A Primer in Econometric Theory

Lecture 1: Vector Spaces

John Stachurski
Lectures by Akshay Shanker

May 5, 2017

Overview

Linear algebra is an important foundation for mathematics and, in particular, for Econometrics:

- performing basic arithmetic on data
- solving linear equations using data
- advanced operations such as quadratic minimisation

Focus of this chapter:

1. vector spaces: linear operations, norms, linear subspaces, linear independence, bases, etc.
2. orthogonal projection theorem

Vector Space

The symbol $\mathbb{R}^N$ represents set of all vectors of length N, or N vectors

An N-vector $\mathbf{x}$ is a tuple of N real numbers:

$$\mathbf{x} = (x_1, \ldots, x_N) \quad \text{where} \quad x_n \in \mathbb{R} \text{ for each } n$$

We can also write $\mathbf{x}$ vertically, like so:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

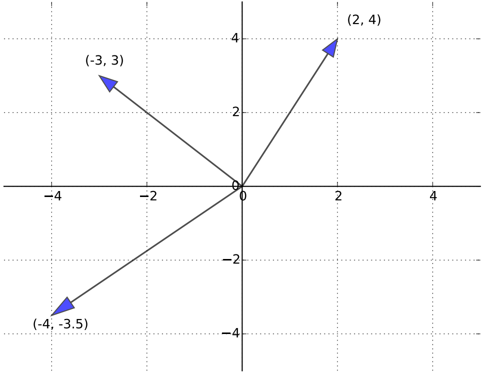


Figure: Three vectors in $\mathbb{R}^2$

The vector of ones will be denoted $\mathbf{1}$

$$\mathbf{1} := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Vector of zeros will be denoted $\mathbf{0}$

$$\mathbf{0} := \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Linear Operations

Two fundamental algebraic operations:

1. Vector addition
2. Scalar multiplication

1. **Sum** of $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{y} \in \mathbb{R}^N$ defined by

$$\mathbf{x} + \mathbf{y} :=: \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} := \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_N + y_N \end{pmatrix}$$

Example 1:

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix} := \begin{pmatrix} 3 \\ 6 \\ 9 \\ 12 \end{pmatrix}$$

Example 2:

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} := \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$$

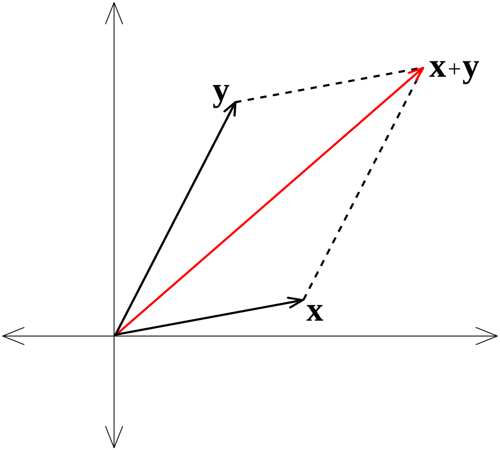


Figure: Vector addition

2. **Scalar product** of $\alpha \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^N$ defined by

$$\alpha \mathbf{x} = \alpha \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} := \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_N \end{pmatrix}$$

Example 1:

$$0.5\begin{pmatrix}1\\2\\3\\4\end{pmatrix} := \begin{pmatrix}0.5\\1.0\\1.5\\2.0\end{pmatrix}$$

Example 2:

$$-1\begin{pmatrix}1\\2\\3\\4\end{pmatrix} := \begin{pmatrix}-1\\-2\\-3\\-4\end{pmatrix}$$

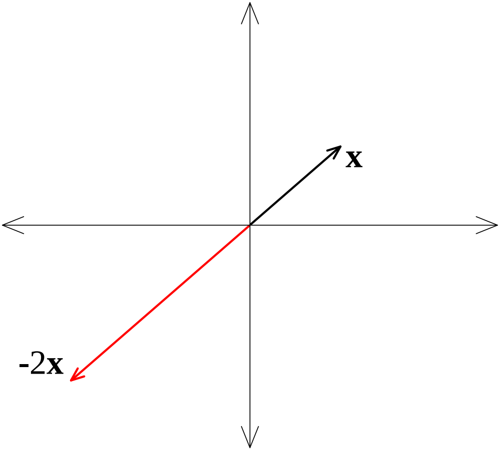


Figure: Scalar multiplication

Subtraction performed element by element, analogous to addition

$$\mathbf{x} - \mathbf{y} := \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_N - y_N \end{pmatrix}$$

Definition can be given in terms of addition and scalar multiplication

$$\mathbf{x} - \mathbf{y} := \mathbf{x} + (-1)\mathbf{y}$$

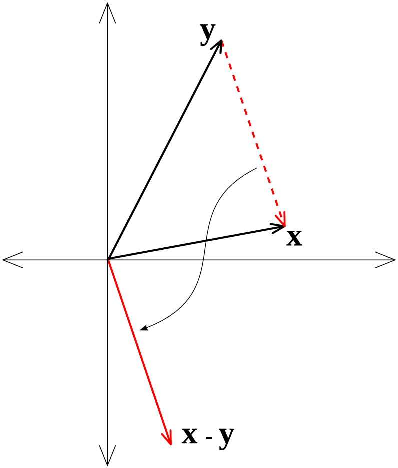


Figure: Difference between vectors

Inner Product

The **inner product** of two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^N$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$, and defined as the sum of the products of their elements:

$$\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{n=1}^{N} x_n y_n$$

Fact. (2.1.2)

For any $\alpha, \beta \in \mathbb{R}$ and any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, the following statements are true:

1. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$,
2. $\langle \alpha\mathbf{x}, \beta\mathbf{y} \rangle = \alpha\beta \langle \mathbf{x}, \mathbf{y} \rangle$, and
3. $\langle \mathbf{x}, \alpha\mathbf{y} + \beta\mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle + \beta \langle \mathbf{x}, \mathbf{z} \rangle$.

