Introduction to Convexity

George Voutsadakis\textsuperscript{1}

\textsuperscript{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 $\lambda$ is a scalar, then
  
  \[
  (x, y, z) + (u, v, w) = (x + u, y + v, z + w);
  \]
  
  \[
  \lambda(x, y, z) = (\lambda x, \lambda y, \lambda z).
  \]
- These equations can be extended in the natural way to define vector addition and scalar multiplication of real $n$-tuples.
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 $\mathbf{x} = (x_1, \ldots, x_n)$ 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 $\mathbf{x}$.

For $n = 1$, we identify the 1-tuple $\mathbf{x} = (x_1)$ 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 $\mathbf{x}, (x, y), (x, y, z)$ 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 $a, b, c, x, y, z$ 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 $\lambda$ is a scalar, then

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

- The vector $(0, \ldots, 0)$ of $\mathbb{R}^n$, all of whose coordinates are 0, is denoted by $0$ 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 $i$th position is denoted by $e_i$ 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 $(-1)x$ is written simply as $-x$.

- **Vector subtraction** is defined by the rule $x - y = x + (-1)y$.

- It is sometimes convenient to write $\frac{x}{\lambda}$ for $\frac{1}{\lambda}x$. 

George Voutsadakis (LSSU)
Convexity
July 2023
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 $x, y, z \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$, then the following relations hold:

1. $x + y = y + x$;
2. $x + (y + z) = (x + y) + z$;
3. $x + 0 = x$;
4. $x + (-x) = 0$;
5. $1 \cdot x = x$;
6. $\lambda (\mu x) = (\lambda \mu) x$;
7. $\lambda (x + y) = \lambda x + \lambda y$;
8. $(\lambda + \mu) x = \lambda x + \mu x$. 
Extending Operations on Sets

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

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

  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 $\lambda A$ 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 $\frac{1}{\lambda}A$. 

George Voutsadakis (LSSU)
Convexity
July 2023
8 / 162
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.
The set \( \{a\} + B \), where \( a \in \mathbb{R}^n \), is often written as \( a + B \) and is called a translate of \( B \) or, more precisely, the translate of \( B \) by \( a \).

It is an easy exercise in set theory to show that

\[
A + B = \bigcup (a + B : a \in A),
\]

i.e., \( A + B \) is the union of all translates of \( B \) by vectors \( a \) 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 $a + B$ is the circular disc with center $a$ and radius 1; $A + B$ is the union of all such discs for $a \in A$.
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 $a$ and $b$ in $\mathbb{R}^n$.
  Then $A + A = \{2a, 2b, a + b\}$, whereas $2A = \{2a, 2b\}$.
Properties of Set Operations

Properties (i)-(viii) above do, however, partially generalize to give the following easily verified results:

1. \( A + B = B + A \);
2. \( A + (B + C) = (A + B) + C \);
3. \( A + 0 = A \);
4. \( 0 \in A + (\neg A) \) when \( A \neq \emptyset \);
5. \( 1A = A \);
6. \( \lambda(\mu A) = (\lambda \mu) A \);
7. \( \lambda(A + B) = \lambda A + \lambda B \);
8. \( (\lambda + \mu)A \subseteq \lambda A + \mu A \).
Subsection 2

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

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

$$x = b + \lambda(a - b) = \lambda a + (1 - \lambda)b.$$ 

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

We note that the subset

$$\{\lambda a + (1 - \lambda)b : 0 \leq \lambda \leq 1\} = \{\lambda a + \mu b : \lambda, \mu \geq 0, \lambda + \mu = 1\}$$

of the line through $a$ and $b$ is the line segment joining $a$ and $b$. 

George Voutsadakis (LSSU)
The line through distinct points \(a\) and \(b\) of \(\mathbb{R}^n\) is the set 
\[ \{ \lambda a + \mu b : \lambda + \mu = 1 \} \].

Clearly this set contains both \(a\) and \(b\), 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 \(\lambda a + \mu b \in A\) whenever \(a, b \in A\) 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 \(\mathbb{R}^3\).
Let $A$ be a flat in $\mathbb{R}^n$ which contains the origin.

Suppose that $a, b \in A$ and $\lambda \in \mathbb{R}$.

Since $A$ is a flat and $a, 0 \in A$, $\lambda a + (1 - \lambda)0 \in A$, i.e., $\lambda a \in A$. Thus $A$ is closed under scalar multiplication.

Since $A$ is a flat and $a, b \in A$, $\frac{1}{2}a + \frac{1}{2}b \in A$. But $A$ is closed under scalar multiplication. So $2(\frac{1}{2}a + \frac{1}{2}b) \in A$, i.e., $a + b \in A$. 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 $\mathbb{R}^n$. 
The Euclidean Space $\mathbb{R}^n$ Flats

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

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

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

Since $A - a$ contains the origin, it must be a subspace of $\mathbb{R}^n$. Hence the non-empty flat $A$ is the translate of the subspace $A - a$ of $\mathbb{R}^n$ by the vector $a$. 

George Voutsadakis (LSSU) Convexity July 2023 18 / 162
Suppose next that $A$ is a subspace of $\mathbb{R}^n$ and that $u \in \mathbb{R}^n$. We show that $A + u$ is a flat. Let $x, y \in A + u$ and $\lambda + \mu = 1$. Then there exist $a, b \in A$ such that $x = a + u$, $y = b + u$. So

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

since $\lambda a + \mu b \in A$, as $A$ is a subspace of $\mathbb{R}^n$. This shows that $A + u$ 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 \( b \in \mathbb{R}^n \) such that \( C = B + b \). Since 0 lies in \( C \), it follows that \(-b\), and hence \( b \), lies in \( B \). Thus \( C = B + b \subseteq B \). By symmetry, \( B \subseteq C \). Hence \( B = C \), and \( A \) is the translate of precisely one subspace of \( \mathbb{R}^n \).
The Euclidean Space $\mathbb{R}^n$

**Flats**

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

Flats

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 $a, b \in \bigcap (A_i : i \in I)$ and $\lambda + \mu = 1$.

Then $a, b \in A_i$. As $A_i$ is a flat, $\lambda a + \mu b \in A_i$, for each $i \in I$.

Thus, $\lambda a + \mu b \in \bigcap (A_i : i \in I)$.

This shows that the intersection is a flat.
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 \( \text{aff} A \subseteq \text{aff} B \).
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 $\lambda a + \mu b \in A$ whenever $a, b \in A$ 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 $a_1, \ldots, a_m$ be points of a flat $A$ in $\mathbb{R}^n$. Let $\lambda_1 + \cdots + \lambda_m = 1$. Then $\lambda_1 a_1 + \cdots + \lambda_m a_m \in A$.

Let $a \in A$. Then the points $a_1 - a, \ldots, a_m - a$ lie in the subspace $A - a$ of $\mathbb{R}^n$, whence so too does the point

$$\lambda_1 (a_1 - a) + \cdots + \lambda_m (a_m - a) = \lambda_1 a_1 + \cdots + \lambda_m a_m - a.$$

Hence $\lambda_1 a_1 + \cdots + \lambda_m a_m \in A$. 

George Voutsadakis (LSSU) Convexity July 2023 25 / 162
Affine Combinations and the Affine Hull

- A point $x$ is said to be an **affine combination** of points $a_1, \ldots, a_m$ in $\mathbb{R}^n$ if there exist scalars $\lambda_1, \ldots, \lambda_m$ with $\lambda_1 + \cdots + \lambda_m = 1$ such that
  
  $$x = \lambda_1 a_1 + \cdots + \lambda_m a_m.$$ 

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

We next show that $B$ is a flat. If $x, y \in B$, then $x = \lambda_1 a_1 + \cdots + \lambda_m a_m$, $y = \mu_1 b_1 + \cdots + \mu_p b_p$, for some $a_1, \ldots, a_m, b_1, \ldots, b_p \in A$, 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

$$\lambda x + \mu y = \lambda \lambda_1 a_1 + \cdots + \lambda \lambda_m a_m + \mu \mu_1 b_1 + \cdots + \mu \mu_p b_p$$

and

$$\lambda \lambda_1 + \cdots + \lambda \lambda_m + \mu \mu_1 + \cdots + \mu \mu_p$$

$$= \lambda (\lambda_1 + \cdots + \lambda m) + \mu (\mu_1 + \cdots + \mu p)$$

$$= \lambda + \mu = 1.$$

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

Corollary

Let $a_1, \ldots, a_m \in \mathbb{R}^n$. Then

$$\text{aff}\{a_1, \ldots, a_m\} = \{\lambda_1 a_1 + \cdots + \lambda_m a_m : \lambda_1 + \cdots + \lambda_m = 1\}.$$ 

Example: Each point $x = (x_1, \ldots, x_n)$ of $\mathbb{R}^n$ can be expressed as an affine combination of the zero vector $0$ and the elementary vectors $e_1, \ldots, e_n$ as follows:

$$x = (1 - x_1 - \cdots - x_n)0 + x_1 e_1 + \cdots + x_n e_n.$$ 

The corollary now shows that $\text{aff}\{0, e_1, \ldots, e_n\} = \mathbb{R}^n$. 

