m.$$

Congruences in $\mathbb{R}^2$

- A congruence transformation in elementary plane geometry is a transformation of the plane which preserves distance.
- Examples of such transformations are reflections, rotations, translations, and combinations of these.
- Algebraically, the congruence transformations of $\mathbb{R}^2$ are precisely those affine transformations $T : \mathbb{R}^2 \to \mathbb{R}^2$ that can be expressed in the form

$$T(x) = \boldsymbol{Qx} + \boldsymbol{q},$$

where $\boldsymbol{Q}$ is a 2×2 orthogonal matrix and $\boldsymbol{q}$ is a 2×1 matrix.

Congruence Transformations

- A mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be a **congruence transformation** of $\mathbb{R}^n$ if
$$\|T(\boldsymbol{x}) - T(\boldsymbol{y})\| = \|\boldsymbol{x} - \boldsymbol{y}\|, \quad \text{for all } \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,$$
i.e., T preserves distance.
- We use a superscript T to denote the transpose of a matrix or a vector.
- Thus, recalling that we identify a point $\boldsymbol{x} = (x_1, \ldots, x_n)$ of $\mathbb{R}^n$ with a column vector in the natural way, we see that $\boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}$ is the 1×1 matrix whose single element is the scalar $x_1^2 + \cdots + x_n^2$.
- We identify this scalar with the matrix $\boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}$ itself, so that we may write
$$\|\boldsymbol{x}\|^2 = x_1^2 + \cdots + x_n^2 = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{x}.$$

Affine Transformations and Congruences

- We now show that an affine transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, which is defined by an equation of the form $T(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{q}$, where $\boldsymbol{Q}$ is an $n \times n$ orthogonal matrix and $\boldsymbol{q}$ is an $n \times 1$ matrix, is a congruence transformation of $\mathbb{R}^n$.
- Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Then

$$\begin{aligned} \|T(\boldsymbol{x}) - T(\boldsymbol{y})\|^2 &= \|\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{y})\|^2 \\ &= (\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{y}))^{\mathsf{T}}(\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{y})) \\ &= (\boldsymbol{x}-\boldsymbol{y})^{\mathsf{T}}\boldsymbol{Q}^{\mathsf{T}}\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{y}) \\ &= (\boldsymbol{x}-\boldsymbol{y})^{\mathsf{T}}(\boldsymbol{x}-\boldsymbol{y}) \\ &= \|\boldsymbol{x}-\boldsymbol{y}\|^2. \end{aligned}$$

  Hence $\|T(\boldsymbol{x}) - T(\boldsymbol{y})\| = \|\boldsymbol{x}-\boldsymbol{y}\|$. This shows that T is a congruence transformation of $\mathbb{R}^n$.

Congruences and Affine Transformations

Theorem

Let $T:\mathbb{R}^n \to \mathbb{R}^n$ be a congruence transformation of $\mathbb{R}^n$. Then there exist an $n\times n$ orthogonal matrix $\boldsymbol{Q}$ and an $n\times 1$ matrix $\boldsymbol{q}$ such that $T(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x}+\boldsymbol{q}$, for all $\boldsymbol{x}$ in $\mathbb{R}^n$.

- Let $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$. Define a mapping $f:\mathbb{R}^n\to\mathbb{R}^n$ by the equation $f(\boldsymbol{x}) = T(\boldsymbol{x}) - T(\boldsymbol{0})$. Since T preserves distance,

$$\|f(\boldsymbol{x})\| = \|T(\boldsymbol{x}) - T(\boldsymbol{0})\| = \|\boldsymbol{x}-\boldsymbol{0}\| = \|\boldsymbol{x}\|.$$

  So f preserves norms.

  Also

$$\|f(\boldsymbol{x})-f(\boldsymbol{y})\|^2 = \|T(\boldsymbol{x})-T(\boldsymbol{y})\|^2 = \|\boldsymbol{x}-\boldsymbol{y}\|^2.$$

  So

$$\|f(\boldsymbol{x})\|^2 - 2f(\boldsymbol{x})\cdot f(\boldsymbol{y}) + \|f(\boldsymbol{y})\|^2 = \|\boldsymbol{x}\|^2 - 2\boldsymbol{x}\cdot\boldsymbol{y} + \|\boldsymbol{y}\|^2.$$

Congruences and Affine Transformations (Cont'd)

- Since $\|f(\boldsymbol{x})\| = \|\boldsymbol{x}\|$ and $\|f(\boldsymbol{y})\| = \|\boldsymbol{y}\|$, we can deduce from the last equation that $f(\boldsymbol{x}) \cdot f(\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$.

  Thus, f preserves inner products.

  It follows that $f(\boldsymbol{e}_1), \ldots, f(\boldsymbol{e}_n)$ is an orthonormal sequence in $\mathbb{R}^n$.

  Hence

  $$f(\boldsymbol{x}) = (f(\boldsymbol{x}) \cdot f(\boldsymbol{e}_1))f(\boldsymbol{e}_1) + \cdots + (f(\boldsymbol{x}) \cdot f(\boldsymbol{e}_n))f(\boldsymbol{e}_n).$$

  Writing $\boldsymbol{x}$ for $(x_1, \ldots, x_n)$ and $\boldsymbol{Q}$ for the $n \times n$ orthogonal matrix whose columns are $f(\boldsymbol{e}_1), \ldots, f(\boldsymbol{e}_n)$, we deduce that

  $$\begin{aligned} f(\boldsymbol{x}) &= (\boldsymbol{x} \cdot \boldsymbol{e}_1)f(\boldsymbol{e}_1) + \cdots + (\boldsymbol{x} \cdot \boldsymbol{e}_n)f(\boldsymbol{e}_n) \\ &= x_1 f(\boldsymbol{e}_1) + \cdots + x_n f(\boldsymbol{e}_n) \\ &= \boldsymbol{Q}\boldsymbol{x}. \end{aligned}$$

  The proof is completed by putting $\boldsymbol{q} = T(\boldsymbol{0})$.

Congruent Subsets

- We have thus identified the congruence transformations $T : \mathbb{R}^n \to \mathbb{R}^n$ of $\mathbb{R}^n$ as being precisely those affine transformations $T : \mathbb{R}^n \to \mathbb{R}^n$ which can be expressed in the form $T(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{q}$, where $\boldsymbol{Q}$ is an $n \times n$ orthogonal matrix and $\boldsymbol{q}$ is an $n \times 1$ matrix.
- Sets A and B in $\mathbb{R}^n$ are said to be **congruent** if there is a congruence transformation T of $\mathbb{R}^n$ such that $T(A) = B$.
- It is easy to verify that congruence is an equivalence relation on the family of all subsets of $\mathbb{R}^n$.
- In elementary geometry, any two points are congruent, any two lines are congruent, and any two planes are congruent.

Congruent Flats

Theorem

Let A and B be r-flats in $\mathbb{R}^n$. Then A and B are congruent.

- We consider the non-trivial cases when $r \geq 1$.
  First we show that the r-flat A is congruent to the r-flat R_r defined by the equation
  $$R_r = \{(x_1, \ldots, x_r, 0, \ldots, 0) : x_1, \ldots, x_r \in \mathbb{R}\}.$$
  Let $\boldsymbol{a} \in A$. Then $A - \boldsymbol{a}$ is an r-dimensional subspace of $\mathbb{R}^n$.
  Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ such that $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}$ is an orthonormal basis for $A - \boldsymbol{a}$. Define a congruence transformation T of $\mathbb{R}^n$ by the equation
  $$T(\boldsymbol{x}) = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]\boldsymbol{x} + \boldsymbol{a}.$$
  Then $T(R_r) = A$. So A and R_r are congruent. Similarly, B and R_r are congruent. Thus A and B are congruent.

