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* $\mathbf{x}_1, \ldots, \mathbf{x}_k$.

For a single $\mathbf{x}_i$, we get a line. For two, if they are co-linear, we still get a line. If there are not, we get a plane. If we're working in $\mathbb{R}^2$, that's all there is. If we're in $\mathbb{R}^n$ for a larger n, there are more possibilities.

Since we are doing economics, there may be an awful lot of possibilities. When dealing with consumers, each good or service needs its own index, leading to a very, very large number of goods, n. A single supermarket has many thousands of goods. Consumers potentially choose from vast numbers of goods.

11.2 The Span of a Set

We define the *span of a set* $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$, $\mathcal{L}\left[\mathbf{x}_1, \ldots, \mathbf{x}_k\right]$, as the set of linear combinations of the points $\mathbf{x}_1, \ldots, \mathbf{x}_k$. In other words,

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

Theorem 11.2.1. *Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are vectors in V and $\mathbf{x}, \mathbf{y} \in \mathcal{L}\left[\mathbf{x}_1, \ldots, \mathbf{x}_k\right]$. Then for $\alpha \in \mathbb{F}$, $\alpha\mathbf{x} + \mathbf{y} \in \mathcal{L}\left[\mathbf{x}_1, \ldots, \mathbf{x}_k\right]$, so the span is a vector subspace of V.*

Proof. In this case there are $s_i, t_i \in \mathbb{F}$ with $\mathbf{x} = \sum_i s_i \mathbf{x}_i$ and $\mathbf{y} = \sum_i t_i \mathbf{x}_i$. Then $\alpha\mathbf{x} + \mathbf{y} = \sum_i (\alpha s_i + t_i)\mathbf{x}_i$ is also in $\mathcal{L}\left[\mathbf{x}_1, \ldots, \mathbf{x}_k\right]$. ∎

11.3 The Span of a Matrix

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 $\mathbf{X}$, so

$$\mathbf{X} = \big(\mathbf{x}_1 \quad | \quad \mathbf{x}_2 \quad | \quad \cdots \quad | \quad \mathbf{x}_k \big).$$

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

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

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

11.4 Spanning Examples

▶ Example 11.4.1: 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 11.4.2: 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 $\mathbb{R}^4$. ◀

11.5 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 whose columns define a coordinate system and suppose $\mathbf{x} = \mathbf{X}\mathbf{t}$ and $\mathbf{x} = \mathbf{X}\mathbf{t}'$, so $\mathbf{t}$ and $\mathbf{t}'$ are both coordinates for $\mathbf{x}$. When can we conclude that $\mathbf{t} = \mathbf{t}'$?

Alternately, when can we conclude that a vector has only one set of coordinates?

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. By Corollary 7.29.1, this will have multiple solutions if and only if $k > \operatorname{rank} \mathbf{X}$, which is equivalent to saying there are free variables.

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 $\mathbf{t}$.

11.6 Theorem on Linear Dependence

Theorem 11.6.1. *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

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

∎

11.7 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

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

11.8 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 $\mathbf{x}_j \in \mathcal{X}$.

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.

11.9 Orthogonal Vectors are Linearly Independent

Orthogonal sets of vectors are automatically independent.

Theorem 11.9.1. *Let $\mathcal{B} = \{\mathbf{b}_i\}_{i=1}^k$ be a set of orthogonal vectors in an inner product space V. Then $\mathcal{B}$ is a linearly independent set.*

Proof. Suppose there are real numbers t_i with

$$\boldsymbol{z} = \sum_{i=1}^{k} t_i \mathbf{b}_i = 0$$

Now consider

$$\begin{aligned} 0 &= \boldsymbol{z} \cdot \mathbf{b}_j \\ &= \left(\sum_{i=1}^{k} t_i \mathbf{b}_i \right) \cdot \mathbf{b}_j \\ &= \left(\sum_{i=1}^{k} t_i \mathbf{b}_i \cdot \mathbf{b}_j \right) \\ &= \left(\sum_{i=1}^{k} t_i \delta_{ij} \right) \\ &= t_j \end{aligned}$$

for every $j = 1, \ldots, k$. Since every $t_j = 0$, the vectors must be linearly independent. ∎

11.10 Too Many Vectors Must Be Dependent

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

Theorem 11.10.1. *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 $\mathbf{Xt} = \mathbf{0}$. Since there are more variables than equations, there is at least one free variable. It follows that $\mathbf{Xt} = \mathbf{0}$ has infinitely many solutions, establishing linear dependence. Alternatively, we could quote Corollary 7.26.1. ∎

► Example 11.10.2: More than n Vectors are Linearly Dependent in $\mathbb{R}^n$. 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: $\mathbf{x}_1 = (1/2)(\mathbf{x}_2 + \mathbf{x}_3 + \mathbf{x}_4)$. ◄

11.11 Spanning Sets

A second issue concerning coordinate systems is whether a given 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 $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subset V$ *spans* a vector space V if $\sum_{j=1}^{k} t_j \mathbf{x}_j = \mathbf{X}\mathbf{t} = \mathbf{y}$ has a solution for every $\mathbf{y} \in V$.

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

11.12 Size of Spanning Sets in $\mathbb{R}^n$

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

Theorem 11.12.1. *If $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is a set of non-zero vectors that spans $\mathbb{R}^n$, then $k \geq n$.*

Proof. If $\mathcal{X}$ spans $\mathbb{R}^n$, construct $\mathbf{X}$ from $\mathcal{X}$ as before. Then $\mathbf{y} = \mathbf{X}\mathbf{t}$ always has a solution. Corollary 7.30.1 tells us that $\operatorname{rank}\mathbf{X} = n$, which implies $k \geq n$. ∎

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

11.13 Examples of Spanning Sets

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. We can use the fact taht

$$\mathbf{X}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = \mathbf{0}$$

to find others.

11.14 Basis of a Vector Space

Larger sets that span will involve some redundancy. Theorem 11.10.1 says they will be linearly dependent, so by Theorem 11.6.1 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.

This brings us to the concept of a basis. A basis is a set of vectors that is big enough to span, but small enough to be linearly independent.

Basis. A set of non-zero vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \subset V$ are a *basis* for a vector space V if

1. $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly independent.
2. $\mathbf{x}_1, \ldots, \mathbf{x}_k$ 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 11.14.1. *Every basis for $\mathbb{R}^n$ has exactly n elements.*

Proof. Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is a basis for $\mathbb{R}^n$. By Theorem 11.12.1, $k \geq n$, and by Theorem 11.10.1, $k \leq n$. Thus $k = n$. ∎

11.15 Bases and Independent Sets II

Theorem 11.15.1. *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

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

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

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

Proof Continues ...

11.16 Bases and Independent Sets I

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

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

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 7.29.1. 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. ∎

11.17 The Dimension of a Vector Space

An important consequence of Theorem 11.15.1 is that every basis of a vector space must be the same size, provided the size if finite. More precisely.

Basis Theorem. *Suppose a vector space V has a basis $\mathcal{B}$ with n elements where n is finite. 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 11.15.1, it cannot have more than n elements.

If there was a basis with fewer than n elements, we could apply Theorem 11.15.1 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 $\mathcal{B}$.

Dimension. 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 $\dim V$.

[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 linear combinations, infinite sums. If we use the metric we previously defined on $\mathbf{s}$, it is possible to show that the partial sums $\sum_{i=1}^{n} x_i \mathbf{e}_i$ converge to $\mathbf{x} \in \mathbf{s}$ for every $\mathbf{x}$. This is an example of a *Schauder basis*.