George Voutsadakis (LSSU)
Linear Hull

Let $A$ be a non-empty set in $\mathbb{R}^n$.

We recall that a point of the form $\lambda_1 \mathbf{a}_1 + \cdots + \lambda_m \mathbf{a}_m$, where $\mathbf{a}_1, \ldots, \mathbf{a}_m \in A$ 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 $A \cup \{0\}$, it follows that $\text{aff}(A \cup \{0\}) \subseteq \text{lin} A$.

On the other hand, $\text{aff}(A \cup \{0\})$ is a subspace of $\mathbb{R}^n$ containing $A$, so $\text{lin} A \subseteq \text{aff}(A \cup \{0\})$.

We conclude that $\text{lin} A = \text{aff}(A \cup \{0\})$.

We define $\text{lin} \emptyset = \{0\}$.

This ensures that $\text{lin} \emptyset$ is the smallest subspace of $\mathbb{R}^n$ which contains $\emptyset$, and that $\text{lin} \emptyset = \text{aff}(\emptyset \cup \{0\})$. 
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 $\alpha$ be a scalar. Then $A + B$ and $\alpha A$ 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 $\alpha A$ 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

$$(\lambda_1 + \cdots + \lambda_m)A \subseteq \lambda_1 A + \cdots + \lambda_m A$$

$$= \lambda(\frac{\lambda_1}{\lambda} A + \cdots + \frac{\lambda_m}{\lambda} A)$$

$$\subseteq \lambda A$$

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

Thus $(\lambda_1 + \cdots + \lambda_m)A = \lambda_1 A + \cdots + \lambda_m A$. 
Subsection 3

Dimension
The set $A$ in $\mathbb{R}^n$ is said to be **affinely dependent** if there exists $a \in A$ such that $a \in \text{aff}(A\{a\})$.

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.
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 $a_1, \ldots, a_m$ of $A$ and scalars $\lambda_1, \ldots, \lambda_m$, not all zero, such that

$$\lambda_1 a_1 + \cdots + \lambda_m a_m = 0 \quad \text{and} \quad \lambda_1 + \cdots + \lambda_m = 0.$$

Suppose that $A$ is affinely dependent. Then there exists $a_1 \in A$ such that $a_1 \in \text{aff}(A \setminus \{a_1\})$. By a previous theorem, there exist (distinct) points $a_2, \ldots, a_m$ of $A \setminus \{a_1\}$ and scalars $\mu_2, \ldots, \mu_m$, such that

$$a_1 = \mu_2 a_2 + \cdots + \mu_m a_m \quad \text{and} \quad \mu_2 + \cdots + \mu_m = 1.$$  

Write $\lambda_1 = -1$, $\lambda_2 = \mu_2$, $\ldots$, $\lambda_m = \mu_m$. Then $\lambda_1, \ldots, \lambda_m$ are not all zero and satisfy the conclusion.
Suppose next that there exist distinct points \( a_1, \ldots, a_m \) of \( A \), and scalars \( \lambda_1, \ldots, \lambda_m \), not all zero, which satisfy the hypothesis. Suppose that \( \lambda_1 \neq 0 \). Then

\[
a_1 = -\frac{1}{\lambda_1} (\lambda_2 a_2 + \cdots + \lambda_m a_m)
\]

and

\[
-\frac{1}{\lambda_1} (\lambda_2 + \cdots + \lambda_m) = 1,
\]

which shows that \( a_1 \) is an affine combination of \( a_2, \ldots, a_m \). Hence \( a_1 \in \text{aff}\{a_2, \ldots, a_m\} \subseteq \text{aff}(A\setminus\{a_1\}) \). So \( A \) is affinely dependent.

**Corollary**

A subset \( \{a_1, \ldots, a_m\} \) of \( \mathbb{R}^n \) is affinely dependent if and only if there exist scalars \( \lambda_1, \ldots, \lambda_m \), not all zero, such that

\[
\lambda_1 a_1 + \cdots + \lambda_m a_m = 0 \quad \text{and} \quad \lambda_1 + \cdots + \lambda_m = 0.
\]
Uniqueness

**Corollary**

Let \( \{a_1, \ldots, a_m\} \) be an affinely independent set in \( \mathbb{R}^n \). Then each point of \( \text{aff}\{a_1, \ldots, a_m\} \) can be expressed uniquely in the form

\[
\lambda_1 a_1 + \cdots + \lambda_m a_m, \quad \text{where} \quad \lambda_1 + \cdots + \lambda_m = 1.
\]

A previous corollary shows that each point of \( \text{aff}\{a_1, \ldots, a_m\} \) can be expressed in the desired form.

To establish the uniqueness, suppose that

\[
\lambda_1 a_1 + \cdots + \lambda_m a_m = \mu_1 a_1 + \cdots + \mu_m a_m, \quad \lambda_1 + \cdots + \lambda_m = \mu_1 + \cdots + \mu_m = 1.
\]
Then

\[(\lambda_1 - \mu_1)a_1 + \cdots + (\lambda_m - \mu_m)a_m = 0\]

with \((\lambda_1 - \mu_1) + \cdots + (\lambda_m - \mu_m) = 0\).

Since \(\{a_1, \ldots, a_m\}\) is affinely independent, the preceding corollary shows that the scalars \(\lambda_1 - \mu_1, \ldots, \lambda_m - \mu_m\) must be zero.

Thus \(\lambda_1 = \mu_1, \ldots, \lambda_m = \mu_m\), and the uniqueness is established.
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 \( \{a_1, \ldots, a_m\} \neq \) in \( \mathbb{R}^n \), where \( m > n + 1 \), is affinely dependent. Let \( \{a_1, \ldots, a_m\} \neq \) be a set in \( \mathbb{R}^n \), where \( m > n + 1 \). Then the system of the \( n + 1 \) linear simultaneous equations

\[
\lambda_1 a_1 + \cdots + \lambda_m a_m = 0, \quad \lambda_1 + \cdots + \lambda_m = 0,
\]

in the \( m \) unknowns \( \lambda_1, \ldots, \lambda_m \) is homogeneous. Since \( m > n + 1 \), it has a non-trivial solution. Hence, \( \{a_1, \ldots, a_m\} \neq \), 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 $a$ is a point of $\mathbb{R}^n$ not lying in $\text{aff} A$. Then the set $A \cup \{a\}$ is affinely independent.

We argue by contradiction. Suppose that $A \cup \{a\}$ is affinely dependent. Then there exist distinct points $a_1, \ldots, a_m$ of $A$ and scalars $\lambda, \lambda_1, \ldots, \lambda_m$, not all zero, such that $\lambda a + \lambda_1 a_1 + \cdots + \lambda_m a_m = 0$ and $\lambda + \lambda_1 + \cdots + \lambda_m = 0$. The scalar $\lambda$ cannot be zero, for then $A$ is affinely dependent. Thus the equation can be used to express $a$ as an affine combination of $a_1, \ldots, a_m$. So $a \in \text{aff}\{a_1, \ldots, a_m\}$. This, however, contradicts the hypothesis that $a \not\in \text{aff} A$. Hence $A \cup \{a\}$ is affinely independent.
Example

In $\mathbb{R}^n$ the set $\{0, e_1, \ldots, e_n\}$ is affinely independent.

To see this, suppose that the scalars $\lambda, \lambda_1, \ldots, \lambda_n$ satisfy

$$\lambda 0 + \lambda_1 e_1 + \cdots + \lambda_n e_n = 0 \quad \text{and} \quad \lambda + \lambda_1 + \cdots + \lambda_n = 0.$$  

The first of these equations shows that $\lambda_1, \ldots, \lambda_n$ are all zero. Hence $\lambda$ must also be zero from the second equation. The corollary now shows that the set $\{0, e_1, \ldots, e_n\}$ 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 $\{0, e_1, e_2, e_3\}$.
- 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 $\{a_0, a_1, \ldots, a_m\}$ be an affinely independent subset of $A$. Since $A$ is a flat, $\text{aff}\{a_0, a_1, \ldots, a_m\} \subseteq A$. Now $A \subseteq \text{aff}\{a_0, a_1, \ldots, a_m\}$, for otherwise there would exist some point $a$ of $A$ not lying in $\text{aff}\{a_0, a_1, \ldots, a_m\}$ and, by a previous corollary, $\{a, a_0, a_1, \ldots, a_m\}$ would be an affinely independent subset of $A$ having $m+2$ elements, so contradicting the definition of $m$. Hence $A = \text{aff}\{a_0, a_1, \ldots, a_m\}$.
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 $\{a_1 - a_0, \ldots, a_m - a_0\}$ of $B$ is a basis for $B$. Let $b \in B$. Then $b = x - a_0$ for some $x \in A$. Thus, there exist scalars $\lambda_0, \lambda_1, \ldots, \lambda_m$ such that

$$x = \lambda_0 a_0 + \lambda_1 a_1 + \cdots + \lambda_m a_m$$

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

$$b = x - a_0 = \lambda_1(a_1 - a_0) + \cdots + \lambda_m(a_m - a_0).$$

This shows that $\{a_1 - a_0, \ldots, a_m - a_0\}$ spans $B$. 