Congruent Copies of a Set

- We now show how, given any r-dimensional set A in $\mathbb{R}^n$ with $1 \leq r \leq n$, it is possible to find a congruent copy of A in the space $\mathbb{R}^r$.
- Moreover, we show that any two such congruent copies of A in $\mathbb{R}^r$ are themselves congruent to one another in $\mathbb{R}^r$.
- Let A be an r-dimensional $(1 \leq r \leq n)$ set in $\mathbb{R}^n$. Then aff$A$ is an r-flat. So by the theorem, there is a congruence transformation of $\mathbb{R}^n$ which maps aff$A$ onto the r-flat

$$R_r = \{(x_1, \ldots, x_r, 0, \ldots, 0) : x_1, \ldots, x_r \in \mathbb{R}\}.$$

It follows that there is a set B in R_r, which is congruent to A. Let $i : R_r \to \mathbb{R}^r$ be the mapping that identifies each point $(x_1, \ldots, x_r, 0, \ldots, 0)$ of R_r with the point $(x_1, \ldots, x_r)$ of $\mathbb{R}^r$. Then $i(B)$ is a set lying in $\mathbb{R}^r$ which is a congruent copy of the set A in $\mathbb{R}^n$.
- In general, there will be an infinite number of such copies. We now see how any two of these copies of A are related.

Congruent Copies of a Set (Cont'd)

- Let $i(B)$ and $i(C)$ be congruent copies of A in $\mathbb{R}^r$, where B and C are congruent to A in $\mathbb{R}^n$ and lie in R_r.
  Then there is a congruence transformation T of $\mathbb{R}^n$ such that $T(B) = C$, and which maps R_r onto itself.
  By considering the images of $\mathbf{0}$ and the elementary vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r$ under T, it follows that T can be expressed in the form
  $$T(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{Q} & \mathbf{0} \\ \mathbf{0} & * \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} \boldsymbol{q} \\ \mathbf{0} \end{bmatrix},$$
  where $\boldsymbol{Q}$ is an $r \times r$ orthogonal matrix, $\boldsymbol{q}$ is an $r \times 1$ matrix, and $\mathbf{0}$ represents zero matrices of suitable shapes and sizes.
  Denote by T_r the congruence transformation of $\mathbb{R}^r$ defined by the equation $T_r(\boldsymbol{x}) = \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{q}$, where $\boldsymbol{x} = (x_1, \ldots, x_r)$.
  Then $T_r(i(B)) = i(C)$. This shows that the congruent copies $i(B)$ and $i(C)$ of A in $\mathbb{R}^r$ are congruent to one another in $\mathbb{R}^r$.

Subsection 7

Open Sets and Closed Sets

Open and Closed Balls

- Let $\boldsymbol{a} \in \mathbb{R}^n$ and $r > 0$.
- Then the **open ball** $B(\boldsymbol{a}; r)$ (**closed ball** $B[\boldsymbol{a}; r]$) with **center** $\boldsymbol{a}$ and **radius** r is the set of all points of $\mathbb{R}^n$ whose distance from $\boldsymbol{a}$ is less than (less than or equal to) r, i.e.,

$$\begin{aligned} B(\boldsymbol{a}; r) &= \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{a}\| < r\}; \\ B[\boldsymbol{a}; r] &= \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{a}\| \leq r\}. \end{aligned}$$

- In $\mathbb{R}^1$ the open (closed) ball with center a and radius r is the open (closed) **interval** $(a - r, a + r)$ ($[a - r, a + r]$).
- In $\mathbb{R}^2$ open (closed) balls are referred to as **open (closed) discs**.

Open and Closed Unit Balls

- The balls $B(\mathbf{0};1)$ and $B[\mathbf{0};1]$ in $\mathbb{R}^n$ are called, respectively, the **open unit ball** and the **closed unit ball**.
- If we denote them, respectively, by V and U, then

$$V = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| < 1\} \quad \text{and} \quad U = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| \leq 1\}.$$

- It follows that $B(\boldsymbol{a};r) = \boldsymbol{a} + rV$ and $B[\boldsymbol{a};r] = \boldsymbol{a} + rU$.
- We adopt U as the standard notation for the closed unit ball.

Open Sets

- A point $\boldsymbol{a}$ of a set A in $\mathbb{R}^n$ is said to be an **interior point** of A if it is the center of some open ball which lies in A, i.e. if there exists some $r > 0$ such that $B(\boldsymbol{a}; r) \subseteq A$.
- The set of interior points of A is called the **interior** of A and is denoted by $\text{int}A$.
- Clearly $\text{int}B \subseteq \text{int}A$ when $B \subseteq A$.
- A set in $\mathbb{R}^n$, each of whose points is an interior point of the set, is said to be **open**.
- Since $\text{int}A \subseteq A$ is always true, A is open if and only if $\text{int}A = A$.
- Clearly the sets $\emptyset$ and $\mathbb{R}^n$ are open.

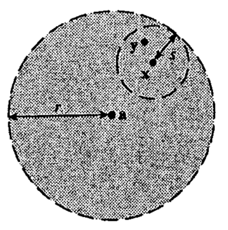
Balls, Halfspaces, Hyperplanes

Theorem

In $\mathbb{R}^n$ open balls and open halfspaces are open, and hyperplanes have empty interiors.

- Consider the open ball $B(\boldsymbol{a}; r)$, where $\boldsymbol{a} \in \mathbb{R}^n$ and $r > 0$. Let $x \in B(\boldsymbol{a}; r)$. We prove that $B(\boldsymbol{a}; r)$ is open by showing that $B(\boldsymbol{x}; s) \subseteq B(\boldsymbol{a}; r)$, where s is the positive number $r - \|\boldsymbol{x} - \boldsymbol{a}\|$.

  Let $\boldsymbol{y} \in B(\boldsymbol{x}; s)$. Then $\|\boldsymbol{y} - \boldsymbol{x}\| < s$.
  So by the triangle inequality



$$\begin{array}{rcl} \|\boldsymbol{y} - \boldsymbol{a}\| & \leq & \|\boldsymbol{y} - \boldsymbol{x} + \boldsymbol{x} - \boldsymbol{a}\| \\ & \leq & \|\boldsymbol{y} - \boldsymbol{x}\| + \|\boldsymbol{x} - \boldsymbol{a}\| \\ & < & s + \|\boldsymbol{x} - \boldsymbol{a}\| = r. \end{array}$$

  Thus $\boldsymbol{y} \in B(\boldsymbol{a}; r)$. So $B(\boldsymbol{x}; s) \subseteq B(\boldsymbol{a}; r)$.

Balls, Halfspaces, Hyperplanes (Cont'd)

- Consider the open halfspace A in $\mathbb{R}^n$ which is defined by the inequality $c_0 + \boldsymbol{c} \cdot \boldsymbol{x} > 0$, where $\boldsymbol{c}$ is a unit vector. Let $\boldsymbol{a} \in A$. We prove that A is open by showing that $B(\boldsymbol{a}; r) \subseteq A$, where r is the positive number $c_0 + \boldsymbol{c} \cdot \boldsymbol{a}$. Let $\boldsymbol{y} \in B(\boldsymbol{a}; r)$. Then $\|\boldsymbol{y} - \boldsymbol{a}\| < r$. Moreover,

$$c_0 + \boldsymbol{c} \cdot \boldsymbol{y} = c_0 + \boldsymbol{c} \cdot \boldsymbol{a} + \boldsymbol{c} \cdot (\boldsymbol{y} - \boldsymbol{a}) = r + \boldsymbol{c} \cdot (\boldsymbol{y} - \boldsymbol{a}) > 0,$$

  since, by the Cauchy-Schwarz Inequality, $|\boldsymbol{c} \cdot (\boldsymbol{y} - \boldsymbol{a})| \leq \|\boldsymbol{y} - \boldsymbol{a}\| < r$. Thus $\boldsymbol{y} \in A$. So $B(\boldsymbol{a}; r) \subseteq A$.
- Consider the hyperplane H in $\mathbb{R}^n$ with equation $c_0 + \boldsymbol{c} \cdot \boldsymbol{x} = 0$, where $\boldsymbol{c}$ is a unit vector. We show that no point $\boldsymbol{a}$ of H is an interior point of H. Let $r > 0$. Then $\boldsymbol{a} + \frac{1}{2} r \boldsymbol{c} \notin H$ and $\|\boldsymbol{a} + \frac{1}{2} r \boldsymbol{c} - \boldsymbol{a}\| = \frac{1}{2} r$. Therefore, $\boldsymbol{a} + \frac{1}{2} r \boldsymbol{c} \in B(\boldsymbol{a}; r)$ and $B(\boldsymbol{a}; r) \nsubseteq H$. Hence, $\boldsymbol{a}$ is not an interior point of H. So H has an empty interior.

