Lecture Notes: Matrix Algebra
Part A: Vectors

Peter J. Hammond
My email is p.j.hammond@warwick.ac.uk

revised 2023 September 15th;
typeset from matrixAlgA23.tex

Matrix Algebra in Six Segments

A. Introduction to Vectors

B. Introduction to Matrices

C. Determinants and Pivoting

D. Determinants, Inverses, and Rank

D, Appendix. Determinants and Volume

E. Quadratic Forms and Their Definiteness

F. Eigenvectors and Diagonalization

The lecture will overrun into the following day.

There will be time to catch up later.

Outline

Example of Two Equations in Two Unknowns

It is easy to check that

$$\left.\begin{aligned} x + y &= 10 \\ x - y &= 6 \end{aligned}\right\} \Longrightarrow x = 8, \; y = 2$$

More generally, one can:

1. add the two equations, to eliminate y;
2. subtract the second equation from the first, to eliminate x.

This leads to the following transformation

$$\left.\begin{aligned} x + y &= b_1 \\ x - y &= b_2 \end{aligned}\right\} \Longrightarrow \begin{cases} 2x = b_1 + b_2 \\ 2y = b_1 - b_2 \end{cases}$$

of the two equation system with general right-hand sides.

Obviously the solution is

$$x = \tfrac{1}{2}(b_1 + b_2), \; y = \tfrac{1}{2}(b_1 - b_2)$$

Using Matrix Notation, I

Matrix notation allows the two simultaneous equations

$$1x + 1y = b_1$$
$$1x - 1y = b_2$$

to be expressed as the single matrix equation

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

This can be written as $\mathbf{Az} = \mathbf{b}$, where

$$\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Here $\mathbf{A}, \mathbf{z}, \mathbf{b}$ are respectively: (i) the coefficient matrix; (ii) the vector of unknowns; (iii) the vector of right-hand sides.

Using Matrix Notation, II

We are considering the matrix equation

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

Its solution $x = \frac{1}{2}(b_1 + b_2),\ y = \frac{1}{2}(b_1 - b_2)$
can be expressed as

$$\begin{aligned} x &= \tfrac{1}{2} b_1 + \tfrac{1}{2} b_2 \\ y &= \tfrac{1}{2} b_1 - \tfrac{1}{2} b_2 \end{aligned}$$

Using matrix notation, this is

$$\mathbf{z} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \mathbf{Cb}, \quad \text{where} \quad \mathbf{C} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

Later we will see that $\mathbf{C}$ is the inverse
of the coefficient matrix $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

Two General Equations

Consider the general system of two equations

$$\begin{aligned} ax + by &= u = 1u + 0v \\ cx + dy &= v = 0u + 1v \end{aligned}$$

in two unknowns x and y, filled in with some extra 1s and 0s.

In matrix form, these equations can be written as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

Throughout the subsequent analysis we assume, for simplicity, that all four coefficients a, b, c, d of the matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ on the left-hand side are $\neq 0$.

Three Different Cases

We are considering the two equations

$$ax + by = u \quad \text{and} \quad cx + dy = v$$

They correspond to the two straight lines

$$y = (u - ax)/b \quad \text{and} \quad y = (v - cx)/d$$

The two lines have respective slopes $-a/b$ and $-c/d$.

These two slopes are equal iff $a/b = c/d$, or iff $ad = bc$, or iff $D := ad - bc = 0$.

We will distinguish three cases:

(A) If $D \neq 0$, the two lines have different slopes, so their intersection consists of a single point.

(B) If $D = 0$ and $u/b \neq v/d$, then the two lines are parallel but distinct, so their intersection is empty.

(C) If $D = 0$ and $u/b = v/d$, then the two lines are identical, so their intersection consists of all the points on either line.

First Steps

We have assumed that $a \neq 0$ in the matrix equation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

So we can eliminate x from the second equation
by "pivoting" about the a element of the matrix.

This requires us to add $-c/a$ times the first row to the second.

Letting $D = ad - bc$, this results in the new equality

$$\begin{pmatrix} a & b \\ 0 & D/a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c/a & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Now multiply the second row by a to obtain

$$\begin{pmatrix} a & b \\ 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Two General Equations: Case A, unique intersection, I

The original matrix equation has been reduced to

$$\begin{pmatrix} a & b \\ 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Consider **Case A** when $D := ad - bc \neq 0$.

Then we can add $-b/D$ times the second row to the first, yielding

$$\begin{pmatrix} a & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 + (bc/D) & -ab/D \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Recognizing now that $1 + (bc/D) = (D + bc)/D = ad/D$ allows us to rewrite these equations as

$$\begin{pmatrix} a & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ad/D & -ab/D \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Two General Equations: Case A, unique intersection, II

With $D = ad - bc \neq 0$, we have transformed the equations to

$$\begin{pmatrix} a & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ad/D & -ab/D \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Dividing the first row/equation by a and the second by D yields

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{D} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

This implies the unique solution

$$x = (1/D)(du - bv) \quad \text{and} \quad y = (1/D)(-cu + av)$$

Exercise

Check that this really is a solution of the two equations.

Two General Equations: Cases B and C

Except in Case A, we have $D := ad - bc = 0$.

Then the multiplier $-ab/D$ is undefined, and the system

$$\begin{pmatrix} a & b \\ 0 & D/a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -c/a & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

collapses to $\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u \\ v - cu/a \end{pmatrix}$.

This leaves us with two cases:

Case B If $cu \neq av$, there is no solution.

We were trying to find the intersection of two distinct parallel lines!

Case C If $cu = av$, then the second equation reduces to $0 = 0$.

There is a continuum of solutions
satisfying the one remaining equation $ax + by = u$,
or $x = (u - by)/a$ where y is any real number.

Outline

The Cartesian Product of Two Sets

Let S and T be any two sets.

Given any pair $s \in S$ and $t \in T$,
let (s, t) denote the associated ordered pair.

This definition implies that $(s, t) \neq (t, s)$
except when there exists $r \in S \cap T$ such that $r = s = t$.

The Cartesian product $S \times T$ of the two sets S and T
is defined as the set of all ordered pairs of members
of those two sets.

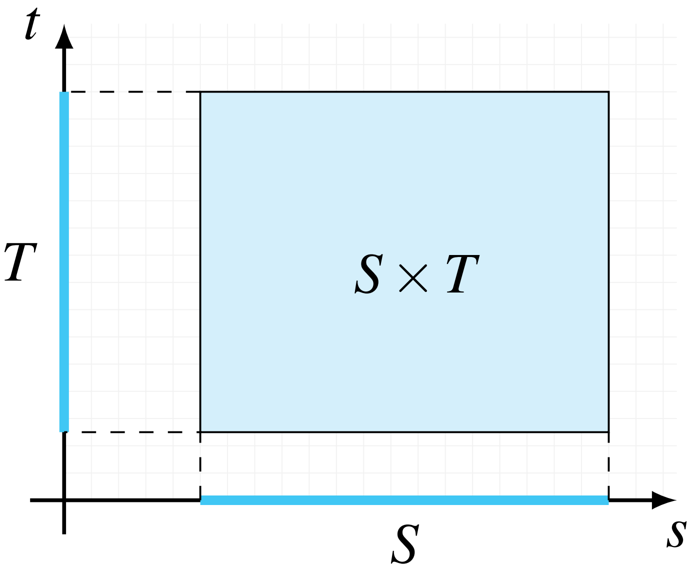
That is $S \times T := \{(s, t) \mid s \in S \text{ and } t \in T\}$.