George Voutsadakis (LSSU)
Convexity
July 2023 44 / 162
Finally, suppose that $\mu_1, \ldots, \mu_m$ satisfy

$$\mu_1(a_1 - a_0) + \cdots + \mu_m(a_m - a_0) = 0.$$ 

Then

$$-(\mu_1 + \cdots + \mu_m)a_0 + \mu_1 a_1 + \cdots + \mu_m a_m = 0,$$

$$-(\mu_1 + \cdots + \mu_m) + \mu_1 + \cdots + \mu_m = 0.$$ 

But $\{a_0, a_1, \ldots, a_m\} \neq \emptyset$ is affinely independent. So all of $\mu_1, \ldots, \mu_m$ are zero. Thus $\{a_1 - a_0, \ldots, a_m - a_0\}$ is linearly independent. We conclude that $\{a_1 - a_0, \ldots, a_m - a_0\}$ is a basis for $B$.

Hence, $m$ is the dimension of $B$, and so is uniquely determined by $A$. 
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 $\{0, e_1, \ldots, e_n\}$, whence $\mathbb{R}^n$ has dimension $n$. 
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 $\mathbb{R}^n$. 
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 $c \in A \cap B$. Then $A - c$ and $B - c$ are subspaces of $\mathbb{R}^n$. So, by the dimension theorem of elementary linear algebra,

$$\dim((A - c) + (B - c)) + \dim((A - c) \cap (B - c)) = \dim(A - c) + \dim(B - c),$$

that is,

$$\dim(A + B - 2c) + \dim((A \cap B) - c) = \dim(A - c) + \dim(B - c).$$

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:** \( \{0, e_1, \ldots, e_n\} \) 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 $\text{aff}A$ that lies in $A$ and contains $B$.

Consider the non-empty family $\mathcal{F}$ 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 $\mathcal{F}$ that is not properly contained in any other member of $\mathcal{F}$. Since $C$ is a subset of $A$, we have $\text{aff}C \subseteq \text{aff}A$. We claim that $\text{aff}C = \text{aff}A$.

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

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

- Let $\{a_0, \ldots, a_r\}$ be an affine basis for a non-empty $r$-flat $A$ in $\mathbb{R}^n$.
- Then, by a previous corollary, each point $x$ of $A$ can be expressed uniquely in the form

$$x = \lambda_0 a_0 + \cdots + \lambda_r a_r, \quad \text{where} \quad \lambda_0 + \cdots + \lambda_r = 1.$$

- The scalars $\lambda_0, \ldots, \lambda_r$ are called the barycentric coordinates of $x$ relative to (the ordered affine basis) $a_0, \ldots, a_r$.
- A previous example shows that the barycentric coordinates of a point $x = (x_1, \ldots, x_n)$ of $\mathbb{R}^n$ relative to $0, e_1, \ldots, e_n$ are $1 - x_1 - \cdots - x_n$, $x_1, \ldots, x_n$. 
Scalors of Point Relative to Affine Basis

Theorem

Let \( \{a_0, \ldots, a_r\} \) 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 \( x = (x_1, \ldots, x_n) \) of \( A \) relative to \( a_0, \ldots, a_r \). 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, \( \{a_0, \ldots, a_r\} \) to an affine basis \( \{a_0, \ldots, a_n\} \) for \( \mathbb{R}^n \). Each point \( x = (x_1, \ldots, x_n) \) of \( \mathbb{R}^n \) can be written uniquely in the form

\[
x = \lambda_0 a_0 + \cdots + \lambda_n a_n, \quad \text{where} \quad \lambda_0 + \cdots + \lambda_n = 1.
\]

In particular, each of the points \( 0, e_1, \ldots, e_n \) can be so expressed. Write \( e_0 = 0 \).
Then there are scalars $b_{ij}$ ($i = 0, \ldots, n; j = 0, \ldots, n$) such that, for $i = 0, \ldots, n$,

$$e_i = b_{0i}a_0 + \cdots + b_{ni}a_n \quad \text{and} \quad b_{0i} + \cdots + b_{ni} = 1.$$ 

Write $x = (1 - x_1 - \cdots - x_n)e_0 + x_1e_1 + \cdots + x_ne_n$.

Then $x = \mu_0a_0 + \cdots + \mu_na_n$, 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 $x$ 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 $x \in A$ 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 $a, b$ be distinct points of $L$, and let $c, d$ be distinct points of $M$. Since $\{a, b, c\}$ is affinely independent, it will form an affine basis for $A$. Thus $d = \alpha a + \beta b + \gamma c$ for some $\alpha, \beta, \gamma \in \mathbb{R}$ with $\alpha + \beta + \gamma = 1$. A typical point on $M$, the line joining $c$ and $d$, has the form

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

for some $\theta \in \mathbb{R}$. Since the latter point does not lie on $L$ for any $\theta$, we must have $\gamma = 1$ and $d = \alpha(a - b) + c$. Hence $d - (c - a) = \alpha(a - b) + c - c + a = (\alpha + 1)a - \alpha b \in L$. Thus, $M - (c - a) \subseteq L$. Since $M - (c - a)$ is a line, we must have $M - (c - a) = L$. Thus $L$ and $M$ are parallel.
Consider the following system of $m$ linear equations in $n$ real variables $x_1, \ldots, x_n$:

$$
\begin{align*}
    a_{11}x_1 + \cdots + a_{1n}x_n &= a_{10} \\
    \vdots \\
    a_{m1}x_1 + \cdots + a_{mn}x_n &= a_{m0}
\end{align*}
$$

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.
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

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

Let $H = \{(x_1, \ldots, x_n) : c_0 + c_1x_1 + \cdots + c_nx_n = 0\}$, 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 $u = (u_1, \ldots, u_n)$, $v = (v_1, \ldots, v_n)$ lie in $H$ and let $\lambda + \mu = 1$. Then

$$c_0 + c_1(\lambda u_1 + \mu v_1) + \cdots + c_n(\lambda u_n + \mu v_n)$$

$$= \lambda(c_0 + c_1u_1 + \cdots + c_nu_n) + \mu(c_0 + c_1v_1 + \cdots + c_nv_n)$$

$$= \lambda 0 + \mu 0 = 0.$$

Thus $\lambda u + \mu v \in H$ and $H$ is a flat.
Define points $a_1, \ldots, a_n$ of $H$ by the equations $a_1 = (-\frac{c_0}{c_1}, 0, 0, \ldots, 0)$ and $a_2 = (-\frac{c_0 + c_2}{c_1}, 1, 0, \ldots, 0), \ldots, a_n = (-\frac{c_0 + c_n}{c_1}, 0, 0, \ldots, 1)$. Since $H$ is a flat, $\text{aff}\{a_1, \ldots, a_n\} \subseteq H$.

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

$$x = (x_1, \ldots, x_n) = (1 - x_2 - \cdots - x_n)a_1 + x_2a_2 + \cdots + x_na_n$$

express $x$ as an affine combination of $a_1, \ldots, a_n$. So $x \in \text{aff}\{a_1, \ldots, a_n\}$

Hence, $H \subseteq \text{aff}\{a_1, \ldots, a_n\}$ and, therefore, $H = \text{aff}\{a_1, \ldots, a_n\}$.
To show that the set \( \{a_1, \ldots, a_n\} \) is affinely independent, suppose that \( \lambda_1, \ldots, \lambda_n \) satisfy

\[
\lambda_1 a_1 + \cdots + \lambda_n a_n = 0 \quad \text{and} \quad \lambda_1 + \cdots + \lambda_n = 0.
\]

Comparing the \( i \)th 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 \( \lambda_1 \), from the second equation.

So, \( \{a_1, \ldots, a_n\} \) is affinely independent.

But \( H = \text{aff}\{a_1, \ldots, a_n\} \), and so \( H \) is an \((n-1)\)-dimensional flat, i.e., \( H \) is a hyperplane.
Conversely, suppose that $H$ is a hyperplane in $\mathbb{R}^n$. Let $\{b_1, \ldots, b_n\}$ be an affine basis for $H$. Extend this to an affine basis $\{b_0, b_1, \ldots, b_n\}$ for $\mathbb{R}^n$. Then each $x = (x_1, \ldots, x_n)$ in $\mathbb{R}^n$ can be written uniquely in the form

$$x = \lambda_0 b_0 + \lambda_1 b_1 + \cdots + \lambda_n b_n,$$

where $\lambda_0 + \lambda_1 + \cdots + \lambda_n = 1$.

