10. Euclidean Spaces

Many of the spaces used in traditional consumer, producer, and general equilibrium theory will be Euclidean spaces—spaces where Euclid's geometry rules.[1]

At this point, we have to start being a little more careful how we write things. We will start with the space $\mathbb{R}^n$, the space of n-vectors, n-tuples of real numbers. When we are being picky, we write them vertically as if they are $n \times 1$ matrices.

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

This is the same way we wrote them when writing linear systems as matrix products, $\mathbf{Ax} = \mathbf{b}$. The x_i are referred to as the *coordinates* of the vector $\mathbf{x}$.

[1] Euclid (ca. 300 BC) was a Greek mathematician known for his treatise on geometry, the *Elements*. Although we don't even know when he was born or died, his treatise on geometry dominated the subject for over 2000 years. Euclidean geometry is still taught in high school. The *Elements* included both his own work and that of previous Greek geometers.

10.1 Vectors Vertical and Covectors Horizontal

Sometimes, especially in text, we will be informal and write vectors horizontally, but this is not strictly correct. I will sometimes attach a transpose symbol to remind you they should be vertical. Truly horizontal "vectors" are called *co-vectors* or *covariant vectors*. They differ from vectors in how they transform if you change coordinates. Ordinary vectors are sometimes called *contravariant vectors*.

We can see this distinction when thinking of commodity vectors and their associated price vectors. Suppose we decide to measure milk in quarts rather than gallons. The quantities of all the different types of milk (skim, 2% milkfat, whole, chocolate, etc.) would all have to be multiplied by 4 to reflect the change in measurement because there are 4 quarts in a gallon.

This is not how this measurement change affects prices. No, no, no! Price "vectors" are actually covectors, and if milk costs 2.40 per gallon, that's 0.60 per quart. We have to **divide** the prices by 4 rather than **multiplying** by 4. This is the essence of the distinction between vectors and covectors.

If we were being really picky, which we won't, the vectors would use superscripts for their coordinates and the covectors would use subscripts. Although that is a useful convention in geometry and physics, it conflicts with other useful conventions in economics. One such is to use superscripts to indicate ownership—which consumer a commodity vector belongs to or which firm is using inputs and producing outputs.

10.2 Vectors

In $\mathbb{R}^2$, a vector $\boldsymbol{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ looks like this:

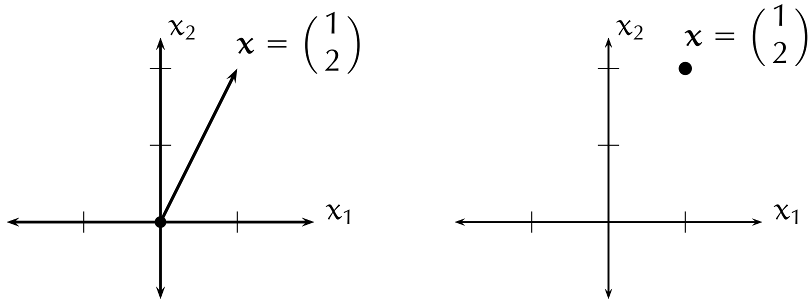


Figure 10.2.1:

Although we will mostly treat vectors as points in the plane, as in the right panel of Figure 10.2.1, it is often useful to think of them as indicating a direction, which $\boldsymbol{x}$ does in the left panel of Figure 10.2.1. In mathematics this is sometimes done by using *vector bundles*, which include the starting point as well as the direction.

10.3 Vector Addition

Algebraically, vector addition is just matrix addition. We add the components.

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

When we add vectors on a diagram, we add them nose to tail, placing the starting point of the second vector at the end of the first. Because vector addition commutes, it doesn't matter which order we use. We obtain the same vector for the sum.

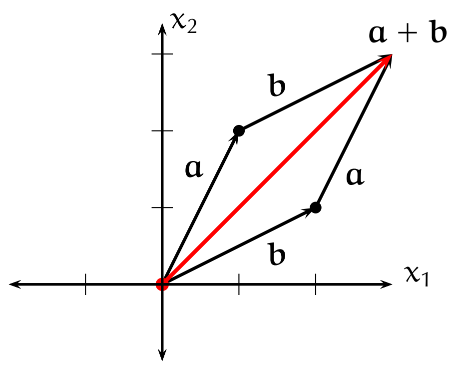


Figure 10.3.1: Two ways to add vectors: Here we add $\mathbf{a} = (1, 2)^{\mathsf{T}}$ to $\mathbf{b} = (2, 1)^{\mathsf{T}}$. The upper combination is $\mathbf{a} + \mathbf{b}$ and the lower $\mathbf{b} + \mathbf{a}$. Of course, both end at the same point, $(3, 3)^{\mathsf{T}}$, because matrix and thus vector addition are commutative. The red vector is the sum $(\mathbf{a} + \mathbf{b})$. In fact, we can read off the diagram that $\mathbf{a} = (1, 2)$ and $\mathbf{b} = (2, 1)$.

10.4 Scalar Multiplication

We can also multiply vectors by scalars. We still use the rules of matrix algebra so

$$\alpha \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{pmatrix}.$$

The diagram illustrates this graphically.

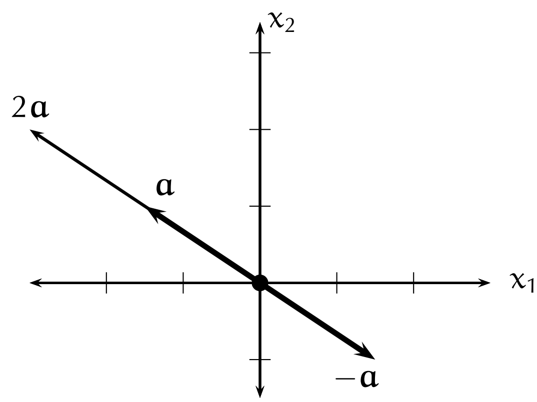


Figure 10.4.1: Vector Multiplication: Here $\mathbf{a} = (-1.5, 1)^T$ is multiplied by 2. Multiplying by a larger number would extend the line further. Multiplying by a smaller number would shrink toward the origin. Finally, multiplying by a negative number goes in the opposite direction as illustrated by $-\mathbf{a}$.

10.5 Coordinate Vectors in $\mathbb{R}^n$

The *standard basis vectors in* $\mathbb{R}^n$, $\mathbf{e}_k$, are defined by $\mathbf{e}_k = (\delta_{ik})$, so

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

These vectors are also referred to as the *canonical basis vectors*.

We can write any vector $\mathbf{x}$ in $\mathbb{R}^n$ as a sum of the canonical basis vectors,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n = \sum_{i=1}^{n} x_i\mathbf{e}_i.$$

This sum can also be written as a matrix product.

$$\begin{aligned} \mathbf{x} &= \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= \mathbf{I}_n\mathbf{x}. \end{aligned}$$

10.6 Linear Transformations

We showed in Theorem 13.6.1 that if any $m \times n$ matrix $\boldsymbol{A}$ defines a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ by $T_{\boldsymbol{A}}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$. We now show the converse, that any linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ can be written in matrix form.

Theorem 10.6.1. *Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there is an $m \times n$ matrix $\boldsymbol{A}$ with $T = T_{\boldsymbol{A}}$.*

Proof. We can write $\boldsymbol{x} = \sum_{j=1}^n x_j \boldsymbol{e}_j$. Then $T(\boldsymbol{x}) = \sum_{j=1}^n x_j T(\boldsymbol{e}_j)$ for any vector $\boldsymbol{x}$.

The vector $T(\boldsymbol{e}_j)$ is in $\mathbb{R}^m$. We denote its i^{th} component by $T(\boldsymbol{e}_j)_i$, $i = 1, \ldots, m$. Define the $m \times n$ matrix $\boldsymbol{A}$ by setting $a_{ij} = T(\boldsymbol{e}_j)_i$. Thus $\boldsymbol{A} = \left[T(\boldsymbol{e}_j)_i\right]$.

With this definition of $\boldsymbol{A}$,

$$T(\boldsymbol{x})_i = T\left(\sum_{j=1}^n x_j \boldsymbol{e}_j\right)_i = \sum_{j=1}^n x_j T(\boldsymbol{e}_j)_i = \sum_{j=1}^n a_{ij} x_j,$$

so $T(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$. This shows that $T = T_{\boldsymbol{A}}$. ∎

This means that the matrix

$$\boldsymbol{A} = \begin{pmatrix} T(\boldsymbol{e}_1) & T(\boldsymbol{e}_2) & \cdots & T(\boldsymbol{e}_n) \end{pmatrix}$$

represents T.

10.7 Matrix Representations of Linear Transformations

▶ Example 10.7.1: Linear Transformation as Matrix. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation with

$$T(\mathbf{e}_1) = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} \text{ and } T(\mathbf{e}_2) = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}.$$

Then it can be represented by the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 3 & 2 \end{pmatrix}$$

so that

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \begin{pmatrix} x_1 - x_2 \\ x_2 \\ 3x_1 + 2x_2 \end{pmatrix} \\ &= x_1 \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} \\ &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2). \end{aligned}$$

◀

10.8 Row Vectors vs. Column Vectors

I'm insisting on writing vectors as columns. One reason I gave earlier was that we can then write our linear systems with coefficient matrix $\mathbf{A}$ as $\mathbf{A}\mathbf{x} = \mathbf{b}$.

But you might object that you could just take the transpose, obtaining $\mathbf{x}^\mathsf{T}\mathbf{A}^\mathsf{T} = \mathbf{b}^\mathsf{T}$, and redefine the coefficient matrix to be $\mathbf{A}^\mathsf{T}$. That may seen slightly unnatural because we would always put the variable on the left, but equations are sometimes written that way, particular in some areas of algebra.

The real issue is that we need to maintain a distinction between the two types of vectors.

10.9 Linear Functionals

Suppose we had a linear function f that maps $\mathbb{R}^n$ into $\mathbb{R}$. Such a linear function is called a *linear functional*. As we just saw, that kind of linear function can be represented by a $1 \times n$ matrix

$$\mathbf{A} = \begin{pmatrix} f(\boldsymbol{e}_1) & f(\boldsymbol{e}_2) & \cdots & f(\boldsymbol{e}_n) \end{pmatrix}.$$

If we write the vectors as columns, the linear functionals become row vectors.