Properties of the Interior

Corollary

Let A be a set in $\mathbb{R}^n$. Then $\text{int}A$ is open and $\text{int}(\text{int}A) = \text{int}A$.

- If $\boldsymbol{a} \in \text{int}A$, then there exists $r > 0$ such that $B(\boldsymbol{a}; r) \subseteq A$. Since $B(\boldsymbol{a}; r)$ is open,

$$B(\boldsymbol{a}; r) = \text{int}(B(\boldsymbol{a}; r)) \subseteq \text{int}A.$$

  Hence, $\boldsymbol{a} \in \text{int}(\text{int}A)$. So $\text{int}A \subseteq \text{int}(\text{int}A)$. Thus, $\text{int}A$ is open and $\text{int}(\text{int}A) = \text{int}A$.

Properties of Open Sets

Theorem

In $\mathbb{R}^n$ every union and every finite intersection of open sets is open.

- Let A be the union of a family $(A_i : i \in I)$ of open sets in $\mathbb{R}^n$. If $\boldsymbol{a} \in A$, then $\boldsymbol{a} \in A_i$, for some $i \in I$. Since A_i is open, there is an $r > 0$ such that $B(\boldsymbol{a}; r) \subseteq A_i$. Hence, $B(\boldsymbol{a}; r) \subseteq A$. Thus, A is open.
- Let A be the intersection of the open sets $A_1, \ldots, A_m$ in $\mathbb{R}^n$. If $\boldsymbol{a} \in A$, then $\boldsymbol{a} \in A_1, \ldots, \boldsymbol{a} \in A_m$. Since $A_1, \ldots, A_m$ are open, there exist $r_1, \ldots, r_m > 0$ such that $B(\boldsymbol{a}; r_1) \subseteq A_1, \ldots, B(\boldsymbol{a}; r_m) \subseteq A_m$. Let $r = \min\{r_1, \ldots, r_m\}$. Then $r > 0$ and

$$B(\boldsymbol{a}; r) \subseteq B(\boldsymbol{a}; r_1) \cap \cdots \cap B(\boldsymbol{a}; r_m) \subseteq A_1 \cap \cdots \cap A_m = A.$$

  Thus A is open.

Intersections of Open Sets

- An arbitrary intersection of open sets in $\mathbb{R}^n$ need not be open.
- To see this, we note that the intersection of the sequence

$$V, \frac{1}{2}V, \frac{1}{3}V, \ldots, \frac{1}{k}V, \ldots$$

of open balls centered at the origin of $\mathbb{R}^n$ is the singleton set $\{\mathbf{0}\}$, which is not open.

Closure of a Set

- A point $\boldsymbol{a}$ of $\mathbb{R}^n$ is said to be a **closure point** of a set A in $\mathbb{R}^n$ if every open ball with center $\boldsymbol{a}$ meets A, i.e., if for every $r>0$ the ball $B(\boldsymbol{a};r)$ meets A.
- The set of closure points of A is called the **closure** of A and is denoted by $\text{cl}A$.
- Clearly $A \subseteq \text{cl}A$.
- Also $\text{cl}B \subseteq \text{cl}A$ whenever $B \subseteq A$.
- Roughly speaking, the closure of A is the set of all points in $\mathbb{R}^n$ which either lie in A or are arbitrarily close to A.
- Thus, in $\mathbb{R}^1$ the closures of the intervals $(0,1]$, $(0,1)$, $[0,1)$ are all equal to the interval $[0,1]$.
- In $\mathbb{R}^2$ the closures of the discs $B(\boldsymbol{a};r)$ and $B[\boldsymbol{a};r]$ are both equal to the disc $B[\boldsymbol{a};r]$.

Closed Sets

- A set in $\mathbb{R}^n$ each of whose closure points lies in the set is said to be **closed**.
- Thus a set A in $\mathbb{R}^n$ is closed if and only if $\text{cl}A \subseteq A$.
- Since $A \subseteq \text{cl}A$ is always true, A is closed if and only if $\text{cl}A = A$.
- Clearly the sets $\emptyset$ and $\mathbb{R}^n$ are closed.
- Thus the sets $\emptyset$ and $\mathbb{R}^n$ are both open and closed.
- It can be shown that they are the only sets in $\mathbb{R}^n$ with this property.
- A set in $\mathbb{R}^n$ may be neither open nor closed.
- For example, in $\mathbb{R}^1$ the interval $[0,1)$ is such a set.

Closure and Interior

- For each set A in $\mathbb{R}^n$, we denote by A^c the complement of A in $\mathbb{R}^n$, i.e., the set $\mathbb{R}^n \backslash A$.

Theorem

Let A be a set in $\mathbb{R}^n$. Then $\text{cl}A = (\text{int}A^c)^c$.

- If $\boldsymbol{x} \in \text{cl}A$, then each open ball with center $\boldsymbol{x}$ contains a point of A. So $\boldsymbol{x}$ cannot belong to $\text{int}A^c$, i.e., $\boldsymbol{x} \in (\text{int}A^c)^c$.

  If $\boldsymbol{x} \in (\text{int}A^c)^c$, then each open ball with center $\boldsymbol{x}$ must contain a point of A, i.e., $\boldsymbol{x} \in \text{cl}A$.

  Thus $\text{cl}A = (\text{int}A^c)^c$.

Closed and Open Sets

Theorem

A set in $\mathbb{R}^n$ is closed if and only if its complement in $\mathbb{R}^n$ is open.

- Let A be a set in $\mathbb{R}^n$. Suppose first that A is closed. Then $\text{cl}A = A$. It follows from a previous corollary and the preceding theorem that A^c is the open set $\text{int}A^c$. Suppose next that A^c is open. Then $\text{int}A^c = A^c$. It follows from the theorem that $\text{cl}A = A$, i.e., A is closed.

Corollary

Let A be a set in $\mathbb{R}^n$. Then $\text{cl}A$ is closed and $\text{cl}(\text{cl}A) = \text{cl}A$.

- Now $\text{int}A^c$ is open by a previous corollary. Hence by the theorem its complement $\text{cl}A$ is closed.

Properties of Closed Sets

Theorem

In $\mathbb{R}^n$ every intersection and every finite union of closed sets is closed.

- Let $(A_i : i \in I)$ be a family of closed sets in $\mathbb{R}^n$. Then, for each $i \in I$, A_i^c is open. By a previous theorem, $\bigcup(A_i^c : i \in I)$ is open. Hence

$$\bigcap(A_i : i \in I) = (\bigcup(A_i^c : i \in I))^c$$

is closed.

Now let $A_1, \ldots, A_m$ be closed sets in $\mathbb{R}^n$. Then $A_1^c, \ldots, A_m^c$ are open. By a previous theorem, $A_1^c \cap \cdots \cap A_m^c$ is open. Hence

$$A_1 \cup \cdots \cup A_m = (A_1^c \cap \cdots \cap A_m^c)^c$$

is closed.

Closures and Unions

Corollary

Let $A_1, \ldots, A_m$ be sets in $\mathbb{R}^n$. Then

$$\mathrm{cl}(A_1 \cup \cdots \cup A_m) = \mathrm{cl}A_1 \cup \cdots \cup \mathrm{cl}A_m.$$

- Since $A_1 \cup \cdots \cup A_m$ is contained in the closed set $\mathrm{cl}A_1 \cup \cdots \cup \mathrm{cl}A_m$,

$$\mathrm{cl}(A_1 \cup \cdots \cup A_m) \subseteq \mathrm{cl}A_1 \cup \cdots \cup \mathrm{cl}A_m.$$

Trivially,

$$\mathrm{cl}(A_1 \cup \cdots \cup A_m) \supseteq \mathrm{cl}A_1 \cup \cdots \cup \mathrm{cl}A_m.$$

Thus,

$$\mathrm{cl}(A_1 \cup \cdots \cup A_m) = \mathrm{cl}A_1 \cup \cdots \cup \mathrm{cl}A_m.$$

Closed Balls, Closed Halfspaces, Flats

Theorem

In $\mathbb{R}^n$ closed balls, closed halfspaces and flats are closed.