Thus $\lambda_0, \lambda_1, \ldots, \lambda_n$ are the barycentric coordinates of $x$ relative to the (ordered) affine basis $b_0, b_1, \ldots, b_n$. 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 $x \in H$ 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.
The Euclidean Space $\mathbb{R}^n$  

Hyperplanes

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 $x_1 = 0, \ldots, x_n = 0, x_1 + \cdots + x_n = 1$.
- 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 $\{a_0, \ldots, a_r\}$ be an affine basis for $A$. Extend this to an affine basis $\{a_0, \ldots, a_n\}$ for $\mathbb{R}^n$. Then each $x$ in $\mathbb{R}^n$ can be expressed uniquely in the form

$$x = \lambda_0 a_0 + \cdots + \lambda_n a_n, \quad \text{where} \quad \lambda_0 + \cdots + \lambda_n = 1.$$
Now $A$ is the set in $\mathbb{R}^n$ consisting precisely of those $x$’s whose barycentric coordinates $\lambda_{r+1}, ..., \lambda_n$ are all zero. But each of the sets $\{x : \lambda_i = 0\}$ is the hyperplane

$$\text{aff}\{a_0, ..., a_{i-1}, a_{i+1}, ..., a_n\}.$$ 

It now follows that $A$ is the intersection of the $n-r$ hyperplanes with equations $\lambda_{r+1} = 0, ..., \lambda_n = 0$. 
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 $a_1, \ldots, a_n$ be the points of $H$ specified as in the first part of the proof of the preceding theorem.
Substituting the coordinates of the $a_i$ 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 $d_0 = \theta c_0, d_1 = \theta c_1, \ldots, d_n = \theta c_n$.

Thus the hyperplane $H$ determines the scalars $c_0, c_1, \ldots, c_n$ to within a common non-zero scalar multiple.
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 $c_0 + c_1 x_1 + \cdots + c_n x_n = 0$ divides $\mathbb{R}^n$ into the two halfspaces determined by the inequalities

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

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

$$\{(x_1, \ldots, x_n) : c_0 + c_1 x_1 + \cdots + c_n x_n \leq 0\} \quad \text{or} \quad \{(x_1, \ldots, x_n) : c_0 + c_1 x_1 + \cdots + c_n x_n \geq 0\}$$

is called a closed halfspace in $\mathbb{R}^n$. 
Halfspaces Determined by a Hyperplane

- A set of either of the forms
  \[
  \{(x_1, \ldots, x_n) : c_0 + c_1 x_1 + \cdots + c_n x_n < 0\} \quad \text{or} \quad \{(x_1, \ldots, x_n) : c_0 + c_1 x_1 + \cdots + c_n x_n > 0\}
  \]

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 \(c_0 + c_1 x_1 + \cdots + c_n x_n = 0\), 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 \( c_0 + c_1 x_1 + \cdots + c_n x_n = 0 \), and that the points \( a = (a_1, \ldots, a_n) \) and \( b = (b_1, \ldots, b_n) \) lie in opposite halfspaces determined by \( H \).
- Omitting the trivial case when either of \( a \) or \( b \) lies on \( H \),

\[
c_0 + c_1 a_1 + \cdots + c_n a_n = \alpha < 0 \quad \text{and} \quad c_0 + c_1 b_1 + \cdots + c_n b_n = \beta > 0.
\]
- The points on the line \( L \) through \( a \) and \( b \) 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 \( \lambda \) 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 \( \lambda \) satisfies \( 0 \leq \lambda \leq 1 \). So the portion of \( L \) lying, between \( a \) and \( b \), the so-called line segment joining \( a \) and \( b \), 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 $\theta$ such that $c'_1 = \theta c_1$, $\ldots$, $c'_n = \theta c_n$.

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

$$c'_0 + c'_1(x_1 + a_1) + \cdots + c'_n(x_n + a_n) = c'_0 + c'_1 a_1 + \cdots + c'_n a_n + c'_1 x_1 + \cdots + c'_n x_n = 0.$$ 

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

George Voutsadakis (LSSU)
Convexity
July 2023
Characterization of Parallel Hyperplanes

Suppose next that $c'_1 = \theta c_1, \ldots, c'_n = \theta c_n$, where $\theta$ is a (non-zero) scalar. Then, for $d_0 = \frac{c'_0}{\theta}$, $H'$ is represented by the equation

$$d_0 + c_1x_1 + \cdots + c_nx_n = 0.$$ 

Let $b = (b_1, \ldots, b_n)$ satisfy

$$c_1b_1 + \cdots + c_nb_n = c_0 - d_0.$$ 

Then $H'$ also has the equation

$$c_0 + c_1(x_1 - b_1) + \cdots + c_n(x_n - b_n) = 0.$$ 

Thus $x = (x_1, \ldots, x_n) \in H'$ if and only if $x - b \in H$. Hence $H' = H + b$. This shows that $H$ and $H'$ are parallel.
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 $\theta$ is a non-zero scalar. If $c'_0 = \theta c_0$, then $H$ and $H'$ are identical. Otherwise they are disjoint.
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 $\theta$ 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 $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called an affine transformation if $T(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda T(\mathbf{x}) + \mu T(\mathbf{y})$ whenever $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda + \mu = 1$.

A simple example of an affine transformation is the mapping $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ 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 $\mathbf{q} \in \mathbb{R}^n$, the mapping $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ defined by the equation $T(\mathbf{x}) = \mathbf{x} + \mathbf{q}$ is an affine transformation called the translation of $\mathbb{R}^n$ through $\mathbf{q}$. 
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.
The Euclidean Space \( \mathbb{R}^n \)  

Affine Transformations

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 \( T(0) = 0 \).

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

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

\[
T(\lambda x) = T(\lambda x + (1 - \lambda)0) = \lambda T(x) + (1 - \lambda)T(0) = \lambda T(x).
\]

Using this last result, we deduce that

\[
T(x + y) = T(2\left(\frac{1}{2}x + \frac{1}{2}y\right)) = 2T\left(\frac{1}{2}x + \frac{1}{2}y\right) = 2\left(\frac{1}{2}T(x) + \frac{1}{2}T(y)\right) = T(x) + T(y).
\]

Thus \( T \) is linear.
The Euclidean Space $\mathbb{R}^n$  Affine Transformations

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

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

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

- The affine transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ determines the matrices $Q$ and $q$ uniquely:
  - The $j$th column of $Q$ must be $T(e_j) - T(0)$;
  - $q$ must be $T(0)$.

- 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, $a_1, \ldots, a_r \in \mathbb{R}^n$ and $\lambda_1 + \cdots + \lambda_r = 1$, then

$$T(\lambda_1 a_1 + \cdots + \lambda_r a_r) = \lambda_1 T(a_1) + \cdots + \lambda_r T(a_r).$$
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 \( x \) lies in \( T(\text{aff}\, A) \) if and only if there exist \( a_1, \ldots, a_r \in A \) and \( \lambda_1, \ldots, \lambda_r \) with \( \lambda_1 + \cdots + \lambda_r = 1 \) such that

\[
x = T(\lambda_1 a_1 + \cdots + \lambda_r a_r) = \lambda_1 T(a_1) + \cdots + \lambda_r T(a_r),
\]

that is, if and only if \( x \in \text{aff}\, T(A) \). 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.
Consider an affine transformation \( T : \mathbb{R}^n \rightarrow \mathbb{R}^n \) of \( \mathbb{R}^n \) into itself.

By the theorem, there exist a real \( n \times n \) matrix \( \mathbf{Q} \) and a real \( n \times 1 \) matrix \( \mathbf{q} \) such that \( T(x) = \mathbf{Q}x + \mathbf{q} \).

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

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 $T^{-1}(y) = Q^{-1}y - Q^{-1}q$ for $y$ in $\mathbb{R}^n$.

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

Let \( \{a_0, \ldots, a_r\} \neq \emptyset \) and \( \{b_0, \ldots, b_r\} \neq \emptyset \) 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 \( T(a_i) = b_i \), for \( i = 0, \ldots, r \).

Extend the sets \( \{a_0, \ldots, a_r\} \neq \emptyset \) and \( \{b_0, \ldots, b_r\} \neq \emptyset \), respectively, to affine bases \( \{a_0, \ldots, a_n\} \) and \( \{b_0, \ldots, b_n\} \) for \( \mathbb{R}^n \). Then each \( x \) in \( \mathbb{R}^n \) can be written uniquely in the form

\[
x = \lambda_0 a_0 + \cdots + \lambda_n a_n, \quad \lambda_0 + \cdots + \lambda_n = 1.
\]

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

\[
T(x) = \lambda_0 b_0 + \cdots + \lambda_n b_n.
\]

It is routine to verify that \( T \) is a bijective affine transformation. Hence \( T \) is a non-singular affine transformation such that \( T(a_i) = b_i \), for \( i = 0, \ldots, r \).
Corollary

Let $A$ and $B$ be flats in $\mathbb{R}^n$ of the same dimension. Then there is a non-singular affine transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 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 $\{a_0, \ldots, a_r\} \neq \emptyset$ and $\{b_0, \ldots, b_r\} \neq \emptyset$, respectively.
- Let $T$ be as in the theorem. Then, by a previous corollary,

$$T(A) = T(\text{aff}\{a_0, \ldots, a_r\}) = \text{aff}\{b_0, \ldots, b_r\} = B.$$
Suppose that $B$ is an $r$-dimensional flat ($r \geq 1$) in $\mathbb{R}^n$ and that $A = \text{aff}\{0, e_1, \ldots, e_r\}$.