11.18 Is the Dimension of a Vector Space Always Finite?

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 11.18.1: Attempted Basis for the Sequence Space. The sequence space $\mathbf{s}$ contains infinite linearly independent sets. For $\mathbf{s}$, define the vectors $\boldsymbol{e}_j$, $j = 1, 2, 3, \dots$ by $(\boldsymbol{e}_j)_i = \delta_{ij}$. Then $\boldsymbol{e}_1 = (1, 0, 0, \dots)$, $\boldsymbol{e}_2 = (0, 1, 0, 0, \dots)$, etc. We now have an infinite set $\mathcal{E} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \dots\}$ that seems 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 $(1, 1, 1, \dots)$ cannot be written as a **finite** sum of the $\boldsymbol{e}_j$.[2] ◄

[2] If we allow infinite sums, the standard basis vectors will do nicely, providing a (Schauder) basis. However, we need to learn more about limits, particularly limits in $\mathbf{s}$, before attempting this. Hamel bases do exist, but methods of showing that are non-constructive. In other words, don't ask what they look like.

11.19 Testing for a Basis 9/15/22

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 11.19.1. *Let $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a collection of vectors in $\mathbb{R}^n$. Form the $n \times n$ matrix $\mathbf{B}$ whose columns 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 rank $\mathbf{B}$ = #cols = n by Corollary 7.30.1. 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 7.31.2, 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 rank $\mathbf{B} = n =$ #cols = #rows, which is equivalent to $\mathbf{B}$ being non-singular (Corollary 7.32.1). Finally, $\det \mathbf{B}$ is non-zero if and only $\mathbf{B}$ is non-singular, completing the proof. ∎

11.20 Finding an Orthonormal Basis

If we have a basis $\mathcal{B}$ for an inner product space, we can use it to construct an orthonormal basis using the *Gram-Schmidt* method.[3]

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for an inner product space V. The Gram-Schmidt method first constructs an orthogonal basis from $\mathcal{B}$, and then normalizes it to obtain an orthonormal basis. Define

$$\begin{aligned} \boldsymbol{w}_1 &= \mathbf{b}_1 \\ \boldsymbol{w}_2 &= \mathbf{b}_2 - \frac{\boldsymbol{w}_1 \cdot \mathbf{b}_2}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1} \boldsymbol{w}_1, \\ \boldsymbol{w}_3 &= \mathbf{b}_3 - \frac{\boldsymbol{w}_1 \cdot \mathbf{b}_3}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1} \boldsymbol{w}_1 - \frac{\boldsymbol{w}_2 \cdot \mathbf{b}_3}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2} \boldsymbol{w}_2, \text{etc.} \\ &\cdots \\ \boldsymbol{w}_n &= \mathbf{b}_n - \sum_{i=2}^{n} \left(\frac{\boldsymbol{w}_{i-1} \cdot \mathbf{b}_n}{\boldsymbol{w}_{i-1} \cdot \boldsymbol{w}_{i-1}} \right) \boldsymbol{w}_{i-1} \end{aligned} \tag{11.20.1}$$

We will show that the set $\mathcal{W} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ is an orthogonal basis.

[3] This can be found on page 624 of Simon and Blume.

11.21 The Gram-Schmidt Vectors are Orthogonal I

Theorem 11.21.1. *If $\mathcal{B}$ is a basis for the inner product space V, then $\mathcal{W} = \{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$ as defined by equation (11.20.1) is also a basis for V.*

Proof. Each of the $\boldsymbol{w}_i$ is defined in terms of the $\mathbf{b}_j$ for $j = 1, \dots, i$, and thus in their span. Alternatively, we can use equation (11.20.1) to write the $\mathbf{b}_i$ in terms of the $\boldsymbol{w}_j$ for $j = 1, \dots, i$. This shows that $\mathcal{B} \subset \mathcal{L}[\mathcal{W}]$. It follows that $V = \mathcal{L}[\mathcal{B}] \subset \mathcal{L}[\mathcal{W}] = V$.

By Theorem 11.9.1, orthogonal vectors are linearly independent. Since they span V, that will imply they are a basis for V. All that remains is to show that $\mathcal{W}$ is an orthogonal set of non-zero vectors.

Note that if any $\boldsymbol{w}_i$ were zero, it would contradict the linear independence of $\mathcal{B}$.

We show that the $\boldsymbol{w}_i$ are orthogonal vectors by inductively showing all of the $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_I\}$ are orthogonal for $I = 1, \dots, n$. In this $I = 1$ case this reduces to $\{\boldsymbol{w}_1\} = \{\mathbf{b}_1\}$, which is trivially an orthogonal set of vectors.

Proof continues ...

11.22 The Gram-Schmidt Vectors are Orthogonal II

Rest of Proof. For the induction step, suppose the set $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_I\}$ is orthogonal for some $I < n$. We must show adding $\boldsymbol{w}_{I+1}$ to the set maintains orthogonality. That means we need to show $\boldsymbol{w}_{I+1} \cdot \boldsymbol{w}_j = 0$ for all $j = 1, \ldots, I$.

Now

$$\begin{aligned}
\boldsymbol{w}_{I+1} \cdot \boldsymbol{w}_j &= \mathbf{b}_{I+1} \cdot \boldsymbol{w}_j - \sum_{i=2}^{I+1} \left(\frac{\boldsymbol{w}_{i-1} \cdot \mathbf{b}_{I+1}}{\boldsymbol{w}_{i-1} \cdot \boldsymbol{w}_{i-1}} \right) (\boldsymbol{w}_{i-1} \cdot \boldsymbol{w}_j) \\
&= \mathbf{b}_{I+1} \cdot \boldsymbol{w}_j - \sum_{i=2}^{I+1} \left(\frac{\boldsymbol{w}_{i-1} \cdot \mathbf{b}_{I+1}}{\boldsymbol{w}_{i-1} \cdot \boldsymbol{w}_{i-1}} \right) \delta_{i-1,j} (\boldsymbol{w}_j \cdot \boldsymbol{w}_j) \\
&= \mathbf{b}_{I+1} \cdot \boldsymbol{w}_j - \boldsymbol{w}_j \cdot \mathbf{b}_{I+1} \\
&= 0.
\end{aligned}$$

Because of the Kronecker delta, only the $i-1 = j$ term remains of the sum in the third line. We have proved the induction step that $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{I+1}\}$ is an orthogonal set of vectors. It follows that $\mathcal{W}$ is an orthogonal set of vectors, and hence a basis. ∎

We can form an orthonormal basis from $\mathcal{W}$ by defining

$$\boldsymbol{v}_i = \frac{\boldsymbol{w}_i}{\|\boldsymbol{w}_i\|}.$$

Then $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is an orthonormal basis derived from $\mathcal{B}$.

11.23 Subspaces and Direct Sums

We can define linear independence of subspaces in a manner similar to independence of vectors.

Independent Subspaces. Let $W_1, \ldots, W_k$ be subspaces of a vector space V. The subspaces are linearly independent if $\boldsymbol{w}_i \in W_i$ for all $i = 1, \ldots, k$ and $\boldsymbol{w}_1 + \boldsymbol{w}_2 + \cdots + \boldsymbol{w}_k = 0$ implies $\boldsymbol{w}_i = 0$ for all $i = 1, \ldots, k$.