- Let A be the closed ball $B[\boldsymbol{a};r]$, where $\boldsymbol{a} \in \mathbb{R}^n$ and $r > 0$.
  We prove that A^c is open. Let $\boldsymbol{x} \in A^c$. Then we show that $B(\boldsymbol{x};s) \subseteq A^c$, where s is the positive number $\|\boldsymbol{x}-\boldsymbol{a}\| - r$.
  Suppose that this is not the case. Then there is some point of A, $\boldsymbol{y}$ say, which lies in $B(\boldsymbol{x};s)$. Now
  $$\|\boldsymbol{x}-\boldsymbol{a}\| = \|\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{y}-\boldsymbol{a}\| < s + r = \|\boldsymbol{x}-\boldsymbol{a}\|,$$
  which is impossible. Hence $B(\boldsymbol{x};s) \subseteq A^c$.
  A previous theorem shows that open halfspaces in $\mathbb{R}^n$ are open. Hence their complements in $\mathbb{R}^n$, i.e., the closed halfspaces, are closed.
  In $\mathbb{R}^n$ each hyperplane is the intersection of two closed halfspaces. So it is closed. By a previous corollary, each flat in $\mathbb{R}^n$ is an intersection of hyperplanes. So it is closed.

Boundaries

- A point $\boldsymbol{a}$ of $\mathbb{R}^n$ is said to be a **boundary point** of a set A in $\mathbb{R}^n$ if every open ball with center $\boldsymbol{a}$ meets both A and its complement A^c.
- The set of boundary points of A is called the **boundary** of A and is denoted by bd$A$.
- Thus a boundary point of a set in $\mathbb{R}^n$ is a point of $\mathbb{R}^n$ which is arbitrarily close both to the set and its complement.
- It follows from the preceding definitions that $\text{bd}A = (\text{cl}A) \cap (\text{cl}A^c)$.
- Hence the boundary of a set in $\mathbb{R}^n$ is always closed, being the intersection of two closed sets.

More Properties of Boundaries

- A boundary point of a set in $\mathbb{R}^n$ may or may not belong to the set itself.
- For example, in $\mathbb{R}^1$ the interval $[0,1)$ contains its boundary point 0, but not its boundary point 1.
- For any set A in $\mathbb{R}^n$, the sets A and A^c have the same boundary.
- Moreover, the sets $\text{int}A$, $\text{bd}A$, $\text{int}A^c$ form a partition of $\mathbb{R}^n$.
- Open (closed) sets in $\mathbb{R}^n$ are characterized by the property that they contain none (all) of their boundary points.

Dependence on Ambient Space

- The above definitions of the interior and the boundary of a set depend upon the space in which the set is embedded.
- For example, a closed line segment in $\mathbb{R}^2$ has an empty interior and is its own boundary.
- The same line segment considered as a subset of $\mathbb{R}^1$ has for its interior the set of all of its points with the exception of its two boundary points, these forming its boundary in $\mathbb{R}^1$.
- The latter interior and boundary, obtained by regarding the one-dimensional line segment as a set in the one-dimensional space $\mathbb{R}^1$, correspond to what may be thought of as the "intrinsic" interior and boundary of the segment.

Relative Interior

- A point $\boldsymbol{a}$ of a set A in $\mathbb{R}^n$ is said to be a **relative interior point** of A if it is the center of some open ball whose intersection with $\operatorname{aff} A$ is contained in A, i.e., if there exists $r > 0$ such that

$$B(\boldsymbol{a}; r) \cap \operatorname{aff} A \subseteq A.$$

- The set of all relative interior points of A is called the **relative interior** of A and is denoted by $\operatorname{ri} A$.
- The relative interior of an n-dimensional set in $\mathbb{R}^n$ coincides with its interior.
- The relative interior of any flat in $\mathbb{R}^n$ is itself.

Relative Boundary

- A point $\boldsymbol{a}$ of $\mathbb{R}^n$ is said to be a **relative boundary point** of a set A in $\mathbb{R}^n$ if it lies in the closure of A but not in its relative interior.
- The set of all relative boundary points of A is called the **relative boundary** of A and is denoted by $\text{rebd}A$.
- The relative boundary of an n-dimensional set in $\mathbb{R}^n$ coincides with its boundary.

Properties of Relative Interior

- It is to be noted that while the inclusion $B \subseteq A$ implies both $\text{int}B \subseteq \text{int}A$ and $\text{cl}B \subseteq \text{cl}A$, it does not in general imply $\text{ri}B \subseteq \text{ri}A$.
- For example, if B is one side of a square A in $\mathbb{R}^2$, then $\text{ri}B$ and $\text{ri}A$ are non-empty but disjoint.
- If, however, $B \subseteq A$ and $\dim B = \dim A$ or, equivalently, $\text{aff}\, B = \text{aff}\, A$, then $\text{ri}B \subseteq \text{ri}A$.

Flats and Relative Boundaries

- Suppose that $\boldsymbol{a}$ is a point of a set A in $\mathbb{R}^n$ and that $\boldsymbol{x}$ is a point of $\operatorname{aff} A$ not lying in A.
- Define a scalar λ_0 by the equation

$$\lambda_0 = \sup\{\lambda \in [0,1] : (1-\lambda)\boldsymbol{a} + \lambda\boldsymbol{x} \in A\}.$$

- Then $(1-\lambda_0)\boldsymbol{a} + \lambda_0\boldsymbol{x}$ is a relative boundary point of A lying between $\boldsymbol{a}$ and $\boldsymbol{x}$.
- It follows that flats are the only sets in $\mathbb{R}^n$ which have an empty relative boundary.

Subsection 8

Convergence and Compactness

Convergence of Sequences

- In $\mathbb{R}^n$ a sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ of points is said to **converge** to a point $\boldsymbol{x}$ if $\|\boldsymbol{x}_k - \boldsymbol{x}\| \to 0$ as $k \to \infty$, i.e., if the distance $\|\boldsymbol{x}_k - \boldsymbol{x}\|$ between $\boldsymbol{x}_k$ and $\boldsymbol{x}$ tends to zero as k tends to infinity.
- We indicate such convergence by writing $\boldsymbol{x}_k \to \boldsymbol{x}$ as $k \to \infty$, or simply $\boldsymbol{x}_k \to \boldsymbol{x}$.
- This convergence for sequences of points in $\mathbb{R}^n$ coincides with that of classical convergence for real sequences.

Properties of Convergence

- The inequality $|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|| \leq \|\boldsymbol{x} - \boldsymbol{y}\|$ proven previously, shows that
$$|\|\boldsymbol{x}_k\| - \|\boldsymbol{x}\|| \leq \|\boldsymbol{x}_k - \boldsymbol{x}\|.$$
Hence $\|\boldsymbol{x}_k\| \to \|\boldsymbol{x}\|$ as $k \to \infty$ whenever $\boldsymbol{x}_k \to \boldsymbol{x}$ as $k \to \infty$.
- The triangle inequality shows that
$$\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \leq \|\boldsymbol{x}_i - \boldsymbol{x}\| + \|\boldsymbol{x} - \boldsymbol{x}_j\|.$$
Hence $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \to 0$ as $i, j \to \infty$ whenever $\boldsymbol{x}_k \to \boldsymbol{x}$ as $k \to \infty$.

Convergence and Coordinate-wise Convergence

- Suppose that $\boldsymbol{x}_k = (x_{k1}, \ldots, x_{kn})$ for $k = 1, 2, \ldots$ and that $\boldsymbol{x} = (x_1, \ldots, x_n)$.
- Then, for $i = 1, \ldots, n$, we have

$$|x_{ki} - x_i|^2 \le (x_{k1} - x_1)^2 + \cdots + (x_{kn} - x_n)^2 = \|\boldsymbol{x}_k - \boldsymbol{x}\|^2.$$

- We also have

$$\begin{array}{rcl} \|\boldsymbol{x}_k - \boldsymbol{x}\|^2 & = & (x_{k1} - x_1)^2 + \cdots + (x_{kn} - x_n)^2 \\ & \le & (|x_{k1} - x_1| + \cdots + |x_{kn} - x_n|)^2. \end{array}$$

- Hence

$$|x_{ki} - x_i| \le \|\boldsymbol{x}_k - \boldsymbol{x}\| \le |x_{k1} - x_1| + \cdots + |x_{kn} - x_n|.$$

- Thus, $\boldsymbol{x}_k \to \boldsymbol{x}$ if and only if $x_{ki} \to x_i$, for $i = 1, \ldots, n$.
- So the convergence of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ to $(x_1, \ldots, x_n)$ is equivalent to the convergence of each of the coordinate sequences $x_{1i}, \ldots, x_{ki}, \ldots$ for $i = 1, \ldots, n$.