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 $x_{r+1} = 0, \ldots, x_n = 0$.

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

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

The following properties of the inner product are immediate consequences of its definition.

For $x, y, z \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$:

(i) $x \cdot x \geq 0$, and $x \cdot x = 0$ if and only if $x = 0$;

(ii) $x \cdot y = y \cdot x$;

(iii) $(\lambda x + \mu y) \cdot z = \lambda (x \cdot z) + \mu (y \cdot z)$. 

The Euclidean Space $\mathbb{R}^n$ Length, Distance and Angle
The Euclidean Space $\mathbb{R}^n$

Length, Distance and Angle

The Norm and the Distance

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

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

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

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

  i.e., the length of the vector $x - y$, or $y - x$. 

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

Length, Distance and Angle

Properties of the Norm

- The norm of the zero vector is 0.
- The norm of each elementary vector $e_i$ 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 $x, y \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$:

(i) $\|x\| \geq 0$, and $\|x\| = 0$ if and only if $x = 0$;
(ii) $\|\lambda x\| = |\lambda| \|x\|$;
(iii) $\|\lambda x + \mu y\|^2 = \lambda^2 \|x\|^2 + 2\lambda \mu x \cdot y + \mu^2 \|y\|^2$. 

George Voutsadakis (LSSU)  
Convexity  
July 2023  
89 / 162
Inequalities Involving the Norm

**Theorem**

Let \( x, y \in \mathbb{R}^n \). Then:

(i) \( |x \cdot y| \leq \| x \| \| y \| \) (Cauchy-Schwarz Inequality);

(ii) \( \| x + y \| \leq \| x \| + \| y \| \) (Triangle Inequality);

(iii) \( \| x \| - \| y \| \leq \| x - y \| \);

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

We only prove (iv), since (i), (ii), and (iii) are standard results.

Let \( \alpha > 0 \) be such that \( \| x + \lambda y \| \geq \| x \| \) whenever \( 0 < \lambda < \alpha \). Then, whenever \( 0 < \lambda < \alpha \),

\[
\| x \|^2 \leq \| x + \lambda y \|^2 = \| x \|^2 + 2\lambda x \cdot y + \lambda^2 \| y \|^2.
\]

Hence \( x \cdot y + \frac{1}{2} \lambda \| y \|^2 \geq 0 \). Letting \( \lambda \to 0^+ \) in the last inequality, we deduce that \( x \cdot y \geq 0 \).
The Cauchy-Schwarz inequality allows us to introduce the concept of angle into $\mathbb{R}^n$.

The angle between non-zero vectors $x$ and $y$ of $\mathbb{R}^n$ is the unique real number $\theta$ satisfying the conditions

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

This definition accords with the usual one of elementary geometry.

The angle between $x$ and $y$ is called \textbf{acute} or \textbf{obtuse} according as $x \cdot y$ is positive or negative.

Vectors $x$ and $y$, whether zero or not, are said to be \textbf{orthogonal} if $x \cdot y = 0$. 


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 $c_0 + \mathbf{c} \cdot \mathbf{x} = 0$, where $\mathbf{c}$ is the non-zero vector $(c_1, \ldots, c_n)$ and $\mathbf{x}$ is $(x_1, \ldots, x_n)$.

Such a vector $\mathbf{c}$ 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 $\lambda \mathbf{c}$ for some non-zero scalar $\lambda$.

Thus $H$ has exactly two unit normal vectors, namely $\pm \frac{\mathbf{c}}{\|\mathbf{c}\|}$.

Hence, given any hyperplane $H$ in $\mathbb{R}^n$, it may be assumed that it has an equation of the form $c_0 + \mathbf{c} \cdot \mathbf{x} = 0$, where $\mathbf{c}$ 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 \( \mathbf{v} \) and \( \mathbf{w} \) lie in a hyperplane \( H \) in \( \mathbb{R}^n \) with equation \( c_0 + \mathbf{c} \cdot \mathbf{x} = 0 \). Then \( c_0 + \mathbf{c} \cdot \mathbf{v} = 0 \) and \( c_0 + \mathbf{c} \cdot \mathbf{w} = 0 \). So \( \mathbf{c} \cdot (\mathbf{w} - \mathbf{v}) = 0 \).
- This shows that \( \mathbf{c} \) is orthogonal to every vector which is the difference of two vectors in \( H \).
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.,

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

It follows easily from this definition that $A^\perp$ is a subspace of $\mathbb{R}^n$ which intersects $A$ in the set $\{0\}$.

A standard result of linear algebra asserts that each vector of $\mathbb{R}^n$ can be expressed uniquely in the form $a + b$, where $a \in A$ and $b \in A^\perp$.

Thus $A + A^\perp = \mathbb{R}^n$. 
Orthonormal Sequences

- A sequence \( u_1, \ldots, u_m \) of vectors in \( \mathbb{R}^n \) is said to be an orthonormal sequence if \( u_i \cdot u_j \) is 1 or 0 according as \( i = j \) or \( i \neq j \).
- The simplest example of such a sequence is the sequence \( e_1, \ldots, e_n \) 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 \( u_1, \ldots, u_m \) in \( \mathbb{R}^n \) form a linearly independent set \( \{u_1, \ldots, u_m\} \).

To see this, suppose that scalars \( \lambda_1, \ldots, \lambda_m \) are such that
\[
\lambda_1 u_1 + \cdots + \lambda_m u_m = 0.
\]
Then, for \( i = 1, \ldots, m \),
\[
\lambda_i = (\lambda_1 u_1 + \cdots + \lambda_m u_m) \cdot u_i = 0 \cdot u_i = 0.
\]
This shows that \( \{u_1, \ldots, u_m\} \) is linearly independent.

Hence \( \{u_1, \ldots, u_m\} \) is an orthonormal basis for the subspace \( \text{lin}\{u_1, \ldots, u_m\} \) of \( \mathbb{R}^n \).
Thus each point \( \mathbf{x} \) of \( \text{lin}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \) can be written uniquely as a linear combination of \( \mathbf{u}_1, \ldots, \mathbf{u}_m \), say

\[
\mathbf{x} = \mu_1 \mathbf{u}_1 + \cdots + \mu_m \mathbf{u}_m.
\]

Then, for \( i = 1, \ldots, m \),

\[
\mathbf{x} \cdot \mathbf{u}_i = (\mu_1 \mathbf{u}_1 + \cdots + \mu_m \mathbf{u}_m) \cdot \mathbf{u}_i = \mu_i.
\]

We conclude that

\[
\mathbf{x} = (\mathbf{x} \cdot \mathbf{u}_1) \mathbf{u}_1 + \cdots + (\mathbf{x} \cdot \mathbf{u}_m) \mathbf{u}_m.
\]
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

\[
T(x) = Qx + q,
\]

where \( Q \) is a \( 2 \times 2 \) orthogonal matrix and \( q \) is a \( 2 \times 1 \) matrix.
A mapping $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is said to be a **congruence transformation** of $\mathbb{R}^n$ if

$$\| T(x) - T(y) \| = \| x - y \|, \quad \text{for all } x, y \in \mathbb{R}^n,$$

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 $x = (x_1, \ldots, x_n)$ of $\mathbb{R}^n$ with a column vector in the natural way, we see that $x^T x$ is the $1 \times 1$ matrix whose single element is the scalar $x_1^2 + \cdots + x_n^2$.

We identify this scalar with the matrix $x^T x$ itself, so that we may write

$$\| x \|^2 = x_1^2 + \cdots + x_n^2 = x^T x.$$
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 $T(x) = Qx + q$, where $Q$ is an $n \times n$ orthogonal matrix and $q$ is an $n \times 1$ matrix, is a congruence transformation of $\mathbb{R}^n$.

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

$$\| T(x) - T(y) \|^2 = \| Q(x - y) \|^2$$

$$= (Q(x - y))^T Q(x - y)$$

$$= (x - y)^T Q^T Q(x - y)$$

$$= (x - y)^T (x - y)$$

$$= \| x - y \|^2.$$

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

George Voutsadakis (LSSU)
The Euclidean Space $\mathbb{R}^n$

Length, Distance and Angle

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 $Q$ and an $n \times 1$ matrix $q$ such that $T(x) = Qx + q$, for all $x$ in $\mathbb{R}^n$.

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

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

So $f$ preserves norms.

Also

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

So

$$
\|f(x)\|^2 - 2f(x) \cdot f(y) + \|f(y)\|^2 = \|x\|^2 - 2x \cdot y + \|y\|^2.
$$
Since \( \|f(x)\| = \|x\| \) and \( \|f(y)\| = \|y\| \), we can deduce from the last equation that \( f(x) \cdot f(y) = x \cdot y \).

Thus, \( f \) preserves inner products.

It follows that \( f(e_1), \ldots, f(e_n) \) is an orthonormal sequence in \( \mathbb{R}^n \).