Properties easy to check using definitions of scalar multiplication and inner product

For example, to show 2., pick any $\alpha, \beta \in \mathbb{R}$ and any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$:

$$\langle \alpha\mathbf{x}, \beta\mathbf{y} \rangle = \sum_{n=1}^{N} \alpha x_n \beta y_n = \alpha\beta \sum_{n=1}^{N} x_n y_n = \alpha\beta \langle \mathbf{x}, \mathbf{y} \rangle$$

Norms and Distance

The (Euclidean) **norm** of $\mathbf{x} \in \mathbb{R}^N$ is defined as

$$\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

Interpretation:

- $\|\mathbf{x}\|$ represents the "length" of $\mathbf{x}$
- $\|\mathbf{x} - \mathbf{y}\|$ represents distance between $\mathbf{x}$ and $\mathbf{y}$

Fact. (2.1.2) For any $\alpha \in \mathbb{R}$ and any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, the following statements are true:

1. $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (**triangle inequality**)
4. $|\mathbf{x}'\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$ (**Cauchy-Schwarz inequality**)

Properties 1. and 2. are straight-forward to prove (exercise)

Property 4. is addressed in ET exercise 3.5.33

To show property 3, by properties of the inner product in fact 2.1.1

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle \\ &= \langle \mathbf{x}, \mathbf{x}\rangle + 2\langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{y}, \mathbf{y}\rangle \\ &\leq \langle \mathbf{x}, \mathbf{x}\rangle + 2|\langle \mathbf{x}, \mathbf{y}\rangle| + \langle \mathbf{y}, \mathbf{y}\rangle\end{aligned}$$

Apply the Cauchy–Schwarz inequality

$$\|\mathbf{x}+\mathbf{y}\|^2 \leq (\|\mathbf{x}\|+\|\mathbf{y}\|)^2$$

Taking the square root gives the triangle inequality

A **linear combination** of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_K$ in $\mathbb{R}^N$ is a vector

$$\mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_K \mathbf{x}_K$$

where $\alpha_1, \ldots, \alpha_K$ are scalars

Example.

$$0.5 \begin{pmatrix} 6.0 \\ 2.0 \\ 8.0 \end{pmatrix} + 3.0 \begin{pmatrix} 0 \\ 1.0 \\ -1.0 \end{pmatrix} = \begin{pmatrix} 3.0 \\ 4.0 \\ 1.0 \end{pmatrix}$$

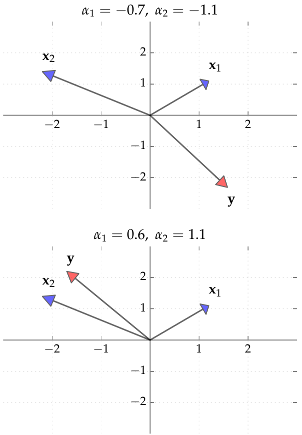


Figure: Linear combinations of $\mathbf{x}_1, \mathbf{x}_2$

Span

Let $X \subset \mathbb{R}^N$ be any nonempty set

Set of all possible linear combinations of elements of X is called the **span** of X, denoted by $\operatorname{span}(X)$

For finite $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ the span can be expressed as

$$\operatorname{span}(X) := \left\{ \text{all } \sum_{k=1}^{K} \alpha_k \mathbf{x}_k \text{ such that } (\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^K \right\}$$

Example. The four vectors labeled $\mathbf{y}$ in the previous figure lie in the span of $X = \{\mathbf{x}_1, \mathbf{x}_2\}$

Can *any* vector in $\mathbb{R}^2$ be created as a linear combination of $\mathbf{x}_1, \mathbf{x}_2$?

The answer is affirmative. We'll prove this in §2.1.5

Example. Let $X = \{\mathbf{1}\} \subset \mathbb{R}^2$, where $\mathbf{1} := (1, 1)$

The span of X is all vectors of the form

$$\alpha\mathbf{1} = \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} \quad \text{with} \quad \alpha \in \mathbb{R}$$

Constitutes a line in the plane that passes through

- the vector $\mathbf{1}$ (set $\alpha = 1$)
- the origin $\mathbf{0}$ (set $\alpha = 0$)

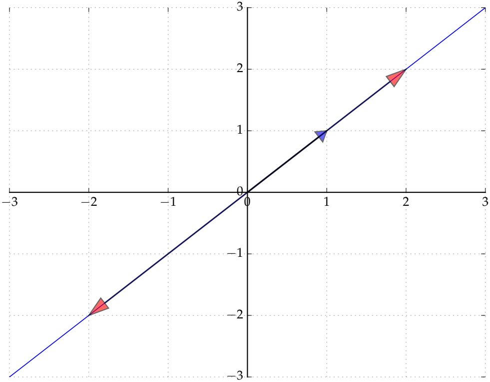


Figure: The span of $\mathbf{1} := (1, 1)$ in $\mathbb{R}^2$

Example. The set of canonical basis vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ is linearly independent in $\mathbb{R}^N$

Proof. Let $\alpha_1, \ldots, \alpha_N$ be coefficients such that $\sum_{k=1}^{N} \alpha_k \mathbf{e}_k = \mathbf{0}$

Equivalently,

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \sum_{k=1}^{N} \alpha_k \mathbf{e}_k = \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

In particular, $\alpha_k = 0$ for all k

Example. Let $\mathbf{x}_1 = (3, 4, 2)$ and let $\mathbf{x}_2 = (3, -4, 0.4)$

By definition, the span is all vectors of the form

$$\mathbf{y} = \alpha \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} + \beta \begin{pmatrix} 3 \\ -4 \\ 0.4 \end{pmatrix} \quad \text{where} \quad \alpha, \beta \in \mathbb{R}$$

This is a plane that passes through

- the vector $\mathbf{x}_1$
- the vector $\mathbf{x}_2$
- the origin $\mathbf{0}$

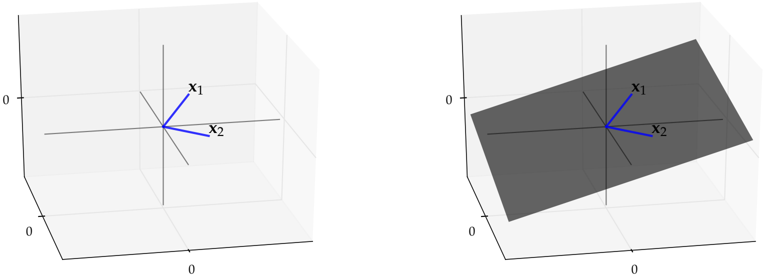


Figure: Span of $\mathbf{x}_1, \mathbf{x}_2$

Example. Consider the vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\} \subset \mathbb{R}^N$, where

$$\mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 := \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \mathbf{e}_N := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

That is, $\mathbf{e}_n$ has all zeros except for a 1 as the n-th element

Vectors $\mathbf{e}_1, \ldots, \mathbf{e}_N$ are called the **canonical basis vectors** of $\mathbb{R}^N$

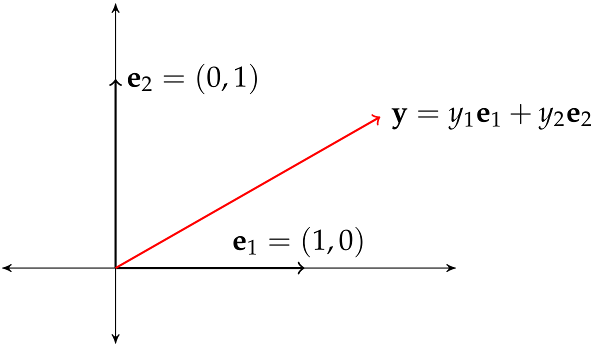


Figure: Canonical basis vectors in $\mathbb{R}^2$

Example. (cont.)

