10. Euclidean Spaces

9/8/20

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

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.

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

as we did when writing linear systems as matrix products, $\mathbf{Ax} = \mathbf{b}$. The x_i are referred to as the *coordinates* of the vector.

Sometimes, especially in text, we will be informal and write them 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 which consumer a commodity vector belongs or which firm is using inputs and producing outputs.

10.1 Vectors

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

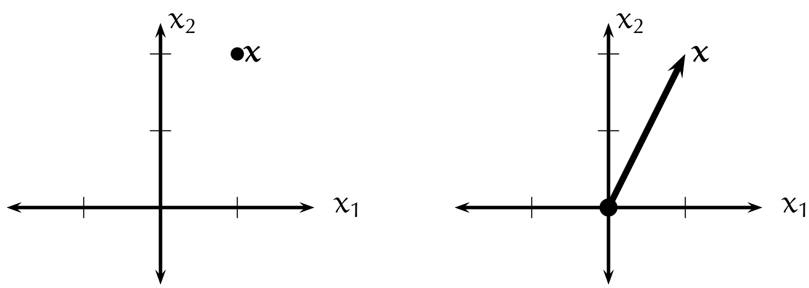


Figure 10.1.1:

Although we will mostly treat vectors as points in the plane, as in the left panel of Figure 10.1.1, it is often useful to think of them as indicating a direction, which $\mathbf{x}$ does in the right panel of Figure 10.1.1. In mathematics this is sometimes indicated by using *vector bundles*. A vector bundle indicates the starting point as well as the direction. We will not be using vector bundles.

10.2 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.

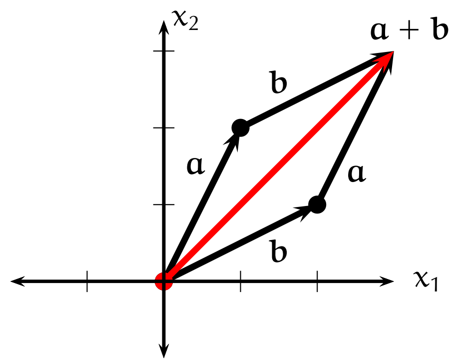


Figure 10.2.1: Two ways to add vectors: Here we add $\mathbf{a} = (1,2)^{\mathsf{T}}$ to $\mathbf{b} = (2,1)^{\mathsf{T}}$. The upper combination is $\mathbf{a} + \mathbf{b}$ and the lower $\mathbf{b} + \mathbf{a}$. Of course, both end at the same point, $(3,3)^{\mathsf{T}}$, because matrix and thus vector addition are commutative. The red vector is the sum $(\mathbf{a} + \mathbf{b})$.

10.3 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.

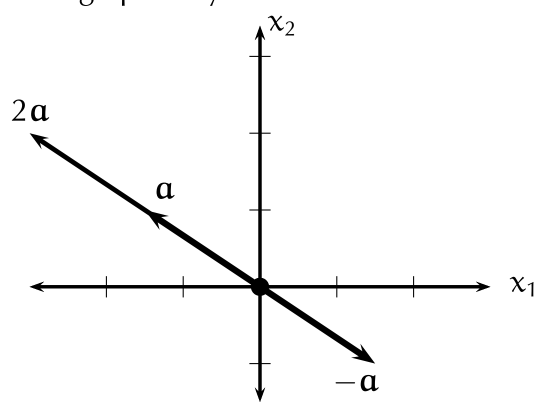


Figure 10.3.1: Vector Multiplication: Here $\mathbf{a} = (-1.5, 1)^{\mathsf{T}}$ is multiplied by 2. Multiplying by a larger number would extend the line further. Multiplying by a smaller number would shrink toward the origin. Finally, multiplying by a negative number goes in the opposite direction as illustrated by $-\mathbf{a}$.

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

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

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

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

We can write any vector $\boldsymbol{x}$ as a sum of basis vectors,

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

The sum can also be written as a matrix product.

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

10.5 Linear Transformations

We are now ready to show that any linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ can be written in matrix form.

Theorem 10.5.1. *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 $T(\boldsymbol{x}) = \sum_{j=1}^n x_j T(\boldsymbol{e}_j)$. Now $T(\boldsymbol{e}_j) \in \mathbb{R}^m$. We denote its i^{th} component by $T(\boldsymbol{e}_j)_i$, $i = 1, \dots, m$. Set $a_{ij} = T(\boldsymbol{e}_j)_i$. We can now define an $m \times n$ matrix $\boldsymbol{A}$ by $\boldsymbol{A} = [a_{ij}]$.

With this definition of $\boldsymbol{A}$, $T(\boldsymbol{x})_i = \sum_{j=1}^n a_{ij} x_j$, 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.

▶ **Example 10.5.2: Linear Transformation as Matrix.** Suppose $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation with

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

Then it can be represented by the matrix

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

so that

$$\boldsymbol{A}\boldsymbol{x} = \begin{pmatrix} x_1 - x_2 \\ x_2 \\ 3x_1 + 2x_2 \end{pmatrix}.$$

◀

10.6 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 math is sometimes done that way, particular in some areas of algebra.

The real issue is that we need to maintain a distinction between the row vectors and column vectors.

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

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

Which one is which is completely arbitrary. But the weight of mathematical tradition, stretching back at least to the 19^{th} 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.

10.7 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 $\mathbf{x}, \mathbf{y} \in V$; and scalar multiplication, which defines $\alpha\mathbf{x}$ for any *scalar* α (any $\alpha \in \mathbb{F}$) 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 is commutative).
2. For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ (addition is associative).
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 all $\mathbf{x} \in V$, 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 is associative).
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).

A vector space can be defined over any number field $\mathbb{F}$, such as the rational numbers, the real numbers, or the complex numbers.[1] 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 $\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n)^\mathsf{T}$.

[1] There are many other number fields and many other types of vector spaces.

10.8 More Vector Spaces and Subspaces

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.

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

We say that 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. 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 of our vector spaces will be subspaces of $\mathbb{R}^n$.

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

Theorem 10.8.1. *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. ∎

10.9 Sequence Spaces

► Example 10.9.1: Sequence Spaces. Another commonly used vector space is $\mathbf{s}$, the space of sequences of real numbers. Elements of $\mathbf{s}$ are of the form $(x_1, x_2, x_3, \ldots)$ 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.

Some vector subspaces of $\mathbf{s}$ include $\mathbf{c}$, the space of convergent sequences, $\mathbf{c}_0$, the space of sequences that converge to 0 ($\lim_n x_n = 0$), and $\mathbf{c}_{00}$, the space of sequences that are zero except for finitely many terms (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 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 arbitrary growth rates is why we cannot find a norm for $\mathbf{s}$.

The sequences

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

are respectively in $\mathbf{c}$, but not $\mathbf{c}_0$; $\mathbf{c}_0$ but not $\mathbf{c}_0 0$; $\mathbf{c}_0 0$. 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 convegent is another sequence space that sits between $\mathbf{c}$ and $\mathbf{s}$. All of the 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. ◄

10.10 The Kernel of a Linear System

We start by considering 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 $T_\mathbf{A}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Its solution set, $\{\mathbf{x} : T_\mathbf{A}(\mathbf{x}) = \mathbf{0}\}$, is a vector subspace of $\mathbb{R}^n$, called the *null space* or *kernel* of $\mathbf{A}$. We denote it by $\ker \mathbf{A}$.[2]

Theorem 10.10.1. *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 10.10.1 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 $\{\mathbf{x}_0\} + \ker \mathbf{A}$.

[2] Simon and Blume prefer to call it the nullspace and use the symbol $\mathcal{N}(\mathbf{A})$.

10.11 Row and Column Spaces

Theorem 10.11.1. *Let $\mathbf{A}$ be an $m \times n$ matrix and define the linear function $T_{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^m$ by $T_{\mathbf{A}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Then $\operatorname{ran} T_{\mathbf{A}}$ is a vector subspace of $\mathbb{R}^m$.*

Proof. Clearly $\operatorname{ran} T_{\mathbf{A}} \subset \mathbb{R}^m$. Let $\mathbf{y}, \mathbf{y}' \in \operatorname{ran} T_{\mathbf{A}}$. Then there are $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^n$ with $T_{\mathbf{A}}(\mathbf{x}) = \mathbf{y}$ and $T_{\mathbf{A}}(\mathbf{x}') = \mathbf{y}'$. Let α be a scalar. Then $T_{\mathbf{A}}(\alpha\mathbf{x} + \mathbf{x}') = \alpha T_{\mathbf{A}}(\mathbf{x}) + T_{\mathbf{A}}(\mathbf{x}') = \alpha\mathbf{y} + \mathbf{y}'$, showing that $\alpha\mathbf{y} + \mathbf{y}' \in \operatorname{ran} T_{\mathbf{A}}$. This shows that $\operatorname{ran} T_{\mathbf{A}}$ is a subspace of $\mathbb{R}^m$. ∎

Let $\mathbf{A}$ be an $m \times n$ matrix. The *column space of* $\mathbf{A}$, $\operatorname{Col}(\mathbf{A})$, is the set of all $\mathbf{A}\mathbf{x}$ for $\mathbf{x} \in \mathbb{R}^n$. Let $\mathbf{a}_j$ denote the j^{th} column of $\mathbf{A}$. Then $\mathbf{A}\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{a}_j$. The column space is the set of all such vectors. It is a vector subspace of $\mathbb{R}^m$. In other words, $\operatorname{Col}(\mathbf{A}) = \operatorname{ran} T_{\mathbf{A}}$.

We define the *row space of* $\mathbf{A}$, $\operatorname{Row}(\mathbf{A})$, as the set of all products $\mathbf{A}^T\mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^m$. It is a subspace of $\mathbb{R}^n$. In fact, $\operatorname{Row}(\mathbf{A}) = \operatorname{ran} T_{\mathbf{A}^T}$.

Corollary 10.11.2. *Let $\mathbf{A}$ be an $m \times n$ matrix with $T_{\mathbf{A}}$ as above. Then $\operatorname{Col}(\mathbf{A}) = \operatorname{ran} T_{\mathbf{A}}$ is a vector subspace of $\mathbb{R}^m$ and $\operatorname{Row}(\mathbf{A}) = \operatorname{ran} T_{\mathbf{A}^T}$ is a vector subspace of $\mathbb{R}^n$.*

Proof. Just apply Theorem 10.11.1. Don't forget that $T_{\mathbf{A}^T}$ maps to $\mathbb{R}^n$. ∎

10.12 The Geometry of Vector Spaces: Norms

Euclidean space is distinguished by its geometry. Part of geometry is measuring distances between points. In vector spaces this entails measuring lengths of vectors. 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 $\|\mathbf{x}\|$ such that:

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. Absolute Homogeneity 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 $\mathbf{x}, \mathbf{y} \in V$, $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

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

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

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

Figure 10.14.1 illustrates all of the unit vectors for three different norms.

10.13 ℓ_p^n Spaces

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

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

while

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

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

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. We use the norm

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

To see that distances change under different norms, consider the distance between $(1,3)$ and $(4,7)$. It is 7 in ℓ_1^2, 5 in ℓ_2^2, approximately 4.17 in ℓ_5^2, and 4 in ℓ_∞^2. In contrast, the distances between $(1,3)$ and $(1,4)$ are 1 in each of the ℓ_p norms. The fact that some distances change while others don't tells us that the geometry itself has changed.

10.14 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 10.14.1 shows all vectors $\mathbf{x}$ 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.

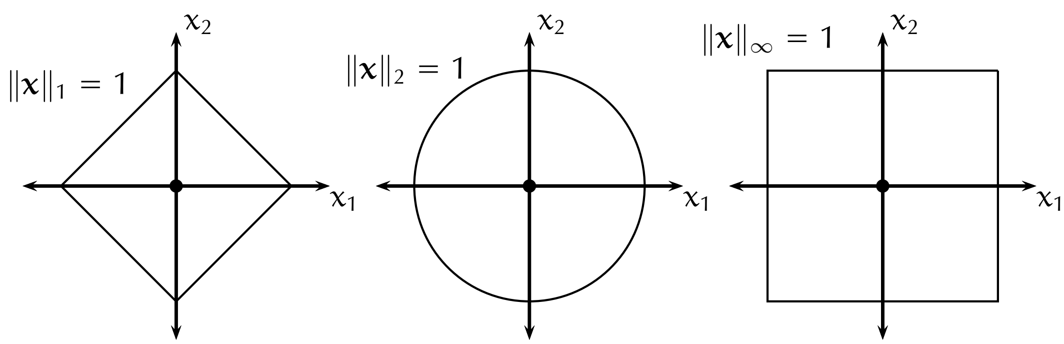


Figure 10.14.1: The left diagram shows the unit vectors, the vectors of length 1 in the ℓ_1 norm. The center diagram shows the unit vectors in the ℓ_2 norm. The right diagram shows the unit vectors in the ℓ_∞ norm. Although the distances along the axes stay the same, the other vectors get farther out 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 distance according to Euclidean geometry, and the points at distance one from the origin form a circle.

10.15 More Normed Spaces

Norms can be defined on 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, $L^2(A)$, 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, it is a generalization of the ℓ_2 norm.

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 10.15.1: Vectors in L^p Spaces.** The function $f(x) = x^{-1/2}$ is in L^1 when $A = [0,1]$ because

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

but it is not in $L^2(0,1)$ because

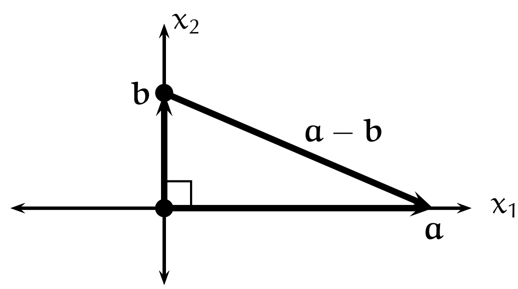
$$\int_0^1 |x^{-1/2}|^2\,dx = \ln x\Big|_0^1 = +\infty.$$

◄

10.16 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. 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 $\mathbf{0}$–$\mathbf{a}$ and $\mathbf{0}$–$\mathbf{b}$ 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 $p = \infty$. The $\mathbf{b}$ side also has length b. This is illustrated in the diagram.



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 $1 \leq p \leq \infty$. We don't worry about $n = 1$ because there is no room for right triangles there. That means that ℓ_p^n is not Euclidean when $p \neq 2$.