The next slide illustrates the case
when S and T are (one-dimensional) intervals of the real line $\mathbb{R}$.

Then $S \times T$ is a (two-dimensional) rectangle.

It becomes the square $S^2 = T^2$ in the special case when $S = T$.

The Cartesian Product of Two Real Line Intervals



The Cartesian Product of Three Sets

Given any three sets S, T, U, the definition for pairs implies that

$$(S \times T) \times U = \{((s,t),u) \mid (s,t) \in S \times T \text{ and } u \in U\}$$
$$S \times (T \times U) = \{(s,(t,u)) \mid s \in S \text{ and } (t,u) \in T \times U\}$$

Evidently both these two sets are equal
to the three-fold product which is well defined by the equation

$$S \times T \times U := \{(s,t,u) \mid s \in S, t \in T, u \in U\}$$

In case the three sets S, T, U are intervals of the real line $\mathbb{R}$,
the product $S \times T \times U$ is a (three-dimensional) cuboid.
It becomes the cube $S^3 = T^3 = U^3$
in the special case when $S = T = U$.

The Cartesian Product of Many Sets

Consider a set I
which is the domain of possible values of an index i.

Suppose that the set S_i is defined for each $i \in I$.

Then the collection $\{S_i \mid i \in I\}$ is an indexed family of sets.

The Cartesian product of all the sets in this indexed family,
which is denoted by $\prod_{i \in I} S_i$, is defined by

$$\prod_{i \in I} S_i = \{\langle s_i \rangle_{i \in I} \mid \forall i \in I : s_i \in S_i\}$$

That is, the members of $\prod_{i \in I} S_i$ are all the (ordered) lists $\langle s_i \rangle_{i \in I}$
whose components satisfy $s_i \in S_i$ for all $i \in I$.

Special Case: The n-fold Cartesian Product S^n

Consider the case when there exists a common set S
and a natural number $n \in \mathbb{N}$
such that the indexed family $\{S_i \mid i \in I\}$
has $I = \{1, 2, \ldots, n\} = \mathbb{N}_n$ and $S_i = S$ for all $i \in \mathbb{N}_n$.

Then the product set $\prod_{i \in I} S_i$
is the Cartesian product of n copies of S.

In this case we use S^n to denote the product,
which we describe as the n-fold Cartesian product of S with itself.

Furthermore, each element $\langle s_i \rangle_{i \in \mathbb{N}_n} = \langle s_i \rangle_{i=1}^n \in S^n$
can be equivalently described as a mapping $\mathbb{N}_n \ni i \mapsto s_i \in S$.

Outline

Definition of Algebraic Field

Definition

An algebraic field $(\mathbb{F}, +, \cdot)$ of scalars is a set $\mathbb{F}$ that, together with the two binary operations $+$ of addition and $\cdot$ of multiplication, satisfies the following axioms for all $a, b, c \in \mathbb{F}$:

1. $\mathbb{F}$ is closed under $+$ and $\cdot$:
   — i.e., both $a + b$ and $a \cdot b$ are in $\mathbb{F}$.
2. $+$ and $\cdot$ are associative:
   both $a + (b + c) = (a + b) + c$ and $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
3. $+$ and $\cdot$ both commute:
   both $a + b = b + a$ and $a \cdot b = b \cdot a$.
4. There are identity elements $0, 1 \in \mathbb{F}$ for $+$ and $\cdot$ respectively, with $0 \neq 1$, such that: (i) $a + 0 = a$; (ii) $1 \cdot a = a$.
5. There are inverse operations $-$ for $+$ and $^{-1}$ for $\cdot$ such that: (i) $a + (-a) = 0$; (ii) provided $a \neq 0$, also $a \cdot a^{-1} = 1$.
6. The distributive law: $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

Three Examples of Real Algebraic Fields

Exercise

Verify that the following well known sets are algebraic fields:

- *the set* $\mathbb{R}$ *of all real numbers,*
  with the usual operations of addition and multiplication;
- *the set* $\mathbb{Q}$ *of all rational numbers*
  — i.e., those that can be expressed as the ratio $r = p/q$
  of integers $p, q \in \mathbb{Z}$ *with* $q \neq 0$*.*
  (Check that $\mathbb{Q}$ *is closed*
  under the usual operations of addition and multiplication,
  and that each non-zero rational
  has a rational multiplicative inverse.)
- *the set* $\mathbb{Q} + \sqrt{2}\mathbb{Q} := \{r_1 + \sqrt{2}r_2 \mid r_1, r_2 \in \mathbb{Q}\} \subset \mathbb{R}$
  of all real numbers that can be expressed as the sum of:
  (i) a rational number;
  (ii) a rational multiple of the irrational number $\sqrt{2}$*.*

Imaginary and Complex Numbers

Background: See FMEA2, Appendix B3 (and many other places).

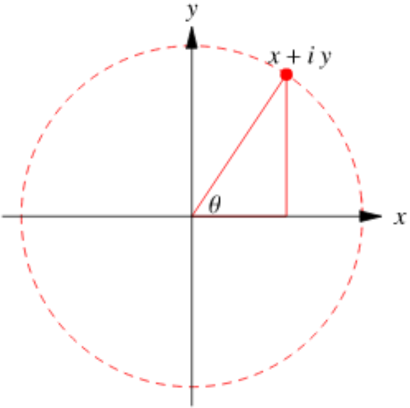
Let $\mathrm{i} \notin \mathbb{R}$ denote the basic basic imaginary number defined to satisfy $\mathrm{i}^2 = -1$.

Let $\mathrm{i}\,\mathbb{R} := \{\mathrm{i}\, b | b \in \mathbb{R}\}$ denote the set of all imaginary numbers that can be expressed as a real multiple of i.

Let $\mathbb{C} := \mathbb{R} + \mathrm{i}\,\mathbb{R}$, the set of all complex numbers that can be expressed as $c = a + \mathrm{i}\, b$ where $a, b \in \mathbb{R}$.

Thus, we define $\mathbb{C}$ so that
the mapping $\mathbb{C} \ni c = a + \mathrm{i}\, b \leftrightarrow (a, b) \in \mathbb{R}^2$ is a bijection
(that is, both a one-to-one "injection" and an onto "surjection").

Argand Diagram



Let $r = \sqrt{x^2 + y^2}$ denote the radius of the circle.

Then the complex number shown is $z = x + \mathrm{i}\, y$,
where $x, y \in \mathbb{R}$ with $x = r \cos \theta$ and $y = r \sin \theta$.

Argand Diagram and Polar Coordinates

Using polar coordinates, we have the bijection

$$\mathbb{C} \setminus \{0\} \ni z = x + \mathrm{i}\, y \leftrightarrow (r, \theta) \in \mathbb{R}_{++} \times [0, 2\pi)$$

where:

1. $x = r \cos\theta$ is the real part of z;
2. $y = r \sin\theta$ is the imaginary part of z;
3. $r = |z| = \sqrt{x^2 + y^2}$ is the modulus of z;
4. for each pair $(x, y) \in \mathbb{R}^2 \setminus (0, 0)$,
   the unique angle $\theta \in [0, 2\pi)$
   that satisfies $x = r \cos\theta$ and $y = r \sin\theta$
   is the argument of $z = x + \mathrm{i}\, y$.