The span of $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ is equal to all of $\mathbb{R}^N$

Proof for $N = 2$:

Pick any $\mathbf{y} \in \mathbb{R}^2$, we have

$$\mathbf{y} := \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ y_1 \end{pmatrix}$$

$$= y_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = y_1 \mathbf{e}_1 + y_2 \mathbf{e}_2$$

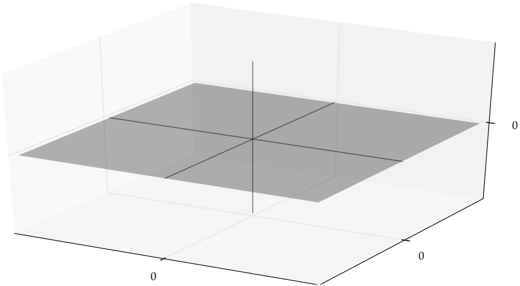
Thus, $\mathbf{y} \in \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\}$

Since $\mathbf{y}$ arbitrary, we have shown $\operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\} = \mathbb{R}^2$

Example. Consider the set

$$P := \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$$

Graphically, $P =$ flat plane in $\mathbb{R}^3$, where height coordinate $= 0$



Example. (cont.)

If $\mathbf{e}_1 = (1, 0, 0)$ and $\mathbf{e}_2 = (0, 1, 0)$, then $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\} = P$

To verify the claim, let $\mathbf{x} = (x_1, x_2, 0)$ be any element of P. We can write $\mathbf{x}$ as

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$$

In other words, $P \subset \text{span}\{\mathbf{e}_1, \mathbf{e}_2\}$

Conversely we have $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\} \subset P$ (why?)

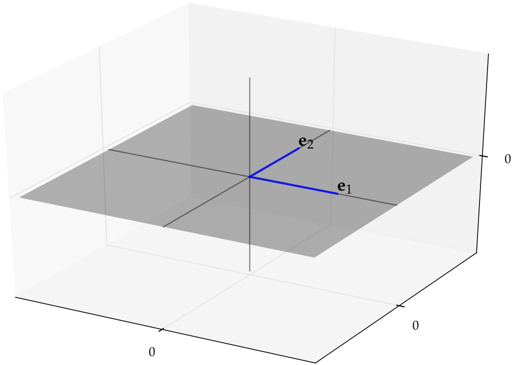


Figure: $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\} = P$

Fact. (2.1.3) If X and Y are non-empty subsets of $\mathbb{R}^N$ and $X \subset Y$, then $\text{span}(X) \subset \text{span}(Y)$

Proof. Pick any nonempty $X \subset Y \subset \mathbb{R}^N$

Let $\mathbf{z} \in \text{span}(X)$, we have

$$\mathbf{z} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k \text{ for some } \mathbf{x}_1, \ldots, \mathbf{x}_K \in X,\ \alpha_1, \ldots, \alpha_K \in \mathbb{R}$$

Proof.(cont.) Since $X \subset Y$, each $\mathbf{x}_k$ is also in Y, giving us

$$\mathbf{z} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k \text{ for some } \mathbf{x}_1, \ldots, \mathbf{x}_K \in Y,\ \alpha_1, \ldots, \alpha_K \in \mathbb{R}$$

Hence $\mathbf{z} \in \mathrm{span}(Y)$

Linear Independence

Important applied questions:

- When is a matrix invertible?
- When do regression arguments suffer from collinearity?
- When does a set of linear equations have a solution?

All of these questions closely related to linear independence

Definition

A nonempty collection of vectors $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\} \subset \mathbb{R}^N$ is called **linearly independent** if

$$\sum_{k=1}^{K} \alpha_k \mathbf{x}_k = \mathbf{0} \implies \alpha_1 = \cdots = \alpha_K = 0$$

Informally, linearly independent sets span large spaces

Example. Consider the two vectors $\mathbf{x}_1 = (1.2, 1.1)$ and $\mathbf{x}_2 = (-2.2, 1.4)$

Suppose α_1 and α_2 are scalars with

$$\alpha_1 \begin{pmatrix} 1.2 \\ 1.1 \end{pmatrix} + \alpha_2 \begin{pmatrix} -2.2 \\ 1.4 \end{pmatrix} = \mathbf{0}$$

This translates to a linear, two-equation system, where the unknowns are α_1 and α_2

The only solution is $\alpha_1 = \alpha_2 = 0$

$\{\mathbf{x}_1, \mathbf{x}_2\}$ is linearly independent

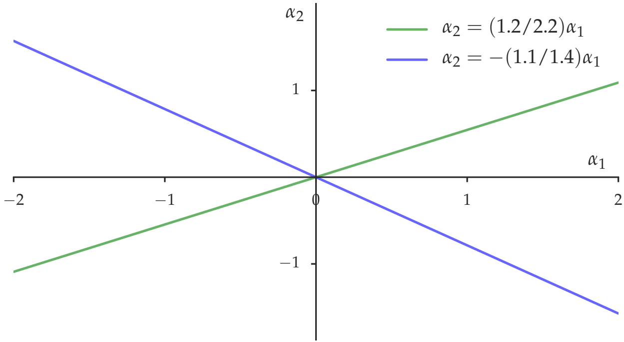


Figure: The only solution is $\alpha_1 = \alpha_2 = 0$

Example. The set of canonical basis vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ is linearly independent in $\mathbb{R}^N$

To see this, let $\alpha_1, \ldots, \alpha_N$ be coefficients such that $\sum_{k=1}^N \alpha_k \mathbf{e}_k = \mathbf{0}$. We have

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \sum_{k=1}^N \alpha_k \mathbf{e}_k = \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

In particular, $\alpha_k = 0$ for all k

Hence $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ linearly independent

Theorem. (2.1.1) Let $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\} \subset \mathbb{R}^N$. For $K > 1$, the following statements are equivalent:

1. X is linearly independent
2. X_0 is a proper subset of $X \implies \operatorname{span} X_0$ is a proper subset of $\operatorname{span} X$
3. No vector in X can be written as a linear combination of the others

Proof is an exercise. See ET ex. 2.4.15 and solution

Example. Dropping any of the canonical basis vectors reduces span

Consider the $N = 2$ case

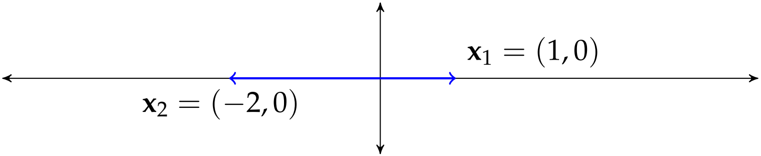
We know $\mathrm{span}\{\mathbf{e}_1, \mathbf{e}_2\} = \mathbb{R}^2$

- removing either element of $\mathrm{span}\{\mathbf{e}_1, \mathbf{e}_2\}$ reduces the span to a line

However, let $\mathbf{x}_1 = (1, 0)$ and $\mathbf{x}_2 = (-2, 0)$

The pair are not linearly independent since $\mathbf{x}_2 = -2\mathbf{x}_1$

- dropping either vector does not change the span—the span remains the horizontal axis
- we have $\mathbf{x}_2 = -2\mathbf{x}_1$, which means that each vector can be written as a linear combination of the other



Fact. (2.1.4) If $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ is linearly independent, then

1. every subset of X is linearly independent,
2. X does not contain $\mathbf{0}$, and
3. $X \cup \{\mathbf{x}\}$ is linearly independent for all $\mathbf{x} \in \mathbb{R}^N$ such that $\mathbf{x} \notin \operatorname{span} X$.