10.17 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 $\mathbf{x}\cdot\mathbf{y}$ on V. The *inner product* is a mapping $V \times V$ to $\mathbb{R}$ or $\mathbb{C}$, denoted $\mathbf{x}\cdot\mathbf{y}$, obeying:

1. (Conjugate) Symmetry. For all $\mathbf{x}, \mathbf{y} \in V$, $\mathbf{x}\cdot\mathbf{y} = \overline{\mathbf{y}\cdot\mathbf{x}}$.[3]
2. Linearity. For all vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$ and scalars α, β, $\mathbf{x}\cdot(\alpha\mathbf{y} + \mathbf{z}) = \alpha(\mathbf{x}\cdot\mathbf{y}) + (\mathbf{y}\cdot\mathbf{z})$.[4]
3. Positive Definite. For all $\mathbf{x} \in V$, $\mathbf{x}\cdot\mathbf{x} \geq 0$ and $\mathbf{x}\cdot\mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

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

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 $\mathbb{R}^n$,

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

but in $\mathbb{C}^n$,

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

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

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

[4] 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.

10.18 Euclidean and Other Inner Products

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

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

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

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

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

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* $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} x_\ell^2}$. When we use the Euclidean norm on $\mathbb{R}^n$, the resulting space is called n-dimensional Euclidean space, ℓ_2^n.

10.19 Cauchy-Schwartz Inequality

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

Cauchy-Schwartz 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. The inequality clearly holds if $x = 0$ as both sides are then zero. We restrict our attention to the case $x \neq 0$. Since we wish to include the complex case, we will sometimes write $x^* y$ for $x \cdot y$.

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

Since $\| \cdot \|$ is positive definite, $|x \cdot y|^2 \leq \|x\|^2 \|y\|^2$, establishing the Cauchy-Schwartz Inequality.

If $|x \cdot y| = \|x\| \, \|y\|$ for non-zero x and y, the preceding equations tell us

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

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

10.20 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-Schwartz inequality tells us it does.

Proposition 10.20.1. *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* $\alpha \in F$ *and* $x \in V$, $\|\alpha x\| = |\alpha|\|x\|$ *(absolute homogeneity of degree one).*
2. $\|x\| \geq 0$ *and* $\|x\| = 0$ *if and only if* $x = 0$ *(positive definite).*
3. *For all* $x, y \in V$, $\|x + y\| \leq \|x\| + \|y\|$ *(triangle inequality).*

This shows that $\|\cdot\|$ *is a norm.*

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

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

For (3),

$$\begin{aligned}
\|x + y\|^2 &= \|x\|^2 + x \cdot y + y \cdot x + \|y\|^2 \\
&\leq \|x\|^2 + 2|x \cdot y| + \|y\|^2 \\
&\leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 \\
&\leq \big(\|x\| + \|y\|\big)^2
\end{aligned}$$

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

10.21 Polarization Identities

9/10/20

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

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

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.

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 $x\cdot y$,

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

10.22 Perpendicular Vectors

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

Theorem 10.22.1. *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. **Only if case:** 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 that $x \cdot y = 0$.

If case: 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. ∎

10.23 Orthogonal and Orthonormal Vectors

The standard basis vectors $\mathcal{E}$ are perpendicular unit vectors because

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

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*.

The standard basis makes it easy to write the coordinates of $\boldsymbol{x}$ in terms of the inner product

$$x_i = \boldsymbol{x} \cdot \boldsymbol{e}_i = \boldsymbol{e}_i \cdot \boldsymbol{x}.$$

► Example 10.23.1: Orthonormal Basis for $\mathbb{R}^3$. Orthonormal bases don't have to 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 $\mathbb{R}^3$. Easily calculations show the vectors are perpendicular to each other, and that they all have norm 1. ◄

10.24 Angles and Inner Products I

Theorem 10.24.1. *If x and y are non-zero vectors in Euclidean $\mathbb{R}^n$, and θ is the angle between them, then*

$$\cos\theta = \frac{x\cdot y}{\|x\|\,\|y\|}. \tag{10.24.1}$$

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 10.24.2) and obtuse (Figure 10.24.3). Notice that we always measure the angle the short way round (or π for 180°).

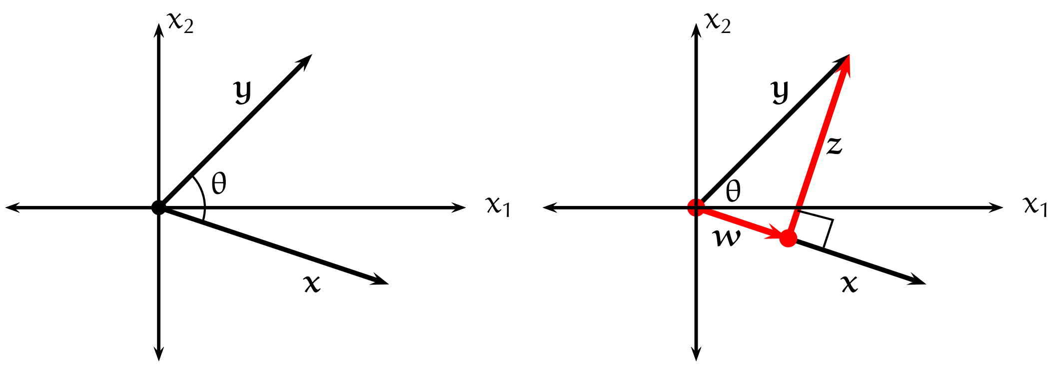


Figure 10.24.2: The acute case is shown in the left-hand diagram. In the right-hand diagram $y = w + z$ where z is perpendicular to x and w is parallel to x. Here $\|w\| = \|y\||\cos\theta|$. As $\cos\theta > 0$, w points in the same direction as x.

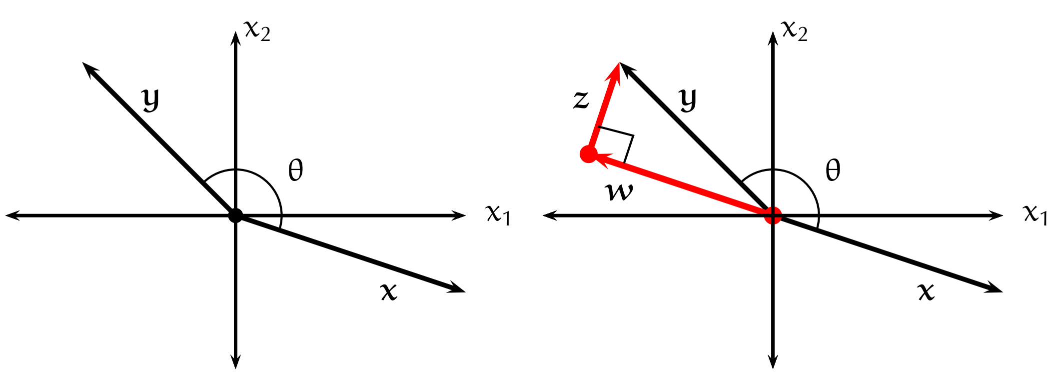


Figure 10.24.3: The obtuse case is shown in the left-hand diagram. In the right-hand diagram $y = w + z$ where z is perpendicular to x and w is parallel to x. Here $\|w\| = \|y\||\cos\theta|$. As $\cos\theta < 0$, w points in the opposite direction as x.

10.25 Angles and Inner Products II

Proof continues. 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

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

Of course, $\cos\theta$ is positive when the angle is acute and negative when it is obtuse. As a result, the formula works for any θ, $0 \le \theta \le \pi$. Since z is perpendicular to x, $x{\cdot}z = 0$ by Theorem 10.22.1. 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 (10.24.1). ∎

In any inner product space, we can use equation (10.24.1) to **define** the angle between any two non-zero vectors.

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

10.26 Metric Spaces

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 $\mathbb{R}_+$ such that:

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 $x, y, z \in X$, $d(x, z) \leq (x, y) + d(y, z)$.

Compared to a norm, we have lost the absolute homogeneity. The 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.

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 non-negative, and zero if and only $x = y$. Finally, the triangle inequality for d follows from

$$d(x, z) = \|x - z\| \leq \|x - y\| + \|y - z\| = d(x, y) + (y, z).$$

► **Example 10.26.1: 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 inquality because the left-hand side is either 0 or 1. 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).

10.27 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 if $a \leq b + c$, for $a, b, c \geq 0$, then

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

Here's how it works:

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

We divided by $(1 + a)(1 + b)(1 + c)$ in the last line.

Finish by setting $a = |x - z|$, $b = |x - y|$, and $c = |y - z|$ to obtain the triangle inequality.

10.28 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 each term is no more than 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. The following diagram applies this to $\mathbb{R}^2$, with the same weighting of $1/2^i$. This means

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

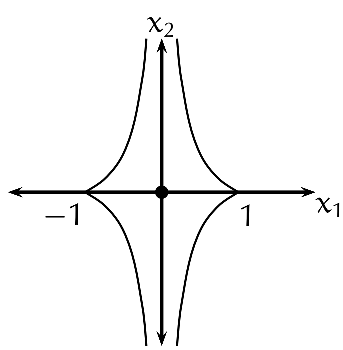


Figure 10.28.1: Although the sequence metric is bounded, that cannot be said about the points at distance 1/4 from zero. The diagram illustrates points $\mathbf{x}$ with $d(\mathbf{x}, \mathbf{0}) = 1/4$ in $\mathbb{R}^2$. The entire vertical axis has $d(\mathbf{x}, \mathbf{0}) < 1/4$, with $d\big((0, x_2), 0\big) \to 1/4$ 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 10.14.1.

As noted in the Figure 10.28.1 the entire vertical axis ends up being distance less than 1/4 from the origin. Here's the calculation:

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

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

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

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

10.29 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

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

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

$$\mathbf{x}(t) = \begin{pmatrix} 0 \\ b \end{pmatrix} + t \begin{pmatrix} 1 \\ m \end{pmatrix}. \tag{10.29.2}$$

The line is specified by a point on the line, $(0, b)^{\mathsf{T}}$, and a direction $(1, m)^{\mathsf{T}}$.

10.30 Lines in Euclidean Space: Parametric Form

We can easily generalize the form in equation (10.29.2) to $\mathbb{R}^n$. We can write a line in $\mathbb{R}^n$ as the points of the form

$$\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{x}_1 \tag{10.30.3}$$

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

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

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.

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

Theorem 10.30.1. *A line* L *can be written* $L = \{\mathbf{x} : \mathbf{x} = t\mathbf{x}^1, t \in \mathbb{R}\}$ *if and only if* $\mathbf{0} \in L$.

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

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

10.31 Perpendiculars and Lines in ℓ_2^2 and ℓ_2^n

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

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, so there are $(n - 1)$ 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 without including the whole space $\mathbb{R}^n$.

The hyperplane $H(\mathbf{a}, b)$ cuts $\mathbb{R}^n$ into two parts whose intersection is $H(\mathbf{a}, b)$, $H^+(\mathbf{a}, b) = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}\cdot\mathbf{x} \geq b\}$ 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 10.31.1: Hyperplane in ℓ_2^2. We examine $H(\boldsymbol{e}_1, 2)$ and its two closed half-spaces.

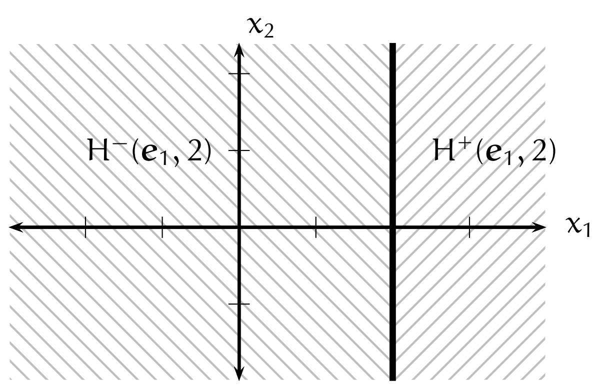


Figure 10.31.2: Here the heavy line/hyperplane $H(\boldsymbol{e}_1, 2)$ separates $\mathbb{R}^2$ into two half-spaces, $H^+(\boldsymbol{e}_1, 2)$ right of the hyperplane, and $H - +(\boldsymbol{e}_1, 2)$ left of the hyperplane.

◄

10.32 Examples of Hyperplanes and Half-Spaces

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

▶ **Example 10.32.1: The Budget Set.** The *budget set* is defined by

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

for $\mathbf{p} \ge \mathbf{0}$ and $m \ge 0$. Here $\mathbb{R}^n_+ = \{\mathbf{x} \in \mathbb{R}^n : x_i \ge 0 \text{ for all } i = 1, \ldots, n\}$. 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 half-spaces $H^+(\mathbf{e}_i, 0)$. Thus

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

◀

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

▶ **Example 10.32.2: Probability Simplex.** The *probability simplex* Δ is defined by

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

where $\mathbf{e} = \sum_{i=1}^n \mathbf{e}_i = (1, \ldots, 1)^T$. 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 $0 \le p_i \le 1$ 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. ◀

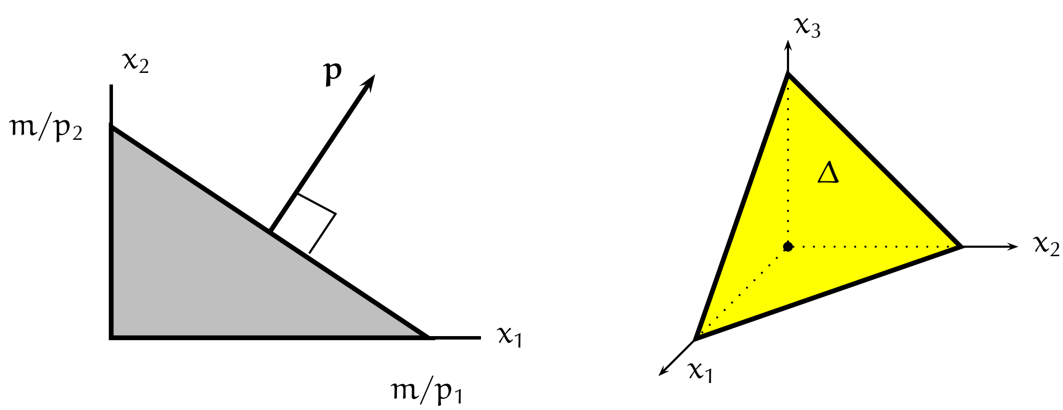


Figure 10.32.3: The left panel shows a budget set. The price vector is perpendicular to the budget line. The intercepts, where all income is spent either on good one or good two, are m/p_1 and m/p_2, respectively.