Complex Conjugates

The complex conjugate of the complex number $z = x + \mathrm{i}\, y$
is the complex number $z^* = x - \mathrm{i}\, y$.

It is the reflection of z in the real axis.

If $z = r(\cos\theta + \mathrm{i}\sin\theta)$,
then $z^* = r(\cos\theta + \mathrm{i}\sin\theta) = r(\cos\eta + \mathrm{i}\sin\eta)$ where $\eta = -\theta$.
Or rather $\eta = 2\pi - \theta \in [0, 2\pi)$ except when $\theta = 0$.

Exercise

Given the two complex numbers $z = x + \mathrm{i}\, y$ and $c = a + \mathrm{i}\, b$, verify that:

1. $(z + c)^* = z^* + c^*$
2. $zz^* = x^2 + y^2$;
3. $(zc)^* = z^* c^*$;
4. $(z^*)^{-1} = (z^{-1})^*$.

Two Examples of Complex Algebraic Fields

Verify that the following well known sets are algebraic fields:

- the set $\mathbb{C}$ of complex numbers $a + b\,\mathrm{i}$ with $a, b \in \mathbb{R}$, together with the operations of:

  (i) addition defined by

  $$(a + b\,\mathrm{i}) + (c + d\,\mathrm{i}) = (a + c) + (b + d)\,\mathrm{i}$$

  (ii) multiplication defined by

  $$(a + b\,\mathrm{i})(c + d\,\mathrm{i}) = (ac - bd) + (ad + bc)\,\mathrm{i}$$

  Hint: Check that $(a + b\,\mathrm{i})^{-1} = (a - b\,\mathrm{i})/(a^2 + b^2)$.
- the set $\mathbb{Q} + \mathrm{i}\,\mathbb{Q}$ of all rational complex numbers that can be expressed as $c = a + \mathrm{i}\,b$ where $a, b \in \mathbb{Q}$.

Outline

Vectors and Inner Products

Let $\mathbf{x} = (x_i)_{i=1}^m \in \mathbb{R}^m$ denote a column m-vector of the form

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}.$$

Its transpose is the row m-vector

$$\mathbf{x}^\top = (x_1, x_2, \ldots, x_m).$$

Given a column m-vector $\mathbf{x}$ and row n-vector $\mathbf{y}^\top = (y_j)_{j=1}^n \in \mathbb{R}^n$ where $m = n$, the dot or scalar or inner product is defined as

$$\mathbf{y}^\top \mathbf{x} := \mathbf{y} \cdot \mathbf{x} := \sum\nolimits_{i=1}^m y_i x_i = \sum\nolimits_{j=1}^n y_j x_j$$

But when $m \neq n$, the scalar product is not defined.

Exercise on Quadratic Forms

Notation

We use the notation $D \ni x \mapsto f(x) \in C$,
which is read as "D owns x, which is mapped to f(x) in C",
in order to indicate a function $f : D \to C$
defined on the domain D*, with co-domain* C.

Exercise

Consider the **quadratic form** $f(\mathbf{w}) := \mathbf{w}^\top \mathbf{w}$
as a function $\mathbb{R}^n \ni \mathbf{w} \mapsto f(\mathbf{w}) := \mathbf{w}^\top \mathbf{w} \in \mathbb{R}$
of the column n-vector $\mathbf{w}$.

Explain why $f(\mathbf{w}) \geq 0$ *for all* $\mathbf{w} \in \mathbb{R}^n$,
with equality if and only if $\mathbf{w} = \mathbf{0}$,
where $\mathbf{0}$ *denotes the zero vector of* $\mathbb{R}^n$.

Net Quantity Vectors

Suppose there are n commodities numbered from $i = 1$ to n.

Each component q_i of the net quantity vector $\mathbf{q} = (q_i)_{i=1}^n \in \mathbb{R}^n$ represents the quantity of the ith commodity.

Often each such quantity is non-negative.

But general equilibrium theory, following Debreu's *Theory of Value*, often uses only the sign of q_i to distinguish between

- a consumer's demands and supplies of the ith commodity;
- or a producer's outputs and inputs of the ith commodity.

This sign is taken to be

positive for demands or outputs;

negative for supplies or inputs.

In fact, q_i is taken to be

- the consumer's net demand for the ith commodity;
- the producer's net supply or net output of the ith commodity.

Then $\mathbf{q}$ is the net quantity vector.

Price Vectors

Each component p_i of the (row) price vector $\mathbf{p}^\top \in \mathbb{R}^n$ indicates the price per unit of commodity i.

Then the scalar product

$$\mathbf{p}^\top \mathbf{q} = \mathbf{p} \cdot \mathbf{q} = \sum\nolimits_{i=1}^n p_i q_i$$

is the total value of the net quantity vector $\mathbf{q}$ evaluated at the price vector $\mathbf{p}$.

In particular, $\mathbf{p}^\top \mathbf{q}$ indicates

- the net profit (or minus the net loss) for a producer;
- the net dissaving for a consumer.

Outline

Definitions

Consider any two n-vectors $\mathbf{x} = (x_i)_{i=1}^n$ and $\mathbf{y} = (y_i)_{i=1}^n$ in $\mathbb{R}^n$.

Their sum $\mathbf{s} := \mathbf{x} + \mathbf{y}$ and difference $\mathbf{d} := \mathbf{x} - \mathbf{y}$ are constructed by adding or subtracting the vectors component by component — i.e., $\mathbf{s} = (s_i)_{i=1}^n$ and $\mathbf{d} = (d_i)_{i=1}^n$ where

$$s_i = x_i + y_i \quad \text{and} \quad d_i = x_i - y_i$$

for $i = 1, 2, \ldots, n$.

The scalar product $\lambda\mathbf{x}$ of any scalar $\lambda \in \mathbb{R}$ and vector $\mathbf{x} = (x_i)_{i=1}^n \in \mathbb{R}^n$ is constructed by multiplying each component of the vector $\mathbf{x}$ by the scalar λ — i.e.,

$$\lambda\mathbf{x} = (\lambda x_i)_{i=1}^n$$

Outline

General Vector Spaces

Definition

A vector (or linear) space V over an algebraic field $\mathbb{F}$ is a combination $\langle V, \mathbb{F}, +, \times \rangle$ of:

- a set V of vectors;
- the field $\mathbb{F}$ of scalars;
- the binary operation $V \times V \ni (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{u} + \mathbf{v} \in V$ of vector addition
- the binary operation $\mathbb{F} \times V \ni (\alpha, \mathbf{u}) \mapsto \alpha \times \mathbf{u} \in V$ of multiplication by a scalar.

These are required to satisfy all of the following eight vector space axioms.

Note: We follow usual notation, which replaces $\alpha \times \mathbf{u}$ by the abbreviation $\alpha\,\mathbf{u}$, without any multiplication sign.