The proof is a solved exercise (ex. 2.4.16 in ET)

Linear Independence and Uniqueness

Linear independence is the key condition for existence *and* uniqueness of solutions to system of linear equations

Theorem. (2.1.2) Let $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ be any collection of vectors in $\mathbb{R}^N$. The following statements are equivalent:

1. X is linearly independent
2. For each $\mathbf{y} \in \mathbb{R}^N$ there exists at most one set of scalars $\alpha_1, \ldots, \alpha_K$ such that

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_K \mathbf{x}_K \tag{1}$$

Proof.(1. $\implies$ 2.)

Let X be linearly independent and pick any $\mathbf{y}$

Suppose by contradiction that (1) holds for more than one set of scalars; we have

$$\exists\ \alpha_1, \ldots, \alpha_K \text{ and } \beta_1, \ldots, \beta_K \ \text{ s.t. } \ \mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k = \sum_{k=1}^{K} \beta_k \mathbf{x}_k$$

$$\therefore \quad \sum_{k=1}^{K} (\alpha_k - \beta_k) \mathbf{x}_k = \mathbf{0}$$

$$\therefore \quad \alpha_k = \beta_k \quad \text{for all} \quad k$$

Proof.(2. $\implies$ 1.)

If 2. holds, then there exists at most one set of scalars such that

$$\mathbf{0} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k$$

Because $\alpha_1 = \cdots = \alpha_k = 0$ has this property, no nonzero scalars yield $\mathbf{0} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k$

We can then conclude X is linearly independent, by the definition of linear independence

Linear Subspaces

A nonempty subset S of $\mathbb{R}^N$ is called a **linear subspace** (or just **subspace**) of $\mathbb{R}^N$ if

$$\mathbf{x}, \mathbf{y} \in S \text{ and } \alpha, \beta \in \mathbb{R} \implies \alpha\mathbf{x} + \beta\mathbf{y} \in S$$

In other words, $S \subset \mathbb{R}^N$ is "closed" under vector addition and scalar multiplication

Example. If X is any nonempty subset of $\mathbb{R}^N$, then $\operatorname{span} X$ is a linear subspace of $\mathbb{R}^N$

Example. $\mathbb{R}^N$ is a linear subspace of $\mathbb{R}^N$

Example. Given any $\mathbf{a} \in \mathbb{R}^N$, the set $A := \{\mathbf{x} \in \mathbb{R}^N : \langle \mathbf{a}, \mathbf{x} \rangle = 0\}$ is a linear subspace of $\mathbb{R}^N$

To see this, let $\mathbf{x}, \mathbf{y} \in A$, let $\alpha, \beta \in \mathbb{R}$ and define $\mathbf{z} := \alpha \mathbf{x} + \beta \mathbf{y} \in A$

We have

$$\langle \mathbf{a}, \mathbf{z} \rangle = \langle \mathbf{a}, \alpha \mathbf{x} + \beta \mathbf{y} \rangle = \alpha \langle \mathbf{a}, \mathbf{x} \rangle + \beta \langle \mathbf{a}, \mathbf{y} \rangle = 0 + 0 = 0$$

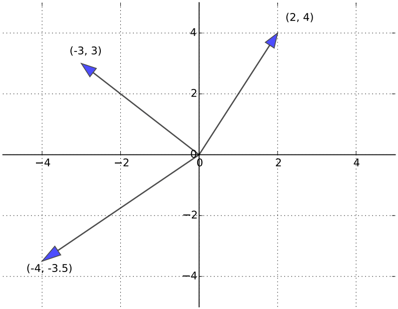
Hence $\mathbf{z} \in A$

Fact. (2.1.5) If S is a linear subspace of $\mathbb{R}^N$, then

1. $\mathbf{0} \in S$
2. $X \subset S \implies \operatorname{span} X \subset S$, and
3. $\operatorname{span} S = S$

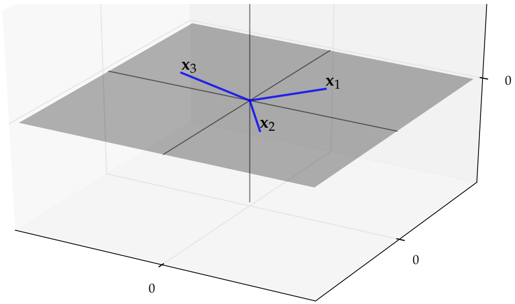
Theorem. (2.1.3) Let S be a linear subspace of $\mathbb{R}^N$. If S is spanned by K vectors, then any linearly independent subset of S has at most K vectors

Example. Recall the canonical basis vectors $\{\mathbf{e}_1, \mathbf{e}_2\}$ spanned $\mathbb{R}^2$. As such, from Theorem 2.1.3, the three vectors below are linearly dependent



Example. The plane $P := \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$ from example 2.1.5 in ET can be spanned by two vectors

By theorem 2.1.3, the three vectors in the figure below are linearly dependent



Bases and Dimension

Theorem. (2.1.4) Let $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ be any N vectors in $\mathbb{R}^N$. The following statements are equivalent:

1. $\operatorname{span} X = \mathbb{R}^N$
2. X is linearly independent

For a proof see page 22 in ET

Let S be a linear subspace of $\mathbb{R}^N$ and let $B \subset S$

The set B is called a **basis** of S if

1. B spans S and
2. B is linearly independent

The plural of basis is **bases**

From theorem 2.1.2, when B is a basis of S, each point in S has exactly one representation as a linear combination of elements of B

From theorem 2.1.4, any N linearly independent vectors in $\mathbb{R}^N$ form a basis of $\mathbb{R}^N$

Example. Recall the plane from the example above

$$P := \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$$

We showed $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\} = P$ for

$$\mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Moreover, $\{\mathbf{e}_1, \mathbf{e}_2\}$ is linearly independent (why?)

Hence $\{\mathbf{e}_1, \mathbf{e}_2\}$ is a basis for P

Theorem. (2.1.5) If S is a linear subspace of $\mathbb{R}^N$ distinct from $\{\mathbf{0}\}$, then

1. S has at least one basis and
2. every basis of S has the same number of elements.

If S is a linear subspace of $\mathbb{R}^N$, then the common number identified in theorem 2.1.5 is called the **dimension** of S, and written as $\dim S$

Example. For $P := \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$, $\dim P = 2$ because

$$\mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

is a basis with two elements

Example. A line $\{\alpha \mathbf{x} \in \mathbb{R}^N : \alpha \in \mathbb{R}\}$ through the origin is one dimensional

In $\mathbb{R}^N$ the singleton subspace $\{\mathbf{0}\}$ is said to have zero dimension

If we take a set of K vectors, then how large will its span be in terms of dimension?