The right panel shows the probability simplex, Δ, as a yellow triangle in $\mathbb{R}^3$ with vertices $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$.

11. Linear Independence

This section focuses on bases for vector spaces. We've seen one basis already, the standard basis for $\mathbb{R}^n$. It allowed us to write any linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ in terms of matrix multiplication.

Bases allow us to define the dimension of a vector space. In the context of solutions to linear systems, the dimension is the number of free variables. It tells us what the solution set looks like.

Finally, the judicious choice of a basis can simplify linear systems, allowing easier interpretation of results. In a dynamic context, this allows us to better understand both short and long-run dynamics of the system.

11.1 Linear Combinations

Let L be a line through the origin. We say $L = \{\mathbf{x} : \mathbf{x} = t\mathbf{x}_1\}$ is the *line generated by* $\mathbf{x}_1$, or the *line spanned by* $\mathbf{x}_1$. It's the set of all scalar multiples of $\mathbf{x}_1$. What if we have more than one generator? What do we get?

Let's try it. Let $\mathbf{x}_1, \ldots, \mathbf{x}_k$ be vectors in a vector space V. A sum of the form

$$t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + \cdots + t_k\mathbf{x}_k = \sum_{j=1}^{k} t_j\mathbf{x}_j$$

for $t_j \in \mathbb{R}$ is called a *linear combination of* $\mathbf{x}_1, \ldots, \mathbf{x}_k$ and we define the *span of* $\mathbf{x}_1, \ldots, \mathbf{x}_k$, $\mathcal{L}[\mathbf{x}_1, \ldots, \mathbf{x}_k]$, as the set of linear combinations of $\mathbf{x}_1, \ldots, \mathbf{x}_k$. The span is a subspace of V, possibly V itself. In other words,

$$\mathcal{L}[\mathbf{x}_1, \ldots, \mathbf{x}_k] = \left\{\mathbf{x} : \mathbf{x} = \sum_{j=1}^{k} t_j\mathbf{x}_j \text{ for some } t_j \in \mathbb{R}\right\}.$$

When $V = \mathbb{R}^n$, we can write the span using a matrix. Form an $n \times k$ matrix $\mathbf{X}$ from the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ by taking $\mathbf{x}_j$ as the j^{th} column of $\mathbf{X}$. The linear combinations of the $\mathbf{x}_1, \ldots, \mathbf{x}_k$ can be written

$$\mathbf{X}\mathbf{t} = \sum_{j=1}^{k} t_j\mathbf{x}_j = t_1\begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{pmatrix} + \cdots + t_k\begin{pmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kn} \end{pmatrix}.$$

Then

$$\mathcal{L}[\mathbf{x}_1, \ldots, \mathbf{x}_k] = \{\mathbf{X}\mathbf{t} : \mathbf{t} \in \mathbb{R}^k\}.$$

When writing $\mathbf{x}$ this way, we can think of the t_j as *coordinates* of $\mathbf{x}$ in the *coordinate system* $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_k]$.

11.2 Span Examples

▶ Example 11.2.1: Standard Basis Vectors. If $k = n$ and $\mathbf{x}_j = \mathbf{e}_j$, then the matrix formed from the standard basis vectors $\mathbf{e}_j$ is the identity matrix and the coordinates of $\mathbf{x}$ are $\mathbf{I}\mathbf{x} = \mathbf{x}$, meaning that the j coordinate is just x_j. ◀

Spans need not resemble the standard basis vectors.

▶ Example 11.2.2: Span of Vectors. Consider the case of three vectors in $\mathbb{R}^4$ given by the columns of

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Then

$$\mathcal{L}[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] = \left\{ \begin{pmatrix} t_1 + t_2 \\ t_2 + t_3 \\ t_1 + t_2 + t_3 \\ t_1 \end{pmatrix} : t_j \in \mathbb{R} \right\}.$$

Since the rank of $\mathbf{X}$ is three, there are vectors cannot be written $\mathbf{x} = \mathbf{X}\mathbf{t}$. These vectors are not in the span, showing that $\mathcal{L}[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ is a proper subspace of $\mathbb{R}^4$. ◀

11.3 When are Linear Combinations Unique?

Linear combinations allow us to write vectors in terms of a particular set of vectors. It lets us set up a coordinate system. Does a vector have a single set of coordinates? Or are there multiple ways to write it in terms of our of vectors?

Let $\mathbf{X}$ be an $n \times k$ matrix and suppose $\mathbf{x} = \mathbf{X}\mathbf{t}$ and $\mathbf{x} = \mathbf{X}\mathbf{t}'$. When can we conclude that $\mathbf{t} = \mathbf{t}'$?

By subtracting, we find $\mathbf{0} = \mathbf{X}(\mathbf{t} - \mathbf{t}')$, so the question is really whether this homogeneous linear system has a unique solution. This will have multiple solutions if and only if there are free variables, which is equivalent to $k > \operatorname{rank} \mathbf{X}$.

This is connected to the idea of linear dependence.

Linear Dependence. Non-zero vectors $\mathbf{x}_1 \ldots, \mathbf{x}_k$ are *linearly dependent* if there are $t_1, \ldots, t_k$, not all zero, with $\sum_{j=1}^{k} t_j \mathbf{x}_j = \mathbf{0}$.

In other words, the vectors are linearly dependent if and only if $\mathbf{X}\mathbf{t} = \mathbf{0}$ has a non-zero solution $\mathbf{t}$.

Theorem 11.3.1. *Suppose $\mathbf{x}_1 \ldots, \mathbf{x}_k$ are linearly dependent vectors. Then there is h so that*

$$\mathbf{x}_h = -\frac{1}{t_h} \sum_{j \neq h} t_j \mathbf{x}_j.$$

Proof. By linear dependence, we can find $t_1, \ldots, t_k$, not all zero, with $\sum_{j=1}^{k} t_j \mathbf{x}_j = \mathbf{0}$. Take h with $t_h \neq 0$. Then

$$t_h \mathbf{x}_h = -\sum_{j \neq h} t_j \mathbf{x}_j$$

implying that $\mathbf{x}_h$ is a linear combination of the other $\mathbf{x}_j$'s. Dividing by t_h, we obtain

$$\mathbf{x}_h = -\frac{1}{t_h} \sum_{j \neq h} t_j \mathbf{x}_j.$$

∎

11.4 Examples of Linear Dependence

The vectors

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 6 \\ 6 \\ 0 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 7 \end{pmatrix}$$

are linearly dependent as $7\mathbf{x}_1 - (7/6)\mathbf{x}_2 - \mathbf{x}_3 = \mathbf{0}$.

Another set of linearly dependent vectors is

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

Here

$$\mathbf{x}_1 - \frac{1}{\sqrt{2}}\left(\mathbf{x}_2 + \mathbf{x}_3\right) = \mathbf{0}.$$

11.5 Linear Dependence and Independence

Linear Independence. We call non-zero vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ *linearly independent* if they are not linearly dependent.

Equivalently, a set of non-zero vectors $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is linearly independent if $\sum_{j=1}^{k} t_j \mathbf{x}_j = \mathbf{0}$ implies $t_1 = t_2 = \cdots = t_k = 0$. Linear independence implies there is at most one vector $\mathbf{t}$ with $\mathbf{x} = \mathbf{X}\mathbf{t}$ where $\mathbf{X}$ is the matrix formed by setting the j^{th} column of $\mathbf{X}$ equal to $\mathbf{x}_j \in \mathcal{X}$.

If there are too many vectors, they must be linearly dependent.

Theorem 11.5.1. *Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are non-zero vectors in $\mathbb{R}^n$ with $k > n$. Then $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly dependent.*

Proof. Consider the equation $\mathbf{X}\mathbf{t} = \mathbf{0}$. Since there are more variables than equations, there is at least one free variable. It follows that $\mathbf{X}\mathbf{t} = \mathbf{0}$ has infinitely many solutions, establishing linear dependence. ∎

The vectors

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

are linearly independent.

Suppose

$$\mathbf{X}\mathbf{t} = t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + t_3\mathbf{x}_3 = \begin{pmatrix} t_1 + t_3 \\ t_1 + t_2 \\ t_2 + t_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Then $t_1 = -t_3$, $t_1 = -t_2$, and $t_2 = -t_3$. Combining these, we find $\mathbf{t} = \mathbf{0}$. Since the only linear combination of $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ that is zero is the zero linear combination, the vector are linearly independent.

► **Example 11.5.2: More than n Vectors are Linearly Dependent.** For example, suppose that in $\mathbb{R}^3$, we have

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \ \mathbf{x}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \ \text{and } \mathbf{x}_4 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

These vectors are linearly dependent because there are too many of them. In fact, $\mathbf{x}_1$ is a linear combination of the others: $\mathbf{x}_1 = (1/2)(\mathbf{x}_2 + \mathbf{x}_3 + \mathbf{x}_4)$. ◄

11.6 Spanning Sets

A second issue concerning coordinate systems is whether our standard set of vectors is big enough to encompass all possible vectors as linear combinations. If so, every vector will have coordinates in our system. If not, there will be vectors outside the coordinate system.

Span. A set of non-zero vectors $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_k\} \subset V$ *spans* a vector space V if $\sum_{j=1}^k t_j\mathbf{x}_j = \mathbf{x}$ has a solution for every $\mathbf{x} \in V$.

Equivalently, $\mathbf{x}_1, \dots, \mathbf{x}_k$ span V if every vector in V is a linear combination of $\mathbf{x}_1, \dots, \mathbf{x}_k$. If the vectors we are using to build a coordinate system span V, then every vector can be written using our coordinate system.

Any set that spans $\mathbb{R}^n$ must contain at least n vectors.

Theorem 11.6.1. *If $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ is a set of non-zero vectors that spans $\mathbb{R}^n$, then $k \ge n$.*

Proof. If $\mathcal{X}$ spans $\mathbb{R}^n$, construct $\mathbf{X}$ as above. Corollary 7.22.1 tells us that rank $\mathbf{X} = n$, which implies $k \ge n$. ∎

More generally, if V is a vector subspace of $\mathbb{R}^n$, $\mathbf{x}_1, \dots, \mathbf{x}_k$ span V if $\mathbf{X}\mathbf{t} = \mathbf{x}$ has a solution for every $\mathbf{x} \in V$.

The set

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

spans $\mathbb{R}^2$. To see it, suppose

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = t_1\mathbf{x}_1 + t_2\mathbf{x}_2 + t_3\mathbf{x}_3 = \begin{pmatrix} t_1 + t_2 + t_3 \\ t_1 - t_2 \end{pmatrix}.$$

This system has infinitely many solutions.

$$\mathbf{t} = \begin{pmatrix} x_2 \\ 0 \\ x_1 - x_2 \end{pmatrix} \quad \text{and} \quad \mathbf{t}' = \begin{pmatrix} x_2 + 1 \\ 1 \\ x_1 - x_2 - 2 \end{pmatrix}$$

are two of them.

Larger sets that span will involve some redundancy. Theorem 11.5.1 says they will be linearly dependent, so by Theorem 11.3.1 at least one can be written as a linear combination of the others. So every time it appears in a linear combination, it can be replaced. It is redundant. It is not needed to span the set.

11.7 Basis of a Vector Space

This brings us to the concept of a basis. A set that both spans and is linearly independent.

Basis. A set of non-zero vectors $\{x_1, \dots, x_k\} \subset V$ are a *basis* for a vector space V if

1. $x_1, \dots, x_k$ are linearly independent.
2. $x_1, \dots, x_k$ span V.

Bases are ideal for building coordinate systems. They are neither too big nor too small. A basis is just right. We can write any vector as a linear combination of the basis vectors, and there is only one way to do it, only one set of coordinates for each vector.

Theorem 11.7.1. *Every basis for $\mathbb{R}^n$ has exactly n elements.*

Proof. Suppose $x_1, \dots, x_k$ is a basis for $\mathbb{R}^n$. By Theorem 11.6.1, $k \geq n$, and by Theorem 11.5.1, $k \leq n$. Thus $k = n$. ∎

We can make a basis for a vector space out of anything.

► **Example 11.7.2: Free Vector Space.** Although our vector spaces usually have meaningful scalar products, it is not necessary. Given a set $\mathcal{B}$, we define the *free vector space over* $\mathcal{B}$ as the set of formal linear combinations of elements of $\mathcal{B}$. For example, if $\mathcal{B} = \{\mathbf{red}, \mathbf{water}, \mathbf{Einstein}\}$, then $V = \{x_1\,\mathbf{red} + x_2\,\mathbf{water} + x_3\,\mathbf{Einstein} : x_i \in \mathbb{R}\}$. Here "formal" means that we don't try to interpret what $x = 1.5\,\mathbf{red} + 2.4\,\mathbf{Einstein}$ actually means. Sometimes, the scalar multiples are written in the form $(\alpha, \mathbf{red})$ to emphasize this. Moreover, we don't need to know what addition means either. We just do calculations concerning vectors according to the rules of vector arithmetic. ◄

We can even make a basis out of nothing, the empty set. Let

$$\mathcal{B}' = \Big\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\Big\}$$

and form the free vector space over $\mathcal{B}'$.

11.8 Bases and Independent Sets

9/15/20

Theorem 11.8.1. *Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. Suppose $S \subset V$ has $m > n$ elements. Then the vectors in S are linearly dependent.*

Proof. Let S be as described. We can write $S = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$. Since $\mathcal{B}$ is a basis for V, and $S \subset V$, we can write each $\mathbf{x}_i$ as a linear combination of the basis vectors $\mathcal{B}$. That means there are a_{ij}, for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, with

$$\mathbf{x}_i = \sum_{j=1}^{n} a_{ij}\mathbf{b}_j.$$

To examine linear independence of the $\mathbf{x}_i$, we consider the equation

$$\sum_{i=1}^{m} t_i\mathbf{x}_i = \mathbf{0}.$$

We will show it has non-zero solutions. We start by rewriting it

$$\mathbf{0} = \sum_{i=1}^{m} t_i \left(\sum_{j=1}^{n} a_{ij}\mathbf{b}_j \right) = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} t_i a_{ij} \right) \mathbf{b}_j.$$

As the $\mathbf{b}_j$ are linearly independent, this implies their coefficients are zero. That means

$$\sum_{i=1}^{m} t_i a_{i1} = 0, \sum_{i=1}^{m} t_i a_{i2} = 0, \ldots, \sum_{i=1}^{m} t_i a_{in} = 0.$$

We have n equations in m unknowns, which we can write in matrix form as $\mathbf{A}^T\mathbf{t} = \mathbf{0}$. This homogeneous system not only has a solution, but must have infinitely many solutions because there are more unknowns (m) than equations (n). See Corollary 7.21.2. It follows that there are $t_1, \ldots, t_m$, not all zero, with

$$\sum_{i=1}^{m} t_i\mathbf{x}_i = \mathbf{0}.$$

In other words, the $\{\mathbf{x}_i\}$ must be linearly dependent. ∎

11.9 Testing for a Basis

We can use our various results on solving equations to construct a test to see if $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ form a basis for $\mathbb{R}^n$. Item (4) of the theorem is the test.