Eight Vector Space Axioms

1. Addition is associative: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
2. Addition is commutative: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
3. Additive identity: There exists a zero vector $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.
4. Additive inverse: For every $\mathbf{v} \in V$, there exists an additive inverse $-\mathbf{v} \in V$ of $\mathbf{v}$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$
5. Multiplication by a scalar is distributive w.r.t. vector addition: $\alpha \times (\mathbf{u} + \mathbf{v}) = (\alpha \times \mathbf{u}) + (\alpha \times \mathbf{v})$
6. Multiplication by a scalar is distributive w.r.t. field addition: $(\alpha + \beta) \times \mathbf{v} = (\alpha \times \mathbf{v}) + (\beta \times \mathbf{v})$
7. Multiplication by a scalar and field multiplication are compatible: $\alpha \times (\beta \times \mathbf{v}) = (\alpha\beta) \times \mathbf{v}$
8. The unit element $1 \in \mathbb{F}$ is an identity element for scalar multiplication: $1 \times \mathbf{v} = \mathbf{v}$.

Multiplication by the Zero Scalar

Remark

From now on we usually suppress the $\times$ sign that has been used to indicate multiplication by a scalar.

In other words, $\alpha \times \mathbf{v}$ becomes just $\alpha\mathbf{v}$.

Exercise

Prove that $0\mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \in V$.

Hint: Which three axioms justify the following chain of equalities

$$0\mathbf{v} = [1 + (-1)]\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v} = \mathbf{v} - \mathbf{v} = \mathbf{0} \text{ ?}$$

A General Class of Finite Dimensional Vector Spaces

Exercise

Given an arbitrary algebraic field $\mathbb{F}$, let $\mathbb{F}^n$ denote the space of all lists $\langle a_i \rangle_{i=1}^n$ of n elements $a_i \in \mathbb{F}$ — i.e., the n-fold Cartesian product of $\mathbb{F}$ with itself.

1. *Show how to construct the respective binary operations*

$$\mathbb{F}^n \times \mathbb{F}^n \ni (\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} + \mathbf{y} \in \mathbb{F}^n$$
$$\mathbb{F} \times \mathbb{F}^n \ni (\lambda, \mathbf{x}) \mapsto \lambda \mathbf{x} \in \mathbb{F}^n$$

   of addition and scalar multiplication so that $(\mathbb{F}^n, \mathbb{F}, +, \times)$ is a vector space.

2. *Show too that subtraction and division by a (non-zero) scalar can be defined by $\mathbf{v} - \mathbf{w} = \mathbf{v} + (-1)\mathbf{w}$ and $\mathbf{v}/\alpha = (1/\alpha)\mathbf{v}$.*

Two Particular Finite Dimensional Vector Spaces

From now on we mostly consider real vector spaces
over the real field $\mathbb{R}$,
and especially the space $(\mathbb{R}^n, \mathbb{R}, +, \times)$ of *n*-vectors over $\mathbb{R}$.

We will consider, however, the space $(\mathbb{C}^n, \mathbb{C}, +, \times)$
of *n*-vectors over $\mathbb{C}$ — the complex plane — when considering:

- eigenvalues and eigenvectors of square matrices that may not be symmetric;
- solutions to systems of simultaneous linear difference and differential equations;
- the characteristic function of a random variable.

Remark

We show later that any eigenvalue of a symmetric real matrix must be real.

Outline

Linear Combinations

Definition

A linear combination of vectors is the weighted sum $\sum_{h=1}^{k} \lambda_h \mathbf{x}^h$, where $\mathbf{x}^h \in V$ and $\lambda_h \in \mathbb{F}$ for $h = 1, 2, \ldots, k$.

Exercise

By induction on k, show that the vector space axioms imply that any linear combination of vectors in V must also belong to V.

Linear Functions

Definition

A function $V \ni \mathbf{u} \mapsto f(\mathbf{u}) \in \mathbb{F}$ is linear provided that

$$f(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda f(\mathbf{u}) + \mu f(\mathbf{v})$$

for every linear combination $\lambda\mathbf{u} + \mu\mathbf{v}$ of two vectors $\mathbf{u}, \mathbf{v} \in V$, with $\lambda, \mu \in \mathbb{F}$.

Exercise

Prove that the function $V \ni \mathbf{u} \mapsto f(\mathbf{u}) \in \mathbb{F}$ is linear if and only if both:

1. *for every vector $\mathbf{v} \in V$ and scalar $\lambda \in \mathbb{F}$ one has $f(\lambda\mathbf{v}) = \lambda f(\mathbf{v})$;*
2. *for every pair of vectors $\mathbf{u}, \mathbf{v} \in V$ one has $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$.*

Show too that property 1 above implies that $f(\mathbf{0}) = 0$.

Key Properties of Linear Functions

Exercise

Use induction on k to show that if the function $f : V \to \mathbb{F}$ is linear, then

$$f\left(\sum_{h=1}^{k} \lambda_h \mathbf{x}^h\right) = \sum_{h=1}^{k} \lambda_h f(\mathbf{x}^h)$$

for all linear combinations $\sum_{h=1}^{k} \lambda_h \mathbf{x}^h$ in V — i.e., f preserves linear combinations.

Exercise

In case $V = \mathbb{R}^n$ and $\mathbb{F} = \mathbb{R}$, show that any linear function is homogeneous of degree 1, meaning that $f(\lambda \mathbf{v}) = \lambda f(\mathbf{v})$ for all $\lambda \in \mathbb{R}$ and all $\mathbf{v} \in \mathbb{R}^n$.

In particular, putting $\lambda = 0$ gives $f(\mathbf{0}) = 0$.

What is the corresponding property in case $V = \mathbb{Q}^n$ and $\mathbb{F} = \mathbb{Q}$?

Affine Functions

Definition

A function $g : V \to \mathbb{F}$ is said to be affine
if there is a scalar additive constant $\alpha \in \mathbb{F}$
and a linear function $f : V \to \mathbb{F}$ such that $g(\mathbf{v}) \equiv \alpha + f(\mathbf{v})$.

Exercise

Under what conditions is an affine function $g : \mathbb{R} \to \mathbb{R}$ linear when its domain $\mathbb{R}$ is regarded as a vector space?

An Economic Aggregation Theorem

Suppose that a finite population of households $h \in H$
with respective non-negative incomes $y_h \in \mathbb{Q}_+$ $(h \in H)$
have non-negative demands $x_h \in \mathbb{R}$ $(h \in H)$
which depend on household income via a function $y_h \mapsto f_h(y_h)$.

Given total income $Y := \sum_h y_h$, under what conditions
can their total demand $X := \sum_h x_h = \sum_h f_h(y_h)$
be expressed as a function $X = F(Y)$ of Y alone?

The answer is an implication of Cauchy's functional equation.

In this context the theorem asserts that this aggregation condition
implies that the functions f_h $(h \in H)$ and F must be co-affine.

This means there exists a common multiplicative constant $\rho \in \mathbb{R}$,
along with additive constants α_h $(h \in H)$ and A, such that

$$f_h(y_h) \equiv \alpha_h + \rho y_h \ (h \in H) \text{ and } F(Y) \equiv A + \rho Y$$

Cauchy's Functional Equation: Proof of Sufficiency

Theorem

Except in the trivial case when H has only one member, Cauchy's functional equation $F(\sum_{h \in H} y_h) \equiv \sum_{h \in H} f_h(y_h)$ is satisfied for functions $F, f_h : \mathbb{Q} \to \mathbb{R}$ if and only if:

1. *there exists a single function $\phi : \mathbb{Q} \to \mathbb{R}$ such that*

   $$F(q) = F(0) + \phi(q) \text{ and } f_h(q) = f_h(0) + \phi(q) \text{ for all } h \in H$$

2. *the function $\phi : \mathbb{Q} \to \mathbb{R}$ is linear,*
   implying that the functions F and f_h are co-affine.