Theorem. (2.1.6) If $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\} \subset \mathbb{R}^N$, then

1. $\dim \operatorname{span} X \leq K$ and
2. $\dim \operatorname{span} X = K$ if and only if X is linearly independent.

For a proof, see exercise 2.4.19 in ET

Fact. (2.1.6) The following statements are true:

1. Let S and S' be K-dimensional linear subspaces of $\mathbb{R}^N$. If $S \subset S'$, then $S = S'$
2. If S is an M-dimensional linear subspace of $\mathbb{R}^N$ and $M < N$, then $S \neq \mathbb{R}^N$

Linear Maps

A function $T \colon \mathbb{R}^K \to \mathbb{R}^N$ is called **linear** if

$$T(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha T\mathbf{x} + \beta T\mathbf{y} \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^K,\ \forall\, \alpha, \beta \in \mathbb{R}$$

Notation:

- Linear functions often written with upper case letters
- Typically omit parenthesis around arguments when convenient

Example. $T\colon \mathbb{R} \to \mathbb{R}$ defined by $Tx = 2x$ is linear

To see this, take any α, β, x, y in $\mathbb{R}$ and observe

$$T(\alpha x + \beta y) = 2(\alpha x + \beta y) = \alpha 2x + \beta 2y = \alpha Tx + \beta Ty$$

Example. The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is nonlinear

To see this, set $\alpha = \beta = x = y = 1$. We then have

$$f(\alpha x + \beta y) = f(2) = 4$$

However, $\alpha f(x) + \beta f(y) = 1 + 1 = 2$

Remark: Thinking of linear functions as those whose graph is a straight line is not correct

Example. Function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 1 + 2x$ is <u>non</u>linear

Take $\alpha = \beta = x = y = 1$. We then have

$$f(\alpha x + \beta y) = f(2) = 5$$

However, $\alpha f(x) + \beta f(y) = 3 + 3 = 6$

This kind of function is called an **affine** function

By definition, if T is linear, then the exchange of order in

$$T[\sum_{k=1}^{K} \alpha_k \mathbf{x}_k] = \sum_{k=1}^{K} \alpha_k T \mathbf{x}_k$$

will be valid whenever $K = 2$

Inductive argument extends this to arbitrary K

Fact. (2.1.7) If $T \colon \mathbb{R}^K \to \mathbb{R}^N$ is a linear map, then

$$\text{rng}(T) = \text{span}(V) \quad \text{where} \quad V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_K\}$$

where $\mathbf{e}_k$ is the k-th canonical basis vector in $\mathbb{R}^K$

Proof. Any $\mathbf{x} \in \mathbb{R}^K$ can be expressed as $\sum_{k=1}^K \alpha_k \mathbf{e}_k$. Hence $\text{rng}(T)$ is the set of all points of the form

$$T\mathbf{x} = T\left[\sum_{k=1}^K \alpha_k \mathbf{e}_k\right] = \sum_{k=1}^K \alpha_k T\mathbf{e}_k$$

as we vary $\alpha_1, \ldots, \alpha_K$ over all combinations. This coincides with the definition of $\text{span}(V)$

The **null space** or **kernel** of linear map $T \colon \mathbb{R}^K \to \mathbb{R}^N$ is

$$\ker(T) := \{\mathbf{x} \in \mathbb{R}^K : T\mathbf{x} = \mathbf{0}\}$$

Fact. (2.1.7) If $T \colon \mathbb{R}^K \to \mathbb{R}^N$ is a linear map, then

$$\operatorname{rng} T = \operatorname{span} V, \quad \text{where } V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_K\}$$

Proofs are straight-forward (complete as exercise)

Linear Independence and Bijections

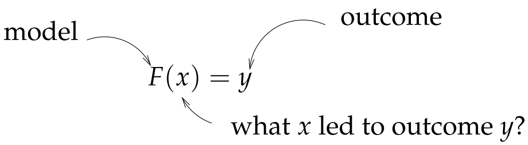
Many scientific and practical problems are “inverse” problems

- we observe outcomes but not what caused them
- how can we work backwards from outcomes to causes?

Examples

- what consumer preferences generated observed market behavior?
- what kinds of expectations led to given shift in exchange rates?

Loosely, we can express an inverse problem as



- does this problem have a solution?
- is it unique?

Answers depend on whether F is one-to-one, onto, etc.

The best case is a bijection

But other situations also arise

Theorem. (2.1.7) If T is a linear function from $\mathbb{R}^N$ to $\mathbb{R}^N$, then all of the following are equivalent:

1. T is a bijection.
2. T is onto.
3. T is one-to-one.
4. $\ker T = \{\mathbf{0}\}$.
5. $V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_N\}$ is linearly independent.
6. $V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_N\}$ forms a basis of $\mathbb{R}^N$.

See exercise 2.4.21 in ET for proof

If any one of these conditions is true, then T is called **nonsingular**. Otherwise T is called **singular**

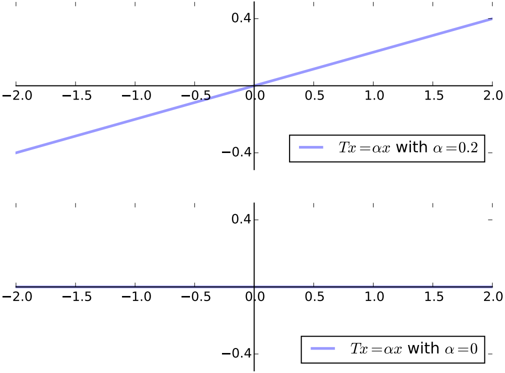


Figure: The case of $N = 1$, nonsingular and singular

If T is nonsingular, then, being a bijection, it must have an inverse function T^{-1} that is also a bijection (fact 15.2.1 on page 410)

Fact. (2.1.9) If $T\colon \mathbb{R}^N \to \mathbb{R}^N$ is nonsingular, then so is T^{-1}.