Hence

\[
f(x) = (f(x) \cdot f(e_1))f(e_1) + \cdots + (f(x) \cdot f(e_n))f(e_n).
\]

Writing \( x \) for \( (x_1, \ldots, x_n) \) and \( Q \) for the \( n \times n \) orthogonal matrix whose columns are \( f(e_1), \ldots, f(e_n) \), we deduce that

\[
f(x) = (x \cdot e_1)f(e_1) + \cdots + (x \cdot e_n)f(e_n)
= x_1f(e_1) + \cdots + x_nf(e_n)
= Qx.
\]

The proof is completed by putting \( q = T(0) \).
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 $T(x) = Qx + q$, where $Q$ is an $n \times n$ orthogonal matrix and $q$ 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.
The Euclidean Space $\mathbb{R}^n$

Length, Distance and Angle

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 $a \in A$. Then $A - a$ is an $r$-dimensional subspace of $\mathbb{R}^n$.

Let $\{u_1, \ldots, u_n\}$ be an orthonormal basis for $\mathbb{R}^n$ such that $\{u_1, \ldots, u_r\}$ is an orthonormal basis for $A - a$. Define a congruence transformation $T$ of $\mathbb{R}^n$ by the equation

$$T(x) = [u_1, \ldots, u_n]x + a.$$ 

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.
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 \( \text{aff} \, A \) is an \( r \)-flat. So by the theorem, there is a congruence transformation of \( \mathbb{R}^n \) which maps \( \text{aff} \, A \) 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 \( i : R_r \rightarrow \mathbb{R}^r \) 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.
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 $0$ and the elementary vectors $e_1, \ldots, e_r$ under $T$, it follows that $T$ can be expressed in the form

$$T(x) = \begin{bmatrix} Q & 0 \\ 0 & * \end{bmatrix} x + \begin{bmatrix} q \\ 0 \end{bmatrix},$$

where $Q$ is an $r \times r$ orthogonal matrix, $q$ is an $r \times 1$ matrix, and $0$ represents zero matrices of suitable shapes and sizes. Denote by $T_r$ the congruence transformation of $\mathbb{R}^r$ defined by the equation $T_r(x) = Qx + q$, where $x = (x_1, \ldots, x_r)$. 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 $\mathbb{R}^r$. 

George Voutsadakis (LSSU) Convexity July 2023 105 / 162
Subsection 7

Open Sets and Closed Sets
The Euclidean Space $\mathbb{R}^n$

Open Sets and Closed Sets

Open and Closed Balls

Let $a \in \mathbb{R}^n$ and $r > 0$.

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

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

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

Open Sets and Closed Sets

Open and Closed Unit Balls

- The balls $B(0; 1)$ and $B[0; 1]$ 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
  
  $$V = \{ \mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < 1 \} \quad \text{and} \quad U = \{ \mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq 1 \}.$$

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

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

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

\[
\|y - a\| \leq \|y - x + x - a\| \\
\leq \|y - x\| + \|x - a\| \\
< s + \|x - a\| = r.
\]

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

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

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

since, by the Cauchy-Schwarz Inequality, $|c \cdot (y - a)| \leq \|y - a\| < r$. Thus $y \in A$. So $B(a; r) \subseteq A$.

Consider the hyperplane $H$ in $\mathbb{R}^n$ with equation $c_0 + c \cdot x = 0$, where $c$ is a unit vector. We show that no point $a$ of $H$ is an interior point of $H$. Let $r > 0$. Then $a + \frac{1}{2}rc \not\in H$ and $\|a + \frac{1}{2}rc - a\| = \frac{1}{2}r$. Therefore, $a + \frac{1}{2}rc \in B(a; r)$ and $B(a; r) \not\subseteq H$. Hence, $a$ 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 $a \in \text{int}A$, then there exists $r > 0$ such that $B(a; r) \subseteq A$. Since $B(a; r)$ is open,

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

Hence, $a \in \text{int}(\text{int}A)$. 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 $a \in A$, then $a \in A_i$, for some $i \in I$. Since $A_i$ is open, there is an $r > 0$ such that $B(a; r) \subseteq A_i$. Hence, $B(a; r) \subseteq A$. Thus, $A$ is open.

Let $A$ be the intersection of the open sets $A_1, \ldots, A_m$ in $\mathbb{R}^n$. If $a \in A$, then $a \in A_1, \ldots, a \in A_m$. Since $A_1, \ldots, A_m$ are open, there exist $r_1, \ldots, r_m > 0$ such that $B(a; r_1) \subseteq A_1, \ldots, B(a; r_m) \subseteq A_m$. Let $r = \min\{r_1, \ldots, r_m\}$. Then $r > 0$ and

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

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 $\{0\}$, which is not open.
Closure of a Set

- A point \( a \) of \( \mathbb{R}^n \) is said to be a \textbf{closure point} of a set \( A \) in \( \mathbb{R}^n \) if every open ball with center \( a \) meets \( A \), i.e., if for every \( r > 0 \) the ball \( B(a; r) \) meets \( A \).

- The set of closure points of \( A \) is called the \textbf{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 \( B(a; r) \) and \( B[a; r] \) are both equal to the disc \( B[a; r] \).
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.
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 $\mathbb{R}^n \setminus A$.

**Theorem**

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

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

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

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

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

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

Trivially,

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

Thus,

$$\text{cl}(A_1 \cup \cdots \cup A_m) = \text{cl}A_1 \cup \cdots \cup \text{cl}A_m.$$
Closed Balls, Closed Halfspaces, Flats

**Theorem**

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

Let $A$ be the closed ball $B[a; r]$, where $a \in \mathbb{R}^n$ and $r > 0$. We prove that $A^c$ is open. Let $x \in A^c$. Then we show that $B(x; s) \subseteq A^c$, where $s$ is the positive number $\|x - a\| - r$.

Suppose that this is not the case. Then there is some point of $A$, $y$ say, which lies in $B(x; s)$. Now

$$\|x - a\| = \|x - y + y - a\| < s + r = \|x - a\|,$$

which is impossible. Hence $B(x; s) \subseteq A^c$.

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 $a$ 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 $a$ 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 $\text{bd}A$.

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

\[
B(a; r) \cap \text{aff} A \subseteq A.
\]

The set of all relative interior points of \(A\) is called the \textbf{relative interior} of \(A\) and is denoted by \(\text{ri}\ A\).

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 \( a \) 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 $\text{ri} B \subseteq \text{ri} A$. 
Suppose that \( a \) is a point of a set \( A \) in \( \mathbb{R}^n \) and that \( x \) is a point of \( \text{aff} A \) not lying in \( A \).

Define a scalar \( \lambda_0 \) by the equation

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

Then \((1 - \lambda_0)a + \lambda_0 x\) is a relative boundary point of \( A \) lying between \( a \) and \( x \).

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 $x_1, \ldots, x_k, \ldots$ of points is said to converge to a point $x$ if $\|x_k - x\| \to 0$ as $k \to \infty$, i.e., if the distance $\|x_k - x\|$ between $x_k$ and $x$ tends to zero as $k$ tends to infinity.

- We indicate such convergence by writing $x_k \to x$ as $k \to \infty$, or simply $x_k \to x$.

- This convergence for sequences of points in $\mathbb{R}^n$ coincides with that of classical convergence for real sequences.
Properties of Convergence

- The inequality $\|x\| - \|y\| \leq \|x - y\|$ proven previously, shows that

  $$\|x_k\| - \|x\| \leq \|x_k - x\|.$$  

  Hence $\|x_k\| \to \|x\|$ as $k \to \infty$ whenever $x_k \to x$ as $k \to \infty$.

- The triangle inequality shows that

  $$\|x_i - x_j\| \leq \|x_i - x\| + \|x - x_j\|.$$  

  Hence $\|x_i - x_j\| \to 0$ as $i,j \to \infty$ whenever $x_k \to x$ as $k \to \infty$. 
Suppose that \( x_k = (x_{k1}, \ldots, x_{kn}) \) for \( k = 1, 2, \ldots \) and that \( x = (x_1, \ldots, x_n) \).

Then, for \( i = 1, \ldots, n \), we have

\[
|x_{ki} - x_i|^2 \leq (x_{k1} - x_1)^2 + \cdots + (x_{kn} - x_n)^2 = \|x_k - x\|^2.
\]

We also have

\[
\|x_k - x\|^2 = (x_{k1} - x_1)^2 + \cdots + (x_{kn} - x_n)^2 \\
\leq (|x_{k1} - x_1| + \cdots + |x_{kn} - x_n|)^2.
\]

Hence

\[
|x_{ki} - x_i| \leq \|x_k - x\| \leq |x_{k1} - x_1| + \cdots + |x_{kn} - x_n|.
\]

Thus, \( x_k \to x \) if and only if \( x_{ki} \to x_i \), for \( i = 1, \ldots, n \).