Theorem 11.9.1. *Let $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a collection of vectors in $\mathbb{R}^n$. Form the $n \times n$ matrix whose columns $\mathbf{B}$ whose columns are the $\mathbf{b}_j$. Then the following are equivalent.*

1. $\mathbf{b}_1, \ldots, \mathbf{b}_n$ *are linearly independent*
2. $\mathbf{b}_1, \ldots, \mathbf{b}_n$ *span* $\mathbb{R}^n$
3. $\mathbf{b}_1, \ldots, \mathbf{b}_n$ *form a basis for* $\mathbb{R}^n$
4. $\det \mathbf{B}$ *is non-zero*

Proof. (1) implies (2). Linear independence means $\mathbf{B}\mathbf{x} = \mathbf{0}$ has at most one solution, so $\operatorname{rank} \mathbf{B} = \#\text{cols} = n$ by Corollary 7.22.1. As $\mathbf{B}$ is $n \times n$, this is also the number of rows, so $\mathbf{B}\mathbf{x} = \mathbf{y}$ always has a solution by Corollary 7.23.2, showing that the vectors span.

(2) implies (3). We do this by showing (2) implies (1). Just use the same arguments in the opposite order. Then the vectors $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ span and are linearly independent, so they are a basis.

(3) clearly implies (1) and (2). So (1), (2), and (3) are equivalent.

(1)-(3) are equivalent to (4). As we saw above, (1), (2), and (3) are equivalent to $\operatorname{rank} \mathbf{B} = n = \#\text{cols} = \#\text{rows}$, which is equivalent to $\mathbf{B}$ being non-singular (Corollary 7.24.1). Finally, $\det \mathbf{B}$ is non-zero if and only $\mathbf{B}$ is non-singular, completing the proof. ∎

11.10 The Dimension of a Vector Space

A consequence of Theorem 11.8.1 is that every basis of a vector space must be the same size. More precisely.

Basis Theorem. *Suppose a vector space V has a basis $\mathcal{B}$ with n elements. Then every other basis of that vector space must also have n elements.*

Proof. Any other basis must be a linearly independent set, so by Theorem 11.8.1, it cannot have more than n elements.

If there was a basis with fewer than n elements, we could apply Theorem 11.8.1 to determine that $\mathcal{B}$ is not a linearly independent set, and so not a basis. This contradicts our hypothesis, so it is impossible. We conclude that any basis has exactly n elements. ∎

The Basis Theorem lets us define the dimension of a vector space, at least when the dimension is finite.[1]

Suppose a vector space V has a finite basis $\mathcal{B}$. The Basis Theorem tells us that any basis for V will have the same number of elements as $\mathcal{B}$. For vector spaces with a finite basis, we define the *dimension of* V as the number of elements of that basis. By the Basis Theorem, the dimension does not depend on which basis we use. We denote the dimension of V by $\dim V$.

Although we started with a basis with a finite number of elements, there are vector spaces with infinite bases. The arguments become trickier then, and we will not consider that case further other than to give an example.

► Example 11.10.1: Attempted Basis for the Sequence Space. The sequence space $\mathbf{s}$ contains infinite linearly independent sets. For $\mathbf{s}$, define the vectors $\boldsymbol{e}_j$, $j = 1, 2, 3, \ldots$ by $(\boldsymbol{e}_j)_i = \delta_{ij}$. Then $\boldsymbol{e}_1 = (1, 0, 0, \ldots)$, $\boldsymbol{e}_2 = (0, 1, 0, 0, \ldots)$, etc. We now have an infinite set $\mathcal{E} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \ldots\}$ that seem like a possible basis.

The set $\mathcal{E}$ is linearly independent in $\mathbf{s}$. However, $\mathcal{E}$ is not a basis for $\mathbf{s}$ because it does not span $\mathbf{s}$. The problem is that linear combinations involve finite sums, and vectors such as $(1, 1, 1, \ldots)$ cannot be written as a **finite** sum of the $\boldsymbol{e}_j$. ◄

[1] When the vector space is infinite, we have two choices. We can continue to use finite linear combinations (*Hamel basis*), or we can allow infinite sums. If we use the metric we previously defined on $\mathbf{s}$, it is possible to show that the partial sums $\sum_{i=1}^{n} x_i \boldsymbol{e}_i$ converge to $\boldsymbol{x} \in \mathbf{s}$ for every $\boldsymbol{x}$. This is an example of a *Schauder basis*.

27. Subspaces Attached to a Matrix

This chapter draws on Chapter 27 of Simon and Blume.

27.1 Row and Column Spaces

We have previously encountered three vector spaces derived from matrices. They are the kernel (null space), row space, and column space of a matrix $\mathbf{A}$. Each of these spaces has a well-defined dimension.

It's not hard to show that elementary column operations do not affect $\dim \mathrm{Col}(\mathbf{A})$ and elementary row operations do not affect $\dim \mathrm{Row}(\mathbf{A})$.

Theorem 27.1.1. *Let $\mathbf{A}$ be an $m \times n$ matrix. The three elementary row operations do not affect the row space of $\mathbf{A}$ and the three elementary column operations do not affect the column space of $\mathbf{A}$.*

Proof. We will prove the row case. The column case is very similar.

Let $\mathbf{a}_1, \dots, \mathbf{a}_m$ be the rows of $\mathbf{A}$. The row space is the set of their linear combinations, $\mathcal{L}\left[\mathbf{a}_1, \dots, \mathbf{a}_m\right]$. We now consider the three elementary row operations in turn.

(1) If we interchange two rows, the set of vectors is completely unchanged, so the span is also unchanged.

(2) If we multiply row i by $r \neq 0$, the linear combination

$$t_1\mathbf{a}_1 + \cdots + t_i\mathbf{a}_i + \cdots + t_m\mathbf{a}_m$$

is the same as

$$t_1\mathbf{a}_1 + \cdots + \frac{t_i}{r}(r\mathbf{a}_i) + \cdots + t_m\mathbf{a}_m,$$

so the span remains the same.

(3) If we add $r \neq 0$ times row i to row j, the linear combination

$$t_1\mathbf{a}_1 + \cdots + t_i\mathbf{a}_i + \cdots + t_j\mathbf{a}_j + \cdots + t_m\mathbf{a}_m$$

is the same as

$$t_1\mathbf{a}_1 + \cdots + (t_i - r)\mathbf{a}_i + \cdots + t_j(\mathbf{a}_j + r\mathbf{a}_i) + \cdots + t_m\mathbf{a}_m$$

and the again span remains the same.

None of the elementary row operations affect the span. ∎

27.2 Dimensions of Row and Column Spaces

Recall that the rank of a matrix $\mathbf{A}$ is the number of basic variables, which is the number of leading ones in the reduced row-echelon form.

Theorem 27.2.1. *Let* $\mathbf{A}$ *be an* $m \times n$ *matrix. Then* $\dim \mathrm{Row}(\mathbf{A}) = \operatorname{rank} \mathbf{A} = \dim \mathrm{Col}(\mathbf{A})$.

Proof. Since each row in reduced row-echelon form has a leading one that is a pivot, the non-zero rows are linearly independent. They also span the row space, which is unchanged by row reduction, and thus form a basis. It follows that

$$\dim \mathrm{Row}(\mathbf{A}) = \operatorname{rank} \mathbf{A}.$$

The column case is similar. ∎

Since $\dim \mathrm{Row}(\mathbf{A})$ ($\dim \mathrm{Col}(\mathbf{A})$) is unaffected by elementary row (column) operations, the rank is independent of the row (column) reduction used.

One consequence is that the rank of the product $\mathbf{AB}$ is no more than the rank of $\mathbf{A}$, and that the ranks are equal when $\mathbf{B}$ is invertible.

Theorem 27.2.2. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be conformable matrices. Then* $\operatorname{rank}(\mathbf{AB}) \leq \operatorname{rank} \mathbf{A}$. *Moreover, if* $\mathbf{B}$ *is invertible,* $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank} \mathbf{A}$.

Proof. We know that $\operatorname{rank} \mathbf{A} = \dim \mathrm{Col}(\mathbf{A})$. Now the columns of $\mathbf{AB}$ are linear combinations of the columns of $\mathbf{A}$. In fact, if $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the columns of $\mathbf{A}$, then the j^{th} column of $\mathbf{AB}$ is $\sum_{i=1}^{n} b_{ij} \mathbf{a}_i$. But then

$$\operatorname{rank}(\mathbf{AB}) = \dim \mathrm{Col}(\mathbf{AB}) \leq \dim \mathrm{Col}(\mathbf{A}) = \operatorname{rank} \mathbf{A}.$$

Now suppose $\mathbf{B}$ is invertible. The previous result tells us that $\operatorname{rank}(\mathbf{AB})\mathbf{B}^{-1} \leq \operatorname{rank}(\mathbf{AB})$. But $(\mathbf{AB})\mathbf{B}^{-1} = \mathbf{A}(\mathbf{BB}^{-1}) = \mathbf{A}$, so $\operatorname{rank} \mathbf{A} \leq \operatorname{rank}(\mathbf{AB})$. Since $\operatorname{rank}(\mathbf{AB}) \leq \operatorname{rank} \mathbf{A}$, they must be equal. Then $\operatorname{rank} \mathbf{AB} = \mathbf{A}$ when $\mathbf{B}$ is invertible. ∎

27.3 The Kernel of a Matrix

Since the solution set of a linear system is not affected by row operations, neither is the dimension of the kernel.

We also saw in section 10.10 that the solution set to $\mathbf{A}\mathbf{x} = \mathbf{b}$ is a translate of $\ker \mathbf{A}$, so the dimension of the kernel tells us what the solution set looks like in general. It is just a translate of the kernel.

So what is the dimension of the kernel?

Let's look at an example. Suppose the reduced row-echelon form of the coefficient matrix is

$$\mathbf{R} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

There are two basic variables (x_1 and x_3) and four free variables (x_2, x_4, x_5, and x_6), marked in red.

The reduced matrix now gives us two equations that define the kernel:

$$\begin{aligned} 0 &= x_1 + x_2 + x_4 + x_5 + x_6 \\ 0 &= x_3 + x_4 + x_6. \end{aligned}$$

We can find a basis for the kernel by successively setting all of the free variables but one to zero. The other can be anything we want. We choose $+1$. Here are the solutions.

$$\mathbf{b}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{b}_3 = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{b}_4 = \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Each of the $\mathbf{b}_i$ is in the kernel, and they are all linearly independent. The point is that row $i = 2, 4, 5, 6$ is non-zero only in one of the vectors. This happens because each vector is generated by considering a case where only one of the free variables is non-zero, which forces the corresponding row to be non-zero.

The result of all this is that

$$\dim \ker \mathbf{A} = \#\text{free vars}.$$

27.4 The Dimension of the Kernel

The example was all very nice, but it is no substitute for a proof. Fortunately, all the proof needs is to add some words of explanation.

Theorem 27.4.1. *Let $\mathbf{A}$ be an $m \times n$ matrix. Then*

$$\dim \ker \mathbf{A} = \#\text{free vars}.$$

Proof. Row reduce $\mathbf{A}$ to a reduced row-echelon form $\mathbf{R}$. Both $\mathbf{A}$ and $\mathbf{R}$ will have the same kernel because elementary row operations do not affect the solution set.

Write down the homogeneous equations corresponding to the reduced row-echelon form $\mathbf{R}$. For each free variable i, set $x_i = 1$, all the other free variables to zero, and solve for the basic variables. This defines a vector $\mathbf{b}_i$ for each free variable i.

Since each $\mathbf{b}_i$ solves the homogeneous equations, $\mathbf{b}_i \in \ker \mathbf{A}$. Now suppose $\mathbf{x} \in \ker \mathbf{A}$. Then $\mathbf{x}$ solves the reduced system for some values x_i of the free variables. We can write

$$\mathbf{x} = \sum_{i \in \text{free vars}} x_i \mathbf{b}_i$$

showing that the $\mathbf{b}_i$ span $\ker \mathbf{A}$.

Since row i is only non-zero in $\mathbf{b}_i$, the $\mathbf{b}_i$ are linearly independent. It follows that they span $\ker \mathbf{A}$, so

$$\dim \ker \mathbf{A} = \#\text{free vars}.$$

∎

27.5 Fundamental Theorem of Linear Algebra

We know that the number of free variables and the number of basic variables add to the number of variables (n). Combining that with Theorem 27.4.1, we obtain

$$\#\text{basic vars} = n - \dim\ker \mathbf{A}.$$

Since $\operatorname{rank}\mathbf{A}$ is the number of basic variables, we can sum this up as follows.

Fundamental Theorem of Linear Algebra. *Let $\mathbf{A}$ be an $m \times n$ matrix. Then*

$$n = \operatorname{rank}\mathbf{A} + \dim\ker\mathbf{A}.$$

31. Transformations and Coordinates

You'll notice there's no Chapter 31 in Simon and Blume. Some of this material is not in Simon and Blume, some is scattered in the text.

31.1 Isomorphic Vector Spaces

It's often helpful to analyze vector space problems based on general considerations rather than being tied to a specific vector space characterized by a particular basis.

We can change bases of vector spaces regardless of whether we think of them as being subspaces of $\mathbb{R}^n$ or something else entirely. However, if two vector spaces have the same finite dimension there will always be a mapping that will allow us to treat them as identical, as far as all vector space constructions are concerned.

So when are two vector spaces the same? If only the vector space properties matter, the answer is that they are the same if they are isomorphic.