Now suppose $W_1, \ldots, W_k$ are independent subspaces of V. The *direct sum of* $W_1, \ldots, W_k$ is $\mathcal{L}\big[W_1, \ldots, W_k\big]$. We write the direct sum as

$$W_1 \oplus \cdots \oplus W_k.$$

By using a direct sum, we indicate there is a unique way to write any $\boldsymbol{x} \in V$ as $\boldsymbol{x} = \boldsymbol{w}_1 + \cdots + \boldsymbol{w}_k$. This type of direct sum, where all of the W_k are contained in the vector space V, is known as an *internal direct sum*. It's easily verified that

$$\dim W_1 \oplus + \cdots + \oplus W_k = \dim W_1 + \cdots + \dim W_k.$$

► **Example 11.23.1: Direct Sum.** In $\mathbb{R}^3$, consider the subspace W_1 spanned by $\boldsymbol{w}_1 = (1,1,1)^\mathsf{T}$ and the subspace W_2 spanned by $\boldsymbol{w}_2 = (1,2,1)^\mathsf{T}$ and $\boldsymbol{w}_3 = (2,1,1)^\mathsf{T}$. Because $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\}$ is a basis for $\mathbb{R}^3$, these are independent subspaces. We can now write $\mathbb{R}^3 = W_1 \oplus W_2$. ◄

11.24 External Direct Sums

Another type of direct sum is the external direct sum. Let $W_1, \dots, W_k$ be vector spaces, and consider

$$V = \{\boldsymbol{v} = (\boldsymbol{w}_1, 0, \dots, 0) + \cdots (0, \dots, 0, \boldsymbol{w}_k) : \text{each } \boldsymbol{w}_i \in W_i\}.$$

Defining vector sums and scalar products in the obvious way makes V a vector space, the *external direct sum of* $W_1, \dots, W_k$,

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_k.$$

We can then write direct sums such as

$$\mathbb{R}^n = \overbrace{\mathbb{R} \oplus \mathbb{R} \oplus \cdots \oplus \mathbb{R}}^{n \text{ times}}$$

or like

$$\mathbb{R}^6 = \mathbb{R}^3 \oplus \mathbb{R}^2 \oplus \mathbb{R}.$$

The only real difference between internal and external direct sums is the starting point. Suppose V is the external direct sum defined in terms of spaces W_k. Setting

$$\begin{aligned}
\hat{W}_1 &= \left\{(\boldsymbol{w}_1, 0, \dots, 0) : \boldsymbol{w}_1 \in W_1\right\}, \\
\hat{W}_2 &= \left\{(0, \boldsymbol{w}_2, 0, \dots, 0) : \boldsymbol{w}_2 \in W_2\right\}, \\
&\ \ \vdots \\
\hat{W}_k &= \left\{(0, \dots, 0, \boldsymbol{w}_k) : \boldsymbol{w}_k \in W_k\right\}
\end{aligned}$$

allows to write V as the internal direct sum $\hat{W}_1 \oplus \cdots \oplus \hat{W}_k$.

11.25 Direct Sums of Linear Transformations

We can also define direct sums of linear transformations. Suppose

$$V = W_1 \oplus \cdots \oplus W_k$$

and that $T_i \colon W_i \to W_i$ are linear transformations for $i = 1, \ldots, k$. We define their direct sum, $T = T_1 \oplus \cdots \oplus T_k$ by

$$T(\boldsymbol{w}_1 \oplus \cdots \oplus \boldsymbol{w}_k) = T_1(\boldsymbol{w}_1) \oplus \cdots \oplus T_k(\boldsymbol{w}_k)$$

for every $\boldsymbol{w}_1 \oplus \cdots \oplus \boldsymbol{w}_k \in W_1 \oplus \cdots \oplus W_k$.

It is easily verified that T is a linear transformation from V to V.

Now suppose each T_i is represented by a matrix $\mathbf{A}_i$ on W_i in some basis $\mathcal{W}_i$. Now $\mathcal{L}[\mathcal{W}_1, \ldots, \mathcal{W}_k]$ is a basis for V and in that basis, the matrix $\mathbf{A}$ for T takes the block diagonal form

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & 0 & 0 & \cdots & 0 \\ 0 & \mathbf{A}_2 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{A}_3 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \cdots & \mathbf{A}_k \end{pmatrix}$$

where row block i has n_i rows and column block i has n_i columns, making the matrix $n \times n$ where $n = n_1 + \cdots + n_k$.

Conversely, any block diagonal matrix can be regarded as a direct sum of matrices.

27. Subspaces Attached to a Matrix

This chapter draws on Chapter 27 of Simon and Blume.

27.1 The Kernel of a Matrix

We saw in Theorem 7.3.1 that the solution set of a linear system is not affected by row operations. When applied to homogeneous systems, that means the kernel is not affected by row operations either. In particular, its dimension is not affected.

We also saw in section 10.16 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?

27.2 Kernel Example

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). Each of the free variable columns have been marked in red.

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

$$\begin{aligned} 0 &= x_1 + x_2 + x_4 + x_5 + x_6 \\ 0 &= x_3 + x_4 + x_6. \end{aligned} \tag{27.2.1}$$

We can find a basis for the kernel by successively setting each of the free variables but one to zero. The non-zero variable can be anything we want. We choose $+1$. Then we use equations (27.2.1) to solve for the basic variables x_1 and x_3. Here are the results.

$$\mathbf{b}_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{b}_4 = \begin{pmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \mathbf{b}_5 = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{b}_6 = \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

We have labelled the vectors by which $x_i = +1$.

Each vector $\mathbf{b}_i$ is in the kernel, and they are all linearly independent. This is clear because each row $i = 2, 4, 5, 6$ is non-zero only for 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

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

27.3 The Dimension of the Kernel

It's nice to have a illustrative example, but it is no substitute for a proof. Fortunately, all the proof needs is to add some words of explanation.

Theorem 27.3.1. *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.

Each $\mathbf{b}_i$ solves the homogeneous equations, so $\mathbf{b}_i \in \ker \mathbf{A}$. Suppose $\mathbf{x} \in \ker \mathbf{A}$. Then $\mathbf{x}$ solves the reduced row-echelon 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 $\ker \mathbf{A}$.

As $\mathbf{b}_i$ is the only one of the $\mathbf{b}_j$ where row i is non-zero, the $\mathbf{b}_j$ are linearly independent. It follows that they form a basis for $\ker \mathbf{A}$, so

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

∎

27.4 Fundamental Theorem of Linear Algebra

Equation (7.19.2) tells us that the number of free variables and the number of basic variables add to the number of variables (n). Combining that with Theorem 27.3.1, 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.

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. In a sense, all finite dimensional real vector spaces can be thought of as $\mathbb{R}^n$.

31.1 Isomorphic Vector Spaces

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

Isomorphic Vector Spaces. Two vector spaces V and W over the same field $\mathbb{F}$ are *isomorphic* if there is a linear transformation $\mathsf{T} : V \to W$ that is one-to-one and onto. Such a mapping is called a *linear isomorphism*.

The fact that the transformation is linear tells us that it preserves the vector space operations. The fact that it is bijective means it has an inverse.

If two isomorphic vector spaces had different underlying fields, we would have difficultly making sense of $\mathsf{T}(\alpha\mathbf{x}) = \alpha\mathsf{T}(\mathbf{x})$ and $\mathsf{T}^{-1}(\alpha\mathbf{y}) = \alpha\mathsf{T}^{-1}(\mathbf{y})$. So we require they be based on the same field, the same set of scalars.

Isomorphism says nothing about how non-vector space structures relate in the two vector spaces. Isomorphism as vector spaces says nothing about whether one space has a norm or metric and the other does not. Such questions require other concepts of isomorphism that preserve those structures, just as linear mappings preserve everything linear.

31.2 The Inverse of a Linear Isomorphism

The inverse of a linear isomorphism is also a linear isomorphism.