Uniqueness and Linearity Properties

- A consequence of coordinate-wise convergence is that a sequence of points in $\mathbb{R}^n$ can converge to at most one point.
- Moreover, if $\boldsymbol{x}_k \to \boldsymbol{x}$, $\boldsymbol{y}_k \to \boldsymbol{y}$ in $\mathbb{R}^n$ and $\lambda_k \to \lambda$, $\mu_k \to \mu$ in $\mathbb{R}$, then

$$\begin{aligned} \boldsymbol{x}_k \cdot \boldsymbol{y}_k \to \boldsymbol{x} \cdot \boldsymbol{y} \quad & \text{in } \mathbb{R}; \\ \lambda_k \boldsymbol{x}_k + \mu_k \boldsymbol{y}_k \to \lambda \boldsymbol{x} + \mu \boldsymbol{y} \quad & \text{in } \mathbb{R}^n. \end{aligned}$$

Boundedness

- We recall that a sequence $x_1, \ldots, x_k, \ldots$ of real numbers is said to be **bounded** if there exists a real number r such that $|x_k| \le r$ for $k = 1, 2, \ldots$.
- Similarly, a sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ of points in $\mathbb{R}^n$ is defined to be **bounded** if there exists a real number r such that $\|\boldsymbol{x}_k\| \le r$ for $k = 1, 2, \ldots$.
- Every convergent sequence of real numbers is bounded, and the same is also true for convergent sequences of points in $\mathbb{R}^n$.

  To see this, suppose that $\boldsymbol{x}_k \to \boldsymbol{x}$ in $\mathbb{R}^n$. By what we proved above, $\|\boldsymbol{x}_k\| \to \|\boldsymbol{x}\|$. So there exists a real number r such that $\|\boldsymbol{x}_k\| \le r$ for $k = 1, 2, \ldots$.

Boundedness and Convergence

- The next theorem generalizes to $\mathbb{R}^n$ the classical result that every bounded sequence of real numbers contains a convergent subsequence.

Theorem

Every bounded sequence of points of $\mathbb{R}^n$ contains a convergent subsequence.

- Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ be a bounded sequence of points in $\mathbb{R}^n$. Then each of the n coordinate sequences associated with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ is bounded in $\mathbb{R}$. In particular, the sequence of the first coordinates of $\boldsymbol{x}_1, \ldots, \boldsymbol{k}_k, \ldots$ is a bounded sequence of real numbers. Thus there exists a subsequence of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ such that the sequence of its first coordinates converges. Similarly, there exists a subsequence of this subsequence of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ such that the sequence of its second coordinates converges.

Boundedness and Convergence (Cont'd)

- After performing this subsequence operation n times in all, we arrive at a subsequence of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ each of whose n coordinate sequences converges.

  I.e., we have found a convergent subsequence of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$.

Closure in Terms of Sequences

Theorem

Let A be a set in $\mathbb{R}^n$. Then $\boldsymbol{x} \in \text{cl}A$ if and only if there exists a sequence of points of A which converges to $\boldsymbol{x}$.

- Suppose first that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ is a sequence of points of A which converges to a point $\boldsymbol{x}$ of $\mathbb{R}^n$. Then, for each $r > 0$, there is some point $\boldsymbol{x}_k$ of the sequence such that $\|\boldsymbol{x}_k - \boldsymbol{x}\| < r$. Hence the open ball $B(\boldsymbol{x}; r)$ meets A. This shows that $\boldsymbol{x} \in \text{cl}A$.

  Suppose next that $\boldsymbol{x} \in \text{cl}A$. Then, for each positive integer k, the ball $B(\boldsymbol{x}; \frac{1}{k})$ meets A. Hence there exists $\boldsymbol{x}_k \in A$ such that $\|\boldsymbol{x}_k - \boldsymbol{x}\| < \frac{1}{k}$. It follows that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ converges to $\boldsymbol{x}$.

Closed Sets in Terms of Sequences

Corollary

Let A be a set in $\mathbb{R}^n$. Then A is closed if and only if each convergent sequence of points of A converges to a point of A.

- The corollary follows from the theorem and the fact that A is closed if and only if $A = \mathrm{cl} A$.

Bounded and Compact Subsets

- The set A in $\mathbb{R}^n$ is said to be **bounded** if there exists a real number r such that $\|\boldsymbol{a}\| \leq r$ for all $\boldsymbol{a} \in A$.
- Clearly, a set in $\mathbb{R}^n$ is bounded if and only if each sequence of its points is bounded.
- In $\mathbb{R}^n$ balls and finite sets are bounded, whereas r-flats $(r \geq 1)$ are not.
- A previous theorem and a corollary, taken together, show that each sequence of points of a closed bounded set in $\mathbb{R}^n$ contains some subsequence which converges to a point of the set.
- A subset of $\mathbb{R}^n$ is said to be **compact**, if each sequence of its points contains some subsequence that converges to a point of the subset.

Characterization of Compact Subsets

Theorem

Let A be a set in $\mathbb{R}^n$. Then A is compact if and only it it is both closed and bounded.

- We know that closed bounded subsets of $\mathbb{R}^n$ are compact.

  Suppose, then, that A is compact. We show first that A is closed. If $\boldsymbol{x} \in \text{cl}A$, then, by a previous theorem, there is a sequence of points of A which converges to $\boldsymbol{x}$. Every subsequence of such a sequence also converges to $\boldsymbol{x}$. The compactness of A and the uniqueness of limits show that $\boldsymbol{x} \in A$. Hence A is closed.

  Suppose next that A is not bounded. Then, for each positive integer k, there must exist a point $\boldsymbol{x}_k$ of A such that $\|\boldsymbol{x}_k\| > k$. The sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ of points of A contains no bounded subsequence, and hence no convergent subsequence, contrary to the hypothesis that A is compact. Hence A is both closed and bounded.

Compactness and Coverings

Theorem

Let A be a non-empty compact set in $\mathbb{R}^n$ and let $r > 0$. Then there exists a finite number of points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ of A such that

$$A \subseteq B(\boldsymbol{a}_1; r) \cup \cdots \cup B(\boldsymbol{a}_m; r).$$

- We argue by contradiction. Suppose that no such finite number of points of A exists. Let $\boldsymbol{x}_1 \in A$. Then $A \nsubseteq B(\boldsymbol{x}_1; r)$. Hence there exists a point $\boldsymbol{x}_2$ of A such that $\|\boldsymbol{x}_2 - \boldsymbol{x}_1\| \geq r$. Now $A \nsubseteq B(\boldsymbol{x}_1; r) \cup B(\boldsymbol{x}_2; r)$. Hence there exists a point $\boldsymbol{x}_3$ of A such that $\|\boldsymbol{x}_3 - \boldsymbol{x}_1\| \geq r$ and $\|\boldsymbol{x}_3 - \boldsymbol{x}_2\| \geq r$. Continuing in this way, we produce a sequence $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k, \ldots$ of points of A with the property that $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \geq r$ whenever $i \neq j$. Clearly such a sequence cannot contain a convergent subsequence. This contradicts the compactness of A.

Balls of Fixed Radius in a Covering

Lemma

Let A be a compact set in $\mathbb{R}^n$ and let $(U_i : i \in I)$ be a family of open sets in $\mathbb{R}^n$ whose union contains A. Then there exists $r > 0$ such that, for each $\boldsymbol{x}$ in A, the open ball $B(\boldsymbol{x}; r)$ is contained in some U_j.