For a proof, see ex. 2.4.20

Maps Across Different Dimensions

Remember that the above results apply to maps from $\mathbb{R}^N$ to $\mathbb{R}^N$

Things change when we look at linear maps across dimensions

The general rules for linear maps are

- maps from lower to higher dimensions cannot be onto
- maps from higher to lower dimensions cannot be one-to-one

In either case they cannot be bijections

Theorem. (2.1.8) For a linear map T from $\mathbb{R}^K \to \mathbb{R}^N$, the following statements are true:

1. If $K < N$, then T is not onto.
2. If $K > N$, then T is not one-to-one.

Proof.(part 1)

Let $K < N$ and let $T \colon \mathbb{R}^K \to \mathbb{R}^N$ be linear

Letting $V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_K\}$, we have

$$\dim(\text{rng}(T)) = \dim(\text{span}(V)) \leq K < N$$

$$\therefore \quad \text{rng}(T) \neq \mathbb{R}^N$$

Hence T is not onto

Proof.(part 2)

Suppose to the contrary that T is one-to-one

Let $\alpha_1, \ldots, \alpha_K$ be a collection of vectors such that

$$\alpha_1 T\mathbf{e}_1 + \cdots + \alpha_K T\mathbf{e}_K = \mathbf{0}$$

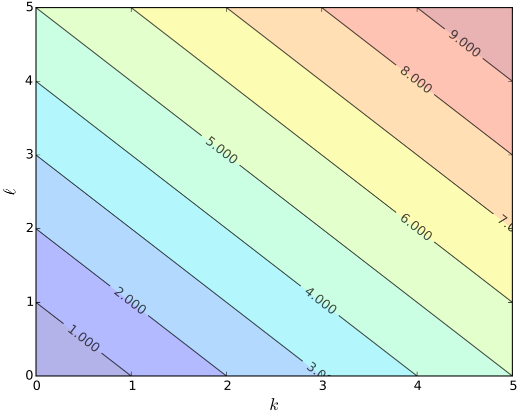
$$\therefore \quad T(\alpha_1 \mathbf{e}_1 + \cdots + \alpha_K \mathbf{e}_K) = \mathbf{0} \qquad \text{(by linearity)}$$

$$\therefore \quad \alpha_1 \mathbf{e}_1 + \cdots + \alpha_K \mathbf{e}_K = \mathbf{0} \qquad \text{(since } \ker(T) = \{\mathbf{0}\}\text{)}$$

$$\therefore \quad \alpha_1 = \cdots = \alpha_K = 0 \qquad \text{(by independence of } \{\mathbf{e}_1, \ldots \mathbf{e}_K\}\text{)}$$

We have shown that $\{T\mathbf{e}_1, \ldots, T\mathbf{e}_K\}$ is linearly independent

But then $\mathbb{R}^N$ contains a linearly independent set with $K > N$ vectors — contradiction



Example. Cost function $c(k, \ell) = rk + w\ell$ cannot be one-to-one

Orthogonal Vectors and Projections

A core concept in the course is orthogonality – not just of vectors, but random variables

Let $\mathbf{x}$ and $\mathbf{z}$ be vectors in $\mathbb{R}^N$

If $\langle \mathbf{x}, \mathbf{z} \rangle = 0$, then we call $\mathbf{x}$ and $\mathbf{z}$ **orthogonal**

Write $\mathbf{x} \perp \mathbf{z}$

In $\mathbb{R}^2$, orthogonal means perpendicular

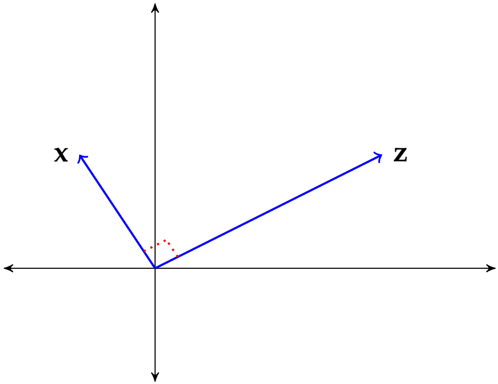


Figure: $\mathbf{x} \perp \mathbf{z}$

Let S be a linear subspace

We say that $\mathbf{x}$ **is orthogonal to** S if $\mathbf{x} \perp \mathbf{z}$ for all $\mathbf{z} \in S$

Write $\mathbf{x} \perp S$

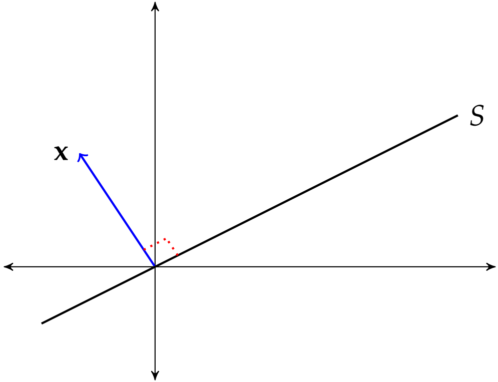


Figure: $\mathbf{x} \perp S$

Fact. (2.2.1) (Pythagorian law)

If $\{\mathbf{z}_1, \ldots, \mathbf{z}_K\}$ is an orthogonal set, then

$$\|\mathbf{z}_1 + \cdots + \mathbf{z}_K\|^2 = \|\mathbf{z}_1\|^2 + \cdots + \|\mathbf{z}_K\|^2$$

Proof is an exercise

Fact. (2.2.2) If $O \subset \mathbb{R}^N$ is an orthogonal set and $\mathbf{0} \notin O$, then O is linearly independent

An orthogonal set $O \subset \mathbb{R}^N$ is called an **orthonormal set** if $\|\mathbf{u}\| = 1$ for all $\mathbf{u} \in O$

An orthonormal set spanning a linear subspace S of $\mathbb{R}^N$ is an **orthonormal basis** of S

- example of an orthonormal basis for all of $\mathbb{R}^N$ is the canonical basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$

Fact. (2.2.3) If $\{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$ is an orthonormal set and $\mathbf{x} \in \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$, then

$$\mathbf{x} = \sum_{k=1}^{K} \langle \mathbf{x}, \mathbf{u}_k \rangle \, \mathbf{u}_k$$

Given $S \subset \mathbb{R}^N$, the **orthogonal complement** of S is

$$S^{\perp} := \{\mathbf{x} \in \mathbb{R}^N : \mathbf{x} \perp S\}$$

Fact. (2.2.4) For any nonempty $S \subset \mathbb{R}^N$, the set $S^{\perp}$ is a linear subspace of $\mathbb{R}^N$

Proof. If $\mathbf{x}, \mathbf{y} \in S^{\perp}$ and $\alpha, \beta \in \mathbb{R}$, then $\alpha\mathbf{x} + \beta\mathbf{y} \in S^{\perp}$ because, for any $\mathbf{z} \in S$

$$\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \beta \langle \mathbf{y}, \mathbf{z} \rangle = \alpha \times 0 + \beta \times 0 = 0$$

Fact. (2.2.5) For $S \subset \mathbb{R}^N$, we have $S \cap S^{\perp} = \{\mathbf{0}\}$

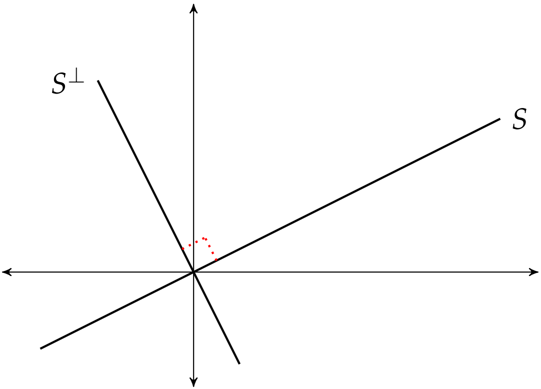


Figure: Orthogonal complement of S in $\mathbb{R}^2$

The Orthogonal Projection Theorem

Problem:

Given $\mathbf{y} \in \mathbb{R}^N$ and subspace S, find closest element of S to $\mathbf{y}$

Formally: Solve for

$$\hat{\mathbf{y}} := \operatorname*{argmin}_{\mathbf{z} \in S} \|\mathbf{y} - \mathbf{z}\| \tag{2}$$

Existence, uniqueness of solution not immediately obvious

Orthogonal projection theorem: $\hat{\mathbf{y}}$ always exists, unique

Also provides a useful characterization

Theorem. (2.2.1) [Orthogonal Projection Theorem I]

Let $\mathbf{y} \in \mathbb{R}^N$ and let S be any nonempty linear subspace of $\mathbb{R}^N$.