Thus the row (covariant) vector $(1, 3, 0, 1)$ defines a linear functional on (contravariant) vectors in $\mathbb{R}^4$ by

$$\begin{pmatrix} 1 & 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_1 + 3x_2 + x_4.$$

10.10 Covariant and Contravariant

Which one is covariant and which is contravariant is arbitrary. But the weight of mathematical tradition, stretching back at least to the 19th century, is to make the regular vectors columns, and the linear functionals rows. It's implicit in the Ricci calculus that Einstein used to develop the general theory of relativity.

In economics, vectors of goods should be represented as columns, but price vectors are properly represented as rows. Thus each set of prices defines a linear functional on the space of goods. If we purchase two different commodity bundles, the cost of them is the sum of the costs, and if we purchase a multiple of a particular commodity bundle under the same price system, its cost is the same multiple of the original. The cost of a bundle is both additive and multiplicative, so it is a linear functional on the commodity space.

10.11 Vector Spaces: Definition

Both $\mathbb{R}^n$ and $\mathbb{C}^n$, the set of complex n-tuples or n-vectors, are examples of vector spaces. But what are vector spaces? They are sets of things called *vectors* that can be added together and multiplied by scalars (numbers). The addition and multiplication has to obey certain rules.

Vector Space. A *vector space* $(V, +, \cdot)$ over a field $\mathbb{F}$ (the set of *scalars*) is a set of *vectors* V with two operations: vector addition, which defines $\mathbf{x} + \mathbf{y} \in V$ for all vectors $\mathbf{x}, \mathbf{y} \in V$; and scalar multiplication, which defines $\alpha\mathbf{x}$ for any scalar $\alpha \in \mathbb{F}$ and any vector $\mathbf{x} \in V$. Scalar multiplication and vector addition obey the following properties:

1. For all $\mathbf{x}, \mathbf{y} \in V$, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (*addition commutes*).
2. For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (*addition associates*).
3. There exists a unique vector $\mathbf{0} \in V$ with $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$ (*additive identity*).
4. For each $\mathbf{x} \in V$, there is a unique vector $-\mathbf{x} \in V$ with $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$ (*additive inverse*).
5. For all $\mathbf{x} \in V$, $1\mathbf{x} = \mathbf{x}$ (*multiplicative identity*).
6. For all $\alpha, \beta \in \mathbb{F}$ and $\mathbf{x} \in V$, $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$ (*scalar multiplication associates*).
7. For all $\alpha \in \mathbb{F}$ and $\mathbf{x}, \mathbf{y} \in V$, $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ (*distributive law I*).
8. For all $\alpha, \beta \in \mathbb{F}$ and $\mathbf{x} \in V$, $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ (*distributive law II*).

10.12 Making Vector Spaces

A vector space can be defined over any field $\mathbb{F}$, such as the rational numbers, the real numbers, or the complex numbers.[2] Vector spaces over the real numbers are called *real vector spaces* and vector spaces over the complex numbers are called *complex vector spaces*. The vector space $\mathbb{R}^n$ is the set of all n-tuples of real numbers, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^\mathsf{T}$. Here vector addition is defined componentwise, with $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)^\mathsf{T}$ and scalar multiplication is given by $\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n)^\mathsf{T}$.

Vector spaces do not have to be n-tuples of numbers. They can be anything that we can add and multiply by a scalar in a way that obeys the vector space axioms. In fact, this can be done in a completely arbitrary fashion. We can define a vector space as linear combinations of $\{\mathrm{Red}, \mathrm{Green}, \mathrm{Blue}\}$ by interpreting linear combinations as sums of the three elements without attaching any concrete meaning to them.[3]

[2] There are many other fields and many other types of vector spaces. There's even a related concept, modules, that use rings instead of fields.

[3] Such constructions are called *free vector spaces*. I presume this is because the linear combinations are free of any actual meaning. Such constructions are used when you want to make sure you aren't smuggling in any special assumptions.

10.13 More Vector Spaces

Spaces of continuous functions are one type of vector space commonly used in economics.

▶ Example 10.13.1: Spaces of Continuous Functions. If A is a set, the set of real-valued continuous functions on A is a vector space with the usual addition of functions and multiplication by real numbers: $(f + g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$ and $(\alpha f)(\mathbf{x}) = \alpha f(\mathbf{x})$.

Interestingly, the set of complex-valued functions on A can be regarded as either a real or complex vector space.

If A is an open set, we can consider the vector space of continuously differentiable, or even k-times continuously differentiable functions on A. These are denoted $\mathcal{C}^k(A)$. The space of continuous functions on A is denoted $\mathcal{C}^0(A)$ or sometimes just $\mathcal{C}(A)$. ◀

Sequence spaces are also commonly used in economics.

▶ Example 10.13.2: A Sequence Space. Another commonly used vector space is $\mathbf{s}$, the space of sequences of real numbers. Elements of $\mathbf{s}$ are of the form $(x_1, x_2, x_3, \dots)$ where each $x_i \in \mathbb{R}$. Vectors are added componentwise, and scalar multiplication is defined componentwise. This and similar spaces often appear in optimal growth models and the dynamic general equilibrium models used in macroeconomics. ◀

10.14 Vector Subspaces

We can make more vector spaces by taking subspaces of existing vector spaces.

Suppose W is a subset of a vector space V. The set W is a *vector subspace* of V if $W \subset V$ is closed under the vector space operations of vector addition and scalar multiplication (that is, $\mathbf{x} + \mathbf{y} \in W$ when $\mathbf{x}, \mathbf{y} \in W$ and $\alpha\mathbf{x} \in W$ when $\mathbf{x} \in W$ and $\alpha \in \mathbb{F}$).

When W is a vector subspace of V it can also be regarded as a vector space in its own right as it inherits all the vector space properties of V. Thus $W = \{\mathbf{x} \in \mathbb{R}^n : x_1 + 3x_2 = 0\}$ is a real vector subspace of $\mathbb{R}^n$ and hence a real vector space. Many vector spaces of interest will be subspaces of larger vector spaces.

The space $\mathcal{C}^k(A)$ of k-times continuously differentiable functions on A is a subspace of $\mathcal{C}^0(A)$ and of any $\mathcal{C}^p(A)$ for integers p, $0 < p < k$.

Both requirements to be in a subspace can be shown simultaneously.

Theorem 10.14.1. *Let V be a vector space over $\mathbb{F}$. If $\alpha\mathbf{x} + \mathbf{y} \in W$ for any $\mathbf{x}, \mathbf{y} \in W \subset V$ and any $\alpha \in \mathbb{F}$, then W is a vector subspace of V.*

Proof. We need only show that any $\mathbf{x} + \mathbf{y} \in W$ and any $\alpha\mathbf{x} \in W$ for $\alpha \in \mathbb{F}$ and $\mathbf{x}, \mathbf{y} \in W$. For the first, set $\alpha = 1$ in the hypothesis. For the second, we must first show that $\mathbf{0} \in W$. Set $\mathbf{x} = \mathbf{y}$ and $\alpha = -1$ to find that $\mathbf{0} = -\mathbf{x} + \mathbf{x} \in W$ Then set $\mathbf{y} = \mathbf{0}$ in the hypothesis. ∎

10.15 Subspaces of Sequence Spaces

▶ Example 10.15.1: Subspaces of Sequence Spaces. Some vector subspaces of $\mathbf{s}$ include:

1. $\mathbf{c}$, the space of convergent sequences, sequences where $\lim_n x_n$ exists.
2. $\mathbf{c}_0$, the space of sequences that converge to 0, meaning $\lim_n x_n = 0$.
3. $\mathbf{c}_{00}$, the space of sequences that are zero except for finitely many terms. That means that for any given $\mathbf{x} \in \mathbf{c}_{00}$, there is some N with $x_n = 0$ for $n \geq N$).

It is straightforward to verify these are all subspaces of $\mathbf{s}$.

The sequences (vectors)

$$\{1, 8, 27, \ldots, n^3, \ldots\} \quad \text{and} \quad \{1, e, e^2, \ldots, e^n, \ldots\}$$

are in $\mathbf{s}$, but not $\mathbf{c}$. The possibility of arbitrarily increasing growth rates is why we cannot find a norm for $\mathbf{s}$.

The sequences

$$\{2, 3/2, 4/3, \ldots, 1 + 1/n, \ldots\}$$
$$\{1, 1/4, 1/9, \ldots, 1/n^2, \ldots\}$$
$$\{1, 2, 3, 4, 0, \ldots, 0, \ldots\}$$

are respectively in $\mathbf{c}$, but not $\mathbf{c}_0$; $\mathbf{c}_0$ but not $\mathbf{c}_{00}$; $\mathbf{c}_{00}$. These spaces are nested, so

$$\mathbf{c}_{00} \subset \mathbf{c}_0 \subset \mathbf{c} \subset \mathbf{s}$$

The set $\mathbf{s}_b$ of sequences that are bounded, but not necessarily convergent is another sequence space that sits between $\mathbf{c}$ and $\mathbf{s}$. All of these spaces except for $\mathbf{s}$ can be normed by $\|\mathbf{x}\|_\infty = \sup_n |x_n|$ where sup denotes the supremum, the least upper bound. ◀

10.16 The Null Space (Kernel) of a Linear System 09/08/22

Another way to generate a subspace is to start with a homogeneous linear system defined by an $m \times n$ coefficient matrix $\mathbf{A}$, $\mathbf{A}\mathbf{x} = \mathbf{0}$. Recall that $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ is given by $T_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$.

The solution set of the homogeneous system,

$$\mathcal{N}(\mathbf{A}) = \ker \mathbf{A} = \{\mathbf{x} : T_{\mathbf{A}}(\mathbf{x}) = \mathbf{0}\},$$

is a vector subspace of $\mathbb{R}^n$. It is called the *null space* (symbol $\mathcal{N}(\mathbf{A})$, preferred by Simon and Blume) or the *kernel* of $\mathbf{A}$ (symbol $\ker \mathbf{A}$).