Theorem 31.2.1. *Suppose* $T\colon V \to W$ *is a linear isomorphism between vector spaces* V *and* W*. Then the inverse transformation* T^{-1} *exists and is also a linear isomorphism.*

Proof. For any $y \in W$, 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.

Now if $x, y \in W$, then for any scalar α,

$$T\left(\alpha T^{-1}(x) + T^{-1}(y)\right) = \alpha x + y,$$

showing that $T^{-1}(\alpha x + y) = \alpha T^{-1}(x) + T^{-1}(y)$. In other words, T^{-1} is linear.

For every $x \in V$, $T^{-1}\big(T(x)\big) = x$, showing that T^{-1} is onto. And if $v, w \in W$ and $T^{-1}(v) = T^{-1}(w)$, then apply T to find $v = w$, showing that T^{-1} is one-to-one. Thus T^{-1} is a linear isomorphism from W to V. ∎

One consequence is that if V and W are isomorphic via T, they are also isomorphic via T^{-1} (in the opposite direction). This means that anything in V is, from a vector space point of view, faithfully reproduced in W, and vice-versa.

31.3 Matrix Isomorphisms

One important result is that a linear isomorphism T from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be written $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$, that $m = n$, and that $\mathbf{A}$ is invertible.

Theorem 31.3.1. *Let* $T\colon \mathbb{R}^n \to \mathbb{R}^m$ *be a linear isomorphism. Then* $m = n$ *and there is an invertible matrix* $\mathbf{A}$ *with* $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$ *for every* $\mathbf{x} \in V$*.*

Proof. Let T be as above. Because the mapping is linear, we can use Theorem 10.6.1 to find an $m \times n$ matrix $\mathbf{A}$ with $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$.

Since T is a linear isomorphism, it must be both one-to-one and onto. The first requires that $\operatorname{rank} \mathbf{A} = n$ by Corollary 7.29.1. By Corollary 7.31.2, the fact that T is onto tells us that $\operatorname{rank} \mathbf{A} = m$.

Combining these shows $m = n = \operatorname{rank} \mathbf{A}$, which means that $\mathbf{A}$ is non-singular, that it must be invertible. ∎

Notice that T^{-1} can be written $T^{-1}(\mathbf{y}) = \mathbf{A}^{-1}\mathbf{y}$ for every $\mathbf{y} \in W$.

31.4 Isomorphic Vector Spaces have the Same Dimension

In fact, any two isomorphic finite dimensional vector spaces, not just $\mathbb{R}^n$, must have the same dimension.

Theorem 31.4.1. *Let* V *and* W *be finite-dimensional vector spaces and* $T\colon V \to W$ *be an isomorphism. Then* $\dim V = \dim W$.

Proof. Let $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ be a linearly independent set in V. I claim that $\{T(\boldsymbol{v}_j)\}_{j=1}^k$ is a linearly independent set in W. **If not**, there are t_j, not all zero, with $\sum_{j=1}^k t_j T(\boldsymbol{v}_j) = \mathbf{0}$. Applying T^{-1}, we find $\sum_{j=1}^k t_j \boldsymbol{v}_j = 0$, **contradicting** the fact that $\mathcal{V}$ is a linearly independent set.

Now let $\mathcal{V}$ be a basis for V. Since $T(\mathcal{V})$ is linearly independent in W, $\dim W \geq \dim V$. Consideration of the isomorphism T^{-1} shows that $\dim V \geq \dim W$, so $\dim V = \dim W$. ∎

31.5 Same Dimension Implies Isomorphic Spaces

It's also easy to show that two vector spaces over a field $\mathbb{F}$ that have the same finite dimension are isomorphic. The key to the proof is to create an isomorphism by mapping basis elements of one to basis elements of the other.

Theorem 31.5.1. *Let V and W be finite-dimensional vector spaces over the same field $\mathbb{F}$. If $\dim V = \dim W$, there is an isomorphism $T\colon V \to W$.*

Proof. Suppose $\dim V = \dim W = n$. Let $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be a basis for V and $\mathcal{W} = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ be a basis for W.

Define $T\colon V \to W$ on the basis elements of V by $T(\boldsymbol{v}_i) = \boldsymbol{w}_i$. Since $\mathcal{V}$ is a basis, we may use linearity to define T on all of V. The resulting transformation T maps V into W. We will show it is bijective.

(1) The linear mapping T maps onto W (surjective). To see that, consider $\boldsymbol{x} \in W$. We can find x_j with $\boldsymbol{x} = \sum_j x_j \boldsymbol{w}_j$. But then $\boldsymbol{x} = T\big(\sum_j x_j \boldsymbol{v}_j\big)$. This shows that T maps onto W.

(2) The mapping T is one-to-one (injective). Suppose $T(\boldsymbol{x}) = T(\boldsymbol{y})$ for some $\boldsymbol{x}, \boldsymbol{y} \in V$. Since $\mathcal{V}$ is a basis for V, we can write $\boldsymbol{x} = \sum_j x_j \boldsymbol{v}_j$ and $\boldsymbol{y} = \sum_j y_j \boldsymbol{v}_j$. Applying T, we find $\sum_j x_j \boldsymbol{w}_j = \sum_j y_j \boldsymbol{w}_j$. As $\mathcal{W}$ is linearly independent, $x_j = y_j$ for all $j = 1, \ldots, n$, showing that $\boldsymbol{x} = \boldsymbol{y}$. It follows that T is an one-to-one.

By (1) and (2), T is an isomorphism. ∎

One corollary is that any n-dimensional real vector space is isomorphic to $\mathbb{R}^n$.

We can combine the last two theorems as follows:

Theorem 31.5.2. *Let V and W be finite-dimensional vector spaces. Then $\dim V = \dim W$ if and only if there is an isomorphism $T\colon V \to W$.*

31.6 Isomorphism Example

▶ Example 31.6.1: An Isomorphism. Let $W = \{\mathbf{x} \in \mathbb{R}^3 : x_1+x_2+x_3 = 0\}$. This is a two-dimensional subspace of $\mathbb{R}^3$ 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 $\mathbf{b}_i$ spans W, it is onto. That makes it an isomorphism. The inverse map $T^{-1}: W \to \mathbb{R}^2$ projects onto $\mathbb{R}^2$. It is

$$T^{-1}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Notice we tossed y_3. Keep in mind that the components of $\mathbf{y}$ must sum to zero by the definition of W, so on W, $y_3 = -y_1 - y_2$. ◀

31.7 Isometric 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, $\|T(x)\|_2 = \|x\|_1$. Such isomorphisms are also called *linear isometries* or *isometric isomorphisms*.

Norm-preserving mappings are often linear. To state this more precisely, define the *midpoint* of two vectors x and y as $(x + y)/2$.

We state the following theorem of Mazur and Ulam without proof

Mazur-Ulam Theorem. *Let* V *and* W *be normed spaces and* T *a mapping of* V *onto* W *with* $\|Tx\|_2 = \|x\|_1$ *and* $T(0) = 0$. *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.

You'll notice that although T maps midpoints to midpoints, that is not generally true of $\|T\|$. What happens is that

$$\left\|T\left(\frac{1}{2}(x + y)\right)\right\| = \left\|\frac{1}{2}\big(T(x) + T(y)\big)\right\| \le \frac{1}{2}\Big(\|T(x)\| + \|T(y)\|\Big).$$

However, the triangle inequality is often strict, so T usually doesn't map midpoints to midpoints.