The following statements are true:

1. The optimization problem (2) has exactly one solution
2. $\hat{\mathbf{y}} \in \mathbb{R}^N$ solves (2) if and only if $\hat{\mathbf{y}} \in S$ and $\mathbf{y} - \hat{\mathbf{y}} \perp S$

The unique solution $\hat{\mathbf{y}}$ is called the **orthogonal projection of $\mathbf{y}$ onto S**

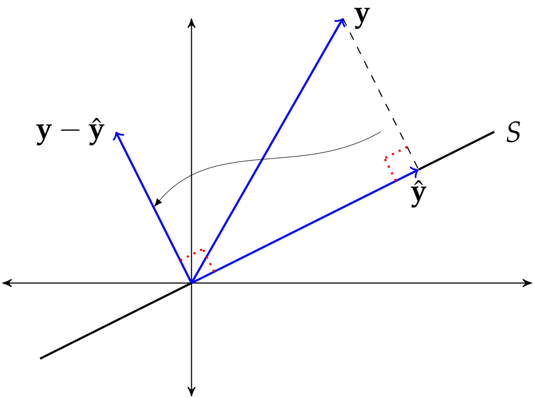


Figure: Orthogonal projection

Proof.(sufficiency of 2.) Let $\mathbf{y} \in \mathbb{R}^N$ and let S be a linear subspace of $\mathbb{R}^N$

Let $\hat{\mathbf{y}}$ be a vector in S satisfying $\mathbf{y} - \hat{\mathbf{y}} \perp S$

Let $\mathbf{z}$ be any point in S. We have

$$\|\mathbf{y} - \mathbf{z}\|^2 = \|(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{z})\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{z}\|^2$$

The second equality follows from $\mathbf{y} - \hat{\mathbf{y}} \perp S$ and the Pythagorian law

Since $\mathbf{z}$ was an arbitrary point in S, we have $\|\mathbf{y} - \mathbf{z}\| \geq \|\mathbf{y} - \hat{\mathbf{y}}\|$ for all $\mathbf{z} \in S$

Example. Let $\mathbf{y} \in \mathbb{R}^N$ and let $\mathbf{1} \in \mathbb{R}^N$ be the vector of ones

Let S be the set of constant vectors in $\mathbb{R}^N$— S is the span of $\{\mathbf{1}\}$

Orthogonal projection of $\mathbf{y}$ onto S is $\hat{\mathbf{y}} := \bar{y}\mathbf{1}$, where $\bar{y} := \frac{1}{N}\sum_{n=1}^N y_n$

Clearly, $\hat{\mathbf{y}} \in S$

To show $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to S, we need to check $\langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{1} \rangle = 0$ (see ex. 2.4.14 on page 36). This is true because

$$\langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{1} \rangle = \langle \mathbf{y}, \mathbf{1} \rangle - \langle \hat{\mathbf{y}}, \mathbf{1} \rangle = \sum_{n=1}^N y_n - \bar{y} \langle \mathbf{1}, \mathbf{1} \rangle = 0$$

Holding subspace S fixed, we have a functional relationship

$$\mathbf{y} \;\mapsto\; \text{its orthogonal projection } \hat{\mathbf{y}} \in S$$

This is a well-defined function from $\mathbb{R}^N$ to $\mathbb{R}^N$

The function is typically denoted by $\mathbf{P}$

- $\mathbf{P}(\mathbf{y})$ or $\mathbf{P}\mathbf{y}$ represents $\hat{\mathbf{y}}$

$\mathbf{P}$ is called the **orthogonal projection mapping onto S** and we write

$$\mathbf{P} = \operatorname{proj} S$$

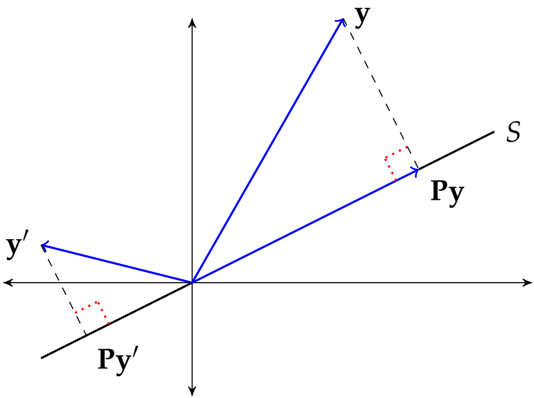


Figure: Orthogonal projection under $\mathbf{P}$

Theorem. (2.2.2) [Orthogonal Projection Theorem II] Let S be any linear subspace of $\mathbb{R}^N$, and let $\mathbf{P} = \operatorname{proj} S$. The following statements are true:

1. $\mathbf{P}$ is a linear function

Moreover, for any $\mathbf{y} \in \mathbb{R}^N$, we have

2. $\mathbf{Py} \in S$,
3. $\mathbf{y} - \mathbf{Py} \perp S$,
4. $\|\mathbf{y}\|^2 = \|\mathbf{Py}\|^2 + \|\mathbf{y} - \mathbf{Py}\|^2$,
5. $\|\mathbf{Py}\| \leq \|\mathbf{y}\|$,
6. $\mathbf{Py} = \mathbf{y}$ if and only if $\mathbf{y} \in S$, and
7. $\mathbf{Py} = \mathbf{0}$ if and only if $\mathbf{y} \in S^{\perp}$.