Proof.

Suppose $f_h(y_h) \equiv \alpha_h + \rho y_h$ for all $h \in H$, and $F(Y) \equiv A + \rho Y$.
Then Cauchy's functional equation $F(\sum_{h \in H} y_h) \equiv \sum_{h \in H} f_h(y_h)$ is obviously satisfied provided that $A = \sum_{h \in H} \alpha_h$. □

Cauchy's Equation: Necessity in the Differentiable Case

Suppose $\#H \geq 2$ and $F(\sum_{h \in H} y_h) \equiv \sum_{h \in H} f_h(y_h)$ where, for each $h \in H$, the function $\mathbb{R} \ni y_h \mapsto f_h(y_h) \in \mathbb{R}$ is differentiable.

For any pair $j, k \in H$, consider the effect of transferring a small amount η from k to j, with y_h fixed for all $h \in H \setminus \{j, k\}$.

Because both $\sum_{h \in H} y_h$ and $\sum_{h \in H \setminus \{j,k\}} f_h(y_h)$ are unchanged, equating the changes to the two sides of the Cauchy equation gives $0 = f_j(y_j + \eta) + f_k(y_k - \eta) - f_j(y_j) - f_k(y_k)$.

Assuming that $f_j'(y_j)$ and $f_k'(y_k)$ both exist, we can differentiate the last equation w.r.t. to η to obtain $0 = f_j'(y_j) - f_k'(y_k)$.

It follows that there exists a constant c such that $f_h'(y) = c$ for all $h \in H$ and all real y, so $f_h(y) = \alpha_h + cy$.

But then, given any $\eta \in \mathbb{R}$, one has

$$F\left(\sum_{h \in H} y_h + \eta\right) - F\left(\sum_{h \in H} y_h\right) = f_j(y_j + \eta) - f_j(y_j) = c\eta$$

It follows that $F(Y)$ is the affine function $A + cY$, for some real A.

Cauchy's Equation: Beginning the Proof of Necessity

Now we prove Cauchy's equation without assuming differentiability.

Lemma

The mapping $\mathbb{Q} \ni q \mapsto \phi(q) := F(q) - F(0) \in \mathbb{R}$ *must satisfy;*

1. $\phi(q) \equiv f_i(q) - f_i(0)$ *for all* $i \in H$ *and* $q \in \mathbb{Q}$*;*
2. $\phi(q + q') \equiv \phi(q) + \phi(q')$ *for all* $q, q' \in \mathbb{Q}$*.*

Proof.

To prove part 1, consider any $i \in H$ and all $q \in \mathbb{Q}$.

Note that Cauchy's equation $F(\sum_h y_h) \equiv \sum_h f_h(y_h)$
implies that $F(q) = f_i(q) + \sum_{h \neq i} f_h(0)$
and also $F(0) = f_i(0) + \sum_{h \neq i} f_h(0)$.

Now define the function $\phi(q) := F(q) - F(0)$ on the domain $\mathbb{Q}$.

Then subtract the second equation from the first to obtain

$$\phi(q) = F(q) - F(0) = f_i(q) - f_i(0) \qquad \square$$

Cauchy's Equation: Continuing the Proof of Necessity

Proof.

To prove part 2, consider any $i, j \in H$ with $i \neq j$,
and any $q, q' \in \mathbb{Q}$.

Note that Cauchy's equation $F(\sum_h y_h) \equiv \sum_h f_h(y_h)$ implies that

$$\begin{aligned} F(q+q') &= f_i(q) + f_j(q') + \textstyle\sum_{h \in H \setminus \{i,j\}} f_h(0) \\ F(0) &= f_i(0) + f_j(0) + \textstyle\sum_{h \in H \setminus \{i,j\}} f_h(0) \end{aligned}$$

Now subtract the second equation from the first,
then use the equation $\phi(q) = F(q) - F(0) = f_i(q) - f_i(0)$
derived in the previous slide, to obtain successively

$$\begin{aligned} \phi(q+q') &= F(q+q') - F(0) \\ &= f_i(q) - f_i(0) + f_j(q') - f_j(0) \\ &= \phi(q) + \phi(q') \end{aligned}$$

□

Cauchy's Equation: Resuming the Proof of Necessity

Because $\phi(q + q') \equiv \phi(q) + \phi(q')$,
for any $k \in \mathbb{N}$ one has $\phi(kq) = \phi\left((k-1)q\right) + \phi(q)$.

As an induction hypothesis, which is trivially true for $k = 2$, suppose that $\phi\left((k-1)q\right) = (k-1)\phi(q)$.

Confirming the induction step, the hypothesis implies that

$$\phi(kq) = \phi\left((k-1)q\right) + \phi(q) = (k-1)\phi(q) + \phi(q) = k\phi(q)$$

So $\phi(kq) = k\phi(q)$ for every $k \in \mathbb{N}$ and every $q \in \mathbb{Q}$.

Putting $q' = kq$ implies that $\phi(q') = k\phi(q'/k)$.

Interchanging q and q', it follows that $\phi(q/k) = (1/k)\phi(q)$.

Cauchy's Equation: Completing the Proof of Necessity

So far we have proved that, for every $k \in \mathbb{N}$ and every $q \in \mathbb{Q}$, one has both $\phi(kq) = k\phi(q)$ and $\phi(q/k) = (1/k)\phi(q)$.

Hence, for every rational $r = m/n \in \mathbb{Q}$ one has $\phi(mq/n) = m\phi(q/n) = (m/n)\phi(q)$ and so $\phi(rq) = r\phi(q)$.

In particular, $\phi(r) = r\phi(1)$, so ϕ is linear on its domain $\mathbb{Q}$ (though not on the whole of $\mathbb{R}$ without additional assumptions such as continuity or monotonicity).

The rest of the proof is routine checking of definitions.

Outline

The Euclidean Norm as Length in $\mathbb{R}^2$

The length $\ell_2(x_1, x_2)$ of the typical 2-vector $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ is defined as the distance in the plane $\mathbb{R}^2$ between the origin $(0, 0)$ and the point (x_1, x_2).

It is the length of the hypotenuse of the right-angled triangle with vertices at $(0, 0)$, $(x_1, 0)$, and (x_1, x_2).

The two sides that meet at right angles at the vertex $(x_1, 0)$ are the two line segments joining: (i) $(0, 0)$ to $(x_1, 0)$; (ii) $(x_1, 0)$ to (x_1, x_2).

The lengths of these two line segments are $|x_1|$ and $|x_2|$.

Pythagoras's theorem implies that the square $[\ell_2(x_1, x_2)]^2$ of the length of the hypotenuse is equal to the sum of the squares of the lengths of the other two sides.

Hence $[\ell_2(x_1, x_2)]^2 = |x_1|^2 + |x_2|^2 = x_1^2 + x_2^2$.