So the convergence of \( x_1, \ldots, x_k, \ldots \) to \( x_1, \ldots, x_n \) is equivalent to the convergence of each of the coordinate sequences \( x_{1i}, \ldots, x_{ki}, \ldots \) for \( i = 1, \ldots, n \).
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 $x_k \to x$, $y_k \to y$ in $\mathbb{R}^n$ and $\lambda_k \to \lambda$, $\mu_k \to \mu$ in $\mathbb{R}$, then

$$x_k \cdot y_k \to x \cdot y \quad \text{in } \mathbb{R};$$
$$\lambda_k x_k + \mu_k y_k \to \lambda x + \mu y \quad \text{in } \mathbb{R}^n.$$
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 \( |x_k| \leq r \) for \( k = 1, 2, \ldots \).

Similarly, a sequence \( x_1, \ldots, x_k, \ldots \) of points in \( \mathbb{R}^n \) is defined to be **bounded** if there exists a real number \( r \) such that \( \|x_k\| \leq r \) 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 \( x_k \to x \) in \( \mathbb{R}^n \). By what we proved above, \( \|x_k\| \to \|x\| \). So there exists a real number \( r \) such that \( \|x_k\| \leq r \) for \( k = 1, 2, \ldots \).
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 $\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots$ be a bounded sequence of points in $\mathbb{R}^n$. Then each of the $n$ coordinate sequences associated with $\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots$ is bounded in $\mathbb{R}$. In particular, the sequence of the first coordinates of $\mathbf{x}_1, \ldots, k_k, \ldots$ is a bounded sequence of real numbers. Thus there exists a subsequence of $\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots$ such that the sequence of its first coordinates converges. Similarly, there exists a subsequence of this subsequence of $\mathbf{x}_1, \ldots, \mathbf{x}_k, \ldots$ such that the sequence of its second coordinates converges.
After performing this subsequence operation \( n \) times in all, we arrive at a subsequence of \( x_1, \ldots, x_k, \ldots \) each of whose \( n \) coordinate sequences converges. I.e., we have found a convergent subsequence of \( x_1, \ldots, x_k, \ldots \).
Theorem

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

- Suppose first that $x_1, \ldots, x_k, \ldots$ is a sequence of points of $A$ which converges to a point $x$ of $\mathbb{R}^n$. Then, for each $r > 0$, there is some point $x_k$ of the sequence such that $\|x_k - x\| < r$. Hence the open ball $B(x; r)$ meets $A$. This shows that $x \in \text{cl}A$.

- Suppose next that $x \in \text{cl}A$. Then, for each positive integer $k$, the ball $B(x; \frac{1}{k})$ meets $A$. Hence there exists $x_k \in A$ such that $\|x_k - x\| < \frac{1}{k}$. It follows that $x_1, \ldots, x_k, \ldots$ converges to $x$. 
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 $A = \text{cl}A$. 
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 $\|a\| \leq r$ for all $a \in A$.
- 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 if 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 $x \in \text{cl}A$, then, by a previous theorem, there is a sequence of points of $A$ which converges to $x$. Every subsequence of such a sequence also converges to $x$. The compactness of $A$ and the uniqueness of limits show that $x \in A$. Hence $A$ is closed.
- Suppose next that $A$ is not bounded. Then, for each positive integer $k$, there must exist a point $x_k$ of $A$ such that $\|x_k\| > k$. The sequence $x_1, \ldots, x_k, \ldots$ 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 $a_1, \ldots, a_m$ of $A$ such that

$$A \subseteq B(a_1; r) \cup \cdots \cup B(a_m; r).$$

We argue by contradiction. Suppose that no such finite number of points of $A$ exists. Let $x_1 \in A$. Then $A \nsubseteq B(x_1; r)$. Hence there exists a point $x_2$ of $A$ such that $\|x_2 - x_1\| \geq r$. Now $A \nsubseteq B(x_1; r) \cup B(x_2; r)$. Hence there exists a point $x_3$ of $A$ such that $\|x_3 - x_1\| \geq r$ and $\|x_3 - x_2\| \geq r$. Continuing in this way, we produce a sequence $x_1, \ldots, x_k, \ldots$ of points of $A$ with the property that $\|x_i - x_j\| \geq r$ whenever $i \neq j$. Clearly such a sequence cannot contain a convergent subsequence. This contradicts the compactness of $A$. 
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 $x$ in $A$, the open ball $B(x; r)$ 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 $a_k$ of $A$ such that $B(a_k; \frac{1}{k})$ is not contained in any $U_i$. Since $A$ is compact, the sequence $a_1, \ldots, a_k, \ldots$ has a subsequence which converges to a point $a$ of $A$. This point $a$ 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 $B(a; 2s) \subseteq U^*$. Since some subsequence of $a_1, \ldots, a_k, \ldots$ converges to $a$, there are infinitely many positive integers $k$ for which $\|a_k - a\| < s$. Choose one of these positive integers, $m$ say, so large that $\frac{1}{m} < s$. Let $x \in B(a_m; \frac{1}{m})$. Then

$$\|x - a\| \leq \|x - a_m\| + \|a_m - a\| < s + s = 2s.$$ 

So $x \in B(a; 2s)$. Thus $B(a_m; \frac{1}{m}) \subseteq B(a; 2s) \subseteq U^*$. This contradicts the assumption that $B(a_m; \frac{1}{m})$ is not contained in any $U_i$. 
The Euclidean Space $\mathbb{R}^n$

Convergence and Compactness

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 $x$ in $A$, the open ball $B(x; r)$ is contained in some $U_i$. By the preceding theorem, there exist points $a_1, \ldots, a_m$ in $A$ such that

$$A \subseteq B(a_1; r) \cup \cdots \cup B(a_m; r).$$

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

George Voutsadakis (LSSU)
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, \(A_{i_0} \subseteq \bigcup (A^c_i : i \in I_0)\). By the theorem, which is applicable since the sets \(A^c_i\) are open, being the complements of closed sets in \(\mathbb{R}^n\), there is a finite subset \(I'\) of \(I_0\) such that \(A_{i_0} \subseteq \bigcup (A^c_i : i \in I')\). 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.
### Theorem

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

1. open when $A$ is open and $\lambda \neq 0$;
2. closed when $A$ is compact and $B$ is closed;
3. bounded when $A$ and $B$ are bounded;
4. compact when $A$ and $B$ are compact.

#### (i)

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

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

This shows that $B(x; |\lambda| r) \subseteq \lambda A + \mu B$. Hence $\lambda A + \mu B$ is open.
(ii) Let $A$ be compact and let $B$ be closed. We consider only the non-trivial case $\mu \neq 0$. If $x \in \text{cl}(\lambda A + \mu B)$, then there exist sequences $a_1, \ldots, a_k, \ldots$ of points of $A$, and $b_1, \ldots, b_k, \ldots$ of points of $B$ such that $\lambda a_k + \mu b_k \to x$ as $k \to \infty$. Since $A$ is compact, there is a subsequence $a_{i_1}, \ldots, a_{i_k}, \ldots$ of $a_1, \ldots, a_k, \ldots$ which converges to some point $a$ of $A$. Thus $\lambda a_{i_k} + \mu b_{i_k} \to x$ and $b_{i_k} \to \frac{x-\lambda a}{\mu}$ as $k \to \infty$. But $B$ is closed, and so $\frac{x-\lambda a}{\mu} \in B$. Hence $x \in \lambda A + \mu B$. Thus $x \in \text{cl}(\lambda A + \mu B)$ implies that $x \in \lambda A + \mu B$. 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 $\|a\| \leq r_1$ and $\|b\| \leq r_2$ whenever $a \in A$ and $b \in B$. If $x \in \lambda A + \mu B$, then $x = \lambda a + \mu b$ for some $a \in A$ and $b \in B$. Hence

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

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

(iv) This follows immediately from (ii) and (iii).
Subsection 9

Continuity
Continuity

Let \( f : A \rightarrow \mathbb{R}^m \) be a mapping, where \( A \) is a non-empty set in \( \mathbb{R}^n \).

Then \( f \) is said to be **continuous at a point** \( a \) of \( A \) if, for each sequence \( a_1, \ldots, a_k, \ldots \) of points of \( A \) that converges to \( a \), the sequence \( f(a_1), \ldots, f(a_k), \ldots \) of points of \( \mathbb{R}^m \) converges to \( f(a) \).

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 \( \| \cdot \| : \mathbb{R}^n \rightarrow \mathbb{R} \) defined by the equation \( \| \cdot \|(x) = \|x\| \) for each point \( x \) of \( \mathbb{R}^n \).

That \( \| \cdot \| \) is continuous follows immediately from the fact that \( \|a_k\| \rightarrow \|a\| \) as \( k \rightarrow \infty \) whenever \( a_k \rightarrow a \) as \( k \rightarrow \infty \).
A mapping \( f : A \rightarrow \mathbb{R}^m \) 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 \( x, y \in A \),

\[
\| f(x) - f(y) \| \leq s \| x - y \|.
\]

If \( f : A \rightarrow \mathbb{R}^m \) satisfies the Lipschitz condition, then it is continuous on \( A \).