31.8 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 $\langle \mathbf{x}, \mathbf{y} \rangle_i, i = 1, 2, \ldots$ to distinguish the inner products. Preserving the inner product means the norm is also preserved, so inner product space isomorphisms are always isometric.

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

Using less precise notation, we may write the inner product condition as

$$T(\mathbf{x}) \cdot T(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}.$$

This notation is a bit dangerous as it leads to easy confusion of the two different inner products that are involved.

31.9 Characterizing Inner Product Isomorphisms

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.

Theorem 31.9.1. *Let* T *be an inner product isomorphism from* $\mathbb{R}^m$ *to* $\mathbb{R}^m$ *and suppose* $\{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ *is an orthonormal basis for* $\mathbb{R}^m$*. Then* $\{T(\mathbf{b}_i)\}$ *is also an orthonormal basis for* $\mathbb{R}^m$*.*

Proof. To see this, we compute

$$T(\mathbf{b}_i)\cdot T(\mathbf{b}_j) = \mathbf{b}_i\cdot\mathbf{b}_j = \delta_{ij}.$$

Since $T(\mathbf{b}_i)\cdot T(\mathbf{b}_j) = \delta_{ij}$ for all $i, j = 1, \ldots, m$, $T(\mathbf{b}_i)$ is an orthonormal set. ∎

In particular, Theorem 31.9.1 applies when we use the standard basis vectors in $\mathbb{R}^m$. They are mapped to an orthonormal set. This means that T is either a rotation, or a rotation together with an reflection.

The following lemma is useful in inner product spaces.

Lemma 31.9.2. *Let* V *be an inner product space and suppose* $\mathbf{x}$ *and* $\mathbf{x}'$ *obey* $\mathbf{x}\cdot\mathbf{y} = \mathbf{x}'\cdot\mathbf{y}$ *for every* $\mathbf{y} \in V$*. Then* $\mathbf{x} = \mathbf{x}'$*.*

Proof. Now $(\mathbf{x} - \mathbf{x}')\cdot\mathbf{y} = 0$ for every $\mathbf{y} \in V$. Set $\mathbf{y} = \mathbf{x} - \mathbf{x}'$ to obtain $\|\mathbf{x} - \mathbf{x}'\|^2 = 0$. It follows that $\mathbf{x} = \mathbf{x}'$. ∎

31.10 Automorphisms

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

Automorphisms are not quite as specialized as you might think. When we have a linear isomorphism, the two vector spaces involved have the same dimension, and the isometry makes them act like they are identical. Actually requiring they be the same space is then a small additional step.

31.11 Characterizing Automorphisms on ℓ_2^n

Suppose a matrix $\mathbf{A}$ defines an automorphism on ℓ_2^n, then $\mathbf{A}^{-1} = \mathbf{A}^\mathsf{T}$.

Theorem 31.11.1. *Let* T *be an automorphism on* ℓ_2^n. *Let* $\mathbf{A}$ *be any matrix representation of* T. *Then* $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{I}$.

Proof. Under these conditions $(\mathbf{A}x)\cdot(\mathbf{A}y) = x\cdot y$ for all $x, y \in \mathbb{R}^n$. Now

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

Since this holds for every $y \in \mathbb{R}^n$, Lemma 31.9.2 implies that $\mathbf{A}^\mathsf{T}\mathbf{A}x = x$. This holds for every x, which implies $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{I}$. ∎

The result can also be written as $\mathbf{A}^{-1} = \mathbf{A}^\mathsf{T}$.

When the vector space is $\mathbb{C}^n$, we use $x\cdot y = x^* y = \sum_{j=1}^n \bar{x}_j y_j$. 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 $\det \mathrm{U} = \pm 1$ to sort out which is which.

31.12 Automorphisms on ℓ_2^n: $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{I}$

Automorphsims on $\mathbb{R}^n$ can be represented by matrices with $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{I}$. Let $\boldsymbol{a}_i$, $i = 1, \dots, n$ denote the columns of $\mathbf{A}$. Then $\boldsymbol{a}_i^\mathsf{T}$ is the i^{th} row of $\mathbf{A}^\mathsf{T}$, and $\boldsymbol{a}_i^\mathsf{T}\boldsymbol{a}_j = \boldsymbol{a}_i \cdot \boldsymbol{a}_j = \delta_{ij}$, showing that the columns form an orthonormal basis for $\mathbb{R}^n$. Since $\mathbf{A}$ maps $\boldsymbol{e}_i \mapsto \boldsymbol{a}_i$, we have either a pure rotation of the coordinates, or a combination of a rotation and reflection.

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. I'll have more to say about this.

▶ Example 31.12.1: A Reflection in $\mathbb{R}^2$. Reflections also obey $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{I}$, but the determinant is -1. For example, reflecting in the horizontal axis maps $(\boldsymbol{e}_1, \boldsymbol{e}_2) \mapsto (\boldsymbol{e}_1, -\boldsymbol{e}_2)$. The new basis matrix is

$$\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

which has determinant -1.

Notice that multiplying again by $\mathbf{B}$ returns us to our original coordinates. This transformation can't possibly be a rotation because the diagonal elements are not the same. We will see in section 31.14 that rotating $\mathbb{R}^2$ by an angle θ makes both diagonal elements of the basis matrix the same: $\cos\theta$. Also, rotations always have determinant $+1$, this has determinant -1. ◀

31.13 Rotations and Reflections in Two Dimensions

Let's start with $\mathbb{R}^2$ and let T be an automorphism of $\mathbb{R}^2$. Equivalently, T is generated by a matrix $\mathbf{A}$ obeying $\mathbf{A}^{\mathsf{T}}\mathbf{A} = \mathbf{I}$.

The fact that $\boldsymbol{a}_1 = T(\boldsymbol{e}_1)$ is perpendicular to $\boldsymbol{a}_2 = T(\boldsymbol{e}_2)$ means that $\boldsymbol{a}_2$ lies on the line perpendicular to $\boldsymbol{a}_1$. Since $\boldsymbol{a}_2$ is a unit vector, there are only two possible places to put it. One is a rotation. The other involves a reflection together with a rotation. This is illustrated in Figure 31.13.1.

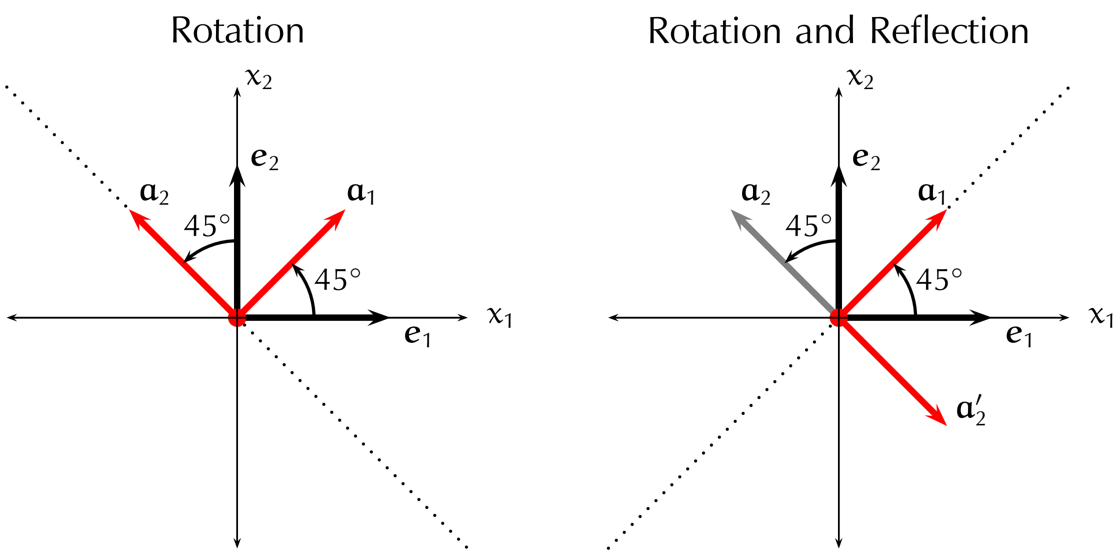


Figure 31.13.1: In the left panel the standard coordinate vectors are rotated counter-clockwise by 45°. The dashed line is perpendicular to $\boldsymbol{a}_1$ and shows the line that $\boldsymbol{a}_2$ must lie in. Then $\boldsymbol{a}_2$ is the only unit vector on that line that is consistent with rotation. Using $-\boldsymbol{a}_2$ would require a combination of both rotation and reflection