Isomorphic Vector Spaces. Two vector spaces V and W are *isomorphic* if there is a linear transformation $T : V \to W$ that is one-to-one and onto.

The fact that the transformation is linear tells us that it preserves the vector space operations. It is also invertible. For $y \in W$, let $T^{-1}(y)$ denote the unique $x \in V$ with $T(x) = y$. Here such x exist because T maps onto W, and x is unique because T is one-to-one.

31.2 Isomorphic Vector Spaces have the Same Dimension

It is fairly easy to show that isomorphic vector spaces have the same dimension. It's also easy to show that two vector spaces of the same finite dimension are isomorphic.

Theorem 31.2.1. *Let* V *and* W *be finite-dimensional vector spaces. Then* $\dim V = \dim W$ *if and only if there is an isomorphism* $T\colon V \to W$.

Proof. **Only if case:** Let $\mathcal{V}$ be a basis for V and $\mathcal{W}$ be a basis for W and let $\dim V = n$ (which is also $\dim W$ in this part). Then set $T(\boldsymbol{v}_i) = \boldsymbol{w}_i$ and use linearity to define T on all of V. The resulting $T\colon V \to W$.

The linear mapping T maps onto W, because if $\boldsymbol{x} = \sum_j x_j \boldsymbol{w}_j \in W$, then $\boldsymbol{x} = T\big(\sum_j x_j \boldsymbol{v}_j\big)$.

The mapping T is one-to-one because if $T(\boldsymbol{x}) = T(\boldsymbol{y})$ for some $\boldsymbol{x}, \boldsymbol{y} \in V$, then $\sum_j x_j \boldsymbol{w}_j = \sum_j y_j \boldsymbol{w}_j$. As $\mathcal{W}$ is linearly independent, $x_j = y_j$ for all $j = 1, \dots, n$, showing that $\boldsymbol{x} = \boldsymbol{y}$. It follows that T is an isomorphism.

If case: Using the same notation, consider the set $T(\mathcal{V})$. Since T maps onto W, the $\mathcal{L}[\mathcal{V}] = W$, showing that $\dim W \leq \dim V$. We next show that $T(\mathcal{V})$ is linearly independent. Suppose there are t_j, $j = 1, \dots, n$ with $\sum_j t_j T(\boldsymbol{v}_j) = \boldsymbol{0}$. Then $T\big(\sum_j t_j \boldsymbol{v}_j\big) = \sum_j t_j T(\boldsymbol{v}_j) = \boldsymbol{0}$. The fact that T is an isomorphism, and so one-to-one, implies $\sum_j t_j \boldsymbol{v}_j = \boldsymbol{0}$. But $\mathcal{V}$ is a basis, so $t_j = 0$ for all j, showing that $T(\mathcal{V})$ is linearly independent. Since $T(\mathcal{V})$ is a linearly independent set in W, $\dim W \geq n$. Combining the results shows $\dim W = n = \dim V$. ∎

You can even use this to show that any n-dimensional vector space is isomorphic to any free vector space of dimension n.

31.3 Isomorphism: One and a half Examples

► Example 31.3.1: An Isomorphism. Let $W = \{\mathbf{x} \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\}$. This is a two-dimensional subspace and should be isomorphic with $\mathbb{R}^2$. We start by finding a basis for W. The vectors

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

will do (as a linearly independent set in a two-dimensional space, they must be a basis).

Define

$$\begin{aligned} T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= x_1\mathbf{b}_1 + x_2\mathbf{b}_2 \\ &= x_1\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + x_2\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \\ &= \begin{pmatrix} x_1 \\ x_2 \\ -x_1 - x_2 \end{pmatrix} \end{aligned}$$

Because $\mathbf{b}_1$ and $\mathbf{b}_2$ are linearly independent, T is one-to-one, and because $\mathbf{b}_i$ span, it is onto. That makes it an isomorphism. The inverse map is

$$T\begin{pmatrix} y_1 \\ y_2 \\ -y_1 - y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Keep in mind that the components of $\mathbf{y}$ must sum to zero, so once we know y_1 and y_2, the value of y_3 is already determined by the need to stay in W. ◄

We could just as easily use the free vector space with basis $\{\mathbf{red}, \mathbf{water}\}$ and set

$$T(x_1\,\mathbf{red} + x_2\,\mathbf{water}) = \begin{pmatrix} x_1 \\ x_2 \\ -x_1 - x_2 \end{pmatrix}.$$

31.4 Isomorphic Normed Spaces

Normed spaces and inner product spaces have to meet higher standards for isomorphism because they have structure beyond their vector space structure that must be preserved.

Isometric Normed Spaces. An *isomorphism* T *between normed spaces* $(V, \|\cdot\|_1)$ and $(W, \|\cdot\|_2)$ is a vector space isomorphism between V and W that preserves the norm, $\|T(\mathbf{x})\|_2 = \|\mathbf{x}\|_1$. Such isomorphisms are also called *linear isometries*.

Isometries are often linear. To state this more precisely, define the *midpoint* of two vectors $\mathbf{x}$ and $\mathbf{y}$ as $(\mathbf{x} + \mathbf{y})/2$.

We state the following theorem of Mazur and Ulam without proof

Mazur-Ulam Theorem. *Let* V *and* W *be normed spaces and* T *map* V *onto* W *with* $\|T\mathbf{x}\| = \|\mathbf{x}\|$ *and* $T(\mathbf{0}) = \mathbf{0}$. *Then* T *maps midpoints to midpoints and is linear as a map over* $\mathbb{R}$.

This means that T is an isometric isomorphism between V and W. The result can fail in complex vector spaces.

31.5 Isomorphic Inner Product Spaces

Inner product spaces have an even higher standard to uphold for isomorphism. The inner product must be preserved, meaning that angles between vectors remain the same. We will use the notation $\langle x, y\rangle_i$ to distinguish the inner products.

Isomorphic Inner Product Spaces. An *isomorphism* T *between inner product spaces* $\big(V, \langle\cdot,\cdot\rangle_1\big)$ and $\big(W, \langle\cdot,\cdot\rangle_2\big)$ is a vector space isomorphism between V and W that preserves the inner product, $\langle T(x), T(y)\rangle_2 = \langle x, y\rangle_1$.

Suppose $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is an inner product isomorphism when both $\mathbb{R}^n$'s have the Euclidean inner product. Such a mapping is also an isometry. The image of the standard basis is not only a basis (guaranteed by the vector space isomorphism), but must be an orthonormal basis. To see this, compute $T(\boldsymbol{e}_i)\cdot T(\boldsymbol{e}_j) = \boldsymbol{e}_i\cdot\boldsymbol{e}_j = \delta_{ij}$.

The mapping of the standard basis vectors to an orthonormal set means that T is either a rotation, or a rotation together with an reflection.

That such a transformation must be a rotation/reflection is pretty clear in $\mathbb{R}^2$. The transformation axes must be perpendicular, and only a rotation, or rotation plus reflection will do that. In $\mathbb{R}^2$, one reflection maps $\{\boldsymbol{e}_1, \boldsymbol{e}_2\} \mapsto \{\boldsymbol{e}_1, -\boldsymbol{e}_2\}$. Notice that reflecting in one axis and then the other maps $\{\boldsymbol{e}_1, \boldsymbol{e}_2\} \mapsto \{-\boldsymbol{e}_1, -\boldsymbol{e}_2\}$, which is a 180° rotation. You don't have to reflect along an axis, any line through the origin will do. E.g., if we use the 45° line, we get $\{\boldsymbol{e}_1, \boldsymbol{e}_2\} \mapsto \{\boldsymbol{e}_2, \boldsymbol{e}_1\}$.

It's also reasonably clear in $\mathbb{R}^3$ that these are all rotations and/or reflections.

31.6 Automorphisms on ℓ_2^n

An *automorphism* on a vector space V is an isomorphism from V to itself, $T\colon V \to V$. Here will will consider automorphisms that preserve the inner product on ℓ_2^n, Euclidean $\mathbb{R}^n$.

Before proceeding, we prove a useful lemma.

Lemma 31.6.1. *Let V be an inner product space and suppose x and x' obey $x{\cdot}y = x'{\cdot}y$ for every $y \in V$. Then $x = x'$.*

Proof. Now $(x - x'){\cdot}y = 0$ for every $y \in V$. Set $y = x - x'$ to obtain $\|x - x'\| = 0$. It follows that $x = x'$. ∎

Let T be an automorphism on ℓ_2^n. Use the standard basis to find a matrix representation $\mathbf{A}$ of T. Then $(\mathbf{A}x){\cdot}(\mathbf{A}y) = x{\cdot}y$ for all $x, y \in \mathbb{R}^n$. Now

$$x{\cdot}y = (\mathbf{A}x){\cdot}(\mathbf{A}y) = (\mathbf{A}y)^{\mathsf{T}}(\mathbf{A}x) = y^{\mathsf{T}}\big((\mathbf{A}^{\mathsf{T}}\mathbf{A})x\big) = (\mathbf{A}^{\mathsf{T}}\mathbf{A})x{\cdot}y.$$

Since this holds for every $x, y \in \mathbb{R}^n$, Lemma 31.6.1 implies that $\mathbf{A}^{\mathsf{T}}\mathbf{A}x = x$, so $\mathbf{A}^{\mathsf{T}}\mathbf{A} = \mathbf{I}$. Then $\mathbf{A}^{-1} = \mathbf{A}^{\mathsf{T}}$. Consideration of the standard basis showed these are rotations and reflections. In fact, the pure rotations all have $\det \mathbf{A} = +1$, while those involving reflections have $\det \mathbf{A} = -1$. As in $\mathbb{R}^2$, an even number of reflections amounts to a rotation. I think the main reason it is less clear in dimensions higher than 3 is that our intuition doesn't work so well there.

When the vector space is $\mathbb{C}^n$, we use $x{\cdot}y = x^*y = \sum_{j=1}^n \overline{x_j}y_j$. In that case, we find that the inner product is preserved when the basis matrix $\mathbf{U}$ obeys $\mathbf{U}^*\mathbf{U} = \mathbf{I}$. Such matrices are called *unitary matrices* and are the complex version of rotations and reflections. We can still use $\det U = \pm 1$ to sort out which is which.

31.7 Example: A Rotation and its Inverse

▶ Example 31.7.1: 45° Rotation: Done and Undone. The matrix

$$\mathbf{B} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

rotates the coordinates of $\mathbb{R}^2$ by 45°, taking $(1,0)^\mathsf{T}$ to $(1/\sqrt{2}, 1/\sqrt{2})^\mathsf{T}$ and $(0,1)^\mathsf{T}$ to $(-1/\sqrt{2}, 1/\sqrt{2})^\mathsf{T}$. Since this is a rotation, its transpose is also its inverse, and we have

$$\mathbf{B}^{-1} = \mathbf{B}^\mathsf{T} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

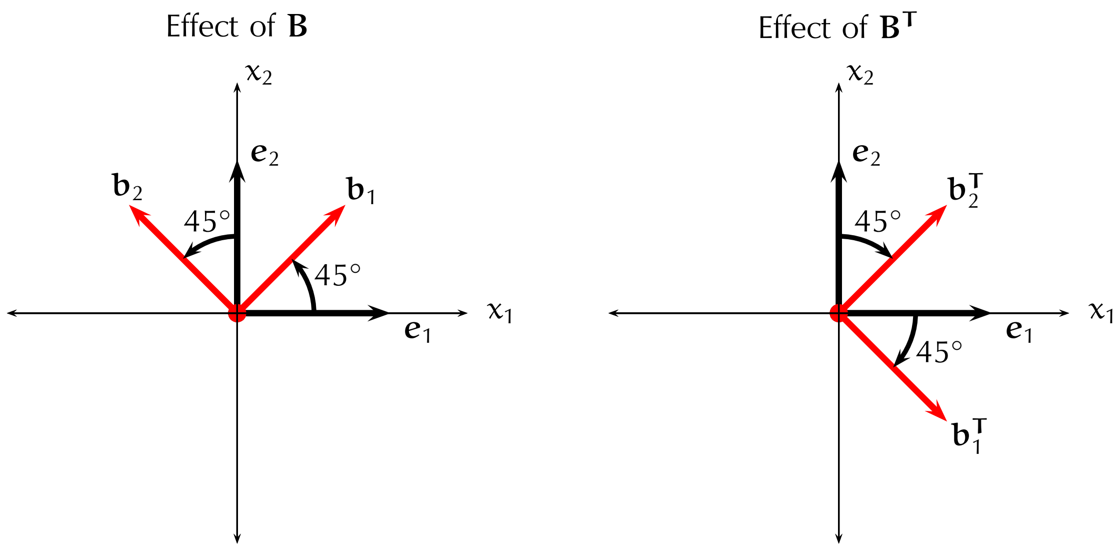


Figure 31.7.2: Here the standard coordinate axes are rotated counter-clockwise by 45° in the left panel. Here $\mathbf{b}_1$ and $\mathbf{b}_2$ are the columns of $\mathbf{B}$. In the right panel, we have the inverse transformation where a 45° clockwise rotation gives us the new coordinate axes. Here $\mathbf{b}_1^\mathsf{T}$ and $\mathbf{b}_2^\mathsf{T}$ are the columns of the matrix $\mathbf{B}^\mathsf{T}$.

◀

31.8 Bases and Coordinates

Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for $\mathbb{R}^n$. Define the basis matrix by lining up the basis vectors in order:

$$\mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix}.$$

Since $\mathcal{B}$ both spans and is a linearly independent set of n vectors in $\mathbb{R}^n$, $\mathbf{B}$ is an $n \times n$ matrix with rank n. That means it is invertible. Given a vector $\mathbf{x} \in \mathbb{R}^n$, we find its vector of coordinates $\mathbf{t_B}$ in the $\mathcal{B}$ basis by solving the equation

$$\mathbf{x} = \mathbf{B}\mathbf{t}_\mathbf{B}.$$

Because $\mathbf{B}$ is invertible, $\mathbf{t_B} = \mathbf{B}^{-1}\mathbf{x}$.

This all applies to the standard basis $\mathcal{E} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. In that case, the basis matrix $\mathbf{E}$ is the $n \times n$ identity matrix. Nonetheless, we use a special name for it to emphasize that we are doing basis calculations. Given a vector $\mathbf{x}$, expressed in the standard coordinates, we find that $\mathbf{t_E} = \mathbf{E}^{-1}\mathbf{x} = \mathbf{I}\mathbf{x} = \mathbf{x}$, meaning that the coordinates are what we think they are.