Theorem 10.16.1. *For any $m \times n$ coefficient matrix $\mathbf{A}$, consider the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$. The set of solutions to this system forms a vector subspace of $\mathbb{R}^n$, $\ker \mathbf{A}$.*

Proof. Let the solution set be V. Suppose $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{R}$. Then $\mathbf{A}(\alpha\mathbf{x} + \mathbf{y}) = \alpha\mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{y} = \mathbf{0}$, so any $\alpha\mathbf{x} + \mathbf{y} \in V$, showing that $V \subset \mathbb{R}^n$ is a vector subspace of $\mathbb{R}^n$. ∎

Theorem 10.16.1 also gives us information about the solutions to the system $\mathbf{A}\mathbf{x} = \mathbf{b}$. If a solution exists, call it $\mathbf{x}_0$. Then $\mathbf{A}\mathbf{x} = \mathbf{b} = \mathbf{A}\mathbf{x}_0$. Subtracting, we find $\mathbf{A}(\mathbf{x} - \mathbf{x}_0) = \mathbf{0}$. In other words, if $\mathbf{x}_0$ is a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$, then every other solution can be written $\mathbf{x} + \mathbf{x}_0$ where $\mathbf{x} \in \ker \mathbf{A}$. The solution set is $\{\mathbf{x}_0\} + \ker \mathbf{A}$.

10.17 The Range of a Linear Transformation

The kernel is not the only subspace associated with a linear transformation. Another is its range.

Theorem 10.17.1. *Let* $T\colon \mathbb{R}^n \to \mathbb{R}^m$ *be a linear transformation. Then* $\operatorname{ran} T$ *is a vector subspace of* $\mathbb{R}^m$.

Proof. Clearly $\operatorname{ran} T \subset \mathbb{R}^m$. Let $\mathbf{y}, \mathbf{y}' \in \operatorname{ran} T$. Then there are $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$ with $T(\mathbf{x}) = \mathbf{y}$ and $T(\mathbf{x}') = \mathbf{y}'$. Let α be a scalar. Then $T(\alpha\mathbf{x} + \mathbf{x}') = \alpha T(\mathbf{x}) + T(\mathbf{x}') = \alpha\mathbf{y} + \mathbf{y}'$, showing that $\alpha\mathbf{y} + \mathbf{y}' \in \operatorname{ran} T$. This shows that $\operatorname{ran} T$ is a subspace of $\mathbb{R}^m$. ∎

In particular, this applies to linear transformations defined by matrices, $T_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$.

Corollary 10.17.2. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix and define the linear function* $T_{\mathbf{A}}\colon \mathbb{R}^n \to \mathbb{R}^m$ *by* $T_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. *Then* $\operatorname{ran} T_{\mathbf{A}}$ *is a vector subspace of* $\mathbb{R}^m$.

10.18 Vector Algebras

Algebra. An *algebra* over a field $\mathbb{F}$ is a vector space A over $\mathbb{F}$ where the product of pairs of vectors $(\mathbf{a}, \mathbf{b})$ is defined. We denote that product $\mathbf{a} \times \mathbf{b}$. Then for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in A$ and $\alpha \in \mathbb{F}$, we have

1. The vector product is *associative*: $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times (\mathbf{b} \times \mathbf{c})$.
2. The vector product *distributes* both ways over vector addition. Both $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$ and $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$.
3. The vector product has a *homogeneity* property: $\alpha(\mathbf{a} \times \mathbf{b}) = (\alpha\mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (\alpha\mathbf{b})$.

Items (2) and (3) are sometimes described by saying the vector product is *bilinear*.

Under matrix multiplication, the $n \times n$ matrices with real entries form an algebra over the real numbers. Similarly, the $n \times n$ matrices with complex entries form an algebra over the complex numbers. Interestingly, they also form an algebra over the real numbers.

Recalling that we can think of the $n \times n$ matrices as corresponding to linear operators on $\mathbb{R}^n$, we can also consider the linear operators on $\mathbb{R}^n$ as an algebra, where the product is composition of the operators.

The set of functions from a set X to $\mathbb{R}$ is also an algebra provided we define the product $f \times g$ as the pointwise product $(f \times g)(x) = f(x)g(x)$. Unlike the case of matrices, the vector product commutes.

10.19 Vector Analysis

Vector analysis refers to traditional vector calculus in $\mathbb{R}^3$ (no wedge products), which involves use of the cross product of vectors, $\boldsymbol{x} \times \boldsymbol{y}$, and the scalar triple product of vectors, $\boldsymbol{x} \cdot (\boldsymbol{y} \times \boldsymbol{z})$.

Using the canonical basis, we can write

$$\begin{aligned}\boldsymbol{x} \times \boldsymbol{y} &= (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1) \\ &= (x_2y_3 - x_3y_2)\boldsymbol{e}_1 + (x_3y_1 - x_1y_3)\boldsymbol{e}_2 + (x_1y_2 - x_2y_1)\boldsymbol{e}_3.\end{aligned}$$

We can write this using the determinant.

$$\boldsymbol{x} \times \boldsymbol{y} = \begin{vmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3. \end{vmatrix}.$$

The cross product makes $\mathbb{R}^3$ an algebra.

10.20 The Vector Cross Product Anti-Commutes

Vector cross products are often expressed using a quaternion inspired notation with $e_1 = \mathbf{i}$, $e_2 = \mathbf{j}$, and $e_3 = \mathbf{k}$.[4] In that case, we have the familiar form of the cross product,

$$\mathbf{x} \times \mathbf{y} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3. \end{vmatrix}.$$

The vector cross product is anti-commutative. That is $\mathbf{x} \times \mathbf{y} = -\mathbf{y} \times \mathbf{x}$. In this notation, $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ obey

$$\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = 0.$$

and

$$\begin{aligned} \mathbf{i} \times \mathbf{j} &= \mathbf{k}, & \mathbf{j} \times \mathbf{i} &= -\mathbf{k} \\ \mathbf{j} \times \mathbf{k} &= \mathbf{i}, & \mathbf{k} \times \mathbf{j} &= -\mathbf{i} \\ \mathbf{k} \times \mathbf{i} &= \mathbf{j}, & \mathbf{i} \times \mathbf{k} &= -\mathbf{j}. \end{aligned}$$

[4] Quaternions are an extension of the complex numbers that were introduced by William Rowan Hamilton in 1843. They have the form $a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$. Unlike real or complex numbers, quaternion multiplication is not commutative. They had a big influence on the development of vector analysis. Rotations are easily modelled using quaternions, which has made them useful in applications ranging from computer graphics to crystallography.

William Rowan Hamilton (1805–1865) was an Irish mathematician, astronomer, and physicist. He's known for the Hamiltonian formulation of mechanics, and the Hamilton-Jacobi equations. His work on quaternions was highly influential. Among other things, it eventually led to the development of vector analysis in the hands of J. Willard Gibbs and Oliver Heaviside in the late 19^{th} century. The words "tensor" and "scalar" were invented by Hamilton. Another important result of his is the Cayley-Hamilton Theorem, that any square matrix obeys its own characteristic polynomial.

10.21 Differential Operators of Vector Analysis

Commonly used differential operators include the gradient

$$\nabla\varphi = \left(\frac{\partial\varphi}{\partial x_1}, \frac{\partial\varphi}{\partial x_2}, \frac{\partial\varphi}{\partial x_3}\right)$$

for scalar functions, and for vector functions $\mathbf{F} = (F_1, F_2, F_3)$, there are the curl, $\operatorname{curl}\mathbf{F} = \nabla \times \mathbf{F}$, and the divergence, $\operatorname{div}\mathbf{F} = \nabla\cdot\mathbf{F}$.

Then

$$\begin{aligned}
\operatorname{curl}\mathbf{F} = \nabla \times \mathbf{F} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ F_1 & F_2 & F_3 \end{vmatrix} \\
&= \left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}\right)\mathbf{i} + \left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1}\right)\mathbf{j} + \left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}\right)\mathbf{k}.
\end{aligned}$$

and

$$\operatorname{div}\mathbf{F} = \nabla\cdot\mathbf{F} = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{\partial F_3}{\partial x_3}.$$

10.22 The Geometry of Vector Spaces: Norms

Euclidean space is distinguished by its geometry. One aspect of geometry is measuring distances between points. The geometry of the surface of the Earth is only approximately Euclidean. It is not completely flat, but curves around at large distances (it is approximately a sphere). Gauss collected data from a geodetic survey of the kingdom of Hanover in an attempt to see if the discrepancies matched what theory predicted. At the scale of the former kingdom of Hanover, the differences are very small.

In vector spaces we measure vectors to determine their length. This can often be accomplished by a norm. There are three basic properties a norm must have.

Norm. A *norm* on a real or complex vector space V is a mapping from V to $\mathbb{R}$, denoted $\|\mathbf{x}\|$ that has three properties:

1. *Positive Definite*. For all $\mathbf{x} \in V$, $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
2. *Absolutely Homogeneous of Degree One*. For all $\alpha \in \mathbb{F}$ and $\mathbf{x} \in V$, $\|\alpha\mathbf{x}\| = |\alpha|\,\|\mathbf{x}\|$.
3. *Triangle Inequality*. For all $\mathbf{x}, \mathbf{y} \in V$, $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

10.23 Normed Vector Spaces

A normed vector space is a vector space where all vectors have a well-defined length.

Normed Vector Space. A *normed vector space* $\bigl((V, +, \cdot), \|\cdot\|\bigr)$ is a vector space $(V, +, \cdot)$ together with a norm $\|\cdot\|$ defined on that space.

We will usually use the abbreviated notation $\bigl(V, \|\cdot\|\bigr)$ for $\bigl((V, +, \cdot), \|\cdot\|\bigr)$

We measure the distance between two points by the length of the vector between them, $\|\mathbf{x} - \mathbf{y}\|$, which is the same as $\|\mathbf{y} - \mathbf{x}\|$ due to absolute homogeneity.

One thing we can do with norms is create *unit vectors*, vectors with norm one in any given direction. If $\mathbf{x} \neq \mathbf{0}$, we define the *unit vector* $\mathbf{u}$ *in direction* $\mathbf{x}$ by $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$. If we compute

$$\|\mathbf{u}\| = \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\| = \frac{\|\mathbf{x}\|}{\|\mathbf{x}\|} = 1,$$

by absolute homogeneity. Since it has norm one, $\mathbf{u}$ is indeed a unit vector. Also, $\mathbf{u}\cdot\mathbf{x} = \|\mathbf{x}\|$.