- We argue by contradiction. Suppose that no such $r > 0$ exists.

  Then, for each positive integer k, there is some point $\boldsymbol{a}_k$ of A such that $B(\boldsymbol{a}_k; \frac{1}{k})$ is not contained in any U_i. Since A is compact, the sequence $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots$ has a subsequence which converges to a point $\boldsymbol{a}$ of A. This point $\boldsymbol{a}$ must belong to one of the U_i's, U^* say.

Balls of Fixed Radius in a Covering (Cont'd)

- Since U^* is open, there is an $s > 0$ such that $B(\boldsymbol{a}; 2s) \subseteq U^*$.

  Since some subsequence of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots$ converges to $\boldsymbol{a}$, there are infinitely many positive integers k for which $\|\boldsymbol{a}_k - \boldsymbol{a}\| < s$.

  Choose one of these positive integers, m say, so large that $\frac{1}{m} < s$.

  Let $\boldsymbol{x} \in B(\boldsymbol{a}_m; \frac{1}{m})$. Then

$$\|\boldsymbol{x} - \boldsymbol{a}\| \le \|\boldsymbol{x} - \boldsymbol{a}_m\| + \|\boldsymbol{a}_m - \boldsymbol{a}\| < s + s = 2s.$$

  So $\boldsymbol{x} \in B(\boldsymbol{a}; 2s)$. Thus $B(\boldsymbol{a}_m; \frac{1}{m}) \subseteq B(\boldsymbol{a}; 2s) \subseteq U^*$. This contradicts the assumption that $B(\boldsymbol{a}_m; \frac{1}{m})$ is not contained in any U_i.

Coverings and Finite Subcoverings

Theorem

Let A be a compact set in $\mathbb{R}^n$ and let $(U_i : i \in I)$ be a family of open sets in $\mathbb{R}^n$ whose union contains A. Then there exists a finite subset I^* of I such that the union of the family $(U_i : i \in I^*)$ contains A.

- We may suppose that A is non-empty. By the lemma, there is an $r > 0$ such that, for each $\boldsymbol{x}$ in A, the open ball $B(\boldsymbol{x}; r)$ is contained in some U_i. By the preceding theorem, there exist points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ in A such that

$$A \subseteq B(\boldsymbol{a}_1; r) \cup \cdots \cup B(\boldsymbol{a}_m; r).$$

  For each $k = 1, \ldots, m$, there exists $i_k \in I$ such that $B(\boldsymbol{a}_k; r) \subseteq U_{i_k}$. We complete the proof by taking I^* to be the set $\{i_1, \ldots, i_m\}$.

Intersection of Families of Compact Sets

Corollary

Let $(A_i : i \in I)$ be a family of compact sets in $\mathbb{R}^n$ whose intersection is empty. Then there exists a finite subset I^* of I such that the intersection of the family $(A_i : i \in I^*)$ is empty.

- Let $i_0 \in I$ and let $I_0 = I \setminus \{i_0\}$. Then, since $\bigcap(A_i : i \in I)$ is empty, $A_{i_0} \subseteq \bigcup(A_i^c : i \in I_0)$. By the theorem, which is applicable since the sets A_i^c are open, being the complements of closed sets in $\mathbb{R}^n$, there is a finite subset I' of I_0 such that $A_{i_0} \subseteq \bigcup(A_i^c : i \in I')$. It follows that, if I^* denotes the finite subset $I' \cup \{i_0\}$ of I, then $\bigcap(A_i : i \in I^*)$ is empty.

Decreasing Sequence of Compact Sets

Corollary

Let $A_1, \ldots, A_k, \ldots$ be a sequence of non-empty compact sets in $\mathbb{R}^n$ such that $A_1 \supseteq \cdots \supseteq A_k \supseteq \cdots$. Then the intersection $\bigcap(A_k : k = 1, 2, \ldots)$ is non-empty.

- The intersection of any finite number of members of the family is itself a member of the family. So it is non-empty.

  Thus, the result follows from the preceding corollary.

Properties of Linear Combinations of Sets

Theorem

Let A and B be sets in $\mathbb{R}^n$ and let $\lambda, \mu \in \mathbb{R}$. Then $\lambda A + \mu B$ is:

(i) open when A is open and $\lambda \neq 0$;
(ii) closed when A is compact and B is closed;
(iii) bounded when A and B are bounded;
(iv) compact when A and B are compact.

(i) Let A be open and let $\lambda \neq 0$. If $\boldsymbol{x} \in \lambda A + \mu B$, then $\boldsymbol{x} = \lambda \boldsymbol{a} + \mu \boldsymbol{b}$ for some $\boldsymbol{a} \in A$ and $\boldsymbol{b} \in B$. Since A is open, there is an $r > 0$ such that $\boldsymbol{a} + rV \subseteq A$, where V is the open unit ball $\{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\| < 1\}$. Thus

$$\boldsymbol{x} + \lambda rV = \lambda \boldsymbol{a} + \mu \boldsymbol{b} + \lambda rV = \lambda(\boldsymbol{a} + rV) + \mu \boldsymbol{b} \subseteq \lambda A + \mu B.$$

This shows that $B(\boldsymbol{x}; |\lambda| r) \subseteq \lambda A + \mu B$. Hence $\lambda A + \mu B$ is open.

Properties of Linear Combinations of Sets (Cont'd)

(ii) Let A be compact and let B be closed. We consider only the non-trivial case $\mu \neq 0$. If $\boldsymbol{x} \in \text{cl}(\lambda A + \mu B)$, then there exist sequences $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots$ of points of A, and $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k, \ldots$ of points of B such that $\lambda \boldsymbol{a}_k + \mu \boldsymbol{b}_k \to \boldsymbol{x}$ as $k \to \infty$. Since A is compact, there is a subsequence $\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_k}, \ldots$ of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots$ which converges to some point $\boldsymbol{a}$ of A. Thus $\lambda \boldsymbol{a}_{i_k} + \mu \boldsymbol{b}_{i_k} \to \boldsymbol{x}$ and $\boldsymbol{b}_{i_k} \to \frac{\boldsymbol{x} - \lambda \boldsymbol{a}}{\mu}$ as $k \to \infty$. But B is closed, and so $\frac{\boldsymbol{x} - \lambda \boldsymbol{a}}{\mu} \in B$. Hence $\boldsymbol{x} \in \lambda A + \mu B$. Thus $\boldsymbol{x} \in \text{cl}(\lambda A + \mu B)$ implies that $\boldsymbol{x} \in \lambda A + \mu B$. This shows that $\lambda A + \mu B$ is closed.

(iii) Let A and B be bounded. Then there exist real numbers r_1 and r_2 such that $\|\boldsymbol{a}\| \leq r_1$ and $\|\boldsymbol{b}\| \leq r_2$ whenever $\boldsymbol{a} \in A$ and $\boldsymbol{b} \in B$. If $\boldsymbol{x} \in \lambda A + \mu B$, then $\boldsymbol{x} = \lambda \boldsymbol{a} + \mu \boldsymbol{b}$ for some $\boldsymbol{a} \in A$ and $\boldsymbol{b} \in B$. Hence

$$\|\boldsymbol{x}\| = \|\lambda \boldsymbol{a} + \mu \boldsymbol{b}\| \leq |\lambda| \|\boldsymbol{a}\| + |\mu| \|\boldsymbol{b}\| \leq |\lambda| r_1 + |\mu| r_2.$$

This shows that $\lambda A + \mu B$ is bounded.

(iv) This follows immediately from (ii) and (iii).

Subsection 9

Continuity

Continuity

- Let $f : A \to \mathbb{R}^m$ be a mapping, where A is a non-empty set in $\mathbb{R}^n$.
- Then f is said to be **continuous at a point $\boldsymbol{a}$** of A if, for each sequence $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots$ of points of A that converges to $\boldsymbol{a}$, the sequence $f(\boldsymbol{a}_1), \ldots, f(\boldsymbol{a}_k), \ldots$ of points of $\mathbb{R}^m$ converges to $f(\boldsymbol{a})$.
- If f is continuous at all points of A, then f is said to be **continuous on** A.
- An important example of a continuous mapping is the norm mapping $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ defined by the equation $\|\cdot\|(\boldsymbol{x}) = \|\boldsymbol{x}\|$ for each point $\boldsymbol{x}$ of $\mathbb{R}^n$.