In the right panel, we make the opposite choice for $\boldsymbol{a}_2$, pointing downward along the 45° line rather than upward. This amounts to first making the rotation shown in the left panel. We then reflect the result about the dotted line through the origin defined by $\boldsymbol{a}_1$. This leaves $\boldsymbol{a}_1$ unchanged, but flips $\boldsymbol{a}_2$ (in gray) to $\boldsymbol{a}_2'$ (in red), yielding the new coordinates.

31.14 General Rotations in $\mathbb{R}^2$ I

Consider a clockwise rotation of the canonical basis vectors in $\mathbb{R}^2$ by an angle $\theta < 0$ (we use negative angles for clockwise rotations, positive for counter-clockwise). This rotates black vectors $\mathbf{e}_i$ into the red ones $\mathbf{b}_i$.

Effect of Rotation by θ

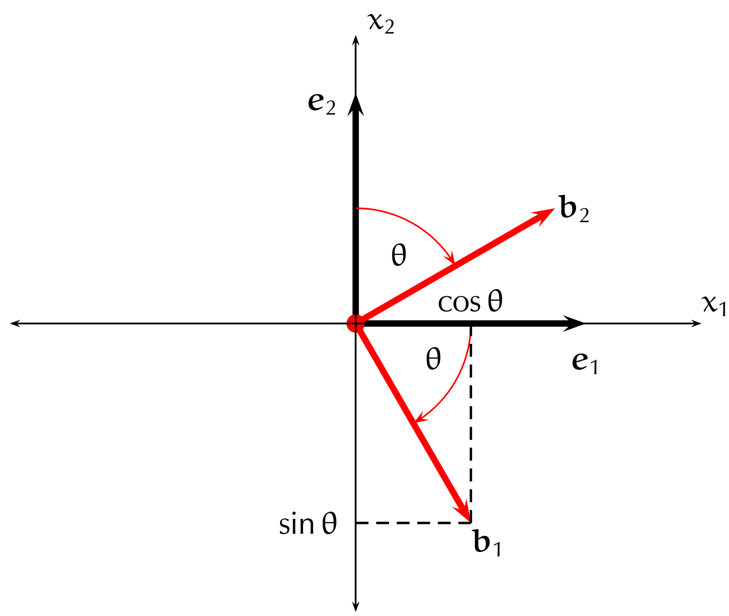


Figure 31.14.1: The standard basis vectors (in black) are rotated clockwise by an angle $\theta < 0$. This yields the new basis vectors (in red). A little trigonometry gives us the coordinates in the old system. Since $\mathbf{b}_i$ is a unit vector, we can read the coordinates in terms of the sine and cosine of θ. Specifically, $\mathbf{b}_1 = (\cos\theta, \sin\theta)$ and $\mathbf{b}_2 = (-\sin\theta, \cos\theta)$.

31.15 General Rotations in $\mathbb{R}^2$ II

As shown in the diagram, we have the following mapping:

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

Note that the new vectors are still an orthonormal basis, as both vectors are rotated by the same angle. The new basis matrix is

$$\mathbf{R}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

which has determinant $\cos^2\theta + \sin^2\theta = 1$, regardless of the angle θ. Rotations always have a determinant of $+1$.

31.16 Example: A Rotation and its Inverse

▶ Example 31.16.1: 45° Rotation: Done and Undone. The matrix

$$\mathbf{B} = \mathbf{R}(45°) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

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

$$\mathbf{B}^{-1} = \mathbf{R}(-45°) = \mathbf{B}^\mathsf{T} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

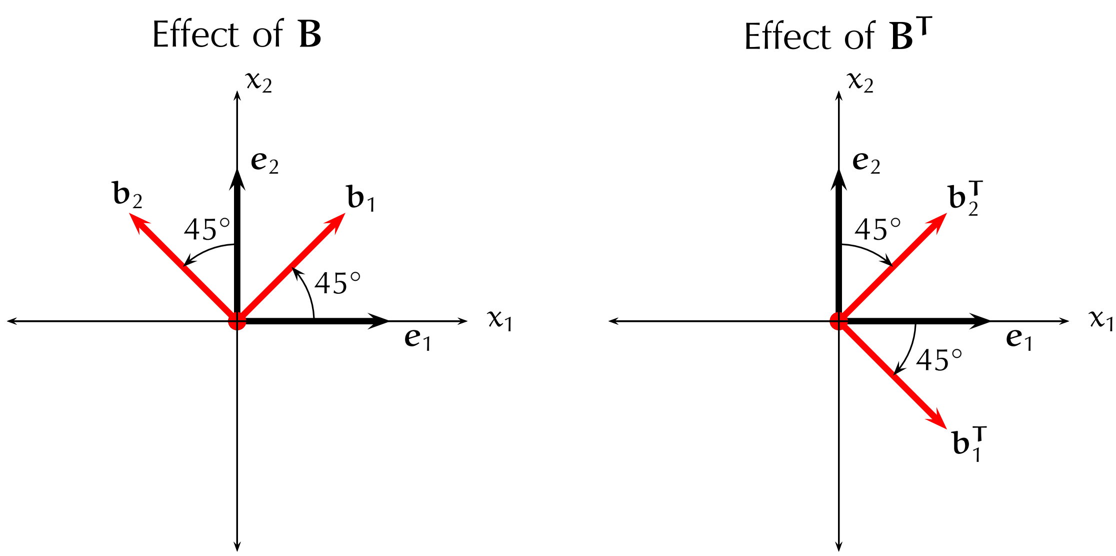


Figure 31.16.2: Here the standard coordinate axes are rotated counterclockwise by 45° in the left panel. Here $\mathbf{b}_1$ and $\mathbf{b}_2$ are the columns of $\mathbf{B}$. In the right panel, we have the inverse transformation where a 45° clockwise rotation gives us the new coordinate axes. Here $\mathbf{b}_1^\mathsf{T}$ and $\mathbf{b}_2^\mathsf{T}$ are the columns of the matrix $\mathbf{B}^\mathsf{T}$.

◀

31.17 Rotations and Reflections in Three Dimensions

In $\mathbb{R}^3$, $T(\boldsymbol{e}_1)$ determines the perpendicular plane that the other $T(\boldsymbol{e}_i)$ lie in. Once we know where $T(\boldsymbol{e}_2)$ goes, there are only two choices for $T(\boldsymbol{e}_3)$. One involves a rotation of $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$, the other combines a reflection and rotation.

That the latter case is possible can be seen by letting your thumb, forefinger, and middle finger represent the basis vectors. Your right hand cannot be rotated to be a left hand, and vice-versa. However, a mirror can turn a right hand into a left hand, which is why a reflection might be needed.