Once we define some more norms, Figure 10.26.1 will illustrate all the unit vectors for three different norms.

10.24 ℓ_p^n Spaces

One family of norms on $\mathbb{R}^n$ are the ℓ_p-norms.[5]. The ℓ_p norm on $\mathbb{R}^n$ is defined for $1 \leq p < \infty$ by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p},$$

while if $p = +\infty$,

$$\|\mathbf{x}\|_\infty = \max_i |x_i|.$$

We use the notation ℓ_p^n to denote $\mathbb{R}^n$ with the ℓ_p norm. Thus ℓ_p^n means $(\mathbb{R}^n, \|\cdot\|_p)$.

The ℓ_p norm can also be defined on sequences of real numbers, when ℓ_p is defined as the set of sequences $\mathbf{x} = \{x_1, x_2, \dots\}$ such that $\sum_i |x_i|^p$ converges. In that case we use the norm

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p}.$$

[5] Called "little el pee" or just "el pee".

10.25 Different Norms, Different Lengths

To see that lengths (and so distances) change under different norms, consider the length of the vector $(3, 4)$ using four different norms.

Space	Length
ℓ_1^2	7
ℓ_2^2	5
ℓ_5^2	≈ 4.17
ℓ_∞^2	4

Interestingly, the length of the vector $(0, 1)$ is unchanged in each of the ℓ_p norms. It's always 1. The fact that some lengths change while others don't tells us that the geometry itself has changed.

10.26 Shapes in ℓ_p

We can get a clue about how ℓ_p geometry changes with p by considering the vectors of length one. Figure 10.26.1 shows all unit vectors $\mathbf{x}$, vectors with $\|\mathbf{x}\| = 1$, for three different values of p. From left to right, the norms used are ℓ_1, ℓ_2, and ℓ_∞.

Although the distances along the coordinate axes are the same in all cases, points in other directions get closer as p increases. As a result, the sets themselves expand in the off-axis directions as p gets larger.

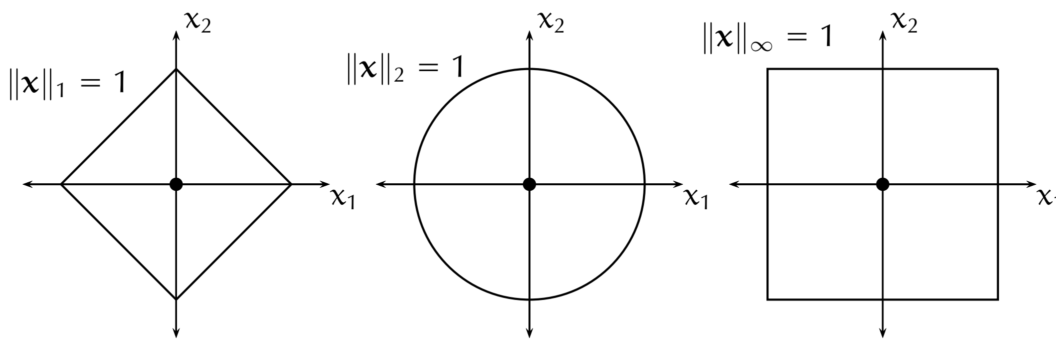


Figure 10.26.1: The left diagram shows the unit vectors, the vectors of length 1 in the ℓ_1 norm. The center diagram shows the unit vectors in the ℓ_2 norm. The right diagram shows the unit vectors in the ℓ_∞ norm. Although the distances along the axes stay the same, the other vectors get farther out as for the same ℓ_p length as p increases.

The ℓ_1 norm is sometimes called the *taxicab norm*. When streets are laid out on a grid, it gives the distance via street between any two locations (no shortcuts through buildings!).

The ℓ_2 norm is the *Euclidean norm*. It measures length according to Euclidean geometry, and the vectors of length one form a circle.

10.27 More Normed Spaces

Norms can be defined for other vector spaces. Consider the set of bounded real-valued continuous functions on a space X, denoted $\mathcal{C}_b(X)$. This space is a vector space when vector addition and scalar multiplication are defined pointwise. That is, $(f+g)(x) = f(x) + g(x)$ and $(tf)(x) = tf(x)$ for all $x \in X$.

Because the functions in $\mathcal{C}_b(X)$ are bounded, the *supremum norm* (or *sup norm*) defined by $\|f\|_\infty = \sup_{x\in X} |f(x)|$ will always be finite. It is easy to show that the sup norm is positive definite, absolutely homogeneous of degree one, and obeys the triangle inequality on $\mathcal{C}_b(X)$.

Another function space with a norm is the vector space of square integrable functions on A, $L^2(A)$.[6] It has norm

$$\|f\|_2 = \left(\int_A |f(x)|^2\,dx\right)^{1/2}.$$

It will turn out that this space is also Euclidean in the sense that the Pythagorean identity is true. In fact, this is a generalization of the ℓ_2 norm.[7]

[6] This space is called "big el two" when we need to distinguish it from one of the ℓ_2 spaces.

[7] Pythagoras of Samos (ca. 570–ca. 495 BC) was an Ionian Greek philosopher. Many discoveries were attributed to Pythagoras in antiquity. It's not clear which really belonged to him. There's certaintly evidence that forms of the Pythagorean identity were known much earlier. Some of these were already known 1500 years earlier, in both Middle Kingdom Egypt and Hammurabi's Babylon.

10.28 L^p Spaces

This generalization can be extended to p, $1 \le p < \infty$ by

$$\|f\|_p = \left(\int_A |f(x)|^p \, dx \right)^{1/p}.$$

▶ **Example 10.28.1: Vectors in L^p Spaces.** The function $f(x) = x^{-1/2}$ is in L^1 when $A = [0, 1]$ because the integral of its absolute value is finite. In fact,

$$\int_0^1 |x^{-1/2}| \, dx = \int_0^1 x^{-1/2} \, dx = 2x^{1/2}\Big|_0^1 = 2.$$

The function is not square integrable (i.e., in $L^2(0, 1)$) because

$$\int_0^1 \left| x^{-1/2} \right|^2 dx = \int_0^1 \frac{1}{x} \, dx = \ln x \Big|_0^1 = +\infty.$$

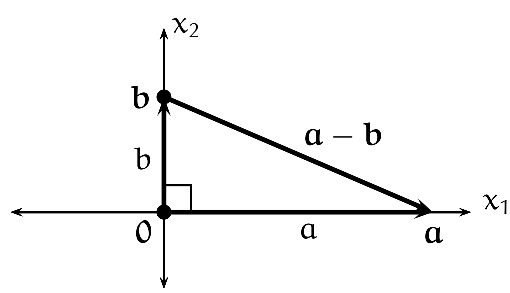
◀

10.29 ℓ_p^n norms and Pythagoras

The ℓ_2 norm is called the *Euclidean norm* because it measures vectors according to Euclidean geometry. The other ℓ_p norms are not Euclidean. The geometry is different.

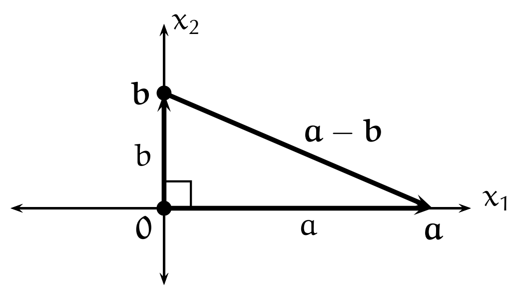
We can see this by measuring a right triangle in $\mathbb{R}^2$. Euclidean geometry requires that the Pythagorean identity hold. If the two short sides of the triangle have lengths a and b, the long side, the hypotenuse, has length $\sqrt{a^2 + b^2}$.

We will use the right triangle defined by $\mathbf{0} = (0, 0)$, $\mathbf{a} = (a, 0)$, and $\mathbf{b} = (0, b)$ with $a, b > 0$ to check this. The two sides of the right angle are $\mathbf{a}$–$\mathbf{0}$ and $\mathbf{b}$–$\mathbf{0}$ while the hypotenuse is $\mathbf{a}$–$\mathbf{b}$. In all of the ℓ_p norms the a side has length $(|a|^p)^{1/p} = a$ for $p < \infty$, and $\max\{0, a\} = a$ for $p = \infty$. Similarly, the $\mathbf{b}$ side has length b. This is illustrated in the diagram.



10.30 When $p \neq 2$, ℓ_p^n is Non-Euclidean

For the Pythagorean identity to be true, the hypotenuse $\mathbf{a} - \mathbf{b} = (a, -b)$ must have length $\sqrt{a^2 + b^2}$. It works fine in the ℓ_2 norm, $\|(a, -b)\|_2 = \sqrt{a^2 + b^2}$. For the ℓ_∞ norm, we have $\|(a, -b)\|_\infty = \max\{|a|, |b|\} < \sqrt{a^2 + b^2}$, failing the test.



For $p \neq \infty$, $\|(a, -b)\|_p = (a^p + b^p)^{1/p}$, which will not be $\sqrt{a^2 + b^2}$ unless $p = 2$. For example, if $a = b = 2$, then the terms are $2^{(p+1)/p}$ and $2^{3/2}$.

Similar calculations apply to all the ℓ_p^n for $n \geq 2$ and $1 \leq p \leq \infty$. That means that ℓ_p^n is not Euclidean when $p \neq 2$.

10.31 Inner Product Spaces

The ℓ_2 norm is special, and one thing that makes it special is that it is based on an inner product.

Inner Product Space. An *inner product space* $(V, \cdot)$ is a real or complex vector space V together with an inner product $x \cdot y$ on V. The *inner product* is a mapping $V \times V$ to $\mathbb{R}$ or $\mathbb{C}$, denoted $x \cdot y$, which is:

1. *(Conjugate) Symmetric*. For all $x, y \in V$, $x \cdot y = \overline{y \cdot x}$.[8]
2. *Linear in* y. For all vectors x, y, z and scalars α, $x \cdot (\alpha y + z) = \alpha(x \cdot y) + (x \cdot z)$.[9]
3. *Positive Definite*. For all $x \in V$, $x \cdot x \geq 0$ and $x \cdot x = 0$ if and only if $x = 0$.