  That $\|\cdot\|$ is continuous follows immediately from the fact that $\|\boldsymbol{a}_k\| \to \|\boldsymbol{a}\|$ as $k \to \infty$ whenever $\boldsymbol{a}_k \to \boldsymbol{a}$ as $k \to \infty$.

Lipschitz Condition

- A mapping $f: A \to \mathbb{R}^m$ defined on a non-empty set A in $\mathbb{R}^n$ is said to satisfy a **Lipschitz condition** on A if there exists a real number s such that, for all $\boldsymbol{x}, \boldsymbol{y} \in A$,

$$\|f(\boldsymbol{x}) - f(\boldsymbol{y})\| \leq s\|\boldsymbol{x} - \boldsymbol{y}\|.$$

- If $f: A \to \mathbb{R}^m$ satisfies the Lipschitz condition, then it is continuous on A.

  To see this, suppose that $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots$ is a sequence of points of A that converges to a point $\boldsymbol{a}$ of A, so that $\|\boldsymbol{a}_k - \boldsymbol{a}\| \to 0$ as $k \to \infty$.
  The Lipschitz condition shows that

$$\|f(\boldsymbol{a}_k) - f(\boldsymbol{a})\| \leq s\|\boldsymbol{a}_k - \boldsymbol{a}\|.$$

  Hence, $\|f(\boldsymbol{a}_k) - f(\boldsymbol{a})\| \to 0$ as $k \to \infty$, i.e., the sequence $f(\boldsymbol{a}_1), \ldots, f(\boldsymbol{a}_k), \ldots$ converges to $f(\boldsymbol{a})$.
  Since f is continuous at an arbitrary $\boldsymbol{a}$ of A, f is continuous on A.

Affine Transformations are Lipschitz Mappings

- The norm mapping $\|\cdot\|:\mathbb{R}^n\to\mathbb{R}$ considered above satisfies the Lipschitz condition.
- Every affine transformation $T:\mathbb{R}^n\to\mathbb{R}^m$ satisfies a Lipschitz condition on $\mathbb{R}^n$.

  Suppose that $\boldsymbol{Q}=[q_{ij}]$ is the real $m\times n$ matrix and $\boldsymbol{q}$ the real $m\times 1$ matrix such that, for each vector $\boldsymbol{x}$ in $\mathbb{R}^n$, considered as a column vector, $T(\boldsymbol{x})=\boldsymbol{Q}\boldsymbol{x}+\boldsymbol{q}$. Let $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$. Write $\boldsymbol{u}=(u_1,\dots,u_n)=\boldsymbol{x}-\boldsymbol{y}$. By the Cauchy-Schwarz inequality, for $i=1,\dots,m$,

  $$(q_{i1}u_1+\cdots+q_{in}u_n)^2\le(q_{i1}^2+\cdots+q_{in}^2)(u_1^2+\cdots+u_n^2).$$

  Setting $s=\sqrt{\sum_{i=1}^m\sum_{j=1}^n q_{ij}^2}$, we get

  $$\begin{aligned}\|T(\boldsymbol{x})-T(\boldsymbol{y})\|^2 &= \|\boldsymbol{Q}\boldsymbol{u}\|^2=\sum_{i=1}^m(q_{i1}u_1+\cdots+q_{in}u_n)^2\\ &\le \sum_{i=1}^m(q_{i1}^2+\cdots+q_{in}^2)(u_1^2+\cdots+u_n^2)\\ &= s^2\|\boldsymbol{u}\|^2.\end{aligned}$$

The Distance Function

- The **distance function** $d_A : \mathbb{R}^n \to \mathbb{R}$ of a non-empty set A in $\mathbb{R}^n$ satisfies a Lipschitz condition.
- This function d_A associates with each point $\boldsymbol{x}$ of $\mathbb{R}^n$ its distance $d_A(\boldsymbol{x})$ from A.
- Formally, d_A is defined by the equation

$$d_A(\boldsymbol{x}) = \inf\{\|\boldsymbol{x} - \boldsymbol{a}\| : \boldsymbol{a} \in A\}, \quad \text{for } \boldsymbol{x} \in \mathbb{R}^n.$$

- If A is the singleton set $\{\boldsymbol{a}\}$, then $d_A(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{a}\|$.
- In particular, if $\boldsymbol{a} = \mathbf{0}$, then $d_A(\boldsymbol{x}) = \|\boldsymbol{x}\|$.
- It follows from the definition of d_A and a previous theorem that a point $\boldsymbol{x}$ of $\mathbb{R}^n$ lies in the closure $\text{cl}A$ of A if and only if its distance $d_A(\boldsymbol{x})$ from A is zero.

The Distance Function is Lipschitz

- Suppose now that $\boldsymbol{x}, \boldsymbol{y}$ lie in $\mathbb{R}^n$.
- Then, for each $\varepsilon > 0$, there exists $\boldsymbol{a}$ in A such that $\|\boldsymbol{x} - \boldsymbol{a}\| < d_A(\boldsymbol{x}) + \varepsilon$.
- By the triangle inequality,

$$d_A(\boldsymbol{y}) \leq \|\boldsymbol{y} - \boldsymbol{a}\| \leq \|\boldsymbol{y} - \boldsymbol{x}\| + \|\boldsymbol{x} - \boldsymbol{a}\| < \|\boldsymbol{y} - \boldsymbol{x}\| + d_A(\boldsymbol{x}) + \varepsilon.$$

- Since $\varepsilon > 0$ is arbitrary, $d_A(\boldsymbol{y}) \leq \|\boldsymbol{y} - \boldsymbol{x}\| + d_A(\boldsymbol{x})$.
- Interchanging $\boldsymbol{x}$ and $\boldsymbol{y}$ in this inequality, $d_A(\boldsymbol{x}) \leq \|\boldsymbol{x} - \boldsymbol{y}\| + d_A(\boldsymbol{y})$.
- Hence d_A satisfies the Lipschitz condition

$$|d_A(\boldsymbol{x}) - d_A(\boldsymbol{y})| \leq \|\boldsymbol{x} - \boldsymbol{y}\|.$$

- It follows that d_A is continuous on $\mathbb{R}^n$.

Remark

- In general, the inf in the definition of d_A cannot be replaced by min.
- To see this, suppose that A is the set $\mathbb{R}^n \setminus \{\mathbf{0}\}$.

  Then $d_A(\mathbf{0}) = 0$, but there is no $\boldsymbol{a} \in A$ such that $\|\mathbf{0} - \boldsymbol{a}\| = 0$.

Distance from Nonempty Closed Sets

Theorem

Let A be a non-empty closed set in $\mathbb{R}^n$ and let $\boldsymbol{x} \in \mathbb{R}^n$. Then there exists $\boldsymbol{a}_0 \in A$ such that $d_A(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{a}_0\|$.

- It follows easily from the definition of $d_A(\boldsymbol{x})$ that there exists a sequence $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots$ of points of A such that $\|\boldsymbol{x} - \boldsymbol{a}_k\| \to d_A(\boldsymbol{x})$ as $k \to \infty$. Since convergent sequences in $\mathbb{R}$ are bounded, there exists a real number r such that $\|\boldsymbol{x} - \boldsymbol{a}_k\| \le r$ for $k = 1, 2, \ldots$. We have

$$\|\boldsymbol{a}_k\| \le \|\boldsymbol{a}_k - \boldsymbol{x}\| + \|\boldsymbol{x}\| \le r + \|\boldsymbol{x}\|, \text{ for } k = 1, 2, \ldots.$$

  So the sequence $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots$ is bounded. Hence it contains some subsequence $\boldsymbol{a}_{i_1}, \ldots, \boldsymbol{a}_{i_k}, \ldots$ which converges to a point $\boldsymbol{a}_0$ of $\mathbb{R}^n$. Since A is closed, $\boldsymbol{a}_0 \in A$. Now $\|\boldsymbol{x} - \boldsymbol{a}_{i_k}\| \to \|\boldsymbol{x} - \boldsymbol{a}_0\|$ as $k \to \infty$. But we already know that $\|\boldsymbol{x} - \boldsymbol{a}_{i_k}\| \to d_A(\boldsymbol{x})$ as $k \to \infty$. The uniqueness of limits in $\mathbb{R}$ shows that $d_A(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{a}_0\|$.
- The point $\boldsymbol{a}_0$ is called a **nearest point** of A to $\boldsymbol{x}$.