In any $\mathbb{R}^n$, there are two orientations of orthogonal axes. One that is a rotation of the standard axes, the other always involves a reflection.

31.18 The Orthogonal Group $\mathbf{O}(n)$

The rotations and reflections in $\mathbb{R}^n$ form a group, the *orthogonal group in* $\mathbb{R}^n$, $\mathbf{O}(n)$. The pure rotations of are distinguished by having positive determinants. They also form a group, the *special orthogonal group in* $\mathbb{R}^n$, $\mathbf{SO}(n)$. The notations $\mathbf{O}(n, \mathbb{R})$ and $\mathbf{SO}(n, \mathbb{R})$ are used when its necessary to distinguish the real and complex cases.

Group. A *group* is a set G together with a binary operation $(a, b) \mapsto ab$ that is associative and has an identity element e. Moreover, each element $g \in G$ has an inverse g^{-1} so that $g^{-1}g = gg^{-1} = e$.

The group operation need not commute. If it does commute, we call the group an *abelian group*.[1] Groups are one of the fundamental concepts in modern algebra.

Groups sometimes involve the manipulations of physical objects. For example, the set of manipulations of Rubik's Cube also forms a group.

The integers together with addition form an abelian group. The positive real numbers under multiplication also form an abelian group. The invertible $n \times n$ matrices form a non-abelian group under matrix multiplication.

As for $\mathbf{O}(n)$, it's the set of $n \times n$ real matrices obeying $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{I}$. The group operation for $\mathbf{O}(n)$ is matrix multiplication, which is associative. If $\mathbf{A}, \mathbf{B} \in \mathbf{O}(n)$, the product $\mathbf{AB}$ is in $\mathbf{O}(n)$ because

$$(\mathbf{AB})^\mathsf{T}(\mathbf{AB}) = \mathbf{B}^\mathsf{T}(\mathbf{A}^\mathsf{T}\mathbf{A})\mathbf{B} = \mathbf{B}^\mathsf{T}\mathbf{B} = \mathbf{I},$$

and because $\mathbf{I} \in \mathbf{O}(n)$ is the identity element.

[1] Named for Neils Henrik Abel (1802–1829), as is the Abel Prize, the mathematicians equivalent of the Nobel.

31.19 $SO(2)$ is Abelian 9/20/22

First, we show $\mathbf{SO}(2)$ is abelian. Recall that the elements of $\mathbf{SO}(2)$ all have the form

$$\mathbf{R}(\theta_i) = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}.$$

Then

$$\begin{aligned}
&\mathbf{R}(\theta_1)\mathbf{R}(\theta_2) \\
&\qquad = \begin{pmatrix} \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 & -\cos\theta_1\sin\theta_2 - \cos\theta_2\sin\theta_1 \\ \sin\theta_1\cos\theta_2 + \sin\theta_2\cos\theta_1 & \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 \end{pmatrix} \\
&\qquad = \begin{pmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{pmatrix} \\
&\qquad = \mathbf{R}(\theta_1+\theta_2).
\end{aligned}$$

Then $\mathbf{R}_1\mathbf{R}_2 = \mathbf{R}(\theta_1+\theta_2) = \mathbf{R}_2\mathbf{R}_1$ for any θ_i, showing that all matrices in $\mathbf{SO}(2)$ commute.

31.20 $SO(n)$ is Not Abelian when $n > 2$

For $n > 2$, $\mathbf{SO}(n)$ is non-abelian. It is enough to show it for $n = 3$, since the same rotations are available for $n > 3$ (holding the extra axes fixed).

To see that $\mathbf{SO}(3, \mathbb{R})$ is not Abelian, consider the rotations

$$\mathbf{R}_1 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{R}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

These do not commute as

$$\mathbf{R}_1\mathbf{R}_2 = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \quad \text{while} \quad \mathbf{R}_2\mathbf{R}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

31.21 Bases and Coordinates

We've used two sets of bases for several pages. It's time to approach the different coordinate systems more systematically.

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

$$\mathbf{B} = \left(\mathbf{b}_1 \quad | \quad \mathbf{b}_2 \quad | \quad \cdots \quad | \quad \mathbf{b}_n \right).$$

By Theorem 11.19.1 (determinant test), $\mathbf{B}$ is invertible. Given a vector $\mathbf{x} \in \mathbb{R}^n$, we find its vector of coordinates $\mathbf{t}_\mathbf{B}$ in the $\mathcal{B}$ basis by solving the equation

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

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

This all applies to the standard basis $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. 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 $\mathbf{t}_\mathbf{E} = \mathbf{E}^{-1}\mathbf{x} = \mathbf{I}\mathbf{x} = \mathbf{x}$, meaning that the coordinates are what we think they are. (That's a relief!)

31.22 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.

31.23 Changing Coordinate Systems

Consider two different bases and 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:

$$\mathbf{t}_\mathbf{B} = \left(\mathbf{B}^{-1}\mathbf{B}'\right)\mathbf{t}_{\mathbf{B}'} \quad \text{and} \quad \mathbf{t}_{\mathbf{B}'} = \left((\mathbf{B}')^{-1}\mathbf{B}\right)\mathbf{t}_\mathbf{B} \tag{31.23.1}$$

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 $(\mathbf{B}')^{-1}$.

31.24 Example: Changing Coordinates in $\mathbb{R}^2$

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

$$\mathbf{t}_{\mathbf{B}} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

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

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

31.25 Linear Transformations and Bases

So how do these coordinate changes 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 10.6, 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 $\mathbf{A}_{\mathcal{E}}\mathbf{B}\mathbf{t}_{\mathbf{B}}$.

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 $(\mathbf{B}^{-1}\mathbf{A}_{\mathcal{E}}\mathbf{B})\mathbf{t}_{\mathbf{B}}$ as the $\mathbf{B}$ coordinates of the transformed vector. Thus

$$\boldsymbol{A}_{\mathcal{B}} = \mathbf{B}^{-1}\mathbf{A}_{\mathcal{E}}\mathbf{B} \quad \text{or} \quad \mathbf{B}\boldsymbol{A}_{\mathcal{B}}\mathbf{B}^{-1} = \boldsymbol{A}_{\mathcal{E}}. \tag{31.25.2}$$

is the matrix for T in $\mathcal{B}$ coordinates. The type of transformation used in equation (31.25.2) is sometimes called a *similarity transformation*.

31.26 Coordinate Change with Arbitrary Bases

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 $\left((\mathbf{B}')^{-1}\mathbf{B}\right)$ to convert $\mathbf{B}$ coordinates to $\mathbf{B}'$ coordinates, apply $\mathbf{A}$, then convert back. The result is:

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

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

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

31.27 Example: Linear Transformation Basis Change

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 $T(\mathbf{b}_1) = 2\mathbf{b}_1$ and $T(\mathbf{b}_2) = \mathbf{b}_2$. In Chapter 23, you will learn how we can sometimes find such a basis from the original matrix.

31.28 Example: Transformations with Complex Numbers

Consider the matrix

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

In section 8.38, we saw that $\mathbf{A}$ is a square root of $-\mathbf{I}$, the negative of the identity matrix. It's a bit like an imaginary number.

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.

31.29 The Dual Space

In section 10.8, 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*.[2] 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 $\mathbf{p} \in (\mathbb{R}^n)^*$, 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).

[2] 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.

