Chapter Three

Maps Between Spaces

I Isomorphisms

In the examples following the definition of a vector space we expressed the intuition that some spaces are “the same” as others. For instance, the space of two-tall column vectors and the space of two-wide row vectors are not equal because their elements — column vectors and row vectors — are not equal, but we feel that these spaces differ only in how their elements appear. We will now make this precise.

This section illustrates a common phase of a mathematical investigation. With the help of some examples we’ve gotten an idea. We will next give a formal definition and then we will produce some results backing our contention that the definition captures the idea. We’ve seen this happen already, for instance in the first section of the Vector Space chapter. There, the study of linear systems led us to consider collections closed under linear combinations. We defined such a collection as a vector space and we followed it with some supporting results.

That wasn’t an end point, instead it led to new insights such as the idea of a basis. Here also, after producing a definition and supporting it, we will get two surprises (pleasant ones). First, we will find that the definition applies to some unforeseen, and interesting, cases. Second, the study of the definition will lead to new ideas. In this way, our investigation will build momentum.

I.1 Definition and Examples

We start with two examples that suggest the right definition.

1.1 Example The space of two-wide row vectors and the space of two-tall column vectors are "the same" in that if we associate the vectors that have the same components, e.g.,

$$\begin{pmatrix}1 & 2\end{pmatrix} \quad\longleftrightarrow\quad \begin{pmatrix}1\\2\end{pmatrix}$$

(read the double arrow as "corresponds to") then this association respects the operations. For instance these corresponding vectors add to corresponding totals

$$\begin{pmatrix}1 & 2\end{pmatrix}+\begin{pmatrix}3 & 4\end{pmatrix}=\begin{pmatrix}4 & 6\end{pmatrix} \quad\longleftrightarrow\quad \begin{pmatrix}1\\2\end{pmatrix}+\begin{pmatrix}3\\4\end{pmatrix}=\begin{pmatrix}4\\6\end{pmatrix}$$

and here is an example of the correspondence respecting scalar multiplication.

$$5\cdot\begin{pmatrix}1 & 2\end{pmatrix}=\begin{pmatrix}5 & 10\end{pmatrix} \quad\longleftrightarrow\quad 5\cdot\begin{pmatrix}1\\2\end{pmatrix}=\begin{pmatrix}5\\10\end{pmatrix}$$

Stated generally, under the correspondence

$$\begin{pmatrix}a_0 & a_1\end{pmatrix} \quad\longleftrightarrow\quad \begin{pmatrix}a_0\\a_1\end{pmatrix}$$

both operations are preserved:

$$\begin{pmatrix}a_0 & a_1\end{pmatrix}+\begin{pmatrix}b_0 & b_1\end{pmatrix}=\begin{pmatrix}a_0+b_0 & a_1+b_1\end{pmatrix} \longleftrightarrow \begin{pmatrix}a_0\\a_1\end{pmatrix}+\begin{pmatrix}b_0\\b_1\end{pmatrix}=\begin{pmatrix}a_0+b_0\\a_1+b_1\end{pmatrix}$$

and

$$r\cdot\begin{pmatrix}a_0 & a_1\end{pmatrix}=\begin{pmatrix}ra_0 & ra_1\end{pmatrix} \quad\longleftrightarrow\quad r\cdot\begin{pmatrix}a_0\\a_1\end{pmatrix}=\begin{pmatrix}ra_0\\ra_1\end{pmatrix}$$

(all of the variables are scalars).