31.9 Example: Coordinates in $\mathbb{R}^3$

Let's see how this works in $\mathbb{R}^3$. The basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$, defined by

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}, \mathbf{b}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

gives us the basis matrix

$$\mathbf{B} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & 1 \\ 3 & 2 & 1 \end{pmatrix}.$$

We now use the formula

$$\mathbf{x} = \mathbf{B}\mathbf{t}_\mathbf{B}$$

To change from $\mathcal{B}$ coordinates to standard coordinates. The vectors with coordinates $\mathbf{t}_\mathbf{B} = (1,1,0)^\mathsf{T}$ and $\mathbf{t}'_\mathbf{B} = (1,0,3)$ yield the vectors $\mathbf{x} = \mathbf{B}\mathbf{t}_\mathbf{B} = (3,2,5)^\mathsf{T}$ and $\mathbf{x}' = \mathbf{B}\mathbf{t}'_\mathbf{B} = (4,5,6)$ in standard coordinates.

Let's see how it works the other way, going from standard basis $\mathcal{E}$ coordinates to $\mathcal{B}$ coordinates. For that, we use the formula

$$\mathbf{t}_\mathbf{B} = \mathbf{B}^{-1}\mathbf{x}.$$

The inverse of $\mathbf{B}$ is

$$\mathbf{B}^{-1} = \begin{pmatrix} -1/2 & 0 & 1/2 \\ 1/4 & -1/2 & 1/4 \\ 1 & 1 & -1 \end{pmatrix}.$$

The vector $\mathbf{x} = (3,2,1)^\mathsf{T}$ then has coordinates $\mathbf{t}_\mathbf{B} = \mathbf{B}^{-1}\mathbf{x} = (-1,0,4)^\mathsf{T}$ in the basis $\mathcal{B}$, meaning that $\mathbf{x} = -\mathbf{b}_1 + 4\mathbf{b}_3$. The vector $\mathbf{x}' = (-1,-1,+5)^\mathsf{T}$ has coordinates $\mathbf{t}'_\mathbf{B} = \mathbf{B}^{-1}\mathbf{x}' = (3, 3/2, -7)$ in the $\mathcal{B}$ basis, so that $\mathbf{x}' = 3\mathbf{b}_1 + (3/2)\mathbf{b}_2 - 7\mathbf{b}_3 = (-1,-1,+5)$ in the standard basis.

31.10 Changing Coordinate Systems

Consider two different bases (aka coordinate systems), $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, and $\mathcal{B}' = \{\mathbf{b}'_1, \ldots, \mathbf{b}'_n\}$, we form the corresponding basis matrices $\mathbf{B}$ and $\mathbf{B}'$.

Given a vector $\mathbf{x}$, we can write it in the two coordinate systems as $\mathbf{x} = \mathbf{B}\mathbf{t}_\mathbf{B}$ and $\mathbf{x} = \mathbf{B}'\mathbf{t}_{\mathbf{B}'}$. Then $\mathbf{B}\mathbf{t}_\mathbf{B} = \mathbf{B}'\mathbf{t}_{\mathbf{B}'}$. Solving for $\mathbf{t}_\mathbf{B}$ and $\mathbf{t}_{\mathbf{B}'}$, we derive the change of coordinates formulas:

$$\mathbf{t}_\mathbf{B} = \left(\mathbf{B}^{-1}\mathbf{B}'\right)\mathbf{t}_{\mathbf{B}'} \quad \text{and} \quad \mathbf{t}_{\mathbf{B}'} = \left((\mathbf{B}')^{-1}\mathbf{B}\right)\mathbf{t}_\mathbf{B}$$

Starting with the $\mathcal{B}'$ coordinates, we multiply by $\mathbf{B}'$ to get the actual vector $\mathbf{x}$, and then multiply by $\mathbf{B}^{-1}$ to put it into the $\mathcal{B}$ coordinate system. Conversely, to convert the $\mathcal{B}$ coordinates to $\mathcal{B}'$ coordinates, we reverse the process, multiplying first by $\mathbf{B}$, and then by $(\mathbf{B}')^{-1}$.

To see how this works, suppose

$$\mathbf{B} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{B}' = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

are the basis matrices. Then

$$\mathbf{B}^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad (\mathbf{B}')^{-1} = \frac{1}{3}\begin{pmatrix} -1 & 2 \\ 2 & -1 \end{pmatrix}.$$

Consider the vector with $\mathcal{B}'$ coordinates $\mathbf{t}_{\mathbf{B}'} = (1, 4)^\mathsf{T}$. Using the formula, we obtain the $\mathcal{B}$ coordinates

$$\mathbf{t}_\mathbf{B} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

Let's check it. Now $\mathbf{t}_{\mathbf{B}'} = (1, 4)^\mathsf{T}$ corresponds to

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 4\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 9 \\ 6 \end{pmatrix}$$

and

$$\mathbf{x} = 3\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 6\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 9 \\ 6 \end{pmatrix}.$$

This shows that both expressions refer to the same vector $\mathbf{x}$, whose standard coordinates are

$$\mathbf{x} = \begin{pmatrix} 9 \\ 6 \end{pmatrix}.$$

31.11 Linear Transformations and Bases

So how do these coordinate change affect linear transformations? Take a linear transformation on $\mathbb{R}^n$, $T\colon \mathbb{R}^n \to \mathbb{R}^n$. As we saw in section 10.5, we can use the standard basis to represent this in matrix form, so that $T(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$.

So what happens if we want to use a different basis? One reason to do such a thing would be to write the transformation in a more convenient form—one that is easier to calculate or interpret.

Let $\mathbf{B}$ be a basis matrix for $\mathcal{B}$. We will write the matrix for T as $\boldsymbol{A}_{\mathcal{E}}$ when it is in standard ($\mathcal{E}$) coordinates and $\boldsymbol{A}_{\mathcal{B}}$ when it is in $\mathcal{B}$ coordinates.

To find $\boldsymbol{A}_{\mathcal{B}}$ from $\boldsymbol{A}_{\mathcal{E}}$, we start with $\mathcal{B}$ coordinates $\mathbf{t}_{\mathbf{B}}$, then convert them to standard coordinates, $\boldsymbol{x} = \mathbf{B}\mathbf{t}_{\mathbf{B}}$. Then we feed this to the matrix in standard coordinates, obtaining $\boldsymbol{A}_{\mathcal{E}}\mathbf{B}\mathbf{t}_{\mathbf{B}}$.

As this is in standard coordinates, we have to convert the result back to the $\mathbf{B}$ coordinates. We do this by multiplying on the left by $\mathbf{B}^{-1}$. That gives us $(\mathbf{B}^{-1}\boldsymbol{A}_{\mathcal{E}}\mathbf{B})\mathbf{t}_{\mathbf{B}}$ as the $\mathbf{B}$ coordinates of the transformed vector. Thus

$$\boldsymbol{A}_{\mathcal{B}} = \mathbf{B}^{-1}\boldsymbol{A}_{\mathcal{E}}\mathbf{B} \text{ or } \mathbf{B}\boldsymbol{A}_{\mathcal{B}}\mathbf{B}^{-1} = \boldsymbol{A}_{\mathcal{E}}. \tag{31.11.1}$$

is the matrix for T in $\mathcal{B}$ coordinates.

Things are a bit more complicated if we had originally used a basis other than the standard basis. If the transformation had been written in $\mathbf{B}'$ coordinates, we would multiply by $\big((\mathbf{B}')^{-1}\mathbf{B}\big)$ to convert $\mathbf{B}$ coordinates to $\mathbf{B}'$ coordinates, apply $\boldsymbol{A}$, then convert back. The result is:

$$\boldsymbol{A}_{\mathcal{B}} = \big(\mathbf{B}^{-1}\mathbf{B}'\big)\boldsymbol{A}_{\mathcal{B}'}\big((\mathbf{B}')^{-1}\mathbf{B}\big).$$

Another way to write this that may make the method clearer is transform them into standard coordinates:

$$\mathbf{B}\boldsymbol{A}_{\mathcal{B}}\mathbf{B}^{-1} = \boldsymbol{A}_{\mathcal{E}} = \mathbf{B}'\boldsymbol{A}_{\mathcal{B}'}(\mathbf{B}')^{-1}.$$

31.12 Example: Linear Transformation Basis Change

9/17/20

Suppose our new basis is

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

with basis matrix

$$\mathbf{B} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

This is an orthogonal basis since the vectors are perpendicular, but not orthonormal because they have length $\sqrt{2}$. This means that $\mathbf{B}^{-1} = (1/2)\mathbf{B}^{\mathsf{T}}$.

Suppose our linear transformation T has matrix

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

in the standard basis.

To find its representation $\mathbf{A}_{\mathcal{B}}$ in the $\mathcal{B}$ basis, we first compute

$$\mathbf{B}^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Then our new transformation matrix is

$$\begin{aligned} \mathbf{A}_{\mathcal{B}} &= \mathbf{B}^{-1}\mathbf{A}\mathbf{B} \\ &= \frac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

As you can see, the transformation has taken a particularly simple form: The transformed matrix $\mathbf{A}_{\mathcal{B}}$ is diagonal. This reflects the fact that $\mathrm{T}(\mathbf{b}_1) = 2\mathbf{b}_1$ and $\mathrm{T}(\mathbf{b}_2) = \mathbf{b}_2$. In Chapter 23, you will learn how we can sometimes find such a basis from the original matrix.

31.13 Example: Transformations with Complex Numbers

Consider the matrix
$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

In section 8.21, we saw that $\mathbf{A}$ is a square root of $-\mathbf{I}$, the negative of the identity matrix.

By using a complex basis, we can see just how true that is. There is a complex basis $\mathbf{B}$ where $\mathbf{A}_{\mathcal{B}}$ is purely imaginary in the weak sense that all of its non-zero elements are purely imaginary. Indeed, the non-zero elements are square roots of -1.

Consider the basis with basis matrix

$$\mathbf{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}.$$

It is a unitary matrix. Its inverse is its Hermitian conjugate

$$\mathbf{B}^{-1} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}.$$

Now

$$\begin{aligned}
\mathbf{A}_{\mathcal{B}} &= \mathbf{B}^{-1}\mathbf{A}\mathbf{B} \\
&= \frac{1}{2}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \\
&= \frac{1}{2}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}\begin{pmatrix} i & -i \\ -1 & -1 \end{pmatrix} \\
&= \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}
\end{aligned}$$

revealing that the matrix $\mathbf{A}$ is more closely connected to the imaginary numbers than we first realized.

If you've had a comprehensive linear algebra course, you may have seen such transformations before. If you had a differential equations class that covered linear differential systems, you may have seen them there too. As was true of the previous page, you will learn how find these transformations like this in Chapter 23.

31.14 The Dual Space

In section 10.6, we defined linear functionals, linear functions from a real vector space $\mathbb{R}^n$ to $\mathbb{R}$. More generally, let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A linear function f from a finite-dimensional vector space V over $\mathbb{F}$ to $\mathbb{F}$ is called a *linear functional*. We saw earlier that any linear functional on $\mathbb{F}^n$ can be represented by a $1 \times n$ matrix, a horizontal vector, a *covector*.[1] The *dual space* of V is the set of all linear functionals on V and is denoted V^*. Since the set of $1 \times n$ matrices is a vector space of dimension $n = \dim V$, $\dim V^* = \dim V$.

The most common duality in economics involves prices and quantities. We think of quantities $\mathbf{x} \in \mathbb{R}^n$ and prices in $(\mathbb{R}^n)^*$, writing $\mathbf{px}$ for cost. Some problems are better studied using functions of quantity (e.g., utility, production), while others are better studied using dual functions of price (cost, expenditure, indirect utility).

To see how duality relates to bases, we treat f as we do any other linear transformation. We express $\mathbf{x}$ using the standard basis and use the linearity of f to write:

$$\begin{aligned} f(\mathbf{x}) &= f\left(\sum_{j=1}^{n} x_j \mathbf{e}_j\right) \\ &= \sum_{j=1}^{n} x_j f(\mathbf{e}_j) \\ &= \begin{pmatrix} f(\mathbf{e}_1) & \cdots & f(\mathbf{e}_n) \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \end{aligned} \tag{31.14.2}$$

So any linear functional f on V defines a $1 \times n$ matrix $\mathbf{v}_f$ by

$$\mathbf{v}_f = \begin{pmatrix} f(\mathbf{e}_1) & \cdots & f(\mathbf{e}_n) \end{pmatrix}.$$

Reading equation (31.14.2) up from the bottom makes it clear that any $1 \times n$ matrix defines a linear functional on V, and vice-versa. We can think of the linear functionals as $1 \times n$ matrices.

In fact, if V is an inner product space, we can identify $\mathbf{x} \in V$ with the dual element $\mathbf{x}^*$ since $\mathbf{y} \mapsto \mathbf{x}\cdot\mathbf{y} = \mathbf{x}^*\mathbf{y}$ is a linear functional on V. However, this mapping is only linear if V is a real vector space. If it is a complex vector space, the mapping is conjugate linear.

[1] When dealing with infinite-dimensional spaces, a distinction is made between linear functions from V to $\mathbb{R}$, sometimes called *linear forms* and continuous linear functions from V to $\mathbb{R}$, called *linear functionals*. We are avoiding these technical issues by restricting ourselves to finite-dimensional spaces.

31.15 The Dual Basis

Given a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ for V, we can define a corresponding dual basis $\mathcal{B}^*$ for V^* by setting $\mathbf{b}_i^*(\mathbf{b}_j) = \delta_{ij}$.

It is easy to verify that this is a basis for V^*.

Theorem 31.15.1. *Let V be a finite-dimensional vector space. The dual basis is a basis for* V^*

Proof. Let f be a linear functional on V. We write $\mathbf{x}$ in the basis $\mathcal{B}$ as $\mathbf{x} = \sum_j x_j \mathbf{b}_j$. Then

$$\mathbf{b}_i^*(\mathbf{x}) = \sum_{j=1}^{n} x_j \mathbf{b}_i^*(\mathbf{b}_j) = \sum_{j=1}^{n} x_j \delta_{ij} = x_i.$$

Now expand $\boldsymbol{v}_f \mathbf{x} = f(\mathbf{x})$:

$$\boldsymbol{v}_f \mathbf{x} = f(\mathbf{x}) = f\left(\sum_{j=1}^{n} x_j \mathbf{b}_j\right) = \sum_{j=1}^{n} x_j f(\mathbf{b}_j) = \sum_{j=1}^{n} f(\mathbf{b}_j)\mathbf{b}_j^*(\mathbf{x}).$$

This shows that $f = \boldsymbol{v}_f = \sum_j f(\mathbf{b}_j)\boldsymbol{e}_j^*$, meaning that $\mathcal{B}^*$ spans V^*.