This implies that $\ell_2(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$.

Or, perhaps less clumsily, that $\ell_2(x_1, x_2) = (x_1^2 + x_2^2)^{1/2}$.

The Euclidean Norm as Length in $\mathbb{R}^3$

The length $\ell_3(x_1, x_2, x_3)$ in the three-dimensional space $\mathbb{R}^3$ of the typical 3-vector $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ is defined as the distance between the origin $(0, 0, 0)$ and the point (x_1, x_2, x_3).

It is the length of the hypotenuse of the right-angled triangle with vertices at $(0, 0, 0)$, $(x_1, x_2, 0)$, and (x_1, x_2, x_3).

The two sides that meet at right angles at the vertex $(x_1, x_2, 0)$ are the two line segments joining: (i) $(0, 0, 0)$ to $(x_1, x_2, 0)$; (ii) $(x_1, x_2, 0)$ to (x_1, x_2, x_3)

The lengths of these two line segments are $\ell_2(x_1, x_2)$ and $|x_3|$.

Pythagoras's theorem implies that the square $[\ell_3(x_1, x_2, x_3)]^2$ of the length of the hypotenuse is equal to the sum of the squares of the lengths of the other two sides.

Hence $[\ell_3(x_1, x_2, x_3)]^2 = [\ell_2(x_1, x_2)]^2 + |x_3|^2 = x_1^2 + x_2^2 + x_3^2$.

This implies that $\ell_3(x_1, x_2, x_3) = \sqrt{x_1^2 + x_2^2 + x_3^2}$.

The Euclidean Norm as Length in $\mathbb{R}^n$ when $n > 3$, I

Length in $\mathbb{R}^n$ can be measured using a ruler when $n \leq 3$,
but for $n > 3$ it cannot.
Given any $\mathbf{x} = (x_1, x_2, x_3, \ldots, x_n) \in \mathbb{R}^n$:

1. in case $x_1 = x_2 = \ldots = x_{n-2} = 0$, it is natural to consider the 2-dimensional subspace $\{(0, 0, \ldots, 0)\} \times \mathbb{R}^2$ and to take $\ell_n(\mathbf{x}) = \ell_n(0, 0, \ldots, 0, x_{n-1}, x_n) = \ell_2(x_{n-1}, x_n) = \sqrt{x_{n-1}^2 + x_n^2}$;
2. in case $x_n = 0$, it is natural to consider the $(n-1)$-dimensional subspace $\mathbb{R}^{n-1} \times \{0\}$ and to take $\ell_n(\mathbf{x}) = \ell_n(x_1, x_2, \ldots x_{n-1}, 0) = \ell_{n-1}(x_1, x_2, \ldots x_{n-1})$;
3. otherwise, in the general case when the three n-vectors $\mathbf{0}$, $(x_1, x_2, \ldots, x_{n-1}, 0)$ and $\mathbf{x}$ correspond to three distinct points in $\mathbb{R}^n$ that all belong to a unique 2-dimensional subspace in $\mathbb{R}^n$ and determine a unique triangle in that subspace.

The Euclidean Norm as Length in $\mathbb{R}^n$ when $n > 3$, II

Consider the third case when $\mathbf{0}$, $(x_1, x_2, \ldots, x_{n-1}, 0)$ and $\mathbf{x}$ are three distinct points in $\mathbb{R}^n$.

The line segment that joins $\mathbf{0}$ to $(x_1, x_2, \ldots, x_{n-1}, 0)$
meets the line segment that joins $(x_1, x_2, \ldots, x_{n-1}, 0)$ to $\mathbf{x}$,
and forms a right angle,
at the common point $(x_1, x_2, \ldots, x_{n-1}, 0)$.

Now apply Pythagoras's theorem to this right-angled triangle,
and use the equality $\ell_n(x_1, x_2, \ldots x_{n-1}, 0) = \ell_{n-1}(x_1, x_2, \ldots x_{n-1})$
that we postulated for case 2.

The result is the recurrence relation

$$\begin{aligned} [\ell_n(\mathbf{x})]^2 &= [\ell_n(x_1, x_2, \ldots x_{n-1}, 0)]^2 + |x_n|^2 \\ &= [\ell_{n-1}(x_1, x_2, \ldots x_{n-1})]^2 + x_n^2 \end{aligned}$$

The Euclidean Norm as Length in $\mathbb{R}^n$ for General $n \in \mathbb{N}$

Length in $\mathbb{R}^1$ is given by $\ell_1(x_1) = |x_1| = \sqrt{x_1^2}$.

Length in $\mathbb{R}^2$ is given by $\ell_2(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$.

Length in $\mathbb{R}^3$ is given by $\ell_3(x_1, x_2, x_3) = \sqrt{x_1^2 + x_2^2 + x_3^2}$.

For $n > 3$, length in $\mathbb{R}^n$ satisfies the recurrence relation

$$[\ell_n(x_1, x_2, \ldots x_{n-1}, x_n)]^2 = [\ell_{n-1}(x_1, x_2, \ldots x_{n-1})]^2 + x_n^2$$

It follows by induction on n that the length of the typical n-vector $\mathbf{x} = (x_i)_{i=1}^n \in \mathbb{R}^n$ satisfies

$$[\ell_n(x_1, x_2, \ldots x_{n-1}, x_n)]^2 = \sum\nolimits_{i=1}^n x_i^2$$

Definition

The length of the typical n-vector $\mathbf{x} = (x_i)_{i=1}^n \in \mathbb{R}^n$ is its (Euclidean) norm $\|\mathbf{x}\| := \sqrt{\mathbf{x}^\top \mathbf{x}} = \sqrt{\mathbf{x} \cdot \mathbf{x}}$.

Unit n-Vectors, the Unit Sphere, and Unit Ball

Definition

A unit vector $\mathbf{u} \in \mathbb{R}^n$ is a vector with unit norm
— i.e., its components satisfy $\sum_{i=1}^n u_i^2 = \|\mathbf{u}\|^2 = 1$.

The set of all such unit vectors forms a surface
called the unit sphere of dimension $n - 1$
(one less than n because of the defining equation).

It is defined as the hollow set (like a "soccer" football)

$$S^{n-1} := \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^n x_i^2 = 1 \right\}$$

Other examples are balls used in tennis, squash, "ping pong".

The unit ball $B \subset \mathbb{R}^n$ is the solid set (like a cricket ball or golf ball)

$$B := \left\{ \mathbf{x} \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^n x_i^2 \leq 1 \right\}$$

of all points bounded by the surface of the unit sphere $S^{n-1} \subset \mathbb{R}^n$.

Cauchy–Schwartz Inequality

Theorem

For all pairs $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, *one has* $|\mathbf{a} \cdot \mathbf{b}| \leq \|\mathbf{a}\| \|\mathbf{b}\|$.

Proof.

Define the function $\mathbb{R} \ni \xi \mapsto f(\xi) := \sum_{i=1}^n (a_i \xi + b_i)^2 \in \mathbb{R}$.

Clearly f is the quadratic function $f(\xi) \equiv A\xi^2 + B\xi + C$
whose three coefficients $A := \sum_{i=1}^n a_i^2 = \|\mathbf{a}\|^2$,
along with $B := 2 \sum_{i=1}^n a_i b_i = 2\mathbf{a} \cdot \mathbf{b}$, and $C := \sum_{i=1}^n b_i^2 = \|\mathbf{b}\|^2$.