1.2 Example Another two spaces that we can think of as "the same" are $\mathcal{P}_2$, the space of quadratic polynomials, and $\mathbb{R}^3$. A natural correspondence is this.

$$a_0+a_1x+a_2x^2 \quad\longleftrightarrow\quad \begin{pmatrix}a_0\\a_1\\a_2\end{pmatrix} \qquad\qquad (\text{e.g., } 1+2x+3x^2 \longleftrightarrow \begin{pmatrix}1\\2\\3\end{pmatrix})$$

This preserves structure: corresponding elements add in a corresponding way

$$\begin{array}{r}a_0+a_1x+a_2x^2\\ +\ b_0+b_1x+b_2x^2\\ \hline (a_0+b_0)+(a_1+b_1)x+(a_2+b_2)x^2\end{array} \longleftrightarrow \begin{pmatrix}a_0\\a_1\\a_2\end{pmatrix}+\begin{pmatrix}b_0\\b_1\\b_2\end{pmatrix}=\begin{pmatrix}a_0+b_0\\a_1+b_1\\a_2+b_2\end{pmatrix}$$

and scalar multiplication corresponds also.

$$r \cdot (a_0 + a_1x + a_2x^2) = (ra_0) + (ra_1)x + (ra_2)x^2 \quad \longleftrightarrow \quad r \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} ra_0 \\ ra_1 \\ ra_2 \end{pmatrix}$$

1.3 Definition An *isomorphism* between two vector spaces V and W is a map $f\colon V \to W$ that

(1) is a correspondence: f is one-to-one and onto;*

(2) *preserves structure:* if $\vec{v}_1, \vec{v}_2 \in V$ then

$$f(\vec{v}_1 + \vec{v}_2) = f(\vec{v}_1) + f(\vec{v}_2)$$

and if $\vec{v} \in V$ and $r \in \mathbb{R}$ then

$$f(r\vec{v}) = rf(\vec{v})$$

(we write $V \cong W$, read "V is isomorphic to W", when such a map exists).

"Morphism" means map, so "isomorphism" means a map expressing sameness.

1.4 Example The vector space $G = \{c_1 \cos\theta + c_2 \sin\theta \mid c_1, c_2 \in \mathbb{R}\}$ of functions of θ is isomorphic to $\mathbb{R}^2$ under this map.

$$c_1 \cos\theta + c_2 \sin\theta \stackrel{f}{\longmapsto} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

We will check this by going through the conditions in the definition. We will first verify condition (1), that the map is a correspondence between the sets underlying the spaces.

To establish that f is one-to-one we must prove that $f(\vec{a}) = f(\vec{b})$ only when $\vec{a} = \vec{b}$. If

$$f(a_1 \cos\theta + a_2 \sin\theta) = f(b_1 \cos\theta + b_2 \sin\theta)$$

then by the definition of f

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

from which we conclude that $a_1 = b_1$ and $a_2 = b_2$, because column vectors are equal only when they have equal components. Thus $a_1 \cos\theta + a_2 \sin\theta = b_1 \cos\theta + b_2 \sin\theta$, and as required we've verified that $f(\vec{a}) = f(\vec{b})$ implies that $\vec{a} = \vec{b}$.

*More information on correspondences is in the appendix.

To prove that f is onto we must check that any member of the codomain $\mathbb{R}^2$ is the image of some member of the domain G. So, consider a member of the codomain

$$\begin{pmatrix} x \\ y \end{pmatrix}$$

and note that it is the image under f of $x\cos\theta + y\sin\theta$.

Next we will verify condition (2), that f preserves structure. This computation shows that f preserves addition.

$$\begin{aligned}
f\bigl(\,(a_1\cos\theta + a_2\sin\theta) + (b_1\cos\theta + b_2\sin\theta)\,\bigr)& \\
&= f\bigl(\,(a_1+b_1)\cos\theta + (a_2+b_2)\sin\theta\,\bigr) \\
&= \begin{pmatrix} a_1+b_1 \\ a_2+b_2 \end{pmatrix} \\
&= \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \\
&= f(a_1\cos\theta + a_2\sin\theta) + f(b_1\cos\theta + b_2\sin\theta)
\end{aligned}$$

The computation showing that f preserves scalar multiplication is similar.

$$\begin{aligned}
f\bigl(\,r\cdot(a_1\cos\theta + a_2\sin\theta)\,\bigr) &= f(\,ra_1\cos\theta + ra_2\sin\theta\,) \\
&= \begin{pmatrix} ra_1 \\ ra_2 \end{pmatrix} \\
&= r\cdot\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \\
&= r\cdot\, f(a_1\cos\theta + a_2\sin\theta)
\end{aligned}$$

With both (1) and (2) verified, we know that f is an isomorphism and we can say that the spaces are isomorphic $G \cong \mathbb{R}^2$.