Next, we consider linear independence. Suppose $f = \boldsymbol{v}_f = \sum_{j=1}^{n} x_j \mathbf{b}_j^* = 0$. Then for each $\mathbf{b}_i$, $i = 1, \ldots, n$,

$$0 = f(\mathbf{b}_i) = \boldsymbol{v}_f(\mathbf{b}_i) = \sum_{j=1}^{n} x_j \mathbf{b}_j^*(\mathbf{b}_i) = \sum_{j=1}^{n} x_j \delta_{ij} = x_i.$$

But then $x_i = 0$ for $i = 1, \ldots, n$, showing that $\mathcal{B}^*$ is a linearly independent set, and therefore a basis for V^*. ∎

We can now define the *standard dual basis* by $\{\boldsymbol{e}_1^*, \ldots, \boldsymbol{e}_n^*\}$ for $(\mathbb{R}^n)^*$ by $\boldsymbol{e}_i^*(\boldsymbol{e}_j) = \delta_{ij}$. Thus the dual basis vectors $\boldsymbol{e}_i^*$ are the row vectors $\boldsymbol{e}_1^* = (1, 0, 0, \ldots, 0)$, $\boldsymbol{e}_2^* = (0, 1, 0, \ldots, 0)$, ..., $\boldsymbol{e}_n^* - (0, \ldots, 0, 1)$. This allows us to write any $f \in V^*$ as

$$f = \sum_{i=1}^{n} f(\boldsymbol{e}_i)\boldsymbol{e}_i^* = \big(f(\boldsymbol{e}_1), \ldots, f(\boldsymbol{e}_n)\big).$$

The following corollary is similar to Lemma 31.6.1, but applies to any dual space and doesn't require that V be an inner product space.

Corollary 31.15.2. *Let V be a finite-dimensional vector space. Suppose* $f(\mathbf{x}) = f(\mathbf{y})$ *for all* $f \in V^*$*, then* $\mathbf{x} = \mathbf{y}$

Proof. Let $\mathcal{B}$ be a basis for V. We can write $\mathbf{x} = \sum_j x_j \mathbf{b}_j$ and $\mathbf{y} = \sum_j y_j \mathbf{b}_j$. Since $\mathbf{b}_i^* \in V$, $x_i = \mathbf{b}_i^*(\mathbf{x}) = \mathbf{b}_i^*(\mathbf{y}) = y_i$ for every $i = 1, \ldots, n$. Then $\mathbf{x} = \mathbf{y}$ by linear independence of the $\mathbf{b}_i$. ∎

31.16 Coordinate Change in the Dual Space

Although we have a formula for the dual basis, we still need to fully identify it. Since the dual basis consists of covectors (row vectors), we form the dual basis matrix $\hat{\mathbf{B}}$ by stacking the rows.[2]

$$\hat{\mathbf{B}} = \begin{pmatrix} \mathbf{b}_1^* \\ \mathbf{b}_2^* \\ \vdots \\ \mathbf{b}_n^* \end{pmatrix}.$$

Then the ij coordinate of $\hat{\mathbf{B}}\mathbf{B}$ is $\mathbf{b}_i^*\mathbf{b}_j = \delta_{ij}$, meaning that $\hat{\mathbf{B}}\mathbf{B} = \mathbf{I}$. The dual basis matrix is simply $\mathbf{B}^{-1}$, keeping in mind that we are using the rows of $\mathbf{B}^{-1}$, not the columns. We formalize this as the following theorem.

Theorem 31.16.1. *Let $\mathcal{B}$ be a basis for an n-dimensional real vector space V and $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ be the basis matrix. Then the dual basis $\mathcal{B}^* = \{\mathbf{b}_1^*, \dots, \mathbf{b}_n^*\}$ has basis matrix $\mathbf{B}^{-1}$.*

Of course, to change coordinates, $\mathbf{t}_\mathbf{B} = \mathbf{B}^{-1}\mathbf{x}$ in V. The dual space works a little differently as the basis matrix must multiply the coordinate vector on the right. Thus if we have a linear functional f defined by the covector $\boldsymbol{\nu}_f$ in the standard basis, the coordinates in the basis $\mathcal{B}^*$ are $\mathbf{t}_\mathbf{B}^* = \boldsymbol{\nu}_f(\hat{\mathbf{B}})^{-1} = \boldsymbol{\nu}_f\mathbf{B}$. Since the coordinates vary directly with the basis matrix, the vector is called *covariant*. With ordinary vectors, we use the inverse of the basis matrix, and call them *contravariant* as a result.

It follows that

$$\mathbf{t}_\mathbf{B}^*(\mathbf{t}_\mathbf{B}) = \big(\boldsymbol{\nu}_f\mathbf{B}\big)\big(\mathbf{B}^{-1}\mathbf{x}\big) = \boldsymbol{\nu}_f\big(\mathbf{B}\mathbf{B}^{-1}\big)\mathbf{x} = \boldsymbol{\nu}_f\mathbf{x} = f(\mathbf{x}),$$

showing that $f(\mathbf{x})$ is unaffected by this double change of coordinates, which is what we need.

▶ Example 31.16.2: Gallons vs. Quarts. Going back to the beginning of the Chapter 10 notes, this means if we measures milk in quarts rather than gallons, and milk is good k, then the coordinate change for quantities is given by $\mathbf{B}^{-1} = \text{diag}(1, \dots, 1, 4, 1, \dots, 1)$ where 4 is in the k^{th} row. Then $\mathbf{B} = \text{diag}(1, \dots, 1, 1/4, 1, \dots, 1)$, so the corresponding (dual) price vector must be multiplied by 1/4. ◀

[2] I don't use $\mathbf{B}^*$ because of potential confusion with the Hermitian conjugate.

31.17 Another Duality Example

Rotations and reflections are a little different from the other transformations of bases. For starters, $\mathbf{B}^{-1} = \mathbf{B}^{\mathsf{T}}$. Now when we apply $\mathbf{B}^{-1} = \mathbf{B}^{\mathsf{T}}$ to the coordinates of a vector, we taking sums of the columns of $\mathbf{B}^{\mathsf{T}}$. When we apply $\mathbf{B}$ to a covector, we obtain sums of the rows of $\mathbf{B}$, which are the columns of $\mathbf{B}^{\mathsf{T}}$. The action is the same on both the vectors and covectors. Let's see how this works with an orthonormal basis.

▶ Example 31.17.1: 45° Rotation and Duality. The matrix

$$\mathbf{B} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

rotates the coordinates of $\mathbb{R}^2$ by 45°, mapping

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \mapsto \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mapsto \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Since this is a rotation, its transpose is also its inverse, and we have

$$\mathbf{B}^{-1} = \mathbf{B}^{\mathsf{T}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Now what happens to the dual basis? It is also rotated by 45° as the covector $(1,0) \mapsto (1/\sqrt{2}, 1/\sqrt{2})$ and $(0,1) \mapsto (-1/\sqrt{2}, 1/\sqrt{2})$.

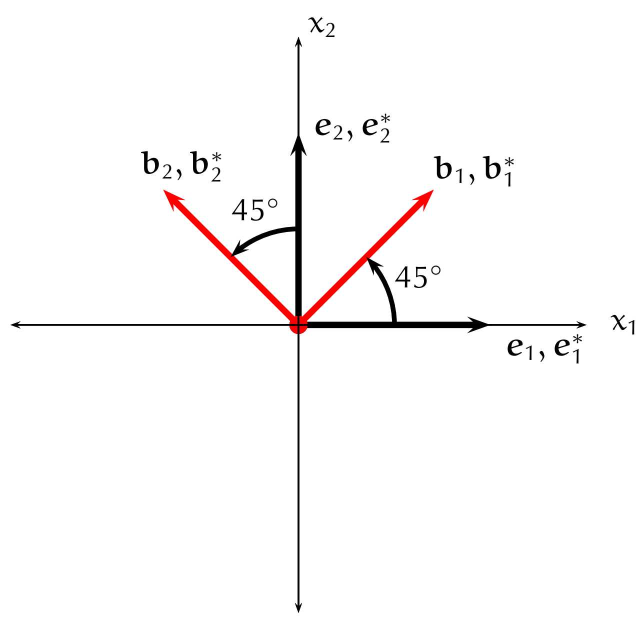
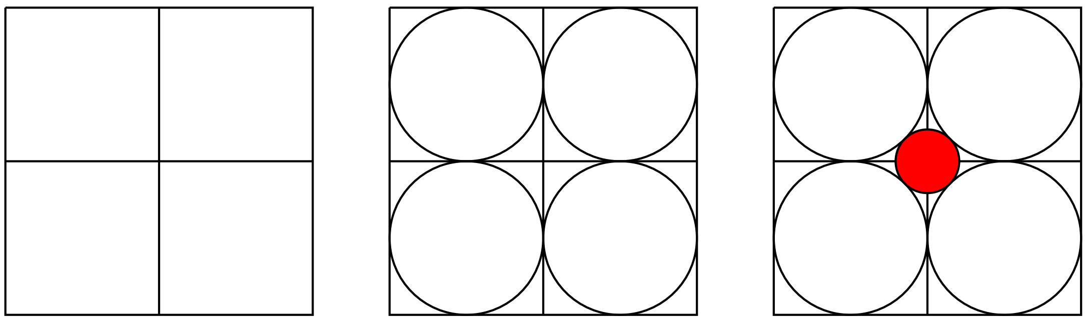


Figure 31.17.2: Here the standard coordinate axes are rotated counter-clockwise by 45°. The standard dual basis lines up with the standard basis itself. Because the new basis is an orthonormal basis, the dual basis must rotate to match.

◀

31.18 $\mathbb{R}^n$ Geometry Puzzle

Consider a square with 2-foot sides. Divide that square into 4 quadrants, with sides 1-foot each. Then inscribe a circle into each quadrant. Finally, inscribe a small circle in the middle of the circles (shown in red below).

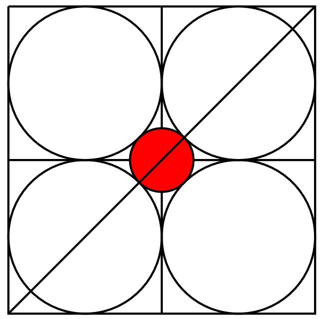


Suppose we try an analogous construction in $\mathbb{R}^3, \mathbb{R}^4, \ldots$. In $\mathbb{R}^3$, we start with a cube with 2-foot sides. We bisect each side with planes, inscribe the 1-foot spheres, then inscribe the red 2-sphere in the middle. In $\mathbb{R}^4$, we have a tesseract with 2-foot sides, we bisect each side with 3-d "planes", inscribe the 1-foot 3-spheres, then the red 3-sphere in the middle. We do this for each $\mathbb{R}^n$ with $n \geq 2$.

Problem: What does the diameter of the red sphere converge to as $n \to \infty$?

(A) 0. (B) 1. (C) 2. (D) ∞.

31.19 The Answer to the Puzzle



As shown in the diagram, we draw the diagonal of the 1-foot square (cube, tesseract, etc.). We are in Euclidean space, so in $\mathbb{R}^2$ the diagonal has length $L_2 = \sqrt{2^2 + 2^2} = 2\sqrt{2}$.

In $\mathbb{R}^n$, the length is $L_n = 2\sqrt{n}$. Examining the diagonal more closely, we find that after subtracting the diameter of 1-foot circles, we have $2\sqrt{n} - 2$. This includes both the diameter of the red circle and the part sticking out of the 1-foot circles at either end. By symmetry, the portion sticking out of the 1-foot circle has the same length as the radius, so the leftover portion ($2\sqrt{n} - 2$) is 4 times the radius, or half the diameter of the red circle.

That means the red circle has diameter $d_n = \sqrt{n} - 1$. When $n = 2$, $d_2 \approx .414$. when $n = 4$, $d_4 = 2 - 1 = 1$. When $n = 9$, $d_9 = 3 - 1 = 2$. At that point the red hypersphere in the middle touches the sides of the large enclosing box. For $n > 9$, the red hypersphere actually pokes out. We can see now that the correct answer was (D) ∞. The inside gets much roomier as the number of dimensions increases, which allows the red hypersphere to partly escape the containment by the other hyperspheres.

32. Tensors and Tensor Products

Here we build up some of the basic facts concerning tensors. We will later use tensors to express the multidimensional version of Taylor's Theorem. They are also useful background for exterior products (a special kind of tensor) and their relation to multidimensional integrals.

32.1 Outer Products

We've spent some time on inner products. So if there's an inner product, is there also an outer product? Indeed there is! There's also something called the exterior product which we will encounter later.

Given $\boldsymbol{x} \in \mathbb{R}^m$ and $\boldsymbol{y} \in \mathbb{R}^n$ the *outer product* $\boldsymbol{x} \otimes \boldsymbol{y}$ is an $m \times n$ matrix defined by $(\boldsymbol{x} \otimes \boldsymbol{y})_{ij} = x_i y_j$. This can also be written $\boldsymbol{x} \otimes \boldsymbol{y} = \boldsymbol{x}\boldsymbol{y}^T$. An immediate consequence is that $(\boldsymbol{x} \otimes \boldsymbol{y})^T = \boldsymbol{y}\boldsymbol{x}^T = \boldsymbol{y} \otimes \boldsymbol{x}$.

For example, if $\boldsymbol{x} \in \mathbb{R}^3$ and $\boldsymbol{y} \in \mathbb{R}^2$,

$$\boldsymbol{x} \otimes \boldsymbol{y} = \begin{pmatrix} x_1 y_1 & x_1 y_2 \\ x_2 y_1 & x_2 y_2 \\ x_3 y_1 & x_3 y_2 \end{pmatrix}.$$

More generally, $\boldsymbol{x} \otimes \boldsymbol{y}$ is an $m \times n$ matrix with ij element $x_j y_i$.