For a discussion of the proof, see page 31 and exercise 2.4.29

The following is a fundamental result

Fact. (2.2.6) If $\{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$ is an orthonormal basis for S, then, for each $\mathbf{y} \in \mathbb{R}^N$,

$$\mathbf{P}\mathbf{y} = \sum_{k=1}^{K} \langle \mathbf{y}, \mathbf{u}_k \rangle \mathbf{u}_k \tag{3}$$

Proof. First, the right-hand side of (3) lies in S since it is a linear combination of vectors spanning S

Next, we know $\mathbf{y} - \mathbf{P}\mathbf{y} \perp S$ if and only if $\mathbf{y} - \mathbf{P}\mathbf{y} \perp \mathbf{u}_j$ for each $\mathbf{u}_j$ in the basis set (exercise ex. 2.4.14)

For any $\mathbf{y} - \mathbf{P}\mathbf{y} \perp \mathbf{u}_j$, the following holds

$$\langle \mathbf{y} - \mathbf{P}\mathbf{y}, \mathbf{u}_j \rangle = \langle \mathbf{y}, \mathbf{u}_j \rangle - \sum_{k=1}^{K} \langle \mathbf{y}, \mathbf{u}_k \rangle \langle \mathbf{u}_k, \mathbf{u}_j \rangle$$

$$= \langle \mathbf{y}, \mathbf{u}_j \rangle - \langle \mathbf{y}, \mathbf{u}_j \rangle = 0$$

This confirms $\mathbf{y} - \mathbf{P}\mathbf{y} \perp S$

Fact. (2.2.7) Let S_i be a linear subspace of $\mathbb{R}^N$ for $i = 1, 2$ and let $\mathbf{P}_i = \operatorname{proj} S_i$. If $S_1 \subset S_2$, then

$$\mathbf{P}_1 \mathbf{P}_2 \mathbf{y} = \mathbf{P}_2 \mathbf{P}_1 \mathbf{y} = \mathbf{P}_1 \mathbf{y} \quad \text{for all } \mathbf{y} \in \mathbb{R}^N$$

The Residual Projection

Project $\mathbf{y}$ onto S, where S is a linear subspace of $\mathbb{R}^N$

- Closest point to $\mathbf{y}$ in S is $\hat{\mathbf{y}} := \mathbf{P}\mathbf{y}$ here $\mathbf{P} = \operatorname{proj} S$
- Unless $\mathbf{y}$ was already in S, some error $\mathbf{y} - \mathbf{P}\mathbf{y}$ remains

Introduce operator $\mathbf{M}$ that takes $\mathbf{y} \in \mathbb{R}^N$ and returns the residual

$$\mathbf{M} := \mathbf{I} - \mathbf{P} \tag{4}$$

where $\mathbf{I}$ is the identity mapping on $\mathbb{R}^N$

For any $\mathbf{y}$ we have $\mathbf{My} = \mathbf{Iy} - \mathbf{Py} = \mathbf{y} - \mathbf{Py}$

In regression analysis $\mathbf{M}$ shows up as a matrix called the "annihilator'

We refer to $\mathbf{M}$ as the **residual projection**

Example. Recall the projection of $\mathbf{y} \in \mathbb{R}^N$ onto span$\{\mathbf{1}\}$ is $\bar{y}\mathbf{1}$

The residual projection is $\mathbf{M}_c\mathbf{y} := \mathbf{y} - \bar{y}\mathbf{1}$

- vector of errors obtained when the elements of a vector are predicted by its sample mean

Fact. (2.2.8) Let S be a linear subspace of $\mathbb{R}^N$, let $\mathbf{P} = \operatorname{proj} S$, and let $\mathbf{M}$ be the residual projection as defined in (4). The following statements are true:

1. $\mathbf{M} = \operatorname{proj} S^{\perp}$
2. $\mathbf{y} = \mathbf{P}\mathbf{y} + \mathbf{M}\mathbf{y}$ for any $\mathbf{y} \in \mathbb{R}^N$
3. $\mathbf{P}\mathbf{y} \perp \mathbf{M}\mathbf{y}$ for any $\mathbf{y} \in \mathbb{R}^N$
4. $\mathbf{M}\mathbf{y} = \mathbf{0}$ if and only if $\mathbf{y} \in S$
5. $\mathbf{P} \circ \mathbf{M} = \mathbf{M} \circ \mathbf{P} = \mathbf{0}$

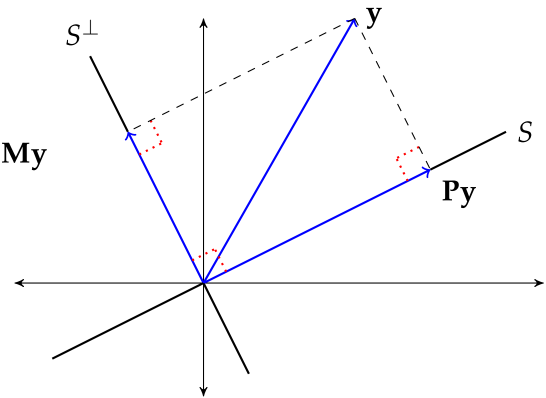


Figure: The residual projection

If S_1 and S_2 are two subspaces of $\mathbb{R}^N$ with $S_1 \subset S_2$, then $S_2^\perp \subset S_1^\perp$

The result in fact 2.2.7 is reversed for $\mathbf{M}$

Fact. (2.2.9) Let S_1 and S_2 be two subspaces of $\mathbb{R}^N$ and let $\mathbf{y} \in \mathbb{R}^N$. Let $\mathbf{M}_1$ and $\mathbf{M}_2$ be the projections onto $S_1^\perp$ and $S_2^\perp$ respectively. If $S_1 \subset S_2$, then

$$\mathbf{M}_1\mathbf{M}_2\mathbf{y} = \mathbf{M}_2\mathbf{M}_1\mathbf{y} = \mathbf{M}_2\mathbf{y}$$

Gram– Schmidt Orthogonalization

Recall we showed every orthogonal subset of $\mathbb{R}^N$ not containing $\mathbf{0}$ is linearly independent – fact 2.2.2

Here is an (important) partial converse

Theorem. (2.2.3) For each linearly independent set $\{\mathbf{b}_1, \ldots, \mathbf{b}_K\} \subset \mathbb{R}^N$, there exists an orthonormal set $\{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$ with

$$\operatorname{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\} = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \quad \text{for } k = 1, \ldots, K$$

Formal proofs are solved as exercises 2.4.34 to 2.4.36

The proof provides an important algorithm for generating the orthonormal set $\{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$

The first step is to construct orthogonal sets $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ with span identical to $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ for each k

The construction of $\{\mathbf{v}_1, \ldots, \mathbf{v}_K\}$ uses the **Gram–Schmidt orthogonalization** procedure:

For each $k = 1, \ldots, K$, let

1. $B_k := \text{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$,
2. $\mathbf{P}_k := \text{proj}\, B_k$ and $\mathbf{M}_k := \text{proj}\, B_k^{\perp}$,
3. $\mathbf{v}_k := \mathbf{M}_{k-1}\mathbf{b}_k$ where $\mathbf{M}_0$ is the identity mapping, and
4. $V_k := \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$.

In step 3. we map each successive element $\mathbf{b}_k$ into a subspace orthogonal to the subspace generated by $\mathbf{b}_1, \ldots, \mathbf{b}_{k-1}$

To complete the argument, define $\mathbf{u}_k$ by $\mathbf{u}_k := \mathbf{v}_k / \|\mathbf{v}_k\|$

The set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is orthonormal with span equal to V_k