1.5 Example Let V be the space $\{c_1x + c_2y + c_3z \mid c_1, c_2, c_3 \in \mathbb{R}\}$ of linear combinations of the three variables under the natural addition and scalar multiplication operations. Then V is isomorphic to $\mathcal{P}_2$, the space of quadratic polynomials.

To show this we must produce an isomorphism map. There is more than one possibility; for instance, here are four to choose among.

$$c_1x + c_2y + c_3z \quad \begin{array}{ll}
\overset{f_1}{\longmapsto} & c_1 + c_2x + c_3x^2 \\
\overset{f_2}{\longmapsto} & c_2 + c_3x + c_1x^2 \\
\overset{f_3}{\longmapsto} & -c_1 - c_2x - c_3x^2 \\
\overset{f_4}{\longmapsto} & c_1 + (c_1+c_2)x + (c_1+c_3)x^2
\end{array}$$

The first map is the more natural correspondence in that it just carries the coefficients over. However we shall do f_2 to underline that there are isomorphisms other than the obvious one. (Checking that f_1 is an isomorphism is Exercise 14.)

To show that f_2 is one-to-one we will prove that if $f_2(c_1x+c_2y+c_3z) = f_2(d_1x+d_2y+d_3z)$ then $c_1x+c_2y+c_3z = d_1x+d_2y+d_3z$. The assumption that $f_2(c_1x+c_2y+c_3z) = f_2(d_1x+d_2y+d_3z)$ gives, by the definition of f_2, that $c_2+c_3x+c_1x^2 = d_2+d_3x+d_1x^2$. Equal polynomials have equal coefficients so $c_2 = d_2$, $c_3 = d_3$, and $c_1 = d_1$. Hence $f_2(c_1x+c_2y+c_3z) = f_2(d_1x+d_2y+d_3z)$ implies that $c_1x+c_2y+c_3z = d_1x+d_2y+d_3z$, and f_2 is one-to-one.

The map f_2 is onto because a member $a+bx+cx^2$ of the codomain is the image of a member of the domain, namely it is $f_2(cx+ay+bz)$. For instance, $2+3x-4x^2$ is $f_2(-4x+2y+3z)$.

The computations for structure preservation are like those in the prior example. The map f_2 preserves addition

$$\begin{aligned}
&f_2\big((c_1x+c_2y+c_3z)+(d_1x+d_2y+d_3z)\big)\\
&\qquad\qquad = f_2\big((c_1+d_1)x+(c_2+d_2)y+(c_3+d_3)z\big)\\
&\qquad\qquad = (c_2+d_2)+(c_3+d_3)x+(c_1+d_1)x^2\\
&\qquad\qquad = (c_2+c_3x+c_1x^2)+(d_2+d_3x+d_1x^2)\\
&\qquad\qquad = f_2(c_1x+c_2y+c_3z)+f_2(d_1x+d_2y+d_3z)
\end{aligned}$$

and scalar multiplication.

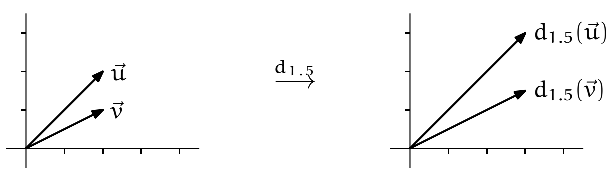
$$\begin{aligned}
f_2\big(r\cdot(c_1x+c_2y+c_3z)\big) &= f_2(rc_1x+rc_2y+rc_3z)\\
&= rc_2+rc_3x+rc_1x^2\\
&= r\cdot(c_2+c_3x+c_1x^2)\\
&= r\cdot\, f_2(c_1x+c_2y+c_3z)
\end{aligned}$$

Thus f_2 is an isomorphism. We write $V \cong \mathcal{P}_2$.