To see this, suppose that \( a_1, \ldots, a_k, \ldots \) is a sequence of points of \( A \) that converges to a point \( a \) of \( A \), so that \( \| a_k - a \| \rightarrow 0 \) as \( k \rightarrow \infty \). The Lipschitz condition shows that

\[
\| f(a_k) - f(a) \| \leq s \| a_k - a \|.
\]

Hence, \( \| f(a_k) - f(a) \| \rightarrow 0 \) as \( k \rightarrow \infty \), i.e., the sequence \( f(a_1), \ldots, f(a_k), \ldots \) converges to \( f(a) \).

Since \( f \) is continuous at an arbitrary \( a \) 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 \( Q = [q_{ij}] \) is the real \( m \times n \) matrix and \( q \) the real \( m \times 1 \) matrix such that, for each vector \( x \) in \( \mathbb{R}^n \), considered as a column vector, \( T(x) = Qx + q \). Let \( x, y \in \mathbb{R}^n \). Write \( u = (u_1, \ldots, u_n) = x - y \).

By the Cauchy-Schwarz inequality, for \( i = 1, \ldots, m \),

\[
(q_{i1}u_1 + \cdots + q_{in}u_n)^2 \leq (q_{i1}^2 + \cdots + q_{in}^2)(u_1^2 + \cdots + u_n^2).
\]

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

\[
\| T(x) - T(y) \|^2 = \| Qu \|^2 = \sum_{i=1}^m (q_{i1}u_1 + \cdots + q_{in}u_n)^2 \\
\leq \sum_{i=1}^m (q_{i1}^2 + \cdots + q_{in}^2)(u_1^2 + \cdots + u_n^2) \\
= s^2 \| u \|^2.
\]
The **distance function** \( d_A : \mathbb{R}^n \rightarrow \mathbb{R} \) of a non-empty set \( A \) in \( \mathbb{R}^n \) satisfies a Lipschitz condition.

This function \( d_A \) associates with each point \( x \) of \( \mathbb{R}^n \) its distance \( d_A(x) \) from \( A \).

Formally, \( d_A \) is defined by the equation

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

If \( A \) is the singleton set \( \{a\} \), then \( d_A(x) = \| x - a \| \).

In particular, if \( a = 0 \), then \( d_A(x) = \| x \| \).

It follows from the definition of \( d_A \) and a previous theorem that a point \( x \) of \( \mathbb{R}^n \) lies in the closure \( \text{cl}A \) of \( A \) if and only if its distance \( d_A(x) \) from \( A \) is zero.
Suppose now that \( x, y \) lie in \( \mathbb{R}^n \).

Then, for each \( \varepsilon > 0 \), there exists \( a \) in \( A \) such that \( \| x - a \| < d_A(x) + \varepsilon \).

By the triangle inequality,

\[
d_A(y) \leq \| y - a \| \leq \| y - x \| + \| x - a \| < \| y - x \| + d_A(x) + \varepsilon.
\]

Since \( \varepsilon > 0 \) is arbitrary,

\[
d_A(y) \leq \| y - x \| + d_A(x).
\]

Interchanging \( x \) and \( y \) in this inequality,

\[
d_A(x) \leq \| x - y \| + d_A(y).
\]

Hence \( d_A \) satisfies the Lipschitz condition

\[
|d_A(x) - d_A(y)| \leq \| x - y \|.
\]

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 $\mathbb{R}^n \setminus \{0\}$.
  
  Then $d_A(0) = 0$, but there is no $a \in A$ such that $\|0 - a\| = 0$. 
Distance from Nonempty Closed Sets

Theorem
Let $A$ be a non-empty closed set in $\mathbb{R}^n$ and let $x \in \mathbb{R}^n$. Then there exists $a_0 \in A$ such that $d_A(x) = \|x - a_0\|$.

- It follows easily from the definition of $d_A(x)$ that there exists a sequence $a_1, \ldots, a_k, \ldots$ of points of $A$ such that $\|x - a_k\| \to d_A(x)$ as $k \to \infty$. Since convergent sequences in $\mathbb{R}$ are bounded, there exists a real number $r$ such that $\|x - a_k\| \leq r$ for $k = 1, 2, \ldots$. We have

  $$\|a_k\| \leq \|a_k - x\| + \|x\| \leq r + \|x\|, \text{ for } k = 1, 2, \ldots.$$  

  So the sequence $a_1, \ldots, a_k, \ldots$ is bounded. Hence it contains some subsequence $a_{i_1}, \ldots, a_{i_k}, \ldots$ which converges to a point $a_0$ of $\mathbb{R}^n$. Since $A$ is closed, $a_0 \in A$. Now $\|x - a_{i_k}\| \to \|x - a_0\|$ as $k \to \infty$. But we already know that $\|x - a_{i_k}\| \to d_A(x)$ as $k \to \infty$. The uniqueness of limits in $\mathbb{R}$ shows that $d_A(x) = \|x - a_0\|$.

- The point $a_0$ is called a **nearest point** of $A$ to $x$.  

George Voutsadakis (LSSU)
Continuity and Compactness

**Theorem**

Let \( f : A \rightarrow \mathbb{R}^n \) 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 \( f(a_1), \ldots, f(a_k), \ldots \) be a sequence of points of \( f(A) \), where \( a_1, \ldots, a_k, \ldots \) is a sequence of points of \( A \).
- Since \( A \) is compact, there is a subsequence \( a_{i_1}, \ldots, a_{i_k}, \ldots \) of \( a_1, \ldots, a_k, \ldots \) which converges to some point \( a \) of \( A \).
- By the continuity of \( f \), the subsequence \( f(a_{i_1}), \ldots, f(a_{i_k}), \ldots \) of \( f(a_1), \ldots, f(a_k), \ldots \) converges to the point \( f(a) \) of \( f(A) \).
- Thus \( f(A) \) is compact.
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 $a, b \in A$ such that

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

The theorem shows that the non-empty set $f(A) = \{f(x) : x \in A\}$ 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 $\text{cl}f(A)$. Hence, since $f(A)$ is closed, they belong to $f(A)$. So there exist $a, b \in A$ such that $f(a) = \inf f(A)$ and $f(b) = \sup f(A)$. 

George Voutsadakis (LSSU)
The Euclidean Space $\mathbb{R}^n$

Continuity

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 $a_0 \in A$, $b_0 \in B$ such that

$$
\|a_0 - b_0\| = \inf \{\|a - b\| : a \in A, b \in B\}.
$$

- 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 $b_0 \in B$ such that $d_A(b_0) = \inf \{d_A(b) : b \in B\}$. By a previous theorem, applicable since $A$ is closed, there exists $a_0 \in A$ such that $d_A(b_0) = \|b_0 - a_0\|$. For each $a \in A$, $b \in B$, we have

$$
\|a - b\| \geq d_A(b) \geq d_A(b_0) = \|a_0 - b_0\|.
$$

Since $a_0 \in A$, $b_0 \in B$, $\|a_0 - b_0\| = \inf \{\|a - b\| : a \in A, b \in B\}$.

- We refer to $a_0$ and $b_0$ as **nearest points** of $A$ and $B$. 

George Voutsadakis (LSSU)

Convexity

July 2023 159 / 162
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 $f : A \rightarrow \mathbb{R}$ be both continuous and positive at some point $a$ of a set $A$ in $\mathbb{R}^n$. Then there exists an $r > 0$ such that $f(x) > 0$ whenever $x \in B(a; r) \cap A$.

Suppose that the stated conclusion does not hold. Then, for each $k = 1, 2, \ldots$ there exists $a_k \in B(a; \frac{1}{k}) \cap A$ such that $f(a_k) \leq 0$. Since $f$ is continuous at $a$ and $a_k \rightarrow a$ as $k \rightarrow \infty$, $f(a_k) \rightarrow f(a)$ as $k \rightarrow \infty$. Because $f(a_k) \leq 0$ for $k = 1, 2, \ldots$, it follows that $f(a) \leq 0$. This contradiction establishes the theorem.
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 \( a_1, \ldots, a_k, \ldots \) be a sequence of points of \( A \) that converges to a point \( a \) of \( A \). Since \( f \) is continuous, the sequence of points \( f(a_1), \ldots, f(a_k), \ldots \) of \( B \) converges to the point \( f(a) \) of \( B \). Since \( g \) is continuous, the sequence \( g(f(a_1)), \ldots, g(f(a_k)), \ldots \) converges to \( g(f(a)) \), i.e., the sequence \( (g \circ f)(a_1), \ldots, (g \circ f)(a_k), \ldots \) converges to \( (g \circ f)(a) \). 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 $a_1, \ldots, a_k, \ldots$ be a sequence of points of $f^{-1}(B)$ that converges to a point $a$ of $\mathbb{R}^n$. The continuity of $f$ shows that the sequence of points $f(a_1), \ldots, f(a_k), \ldots$ of $B$ converges to the point $f(a)$ of $\mathbb{R}^m$. But $B$ is closed. So $f(a) \in B$, i.e., $a \in f^{-1}(B)$. 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.