The *inner product* is also known as the *dot product* or *scalar product*. Various notations are used for the inner product, including $x \cdot y$, (x, y), $\langle x, y \rangle$, and $\langle x | y \rangle$. We will generally use the dot notation, $x \cdot y$.

[8] The conjugate has no effect when V is a real vector space. There we just have ordinary symmetry.

[9] In the complex case, this implies conjugate linearity in the first term. Some authors use the opposite convention with conjugate linearity in the second term.

10.32 Euclidean Inner Product on $\mathbb{R}^n$

On $\mathbb{R}^n$, we define the *Euclidean inner product* by

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i.$$

Another way to write the inner product on $\mathbb{R}^n$ is as a matrix product, $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^\mathsf{T} \boldsymbol{y}$. In $\mathbb{C}^n$, the transpose must be replaced by the Hermitian conjugate, so $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^* \boldsymbol{y}$ which can also be written

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{n} \bar{x}_i y_i.$$

10.33 Bilinear or Sesquilinear?

The complex conjugates only play a role when we have a complex vector space. They can be ignored when the vector space is real.

When we use a notation such as $\langle \mathbf{x} | \mathbf{y} \rangle$ we are writing the inner product as a *bilinear form* or a *sesquilinear form*, where $\langle \mathbf{x} | \mathbf{y} \rangle$ is separately linear in both $\mathbf{x}$ and in $\mathbf{y}$ (bilinear) or linear in one and conjugate linear in the other (sesquilinear). In $\mathbb{R}^n$,

$$(\alpha\mathbf{x} + \mathbf{y})\cdot\mathbf{z} = \alpha\mathbf{x}\cdot\mathbf{y} + \mathbf{y}\cdot\mathbf{z},$$

implying the inner product is bilinear. But in $\mathbb{C}^n$,

$$\begin{aligned}(\alpha\mathbf{x} + \mathbf{y})\cdot\mathbf{z} &= \alpha\mathbf{x}^*\mathbf{z} + \mathbf{y}^*\mathbf{z}\\ &= \bar{\alpha}\mathbf{x}^*\mathbf{z} + \mathbf{y}^*\mathbf{z}\\ &= \bar{\alpha}\mathbf{x}\cdot\mathbf{z} + \mathbf{y}\cdot\mathbf{z}.\end{aligned}$$

The presence of the conjugate is why the inner product is sesquilinear in $\mathbb{C}^n$, not bilinear.

10.34 More Inner Products

There are other inner products on $\mathbb{R}^n$. In fact, whenever $\boldsymbol{A}$ is a symmetric positive definite matrix, we can define an inner product by $\mathbf{x}\cdot\mathbf{y} = \mathbf{x}^\mathsf{T}\boldsymbol{A}\mathbf{y}$. Linearity in the first argument is clear. The symmetry of $\boldsymbol{A}$ ensures that the resulting inner product is symmetric. The fact that $\boldsymbol{A}$ is positive definite, makes the inner product positive definite. This also works in $\mathbb{C}^n$ provided that $\boldsymbol{A}$ is Hermitian.

10.35 Creating a Norm from the Inner Product

Any time we have an inner product $\mathbf{x}\cdot\mathbf{y}$ on a vector space V, we can define an associated *norm* by

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}\cdot\mathbf{x}}.$$

If we use the Euclidean inner product, this becomes the *Euclidean norm*

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} |x_i|^2}.$$

When we use the Euclidean norm on $\mathbb{R}^n$, the resulting space is called n-dimensional Euclidean space, ℓ_2^n.

10.36 Cauchy-Schwarz Inequality

An inner product and its associated norm are closely related. One aspect of this is the Cauchy-Schwarz inequality, which will help us prove that the associated norm obeys the triangle inequality.[10]

Cauchy-Schwarz Inequality. *Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in an inner product space $(V, \cdot)$. Then*

$$|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$$

for all vectors $\mathbf{x}$ and $\mathbf{y}$ in V. Moreover, if $|\mathbf{x}\cdot\mathbf{y}| = \|\mathbf{x}\| \, \|\mathbf{y}\|$ for non-zero $\mathbf{x}$ and $\mathbf{y}$, then $\mathbf{x}$ and $\mathbf{y}$ are proportional.

[10] The oldest version of this inequality is due to Cauchy (1821), for ℓ_2 norms. H.A. Schwarz proved it in 1888 for integrals (L^2), although it had been stated by Bunyakovsky in 1859. You'll notice that the modern proof is quite agnostic about the nature of the inner product involved.

Augustin-Louis Cauchy (1789–1857) was one of the top mathematicians of the early 19^{th} century. He was one of the first to precisely state and rigorously prove a number of theorems in calculus, a project that can be considered a forerunner of the later rebuilding of the foundations of mathematics by Cantor, Weierstrass, and others later in the 19^{th} century. He also founded complex analysis, proving a number of the basic theorems. Although he invented the use of ε (error) and δ (distance), some of his mathematics remained distinctly old-school as he sometimes used infinitesimals in his calculations and proofs.

Viktor Yakovlevich Bunyakovsky (1804–1889) was a Russian mathematician and student of Cauchy's who wrote his dissertation on mathematical physics. He also worked in probability and number theory. In the latter, he's best known for the still-unresolved Bunyakovsky conjecture.

The German mathematician Hermann Amandus Schwarz (1843–1921). Besides the Cauchy-Schwarz inequality, he gave the first rigorous proof of Clairaut's Theorem in 1873 (known as Young's Theorem among economists). Most of his work was in differential geometry, complex analysis, and the calculus of variations. The Schwarz reflection principle for extending complex analytic functions is an example.

10.37 Proof of Cauchy-Schwarz Inequality

Cauchy-Schwarz Inequality. *Let x and y be vectors in an inner product space $(V, \cdot)$. Then*

$$|x \cdot y| \leq \|x\| \, \|y\|$$

for all vectors x and y in V. Moreover, if $|x \cdot y| = \|x\| \, \|y\|$ for non-zero x and y, then x and y are proportional.

Proof. If $x = 0$, both sides of the inequality are zero, and we are done. Otherwise, $x \neq 0$. Since we wish to include the complex case, keep in mind that $x \cdot y = x^* y$ and $y \cdot x = y^* x = \overline{x^* y} = \overline{x \cdot y}$. We calculate using matrix products:

$$\begin{aligned}
0 &\leq \left\| y - \frac{x \cdot y}{\|x\|^2} x \right\|^2 && (10.37.1) \\
&= \left(y^* - \frac{\overline{x \cdot y}}{\|x\|^2} x^* \right) \left(y - \frac{x \cdot y}{\|x\|^2} x \right) \\
&= \|y\|^2 - \frac{\overline{(x \cdot y)}(x \cdot y)}{\|x\|^2} - \frac{\overline{(x \cdot y)}(x \cdot y)}{\|x\|^2} + \frac{\overline{(x \cdot y)}(x \cdot y)}{\|x\|^4} \|x\|^2 \\
&= \|y\|^2 - \frac{\overline{(x \cdot y)}(x \cdot y)}{\|x\|^2} \\
&= \|y\|^2 - \frac{|x \cdot y|^2}{\|x\|^2}. && (10.37.2)
\end{aligned}$$

Since $\|\cdot\|$ is positive definite,

$$|x \cdot y|^2 \leq \|x\|^2 \|y\|^2.$$

We then take the positive square root to establish the Cauchy-Schwarz Inequality.

If $|x \cdot y| = \|x\| \, \|y\|$ for non-zero x and y, eq. 10.37.2 implies that the right-hand side of eq. 10.37.1 is zero, meaning that

$$y = \left(\frac{y \cdot x}{\|x\|^2} \right) x.$$

But then y is proportional to x. Since both x and y are non-zero, they are proportional to each other. ∎

10.38 The Inner Product defines a Norm

Now consider the norm derived from the inner product, $\|x\| = (x\cdot x)^{1/2}$. This is obviously absolutely homogeneous and positive definite. But is it a norm? Does the triangle inequality apply? The Cauchy-Schwarz inequality tells us it does.

Proposition 10.38.1. *Let $(V, \cdot)$ be an inner product space over $\mathbb{F} = \mathbb{C}$ or $\mathbb{F} = \mathbb{R}$ and set $\|x\| = (x\cdot x)^{1/2}$. Then $\|\cdot\|$ obeys:*

1. *For all $\alpha \in \mathbb{F}$ and $x \in V$, $\|\alpha x\| = |\alpha|\,\|x\|$ (absolute homogeneity of degree one).*
2. *$\|x\| \ge 0$ and $\|x\| = 0$ if and only if $x = 0$ (positive definite).*
3. *For all $x, y \in V$, $\|x + y\| \le \|x\| + \|y\|$ (triangle inequality).*

In other words, $\|\cdot\|$ is a norm.

Proof. Now $(\alpha x)\cdot(\alpha x) = (\alpha\bar{\alpha})(x\cdot x) = |\alpha|^2\,\|x\|^2$, taking the positive square root shows $\|\alpha x\| = |\alpha|\,\|x\|$, proving (1).

For (2), $\|x\| \ge 0$ by definition. If $\|x\| = 0$, then $x\cdot x = 0$. Since the inner product is positive definite, $x = 0$. This proves (2).

For (3),

$$\begin{aligned}\|x + y\|^2 &= \|x\|^2 + x\cdot y + y\cdot x + \|y\|^2\\ &= \|x\|^2 + 2(\mathrm{Re}\, x\cdot y) + \|y\|^2\\ &\le \|x\|^2 + 2|x\cdot y| + \|y\|^2\\ &\le \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2\\ &\le \big(\|x\| + \|y\|\big)^2\end{aligned}$$

where the third line uses the Cauchy-Schwarz inequality, $|x\cdot y| \le \|x\|\,\|y\|$. This proves (3). ∎

10.39 Polarization Identity in $\mathbb{R}^n$

If you have a norm derived from an inner product on $\mathbb{R}^n$, it is possible to reconstruct the inner product from the norm using the *polarization identity*

$$x \cdot y = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right).$$

Expanding the right-hand side gives

$$\frac{1}{4}\Big(\|x\|^2 + 2x\cdot y + \|y\|^2 - \|x\|^2 + 2x\cdot y - \|y\|^2\Big) = x\cdot y.$$

Although the expression defined by the polarization identity is always symmetric and positive definite, it will fail to be separately linear in x and y unless the norm obeys the *parallelogram law*

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

which states that the sum of squares of the lengths of the four sides of a parallelogram is equal to the sum of squares of the lengths of the two diagonals. In Euclidean geometry, it follows from the law of cosines. In inner product spaces the parallelogram law follows immediately after expanding the right-hand terms.