31.30 Duality and Bases

To see how duality relates to bases, we treat a linear functional f as we can treat any other linear transformation. We represent it using a matrix by expressing $\boldsymbol{x}$ using the standard basis and using linearity of f to write:

$$\begin{aligned} f(\boldsymbol{x}) &= f\left(\sum_{j=1}^{n} x_j \boldsymbol{e}_j\right) \\ &= \sum_{j=1}^{n} x_j f(\boldsymbol{e}_j) \qquad (31.30.3) \\ &= \begin{pmatrix} f(\boldsymbol{e}_1) & \cdots & f(\boldsymbol{e}_n) \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}. \end{aligned}$$

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

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

Reading equation (31.30.3) 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 $\boldsymbol{x} \in V$ with the dual element $\boldsymbol{x}^*$ since $\boldsymbol{y} \mapsto \boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^* \boldsymbol{y}$ 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.

31.31 The Dual of a 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 $\mathbf{b}_i^*(\mathbf{b}_j) = \delta_{ij}$. The elements of the dual basis can be used to read the $\mathcal{B}$ coordinates of any vector. Just compute

$$\mathbf{b}_i^*(\mathbf{x}) = \mathbf{b}_i^* \left(\sum_j x_j \mathbf{b}_j \right) = \sum_j x_j \mathbf{b}_i^*(\mathbf{b}_j) = \sum_j x_j \delta_{ij} = x_i.$$

When the original basis the canonical basis, its dual basis, the *standard* or *canonical dual basis* is

$$\left\{\boldsymbol{e}_1^*, \ldots, \boldsymbol{e}_n^*\right\} = \left\{\boldsymbol{e}_1^{\mathsf{T}}, \ldots, \boldsymbol{e}_n^{\mathsf{T}}\right\}.$$

This follows because

$$\boldsymbol{e}_i^*(\boldsymbol{e}_j) = \boldsymbol{e}_i^{\mathsf{T}} \boldsymbol{e}_j = \delta_{ij}$$

as required.

This allows us to write any $f \in V^*$ as

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

31.32 Dual Bases are Bases

It is easy to verify that the dual basis $\{\mathbf{b}_i^*\}$ is a basis for V^*.

Theorem 31.32.1. *Let* V *be a finite-dimensional vector space and* $\mathcal{B}$ *be a basis for* V*. Then the dual basis of* $\mathcal{B}$ *is a basis for* V^*

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

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

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

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

This shows that $f = \boldsymbol{v}_f = \sum_j f(\mathbf{b}_j)\mathbf{b}_j^*$, 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^*. ∎

31.33 Linear Functionals Separate Points in V

The following lemma is similar to Lemma 31.9.2, but applies to any dual space and doesn't require that V be an inner product space. It tells us that V^* contains a rich collection of dual elements, rich enough to distinguish between the points of V.

Lemma 31.33.1. *Let V be a finite-dimensional vector space. Suppose $f(\mathbf{x}) = f(\mathbf{y})$ for every $f \in V^*$, 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 $\mathbf{b}_i$. ∎

31.34 The Dual Basis Matrix

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.[3]

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

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

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

31.35 Coordinate Change in the Dual Space

Of course, to change coordinates, $\mathbf{t}_\mathbf{B} = \mathbf{B}^{-1}\boldsymbol{x}$ 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 $\boldsymbol{v}_f$ in the standard basis, the coordinates in the basis $\mathcal{B}^*$ are $\mathbf{t}^*_\mathbf{B} = \boldsymbol{v}_f(\hat{\mathbf{B}})^{-1} = \boldsymbol{v}_f\mathbf{B}$. 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

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

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

31.36 Example: Gallons vs. Quarts

▶ Example 31.36.1: Gallons vs. Quarts Again. Going back to section 10.1, 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

$$\mathbf{B}^{-1} = \mathrm{diag}(1, \ldots, 1, 4, 1, \ldots, 1)$$

where 4 is in the k^{th} row. It follows that

$$\mathbf{B} = \mathrm{diag}(1, \ldots, 1, 1/4, 1, \ldots, 1),$$

so the corresponding (dual) price vector must be multiplied by $1/4$. ◀

31.37 Rotations and Reflections

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.

31.38 Rotations and Reflections: An Example

Let's see how this works with an orthonormal basis.

▶ **Example 31.38.1: 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 $(0,1) \mapsto (-1/\sqrt{2}, 1/\sqrt{2})$.

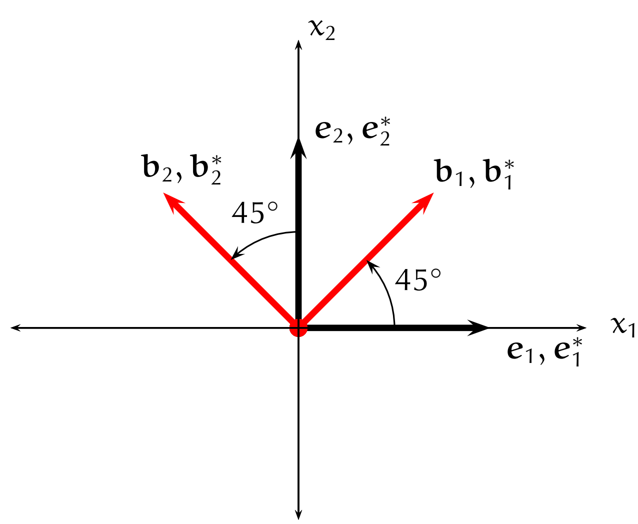


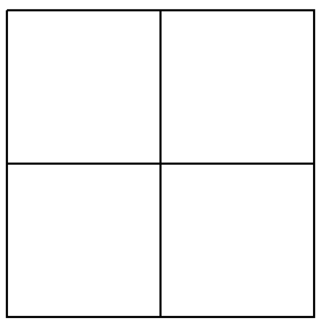
Figure 31.38.2: Here the standard coordinate axes are rotated counterclockwise by 45°. The standard dual basis lines up with the standard basis itself. Because the new basis is an orthonormal basis, the dual basis must rotate to match.

◀

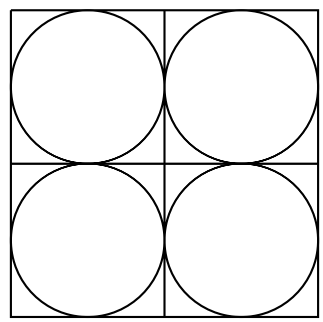
32. $\mathbb{R}^n$ Geometry Puzzle: The Setting

Consider a square with 2-foot sides. Divide that square into 4 quadrants, with sides of 1-foot each (left diagram). Then inscribe a circle into each of the 4 quadrants (right diagram).

Division into Quadrants

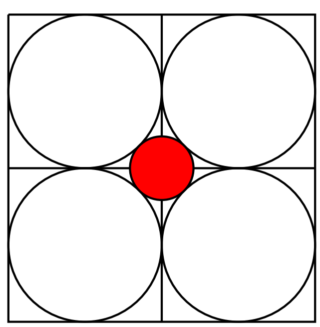


Inscribed Circles



Finally, inscribe a small circle in the middle of the circles (shown in red below).

Central Inscribed Circle



32.1 $\mathbb{R}^n$ Geometry Puzzle: The Question

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 hyperplanes, 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.

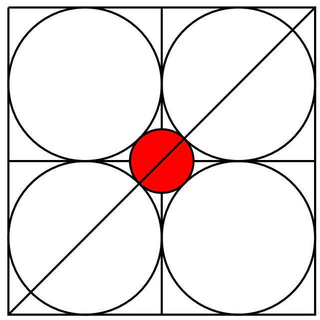
(c) 2.

(d) ∞.

32.2 $\mathbb{R}^n$ Geometry Puzzle: The Answer

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

Main Diagonal



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 twice 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.

September 22, 2022