There is a trivial case when $A = 0$ because $\mathbf{a} = \mathbf{0}$.

Otherwise $A > 0$, so we can complete the square to get

$$f(\xi) \equiv A\xi^2 + B\xi + C = A[\xi + (B/2A)]^2 + C - B^2/4A$$

But the definition of f implies that $f(\xi) \geq 0$ for all $\xi \in \mathbb{R}$,
including $\xi = -B/2A$, so $0 \leq f(-B/2A) = C - B^2/4A$.
Because we are considering the case $A > 0$, we have $\frac{1}{4}B^2 \leq AC$,
implying that $|\mathbf{a} \cdot \mathbf{b}| = \left|\frac{1}{2}B\right| \leq \sqrt{AC} = \|\mathbf{a}\| \|\mathbf{b}\|$. □

Outline

The Angle Between Two Vectors

Consider the triangle in $\mathbb{R}^n$ whose vertices are the three disjoint vectors $\mathbf{x}, \mathbf{y}, \mathbf{0}$.

Its three sides or edges have respective lengths $\|\mathbf{x}\|$, $\|\mathbf{y}\|$, $\|\mathbf{x} - \mathbf{y}\|$, where the last follows from the parallelogram law.

Note that $\|\mathbf{x} - \mathbf{y}\|^2 \lesseqqgtr \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ according as the angle at $\mathbf{0}$ is: (i) acute; (ii) a right angle; (iii) obtuse. But

$$\begin{aligned}\|\mathbf{x} - \mathbf{y}\|^2 - \|\mathbf{x}\|^2 - \|\mathbf{y}\|^2 &= \sum\nolimits_{i=1}^n (x_i - y_i)^2 - \sum\nolimits_{i=1}^n (x_i^2 + y_i^2) \\ &= \sum\nolimits_{i=1}^n (-2x_i y_i) = -2\mathbf{x} \cdot \mathbf{y}\end{aligned}$$

So the three cases (i)–(iii) occur according as $\mathbf{x} \cdot \mathbf{y} \gtreqqless 0$.

Using the Cauchy–Schwartz inequality,
one can define the angle between $\mathbf{x}$ and $\mathbf{y}$
as the unique solution $\theta = \cos^{-1}(\mathbf{x} \cdot \mathbf{y}/\|\mathbf{x}\|\|\mathbf{y}\|) = \arccos(\mathbf{x} \cdot \mathbf{y}/\|\mathbf{x}\|\|\mathbf{y}\|)$
in the interval $[0, \pi)$ of the equation $\cos\theta = \mathbf{x} \cdot \mathbf{y}/\|\mathbf{x}\|\|\mathbf{y}\| \in [-1, 1]$.

Orthogonal and Orthonormal Sets of Vectors

Case (ii) suggests defining two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ as orthogonal just in case $\mathbf{x} \cdot \mathbf{y} = 0$, which is true if and only if the angle $\theta = \arccos 0 = \frac{1}{2}\pi = 90^\circ$.

A set of k vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} \subset \mathbb{R}^n$ is said to be:

- pairwise orthogonal just in case $\mathbf{x}_i \cdot \mathbf{x}_j = 0$ whenever $j \neq i$;
- orthonormal just in case, in addition, each $\|\mathbf{x}_i\| = 1$ — i.e., all k elements of the set are vectors of unit length.

Define the Kronecker delta function

$$\{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\} \ni (i, j) \mapsto \delta_{ij} \in \{0, 1\}$$

on the set of pairs $i, j \in \{1, 2, \ldots, n\}$ by

$$\delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Then the set of k vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} \subset \mathbb{R}^n$ is orthonormal if and only if $\mathbf{x}_i \cdot \mathbf{x}_j = \delta_{ij}$ for all pairs $i, j \in \{1, 2, \ldots, k\}$.

Outline

The Canonical Basis of $\mathbb{R}^n$

Example

A prominent orthonormal set is the canonical basis of $\mathbb{R}^n$.

It is defined as the set of the n different n-dimensional vectors $\mathbf{e}^i$ $(i = 1, 2, \ldots, n)$ whose respective components $(e^i_j)_{j=1}^n$ satisfy $e^i_j = \delta_{ij}$ for all $j \in \{1, 2, \ldots, n\}$.

Exercise

1. *Show that each n-vector* $\mathbf{x} = (x_i)_{i=1}^n$ *is a linear combination*

$$\mathbf{x} = (x_i)_{i=1}^n = \sum\nolimits_{i=1}^n x_i \mathbf{e}^i$$

   of the canonical basis vectors $\{\mathbf{e}^1, \mathbf{e}^2, \ldots, \mathbf{e}^n\}$, *with the multiplier attached to each basis vector* $\mathbf{e}^i$ *equal to the respective component* x_i $(i = 1, 2, \ldots, n)$.

2. *Show that* $\left(\sum_{i=1}^n x_i \mathbf{e}^i\right) \cdot \left(\sum_{i=1}^n y_i \mathbf{e}^i\right) = \sum_{i=1}^n x_i y_i$.

The Canonical Basis in Commodity Space

Example

Consider the case when each vector $\mathbf{x} \in \mathbb{R}^n$ is a quantity vector,
whose components are $(x_i)_{i=1}^n$,
where x_i indicates the net quantity of commodity i.

Then the ith unit vector $\mathbf{e}^i$ of the canonical basis of $\mathbb{R}^n$
represents a commodity bundle
that consists of one unit of commodity i,
but nothing of every other commodity.

Suppose that the row vector $\mathbf{p}^\top \in \mathbb{R}^n$
is a price vector for the same list of n commodities.
Then the value $\mathbf{p}^\top \mathbf{e}^i$ of the ith unit vector $\mathbf{e}^i$ must equal p_i,
the price per unit of the ith commodity.

Linear Functions

Theorem

The function $\mathbb{R}^n \ni \mathbf{x} \mapsto f(\mathbf{x}) \in \mathbb{R}$ *is linear*
if and only if there exists $\mathbf{y} \in \mathbb{R}^n$ *such that* $f(\mathbf{x}) = \mathbf{y}^\top \mathbf{x}$.

Proof.

An easy exercise is to check that the condition is sufficient.

Conversely, for necessity, recall that each vector $\mathbf{x} \in \mathbb{R}^n$ equals the linear combination $\sum_{i=1}^n x_i \mathbf{e}^i$ of the n canonical basis vectors.