10.40 Polarization Identity in $\mathbb{C}^n$

In $\mathbb{C}^n$, the polarization identity takes a somewhat more complicated form:

$$x \cdot y = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right) - \frac{i}{4}\left(\|ix-y\|^2 - \|ix+y\|^2\right).$$

the first term is the real part of $x \cdot y$, given by

$$\operatorname{Re} x \cdot y = \frac{1}{2}\left(x \cdot y + \overline{x \cdot y}\right).$$

The second term is i times the imaginary part of $x \cdot y$,

$$i \operatorname{Im} x \cdot y = \frac{1}{2}\left(x \cdot y - \overline{x \cdot y}\right).$$

Put the two together and you have $x \cdot y$.

10.41 Perpendicular Vectors

One important fact about the Euclidean inner product is that perpendicular vectors have a dot product of zero.

Theorem 10.41.1. *Let x and y be non-zero vectors in Euclidean $\mathbb{R}^n$. Then x is perpendicular to y if and only if $x \cdot y = 0$.*

Proof. If case ($\Leftarrow$): Suppose $x \cdot y = 0$. Then

$$\|y - x\|^2 = \|x\|^2 - 2y \cdot x + \|y\|^2 = \|x\|^2 + \|y\|^2$$

which is the Pythagorean identity. That means that x, y, and $y - x$ form a right triangle. As $y - x$ is the hypotenuse, x and y are perpendicular.

Only if case ($\Rightarrow$): Suppose x is perpendicular to y. Consider the right triangle with sides x, y, and $y - x$. By the Pythagorean identity

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

It follows immediately that $x \cdot y = 0$. ∎

10.42 Orthogonal and Orthonormal Vectors

The standard basis vectors $\mathcal{E} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ are perpendicular unit vectors because for any i, j from 1 to n, we have

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \boldsymbol{e}_j \cdot \boldsymbol{e}_i = \sum_k \delta_{ki}\delta_{kj} = \delta_{ii}\delta_{jj} = \delta_{ij}.$$

So $\boldsymbol{e}_i \cdot \boldsymbol{e}_j = 0$ if $i \neq j$. The basis vectors are perpendicular to one another. A set of vectors that are mutually perpendicular are referred to as *orthogonal vectors*.

Also, $\boldsymbol{e}_i \cdot \boldsymbol{e}_i = \|\boldsymbol{e}_i\|^2 = 1$. The basis vectors are also unit vectors. A set of unit vectors that are also orthogonal are called *orthonormal vectors*.

▶ **Example 10.42.1: An Orthonormal Basis for $\mathbb{R}^3$.** Orthonormal bases don't have to look anything like the standard basis vectors. Let

$$\mathbf{b}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{b}_2 = \frac{1}{\sqrt{3}}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{b}_3 = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

Then $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ is an orthonormal basis for $\mathbb{R}^3$. Easy calculations show the vectors are perpendicular to each other, and that they all have norm 1. ◀

10.43 Coordinates and the Canonical Basis

The standard (or canonical) basis makes it easy to write the coordinates of $\mathbf{x}$ in terms of the inner product.

Using the canonical basis, we can write $\mathbf{x} = \sum_i x_i \mathbf{e}_i$. Now

$$\begin{aligned}
\mathbf{e}_j \cdot \mathbf{x} &= \mathbf{e}_j \cdot \left(\sum_i x_i \mathbf{e}_i \right) \\
&= \sum_i x_i \mathbf{e}_j \cdot \mathbf{e}_i \\
&= \sum_i x_i \delta_{ji} \\
&= x_j.
\end{aligned}$$

That means we can find the standard coordinates of $\mathbf{x}$ by taking the dot product of $\mathbf{x}$ with each $\mathbf{e}_j$.

10.44 Angles and Inner Products I

Theorem 10.44.1. *If x and y are non-zero vectors in Euclidean $\mathbb{R}^n$, and θ is the angle between them, then*

$$\cos\theta = \frac{x \cdot y}{\|x\|\,\|y\|}. \tag{10.44.3}$$

where $x \cdot y$ is the Euclidean inner product.

Proof. We will write $y = w + z$ with w a multiple of x and z perpendicular to x. The diagrams below show how that works when the angle θ between x and y is acute (Figure 10.44.2) or obtuse (Figure 10.45.1). Notice that we always measure the angle the short way round.

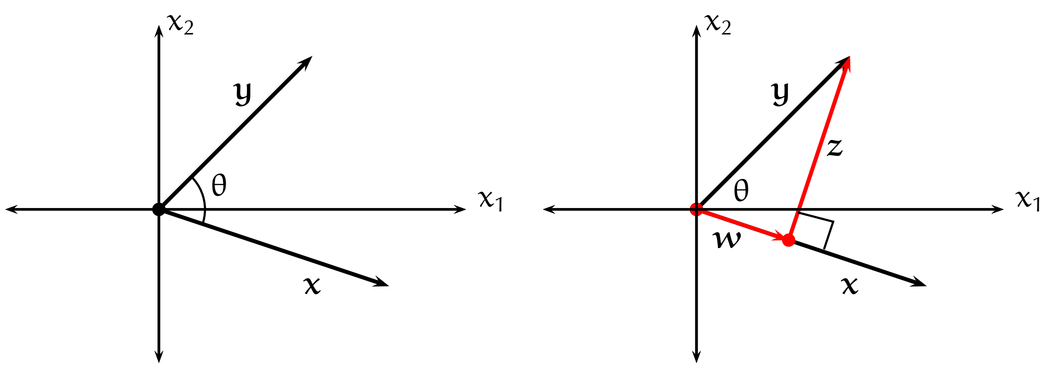


Figure 10.44.2: The acute case is shown in the left-hand diagram. In the right-hand diagram $y = w + z$ where z is perpendicular to x and w is parallel to x. Here $\|w\| = \|y\|\,|\cos\theta|$. As $\cos\theta > 0$, w points in the same direction as x.

Proof continues ...

10.45 Angles and Inner Products II

Proof continues.

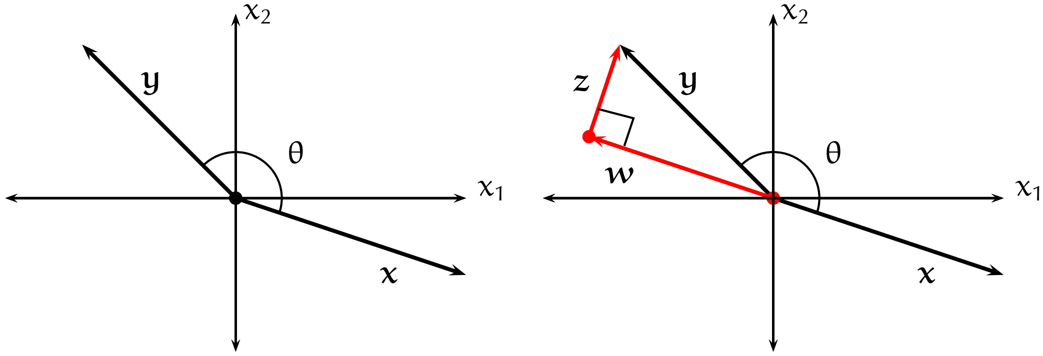


Figure 10.45.1: The obtuse case is shown in the left-hand diagram. In the right-hand diagram $y = w + z$ where z is perpendicular to x and w is parallel to x. Here $\|w\| = \|y\||\cos\theta|$. As $\cos\theta < 0$, w points in the opposite direction as x.

Now $y = w + z$ where w is parallel to x and z is perpendicular to x. Euclidean geometry tells us that w has signed length $\|y\|\cos\theta$, we multiply that by the unit vector in the x direction to find

$$w = \|y\|\cos\theta\frac{x}{\|x\|}.$$

Since $\cos\theta$ is positive when the angle is acute and negative when it is obtuse, the formula works for any θ, $0 \le \theta \le \pi$.

Proof continues ...

10.46 Angles and Inner Products III

Remainder of Proof. Since z is perpendicular to x, $x \cdot z = 0$ by Theorem 10.41.1. Then

$$\begin{aligned} x \cdot y &= x \cdot w + x \cdot z \\ &= x \cdot w \\ &= x \cdot \left(\|y\| \cos\theta \frac{x}{\|x\|} \right) \\ &= \|y\| \cos\theta \frac{\|x\|^2}{\|x\|} \\ &= \|x\|\|y\| \cos\theta. \end{aligned}$$

Divide by $\|x\|\|y\|$ to obtain equation (10.44.3). ∎

In any inner product space, we can use equation (10.44.3) to **define** the angle between any two non-zero vectors.

$$\theta = \arccos\left(\frac{x \cdot y}{\|x\| \, \|y\|} \right).$$

10.47 Metric Spaces 09/13/22

A metric is a way of measuring the distance between two points that is more general than a norm. There are several basic criteria it must satisfy.

Metric Space. Given a set X, a metric d on X is a mapping from $X \times X$ into $\mathbb{R}_+$ that satisfies:

1. *Symmetry.* $d(x, y) = d(y, x)$.
2. *Positive Definite.* $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$.
3. *Triangle Inequality.* For all $x, y, z \in X$, $d(x, z) \leq d(x, y) + d(y, z)$.

Compared to a norm, we have lost the absolute homogeneity. One consequence is that geometry can change with distance.

A metric space is a set with a metric on it.

Metric Space. A *metric space* (X, d) is a set X together with a metric d defined on X.