There are two important relations:

$$(\alpha \boldsymbol{x}) \otimes \boldsymbol{y} = \boldsymbol{x} \otimes (\alpha \boldsymbol{y}) = \alpha(\boldsymbol{x} \otimes \boldsymbol{y})$$

and

$$\begin{aligned} \boldsymbol{x} \otimes (\boldsymbol{y} + \boldsymbol{z}) &= \boldsymbol{x} \otimes \boldsymbol{y} + \boldsymbol{x} \otimes \boldsymbol{z} \\ (\boldsymbol{x} + \boldsymbol{z}) \otimes \boldsymbol{y} &= \boldsymbol{x} \otimes \boldsymbol{y} + \boldsymbol{z} \otimes \boldsymbol{y}. \end{aligned}$$

You can see that some sort of bilinearity has been built into the outer product.

Outer products can be used to represent tensor products, but there is an important difference. The outer product obeys $(\boldsymbol{x} \otimes \boldsymbol{y})^T = \boldsymbol{y} \otimes \boldsymbol{x}$, which often doesn't make sense for tensor products.

32.2 Outer Products and Bilinearity

Tf T is a linear map of $m \times n$ matrices (considered as a vector space) into $\mathbb{R}$, $T(\mathbf{x} \otimes \mathbf{y})$ is a bilinear function of $(\mathbf{x}, \mathbf{y})$.

Theorem 32.2.1. *Let* T *be a linear function on the set of* $m \times n$ *matrices,* $\mathbf{x} \in \mathbb{R}^m$ *and* $\mathbf{y} \in \mathbb{R}^n$*. Then the function* $f(\mathbf{x}, \mathbf{y}) = T(\mathbf{x} \otimes \mathbf{y})$ *is a bilinear function on* $\mathbb{R}^m \times \mathbb{R}^n$.

Proof. To see this, note that

$$T\big((\alpha\mathbf{x}) \otimes \mathbf{y}\big) = T\big(\mathbf{x} \otimes (\alpha\mathbf{y})\big) = T\big(\alpha(\mathbf{x} \otimes \mathbf{y})\big) = \alpha T(\mathbf{x} \otimes \mathbf{y})$$

and

$$\begin{aligned} T\big((\mathbf{x} + \mathbf{y}) \otimes \mathbf{z}\big) &= T(\mathbf{x} \otimes \mathbf{z} + \mathbf{y} \otimes \mathbf{z}) = T(\mathbf{x} \otimes \mathbf{z}) + T(\mathbf{y} \otimes \mathbf{z}) \\ T\big(\mathbf{x} \otimes (\mathbf{y} + \mathbf{z})\big) &= T(\mathbf{x} \otimes \mathbf{y} + \mathbf{x} \otimes \mathbf{z}) = T(\mathbf{x} \otimes \mathbf{y}) + T(\mathbf{x} \otimes \mathbf{z}). \end{aligned}$$

Together, these imply that $f(\mathbf{x}, \mathbf{y}) = T(\mathbf{x} \otimes \mathbf{y})$ is bilinear in $(\mathbf{x}, \mathbf{y})$. ∎

According to Theorem 10.5.1, we can always represent a linear transformation between vector spaces by a matrix. Here T maps the vector space of $m \times n$ matrices into the real numbers. All we need is a basis for the matrices to help us write the transformation in matrix form.

We will take the basis of the $m \times n$ real matrices defined by $\mathbf{b}_{ij}$ is the $m \times n$ matrix with 1 in position ij, and zero elsewhere. It is easy to see this spans the $m \times n$ matrices and is linearly independent. In fact, the ij basis element is $\mathbf{e}_i \otimes \mathbf{e}_j$.

Define an $m \times n$ matrix $\mathbf{A}$ by $a_{ij} = T(\mathbf{e}_i \otimes \mathbf{e}_j)$. We can represent f and T by the matrix $\mathbf{A}$ as

$$f(\mathbf{x}, \mathbf{y}) = T(\mathbf{x} \otimes \mathbf{y}) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j,$$

which is a bilinear form (2-tensor).

32.3 The Vector Space of Outer Products

We can define a basis on the set of $m \times n$ outer products. One way is to use the standard basis of $\boldsymbol{e}_i$. Then $\boldsymbol{e}_1 \otimes \boldsymbol{e}_1, \boldsymbol{e}_1 \otimes \boldsymbol{e}_2, \ldots, \boldsymbol{e}_1 \otimes, \boldsymbol{e}_n, \boldsymbol{e}_2 \otimes \boldsymbol{e}_1, \ldots, \boldsymbol{e}_m \otimes \boldsymbol{e}_n$ is a basis for the space of $m \times n$ outer products.

The way this is normally done is to take the free vector space generated by the pairs $(\boldsymbol{e}_i, \boldsymbol{e}_j)$ and then take equivalence classes so that the result obeys the rules:

$$\boldsymbol{u} \otimes (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \otimes \boldsymbol{v} + \boldsymbol{u} \otimes \boldsymbol{w}$$
$$(\boldsymbol{u} + \boldsymbol{v}) \otimes \boldsymbol{w} = \boldsymbol{u} \otimes \boldsymbol{w} + \boldsymbol{v} \otimes \boldsymbol{w}$$
$$\alpha(\boldsymbol{v} \otimes \boldsymbol{w}) = (\alpha \boldsymbol{v}) \otimes \boldsymbol{w} = \boldsymbol{v} \otimes (\alpha \boldsymbol{w})$$

with $\boldsymbol{x} \otimes \boldsymbol{y}$ representing an equivalence class. Fortunately, we only need to know that we have a vector space with the extra operation, the *outer* or *tensor product*.[1]

The use of the free vector space has purged such notions as taking the transpose of a tensor product as there is no rule for handling it.

However, it does allow us to define the tensor product of linear maps. Suppose V, W, X, Y are vector spaces and we have linear transformations $T\colon V \to X$ and $S\colon W \to Y$. Then we can define $S \otimes T\colon V \otimes W \to X \otimes Y$ by

$$(S \otimes T)(\boldsymbol{v} \otimes \boldsymbol{w}) = S(\boldsymbol{v}) \otimes T(\boldsymbol{w}).$$

[1] The outer product is a way of representing the tensor product, but the two are isomorphic in this case.

32.4 Pure 2-Tensors

Tensors of the form $\boldsymbol{v} \otimes \boldsymbol{w} \in V \otimes W$ are called *pure tensors* or *simple tensors*. Not all tensors can be written that way. Some must be written as linear combinations of pure tensors.

► **Example 32.4.1: Not All Tensors are Pure.** Let $V = W = \mathbb{R}^2$,

$$\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Suppose there are $\boldsymbol{v}$ and $\boldsymbol{w}$ with

$$\boldsymbol{v} \otimes \boldsymbol{w} = \begin{pmatrix} v_1 w_1 & v_1 w_2 \\ v_2 w_1 & v_2 w_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then $v_1 w_1 = 1$, so neither v_1 nor w_1 is zero, and $v_2 w_2 = 1$, so neihter v_2 nor w_2 is zero. But $w_1 v_2 = 0$ and $w_2 v_1 = 0$, so at least of the v_i or w_j must be zero, which is impossible. We can only conclude that $\boldsymbol{e}_1 \otimes \boldsymbol{e}_1 + \boldsymbol{e}_2 \otimes \boldsymbol{e}_2$ is not a simple tensor. ◄

32.5 Compound Tensors

In fact, there is a rich supply of tensors that are not pure if the dimensions of both vector spaces are greater than one.

Theorem 32.5.1. *Suppose both* $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ *and* $\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$ *are linearly independent sets. Then* $\boldsymbol{v}_1 \otimes \boldsymbol{w}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{w}_2$ *is not a pure tensor.*

Proof. We use the outer product to see this. Suppose there are $\boldsymbol{v}_3$ and $\boldsymbol{w}_3$ with

$$\boldsymbol{v}_1 \otimes \boldsymbol{w}_1 + \boldsymbol{v}_2 \otimes \boldsymbol{w}_2 = \boldsymbol{v}_3 \otimes \boldsymbol{w}_3.$$

Since each of the outer products is $m \times n$, this equation can be written as m equations. Let w_{ij} denote the j^{th} component of $\boldsymbol{w}_i$.

$$\begin{aligned} w_{11}\boldsymbol{v}_1 + w_{21}\boldsymbol{v}_2 &= w_{31}\boldsymbol{v}_3 \\ w_{12}\boldsymbol{v}_1 + w_{22}\boldsymbol{v}_2 &= w_{32}\boldsymbol{v}_3 \\ \vdots \qquad\qquad & \quad \vdots \\ w_{1m}\boldsymbol{v}_1 + w_{2m}\boldsymbol{v}_2 &= w_{3m}\boldsymbol{v}_3 \end{aligned}$$

If any of the $w_{3i} = 0$, the corresponding w_{1i} and w_{2i} must be zero due to linear independence of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$. If all of the $w_{3i} = 0$, $\boldsymbol{w}_1 = \boldsymbol{w}_2 = \boldsymbol{0}$, contradicting the linear independence of the $\boldsymbol{w}_j$. If only one of the $w_{3i} \neq 0$, both $w_{kj} = 0$ for $j \neq i$, so either both $\boldsymbol{w}_i$ are zero, or one is proportional to the other, again violating linear independence. Finally, if at least two $w_{3i} \neq 0$, either the non-zero equations are proportional, violating linear independence, or they are not proportional. Then we can write a non-trivial equation of the form $\alpha\boldsymbol{w}_1 + \beta\boldsymbol{w}_2 = 0$, again contradicting linear independence.

A simpler way to see that there are many tensors that are not pure is to consider the rank of the matrix $\boldsymbol{x} \otimes \boldsymbol{y}$. Since $\boldsymbol{x} \otimes \boldsymbol{y} = \boldsymbol{x}\boldsymbol{y}^T$ and the rank of $\boldsymbol{x}$ is either zero or one, the rank of any pure tensor must be zero or one. In fact, the rank of $\boldsymbol{x} \otimes \boldsymbol{y}$ is only zero when it is the zero tensor (aka the zero matrix). Summing up,

Theorem 32.5.2. *Tensors with rank larger than one are not pure tensors.*

Proof. As shown above, non-zero pure tensors have rank one, so tensors with rank larger than one are not pure tensors. ∎

32.6 Tensor Products

To get a taste of how we can free tensors from the use of coordinates, suppose V is a vector space. We can write V^k as a tensor product:

$$\bigotimes_{i=1}^{k} V = V^{\otimes n} = \overbrace{V \otimes V \otimes \cdots \otimes V}^{k \text{ times}}.$$

The tensor product has a natural vector space structure.

We write the *tensor product* of vectors by

$$\boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \otimes \cdots \otimes \boldsymbol{x}_n$$

where each $\boldsymbol{x}_i \in \mathbb{R}^n$. The tensor product itself is k-linear and elements of the tensor product are said to have *order* k. Now

$$\begin{aligned}
(\alpha \boldsymbol{x}_1 + \boldsymbol{y}_1) \otimes \boldsymbol{x}_2 \otimes \cdots \otimes \boldsymbol{x}_k &= \alpha(\boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \otimes \cdots \otimes \boldsymbol{x}_k) + \boldsymbol{y}_1 \otimes \boldsymbol{x}_2 \otimes \cdots \otimes \boldsymbol{x}_k, \\
\boldsymbol{x}_1 \otimes (\alpha \boldsymbol{x}_2 + \boldsymbol{y}_2) \otimes \cdots \otimes \boldsymbol{x}_k &= \alpha(\boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \otimes \cdots \otimes \boldsymbol{x}_k) + \boldsymbol{x}_1 \otimes \boldsymbol{y}_2 \otimes \cdots \otimes \boldsymbol{x}_k, \\
\vdots \qquad\qquad & \qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
\boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \otimes \cdots \otimes (\alpha \boldsymbol{x}_k + \boldsymbol{y}_k) &= \alpha(\boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \otimes \cdots \otimes \boldsymbol{x}_k) + \boldsymbol{x}_1 \otimes \boldsymbol{x}_2 \otimes \cdots \otimes \boldsymbol{y}_k.
\end{aligned}$$

Finally, if A is linear on $\otimes_{i=1}^{k} \mathbb{R}^n$, then $A(\boldsymbol{x}_1 \otimes \cdots \otimes \boldsymbol{x}_k)$ is a k-linear function of $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)$ because the tensor product is k-linear.

If we then needed to use coordinates, we could write $\boldsymbol{x}_i = \sum_{j=1}^{n} x_{ij} \boldsymbol{e}_j$. Expanding A, we obtain

$$A(\boldsymbol{x}_1 \otimes \cdots \otimes \boldsymbol{x}_k) = \sum_{j_1 \cdots j_k}^{n} x_{1j_1} \cdots x_{kj_k} A(\boldsymbol{e}_{j_1} \otimes \cdots \otimes \boldsymbol{e}_{j_k}) = \sum_{j_1 \cdots j_k}^{n} a_{j_1 \cdots j_k} x_{1j_1} \cdots x_{kj_k}$$

where each $a_{j_1 \cdots j_k} = A(\boldsymbol{e}_{j_1} \otimes \cdots \otimes \boldsymbol{e}_{j_k})$.

32.7 Covariance and Contravariance

We will touch briefly on covariant and contravariant tensors, in case you encounter them in the future. Suppose V is a vector space and V^* its dual. A tensor of order (p, q) is in the tensor product

$$\overbrace{V \otimes \cdots \otimes V}^{p \text{ times}} \otimes \overbrace{V^* \otimes \cdots \otimes V^*}^{q \text{ times}} = V^{\otimes p} \otimes (V^*)^{\otimes q}.$$

It is p times contravariant and q times covariant, with total order $p + q$.

Among the things that can be done with a $(1, 1)$-tensor is to form an *evaluation map*, mapping

$$x \otimes f \to f(x).$$

It just evaluates the linear functional f at the vector x and is a special case of a more general operation called *contraction* which reduces a (p, q)-tensor to a $(p - 1, q - 1)$-tensor.

Of course, tensors in $V^{\otimes p} \otimes (V^*)^{\otimes q}$ are often written as with p superscripts (for contravariant coordinates) and q subscripts (for covariant coordinates) and summation occurs over all repeated indices. Thus $x^i f_j$ denotes $x \otimes f$ while $x^i f_i$ is the evaluation map. A more complicated $(2, 3)$-tensor might have the form $T^{ij}_{k\ell m}$ and a contraction on it could be written $T^{ij}_{j\ell m}$, meaning $\sum_j T^{ij}_{j\ell m}$.

November 15, 2020