Hence, linearity of f implies that

$$f(\mathbf{x}) = f\left(\sum_{i=1}^n x_i \mathbf{e}^i\right) = \sum_{i=1}^n x_i f(\mathbf{e}^i) = \sum_{i=1}^n f(\mathbf{e}^i)\, x_i = \mathbf{y}^\top \mathbf{x}$$

where $\mathbf{y}$ is the column vector whose components are $y_i = f(\mathbf{e}^i)$ for $i = 1, 2, \ldots, n$. □

Linear Transformations: Definition

Definition

The vector-valued function

$$\mathbb{R}^n \ni \mathbf{x} \mapsto \mathbf{F}(\mathbf{x}) = (F_i(\mathbf{x}))_{i=1}^m \in \mathbb{R}^m$$

is a linear transformation just in case
each component function $\mathbb{R}^n \ni \mathbf{x} \mapsto \mathbf{F}_i(\mathbf{x}) \in \mathbb{R}$ is linear
— or equivalently, iff $\mathbf{F}(\lambda \mathbf{x} + \mu \mathbf{y}) = \lambda \mathbf{F}(\mathbf{x}) + \mu \mathbf{F}(\mathbf{y})$
for every linear combination $\lambda \mathbf{x} + \mu \mathbf{y}$ of every pair $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Characterizing Linear Transformations

Theorem

The mapping $\mathbb{R}^n \ni \mathbf{x} \mapsto \mathbf{F}(\mathbf{x}) \in \mathbb{R}^m$ *is a linear transformation if and only if there exist vectors* $\mathbf{y}_i \in \mathbb{R}^m$ *for* $i = 1, 2, \ldots, n$ *such that each component function* $\mathbb{R}^n \ni \mathbf{x} \mapsto \mathbf{F}_i(\mathbf{x}) \in \mathbb{R}$ *satisfies* $F_i(\mathbf{x}) = \mathbf{y}_i^\top \mathbf{x}$.

Proof.

Sufficiency is obvious.

Conversely, recall that $\mathbf{x}$ equals the linear combination $\sum_{i=1}^n x_i \mathbf{e}_i$ of the n canonical basis vectors $\{\mathbf{e}_i\}_{i=1}^n$.

Then, because each component function $\mathbf{x} \mapsto F_i(\mathbf{x})$ is linear, one has

$$F_i(\mathbf{x}) = F_i\left(\sum_{j=1}^n x_j \mathbf{e}_j\right) = \sum_{j=1}^n x_j F_i(\mathbf{e}_j) = \mathbf{y}_i^\top \mathbf{x}$$

where, for each $i = 1, 2, \ldots, m$, the row vector $\mathbf{y}_i^\top$ has components $(\mathbf{y}_i)_j = F_i(\mathbf{e}_j)$ for $j = 1, 2, \ldots, n$. □

Representing a Linear Transformation by a Matrix

Definition

A matrix representation
of the linear transformation $\mathbb{R}^n \ni \mathbf{x} \mapsto \mathbf{F}(\mathbf{x}) \in \mathbb{R}^m$
relative to the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$
is an $m \times n$ array whose n columns
are the m-vector images $\mathbf{F}(\mathbf{e}_j) = (F_i(\mathbf{e}_j))_{i=1}^m \in \mathbb{R}^m$
of the n canonical basis vectors $\{\mathbf{e}_j\}_{j=1}^n$ of $\mathbb{R}^n$.

Outline

Linear Combinations and Dependence: Definitions

Definition

A linear combination of the finite set $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k\}$ of n-vectors is the scalar weighted sum $\sum_{h=1}^k \lambda_h \mathbf{x}^h$, where $\lambda_h \in \mathbb{R}$ for $h = 1, 2, \ldots, k$.

Definition

The finite set $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k\}$ of n-vectors is linearly independent just in case the only solution $(\lambda_1, \lambda_2, \ldots, \lambda_k) \in \mathbb{R}^k$ of the equation $\sum_{h=1}^k \lambda_h \mathbf{x}^\mathbf{h} = \mathbf{0}$ in $\mathbb{R}^n$ is the trivial solution $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$.

Alternatively, if the equation $\sum_{h=1}^k \lambda_h \mathbf{x}^\mathbf{h} = \mathbf{0}$ in $\mathbb{R}^n$ has a non-trivial solution $(\lambda_1, \lambda_2, \ldots, \lambda_k) \in \mathbb{R}^k$, then the set of vectors is linearly dependent.

Characterizing Linear Dependence

Theorem

The finite set $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k\}$ *of* k *n-vectors is linearly dependent if and only if at least one of the vectors, say* $\mathbf{x}^1$ *after reordering, can be expressed as a linear combination of the others*
— i.e., there exist scalars α^h $(h = 2, 3, \ldots, k)$
such that $\mathbf{x}^1 = \sum_{h=2}^k \alpha_h \mathbf{x}^h$.

Proof.

If $\mathbf{x}^1 = \sum_{h=2}^k \alpha_h \mathbf{x}^h$, then $(-1)\mathbf{x}^1 + \sum_{h=2}^k \alpha_h \mathbf{x}^h = \mathbf{0}$.

So $\sum_{h=1}^k \lambda_h \mathbf{x}^h = \mathbf{0}$ has a non-trivial solution $(\lambda_1, \lambda_2, \ldots, \lambda_k) \in \mathbb{R}^k$ with $\lambda_1 = -1 \neq 0$.

Conversely, suppose $\sum_{h=1}^k \lambda_h \mathbf{x}^h = \mathbf{0}$ has a non-trivial solution.

After reordering, we can suppose that $\lambda_1 \neq 0$.

Then $\mathbf{x}^1 = \sum_{h=2}^k \alpha_h \mathbf{x}^h$,
where $\alpha_h = -\lambda_h/\lambda_1$ for $h = 2, 3, \ldots, k$. □

Orthogonality Implies Linear Independence

Theorem

If the finite set $S = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k\}$ of k non-zero n-vectors is pairwise orthogonal, then it is linearly independent.

Proof.

Let $\mathbf{s} \in \mathbb{R}^n$ denote the linear combination $\sum_{h=1}^k \alpha_h \mathbf{x}^h$ of vectors in S.

Then for each $j = 1, \ldots, k$ one has $\mathbf{s} \cdot \mathbf{x}^j = \sum_{h=1}^k \alpha_h \mathbf{x}^h \cdot \mathbf{x}^j$.

In case S is pairwise orthogonal, one has $\mathbf{x}^h \cdot \mathbf{x}^j = 0$ for all $h \neq j$, and so $\mathbf{s} \cdot \mathbf{x}^j = \alpha_j \mathbf{x}^j \cdot \mathbf{x}^j$.

So when $\mathbf{s} = \mathbf{0}$, it follows that $\alpha_j \mathbf{x}^j \cdot \mathbf{x}^j = 0$ for all $j = 1, \ldots, k$.

Then, because we assumed that $\mathbf{x}^j \neq \mathbf{0}$, we have $\alpha_j = 0$ for all $j = 1, \ldots, k$.

This proves that S is linearly independent. □

Dimension

Definition

The dimension of a vector space V is the size of any maximal set of linearly independent vectors, if this number is finite.

Otherwise, if there is an infinite set of linearly independent vectors, the dimension is infinite.

Exercise

Show that the canonical basis of $\mathbb{R}^n$ *is linearly independent.*

Example

The previous exercise shows that the dimension of $\mathbb{R}^n$ is at least n.

Later results on the rank of a matrix will imply that any set of $k > n$ vectors in $\mathbb{R}^n$ is linearly dependent.

This implies that the dimension of $\mathbb{R}^n$ is exactly n.

Matrix Algebra in Six Segments

A. Introduction to Vectors: now completed

B. Introduction to Matrices and Determinants

C. Determinants and Gaussian Elimination

D. Determinants and Rank

E. Quadratic Forms and Their Definiteness

F. Eigenvalues and Eigenvectors