10.48 Metrics for Normed Spaces

Normed vector spaces have a natural metric defined by $d(x, y) = \|x - y\|$. It is easy to see that this metric is symmetric by absolute homogeneity of the norm. It is positive definite because the norm is positive definite. Finally, the triangle inequality for d follows from the triangle inequality for norms via the following calculation:

$$\begin{aligned} d(x, z) &= \|x - z\| \\ &\leq \|x - y\| + \|y - z\| \\ &= d(x, y) + d(y, z). \end{aligned}$$

10.49 Discrete Metric

► Example 10.49.1: Discrete Metric. A metric that does not require that X be a normed vector space, or a vector space, or that there be any structure at all on the space X, is the *discrete metric*. It is defined by

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases}$$

The discrete metric is defined on every set X. It is positive definite and symmetric. It also obeys the triangle inequality

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

because the left-hand side is either 0 or 1. If zero, we are done. If it is one, either $d(x, y) = 1$ or $d(y, z) = 1$, or both, so the right-hand side is either 1 or 2, satisfying the triangle inequality. ◄

Unless otherwise noted, we will use the Euclidean norm on $\mathbb{R}^n$ and its subsets. If there is any ambiguity about which metric to use with a space, it should be specified.

Metrics will be important later on when we examine limits, open and closed sets, and continuity (Chapters 12, 29, and 13).

10.50 Bounded Metrics

Unlike a norm, a metric may be bounded. The following metric on the real line is bounded by one.

$$d(x, y) = \frac{|x - y|}{1 + |x - y|} < 1.$$

That d is symmetric and positive definite is pretty obvious. It takes a little work to show the triangle inequality, but it is based on the fact that:

Lemma 10.50.1. *If $a \leq b + c$ for any $a, b, c \geq 0$, then*

$$\frac{a}{1 + a} \leq \frac{b}{1 + b} + \frac{c}{1 + c}.$$

Lemma 10.50.2. *Suppose $d(x, y) = \frac{|x-y|}{1+|x-y|}$ for $x, y \in \mathbb{R}$. Then d obeys the triangle inequality.*

Proof. Set $a = |x - z|$, $b = |x - y|$, and $c = |y - z|$. Then $a \leq b + c$ by the triangle inequality for absolute value. By Lemma 10.50.1, d obeys the triangle inequality. ∎

In fact, Lemma 10.50.1 can be used to show that if d is a metric on X, then

$$d'(\mathbf{x}, \mathbf{y}) = \frac{d(\mathbf{x}, \mathbf{y})}{1 + d(\mathbf{x}, \mathbf{y})}$$

is a metric on X that is bounded by 1.

10.51 Proof of Lemma 10.50.1

Proof of Lemma 10.50.1.

$$\begin{aligned}
a &\le b + c \\
&\le b + c + 2bc + abc \\
a + ab + ac + abc &\le b + ab + bc + abc + c + ac + bc + abc \\
a(1 + b)(1 + c) &\le b(1 + a)(1 + c) + c(1 + a)(1 + b) \\
\frac{a}{1 + a} &\le \frac{b}{1 + b} + \frac{c}{1 + c}.
\end{aligned}$$

We divided by $(1 + a)(1 + b)(1 + c)$ to get the last line. ∎

If you're wondering how anyone came up with such a calculation, it's easy once you know the trick. Work backwards from the end. Then it is straightforward.

10.52 A Metric for the Sequence Space

Although it is not possible to define a norm on the sequence space $\mathbf{s}$, we can define a metric on it by

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{\infty} \frac{1}{2^i} \frac{|x_i - y_i|}{1 + |x_i - y_i|}.$$

Since

$$\frac{1}{2^i} \frac{|x_i - y_i|}{1 + |x_i - y_i|} < \frac{1}{2^i},$$

the sum converges uniformly to a number less than one.

Although the metric is bounded, this doesn't really translate into bounds on the x_i. In fact, that even happens if we apply the same technique to $\mathbb{R}^2$. On $\mathbb{R}^2$, define

$$d(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \cdot \frac{|x_1 - y_1|}{1 + |x_1 - y_1|} + \frac{1}{4} \cdot \frac{|x_2 - y_2|}{1 + |x_2 - y_2|}.$$

Now consider the points on the horizontal axis at distance $\frac{1}{4}$ from the origin. They obey

$$\frac{1}{4} = \frac{1}{2} \frac{|x_1|}{1 + |x_1|}$$

so $|x_1| = 1$. The points are $(\pm 1, 0)$.

10.53 Distances in Sequence Space

Figure 10.53.1 illustrates all the points in $\mathbb{R}^2$ at distance $\frac{1}{4}$ from the origin.

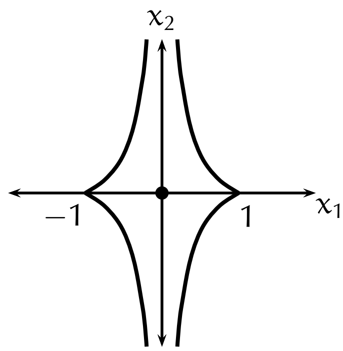


Figure 10.53.1: Although the sequence metric is bounded, that cannot be said about the points at distance 1/4 from zero. The diagram shows the points $\mathbf{x}$ with $d(\mathbf{x}, \mathbf{0}) = 1/4$ in $\mathbb{R}^2$. The entire vertical axis has $d(\mathbf{x}, \mathbf{0}) < 1/4$, with $d\big((0, x_2), \mathbf{0}\big) \to 1/4$ as $x_2 \to \pm\infty$. The geometry is obviously quite different from any of the ℓ_p, some of which were illustrated in Figure 10.26.1.

As noted in Figure 10.53.1 the entire vertical axis ends up being distance less than 1/4 from the origin. Here's the calculation:

$$d\big((0,0),(0,y)\big) = \frac{1}{4}\frac{|y|}{1+|y|} < \frac{1}{4}.$$

Moreover,

$$\lim_{y\to\infty} d\big((0,0),(0,y)\big) = \frac{1}{4}.$$

As a result, when applied to the sequence space, the set

$$\{\mathbf{x} \in \mathbf{s} : d(\mathbf{0}, \mathbf{x}) < r\}$$

is unbounded. Even for $r < 1$ it will typically include the vertical axes for large values of i. When $r \geq 1$, it is the entire sequence space $\mathbf{s}$.

10.54 Lines in Euclidean Space: Slope-Intercept Form

Now that we have measurement of distances and angles under control, via the inner product (angles and length), ℓ_2^n norm (length), and associated metric (distance function), we turn our attention to other aspects of the geometry of $\mathbb{R}^n$. Until perpendicular angles become involved, this will apply to $\mathbb{R}^n$ in general, not just Euclidean $\mathbb{R}^n$, ℓ_2^n.

We start with lines in $\mathbb{R}^2$. As you well know, there's more than one way to write the equation of a line. We will start with the slope-intercept form, where $y = mx + b$. The coordinates are (x, y), the slope is m and b is the vertical intercept. Writing the equation in terms of coordinates, we have

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ mx + b \end{pmatrix} = x \begin{pmatrix} 1 \\ m \end{pmatrix} + \begin{pmatrix} 0 \\ b \end{pmatrix}.$$

10.55 General One-Parameter Equation of a Line

We can think of this as a one-parameter family of coordinates. In order to remove the link between a particular coordinate system and the equation for the line, we rewrite this in terms of a parameter $t \in \mathbb{R}$:

$$\mathbf{x}(t) = \begin{pmatrix} 0 \\ b \end{pmatrix} + t \begin{pmatrix} 1 \\ m \end{pmatrix}. \tag{10.55.4}$$

or even

$$\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{x}_1$$

where $\mathbf{x}_0 = (0, b)^\mathsf{T}$ and $\mathbf{x}_1 = (1, m)^\mathsf{T}$. The line is specified by a point on the line, $\mathbf{x}_0$, and a direction $\mathbf{x}_1$. E.g., $\mathbf{x}_0 = (0, b)^\mathsf{T}$ and $\mathbf{x}_1 = (1, m)^\mathsf{T}$ gives us the slope-intercept form. If $\mathbf{x}_0$ is on the horizontal axis, $\mathbf{x}_0$ is the horizontal intercept and $\mathbf{x}_1$ is a multiple of $(1, m)$.

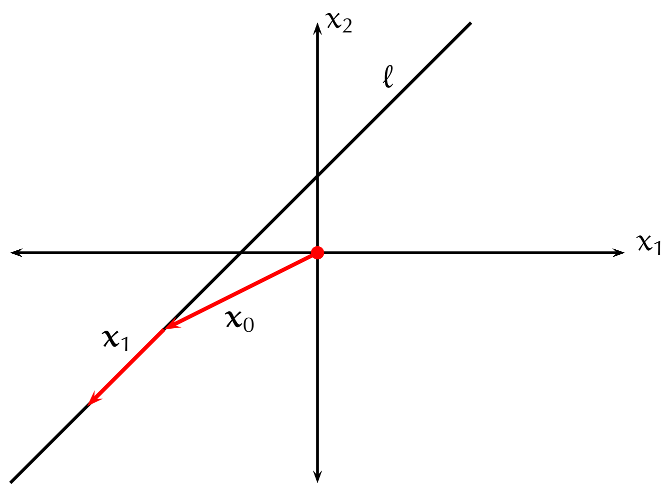


Figure 10.55.1: The line ℓ is $\{\mathbf{x}_0 + t\mathbf{x}_1\}$, with positive values of t continuing SW and negative values moving NE. Both $\mathbf{x}_0$ and $\mathbf{x}_1$ are shown in red on the diagram.

10.56 Lines in Euclidean Space: Parametric Form

We can easily generalize the form in equation (10.55.4) to $\mathbb{R}^n$. We can write a line in $\mathbb{R}^n$ as the points of the form

$$\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{x}_1 \tag{10.56.5}$$

where $\mathbf{x}_0$ is a point on the line, and $\mathbf{x}_1$ is the direction of the line. This is the *parametric form of a line*. Coordinates have been eliminated from the definition. When we need them, we can write the equation using any coordinates we wish.

If L is the line, we can now write

$$L = \left\{\mathbf{x}(t) : \mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{x}_1, t \in \mathbb{R}\right\}.$$

One advantage of representing the line this way is that we can write equations for vertical lines. In $\mathbb{R}^2$, just set $\mathbf{x}_1 = (0,1)^\mathsf{T}$ (or even $(0,-12)^\mathsf{T}$) to get a vertical line through $\mathbf{x}_0$. This is one advantage of a coordinate-free definition. We are not tied to writing y as a function of x, x can be a function of y, or both functions of some other variable, such as the parameter t.