Continuity and Compactness

Theorem

Let $f : A \to \mathbb{R}^n$ be a continuous mapping, where A is a non-empty compact set in $\mathbb{R}^n$. Then $f(A)$ is a compact set in $\mathbb{R}^n$.

- Let $f(\boldsymbol{a}_1),\ldots,f(\boldsymbol{a}_k),\ldots$ be a sequence of points of $f(A)$, where $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k,\ldots$ is a sequence of points of A.

  Since A is compact, there is a subsequence $\boldsymbol{a}_{i_1},\ldots,\boldsymbol{a}_{i_k},\ldots$ of $\boldsymbol{a}_1,\ldots,\boldsymbol{a}_k,\ldots$ which converges to some point $\boldsymbol{a}$ of A.

  By the continuity of f, the subsequence $f(\boldsymbol{a}_{i_1}),\ldots,f(\boldsymbol{a}_{i_k}),\ldots$ of $f(\boldsymbol{a}_1),\ldots,f(\boldsymbol{a}_k),\ldots$ converges to the point $f(\boldsymbol{a})$ of $f(A)$.

  Thus $f(A)$ is compact.

Attainability of Sup and Inf

- Recall from elementary analysis that a continuous function $f : [a, b] \to \mathbb{R}$ is bounded and attains its bounds.

Corollary

Let $f : A \to \mathbb{R}$ be a continuous mapping, where A is a non-empty compact set in $\mathbb{R}^n$. Then there exist $\boldsymbol{a}, \boldsymbol{b} \in A$ such that

$$f(\boldsymbol{a}) = \inf\{f(\boldsymbol{x}) : \boldsymbol{x} \in A\} \quad \text{and} \quad f(\boldsymbol{b}) = \sup\{f(\boldsymbol{x}) : \boldsymbol{x} \in B\}.$$

- The theorem shows that the non-empty set $f(A) = \{f(\boldsymbol{x}) : \boldsymbol{x} \in A\}$ of real numbers is compact, and therefore closed and bounded. Thus $f(A)$ possesses both an infimum and supremum. Moreover, the infimum and supremum of $f(A)$ belong to $\mathrm{cl} f(A)$. Hence, since $f(A)$ is closed, they belong to $f(A)$. So there exist $\boldsymbol{a}, \boldsymbol{b} \in A$ such that $f(\boldsymbol{a}) = \inf f(A)$ and $f(\boldsymbol{b}) = \sup f(A)$.

Attainability of Infimum of Distance

Theorem

Let A and B be non-empty sets in $\mathbb{R}^n$ with A closed and B compact. Then there exist $\boldsymbol{a}_0 \in A$, $\boldsymbol{b}_0 \in B$ such that

$$\|\boldsymbol{a}_0 - \boldsymbol{b}_0\| = \inf\{\|\boldsymbol{a} - \boldsymbol{b}\| : \boldsymbol{a} \in A, \boldsymbol{b} \in B\}.$$

- The distance function d_A of A is continuous on $\mathbb{R}^n$. So, by restriction, it is continuous on B. By the corollary, applicable since B is compact, there exists $\boldsymbol{b}_0 \in B$ such that $d_A(\boldsymbol{b}_0) = \inf\{d_A(\boldsymbol{b}) : \boldsymbol{b} \in B\}$. By a previous theorem, applicable since A is closed, there exists $\boldsymbol{a}_0 \in A$ such that $d_A(\boldsymbol{b}_0) = \|\boldsymbol{b}_0 - \boldsymbol{a}_0\|$. For each $\boldsymbol{a} \in A$, $\boldsymbol{b} \in B$, we have

$$\|\boldsymbol{a} - \boldsymbol{b}\| \geq d_A(\boldsymbol{b}) \geq d_A(\boldsymbol{b}_0) = \|\boldsymbol{a}_0 - \boldsymbol{b}_0\|.$$

  Since $\boldsymbol{a}_0 \in A$, $\boldsymbol{b}_0 \in B$, $\|\boldsymbol{a}_0 - \boldsymbol{b}_0\| = \inf\{\|\boldsymbol{a} - \boldsymbol{b}\| : \boldsymbol{a} \in A, \boldsymbol{b} \in B\}$.
- We refer to $\boldsymbol{a}_0$ and $\boldsymbol{b}_0$ as **nearest points** of A and B.

Continuity and Positivity

- Recall that if a real function is both continuous and positive at some point, then it is positive at all points of its domain sufficiently close to that point.

Theorem

Let the mapping $f : A \to \mathbb{R}$ be both continuous and positive at some point $\boldsymbol{a}$ of a set A in $\mathbb{R}^n$. Then there exists an $r > 0$ such that $f(\boldsymbol{x}) > 0$ whenever $\boldsymbol{x} \in B(\boldsymbol{a}; r) \cap A$.

- Suppose that the stated conclusion does not hold. Then, for each $k = 1, 2, \ldots$ there exists $\boldsymbol{a}_k \in B(\boldsymbol{a}; \frac{1}{k}) \cap A$ such that $f(\boldsymbol{a}_k) \le 0$. Since f is continuous at $\boldsymbol{a}$ and $\boldsymbol{a}_k \to \boldsymbol{a}$ as $k \to \infty$, $f(\boldsymbol{a}_k) \to f(\boldsymbol{a})$ as $k \to \infty$. Because $f(\boldsymbol{a}_k) \le 0$ for $k = 1, 2, \ldots$, it follows that $f(\boldsymbol{a}) \le 0$. This contradiction establishes the theorem.

Continuity of Composition

- Recall that a continuous function of a continuous function is itself continuous.

Theorem

Let $f : A \to \mathbb{R}^m$ and $g : B \to \mathbb{R}^p$ be continuous mappings, where A and B are, respectively, non-empty sets in $\mathbb{R}^n$ and $\mathbb{R}^m$ such that $f(A) \subseteq B$. Then the composite mapping $g \circ f : A \to \mathbb{R}^p$ is continuous.

- Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots$ be a sequence of points of A that converges to a point $\boldsymbol{a}$ of A. Since f is continuous, the sequence of points $f(\boldsymbol{a}_1), \ldots, f(\boldsymbol{a}_k), \ldots$ of B converges to the point $f(\boldsymbol{a})$ of B. Since g is continuous, the sequence $g(f(\boldsymbol{a}_1)), \ldots, g(f(\boldsymbol{a}_k)), \ldots$ converges to $g(f(\boldsymbol{a}))$, i.e., the sequence $(g \circ f)(\boldsymbol{a}_1), \ldots, (g \circ f)(\boldsymbol{a}_k), \ldots$ converges to $(g \circ f)(\boldsymbol{a})$. This shows that $g \circ f$ is continuous.

Inverse Images of Open and of Closed Sets

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous mapping and let B be a closed (open) subset of $\mathbb{R}^m$. Then $f^{-1}(B)$ is closed (open).

- Suppose first that B is closed. Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots$ be a sequence of points of $f^{-1}(B)$ that converges to a point $\boldsymbol{a}$ of $\mathbb{R}^n$. The continuity of f shows that the sequence of points $f(\boldsymbol{a}_1), \ldots, f(\boldsymbol{a}_k), \ldots$ of B converges to the point $f(\boldsymbol{a})$ of $\mathbb{R}^m$. But B is closed. So $f(\boldsymbol{a}) \in B$, i.e., $\boldsymbol{a} \in f^{-1}(B)$. This shows that $f^{-1}(B)$ is closed.

  Suppose next that B is open. Then the complement B^c of B in $\mathbb{R}^m$ is closed. Hence, by what has just been proved, $f^{-1}(B^c)$ is closed in $\mathbb{R}^n$. Thus, the complement $f^{-1}(B) = (f^{-1}(B^c))^c$ in $\mathbb{R}^n$ is open.