10.57 Lines Through the Origin

Before moving on, let's consider lines through the origin, where $\mathbf{0} \in L$. These can be written in a special form.

Theorem 10.57.1. *A line through the origin,* L*, can be written* $L = \{\boldsymbol{x} : \boldsymbol{x} = t\boldsymbol{x}^1, t \in \mathbb{R}\}$ *if and only if* $\mathbf{0} \in L$.

Proof. If case ($\Leftarrow$): Since $\mathbf{0} \in L$, there is a t^0 with $\boldsymbol{x}^0 + t^0\boldsymbol{x}^1 = \mathbf{0}$. Then $\boldsymbol{x}^0 + t\boldsymbol{x}^1 = (t - t^0)\boldsymbol{x}^1$ for any $t \in \mathbb{R}$. Equivalently, $\boldsymbol{x}(t) = \boldsymbol{x}^0 + t\boldsymbol{x}^1 = t'\boldsymbol{x}^1$. Since $t' = t - t^0$ can be any real number, we can write the line as $L = \{\boldsymbol{x} : \boldsymbol{x} = t'\boldsymbol{x}^1, t' \in \mathbb{R}\}$.

Only if case ($\Rightarrow$): If the line L has the specified form, set $t = 0$ to find $\mathbf{0} \in L$. ∎

10.58 Perpendiculars, Lines, and Hyperplanes

We return to the equation $y = mx + b$, this time restricting our attention to 2-dimensional Euclidean space. We previously converted this to an equation involving the direction $(1, m)$. Suppose we think of the line as being defined by its perpendicular direction. Since the line runs in the direction $(1, m)$, we need (x_1, x_2) with $0 = (x_1, x_2)\cdot(1, m) = x_1 + mx_2$. One such vector is $(x_1, x_2) = (-m, 1)$ (any scalar multiple would do).

Now rewrite $y = mx + b$ as $y - mx = b$, or

$$(-m, 1)\cdot(x, y) = b.$$

In ℓ_2^n, we can generalize this to $\mathbf{a}\cdot\mathbf{x} = b$ for $\mathbf{a} \neq \mathbf{0}$. Define

$$H(\mathbf{a}, b) = \big\{\mathbf{x} : \mathbf{a}\cdot\mathbf{x} = b\big\}.$$

Consider the equation defining H,

$$\sum_{i=1}^{n} a_i x_i = b.$$

Since at least one $a_i \neq 0$, the coefficient matrix of this linear system has rank one. That means there $(n - 1)$ free variables. If $n = 2$, there is one free variable and we have a line. If $n = 3$, there are two free variables, and the equation $a_1x_1 + a_2x_2 + a_3x_3 = b$ defines a plane. When $n = 4$, we have a three-dimensional surface in 4-space. We refer to $H(\mathbf{a}, b)$ as a *hyperplane*. It has the most free variables possible, $n-1$, without including the whole space $\mathbb{R}^n$.

10.59 Hyperplanes and Half-Spaces

The hyperplane $H(\mathbf{a}, b)$ cuts $\mathbb{R}^n$ into two parts whose intersection is $H(\mathbf{a}, b)$. The two parts are

$$H^+(\mathbf{a}, b) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}\cdot\mathbf{x} \geq b\} \text{ and}$$
$$H^-(\mathbf{a}, b) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}\cdot\mathbf{x} \leq b\}.$$

These two sets are referred to as *closed half-spaces*. The term closed means that the boundary, $H(\mathbf{a}, b)$ itself, is included (we will formalize terms such as closed and boundary in Chapter 12).

► Example 10.59.1: Hyperplane in ℓ_2^2. We illustrate $H(\boldsymbol{e}_1, 2)$ and its two closed half-spaces.

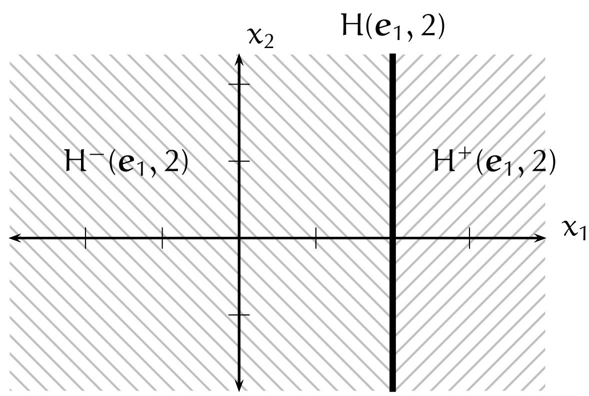


Figure 10.59.2: Here the heavy line/hyperplane $H(\boldsymbol{e}_1, 2)$ separates $\mathbb{R}^2$ into two half-spaces, $H^+(\boldsymbol{e}_1, 2)$ right of the hyperplane, and $H^-(\boldsymbol{e}_1, 2)$ left of the hyperplane.

◄

10.60 Hyperplanes, Half-Spaces, and the Budget Set

One place hyperplanes and half-spaces appear in economics is in the budget set.

► Example 10.60.1: The Budget Set. The *budget set* is defined by

$$B(\mathbf{p}, m) = \{\mathbf{x} \in \mathbb{R}^n_+ : \mathbf{p}\cdot\mathbf{x} \leq m\}$$

for $\mathbf{p} \geq \mathbf{0}$ and $m \geq 0$. Here $\mathbb{R}^n_+ = \{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i = 1, \ldots, n\}$. We can think of this as an intersection of half-spaces. There is the half-space $H^-(\mathbf{p}, m)$, and there are also the n half-spaces $H^+(\boldsymbol{e}_i, 0)$. Thus

$$B(\mathbf{p}, m) = H^-(\mathbf{p}, m) \bigcap \left(\bigcap_{i=1}^{n} H^+(\boldsymbol{e}_i, 0) \right).$$

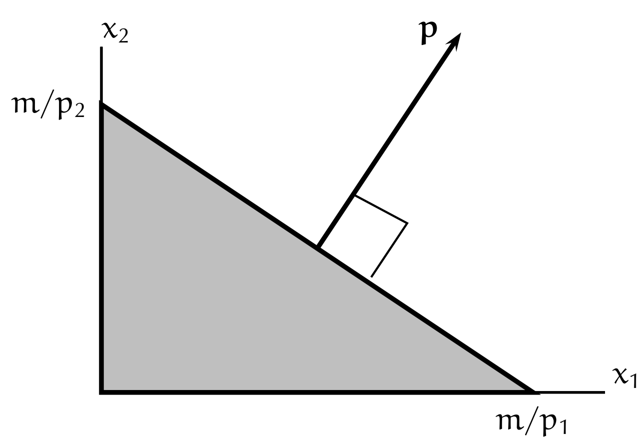


Figure 10.60.2: The figure illustrates a budget set. The price vector is perpendicular to the budget line. The intercepts, where all income is spent either on good one or good two, are m/p_1 and m/p_2, respectively.

◄

10.61 Hyperplanes and the Probability Simplex

A second example of hyperplanes and half-spaces in economics is the probability simplex.

▶ Example 10.61.1: Probability Simplex. The *probability simplex* Δ in $\mathbb{R}^n$ is defined by

$$\Delta = \{\mathbf{p} \in \mathbb{R}^n_+ : \mathbf{p} \cdot \boldsymbol{e} = 1\}$$

where $\boldsymbol{e} = \sum_{i=1}^n \boldsymbol{e}_i = (1, \ldots, 1)^\mathsf{T}$. Thus

$$\Delta = \{\mathbf{p} \in \mathbb{R}^n_+ : \sum_{i=1}^n p_i = 1\}$$

The idea is that $i = 1, \ldots, n$ are mutually exclusive possible events and that p_i is the probability of each event. Since $0 \leq p_i \leq 1$ we can think of the p_i as percentages. The fact that $\sum_i p_i = 1$ tells us that the probabilities add up to 100%. One of the events must happen. ◀

10.62 The $\mathbb{R}^3$ Probability Simplex Illustrated

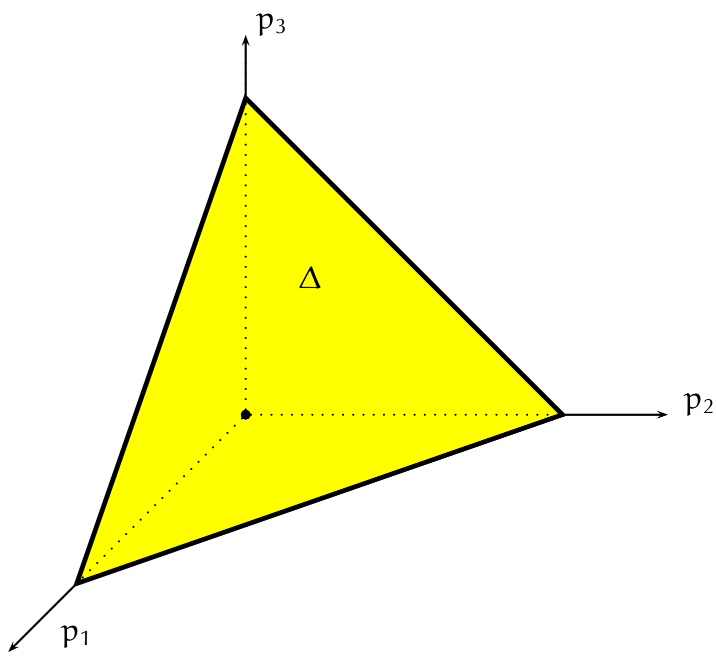


Figure 10.62.1: The probability simplex Δ is the yellow triangle in $\mathbb{R}^3$ with vertices $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. It is perpendicular to the vector $(1, 1, 1)$.

Suppose state i has a payoff of x_i. Then the *expected payoff* at point $\mathbf{p} \in \Delta$ is $\sum_{i=1}^{3} p_i x_i$. If a consumer receives utility $u(x_i)$ from consuming x_i, the *expected utility* is $\sum_{i=1}^{3} p_i u(x_i)$.

Copyright ©2022 by John H. Boyd III: Department of Economics, Florida International University, Miami, FL 33199