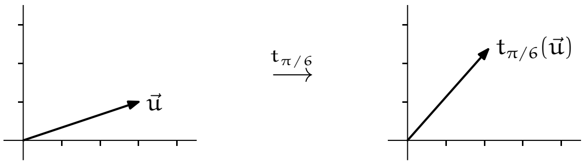
1.6 Example Every space is isomorphic to itself under the identity map. The check is easy.

1.7 Definition An *automorphism* is an isomorphism of a space with itself.

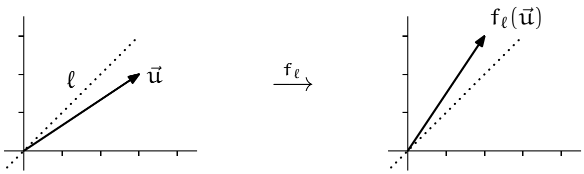
1.8 Example A *dilation* map $d_s\colon \mathbb{R}^2 \to \mathbb{R}^2$ that multiplies all vectors by a nonzero scalar s is an automorphism of $\mathbb{R}^2$.



Another automorphism is a *rotation* or *turning map*, $t_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ that rotates all vectors through an angle θ.



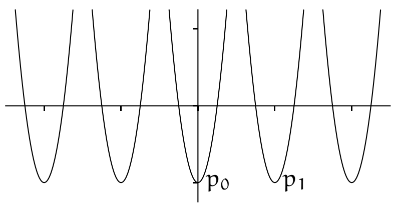
A third type of automorphism of $\mathbb{R}^2$ is a map $f_\ell\colon \mathbb{R}^2 \to \mathbb{R}^2$ that *flips* or *reflects* all vectors over a line ℓ through the origin.



Checking that these are automorphisms is Exercise 31.

1.9 Example Consider the space $\mathcal{P}_5$ of polynomials of degree 5 or less and the map f that sends a polynomial $p(x)$ to $p(x-1)$. For instance, under this map $x^2 \mapsto (x-1)^2 = x^2-2x+1$ and $x^3+2x \mapsto (x-1)^3+2(x-1) = x^3-3x^2+5x-3$. This map is an automorphism of this space; the check is Exercise 23.

This isomorphism of $\mathcal{P}_5$ with itself does more than just tell us that the space is “the same” as itself. It gives us some insight into the space's structure. Below is a family of parabolas, graphs of members of $\mathcal{P}_5$. Each has a vertex at $y=-1$, and the left-most one has zeroes at -2.25 and -1.75, the next one has zeroes at -1.25 and -0.75, etc.



Substitution of $x-1$ for x in any function's argument shifts its graph to the right by one. Thus, $f(p_0)=p_1$, and f's action is to shift all of the parabolas to the right by one. Notice that the picture before f is applied is the same as the picture after f is applied because while each parabola moves to the right, another one comes in from the left to take its place. This also holds true for cubics, etc. So the automorphism f expresses the idea that P_5 has a certain horizontal-homogeneity: if we draw two pictures showing all members of $\mathcal{P}_5$, one picture centered at $x = 0$ and the other centered at $x = 1$, then the two pictures would be indistinguishable.

As described in the opening to this section, having given the definition of isomorphism, we next look to support the thesis that it captures our intuition of vector spaces being the same. First, the definition itself is persuasive: a vector space consists of a set and some structure and the definition simply requires that the sets correspond and that the structures correspond also. Also persuasive are the examples above, such as Example 1.1, which dramatize that isomorphic spaces are the same in all relevant respects. Sometimes people say, where $V \cong W$, that "W is just V painted green" — differences are merely cosmetic.

The results below further support our contention that under an isomorphism all the things of interest in the two vector spaces correspond. Because we introduced vector spaces to study linear combinations, "of interest" means "pertaining to linear combinations." Not of interest is the way that the vectors are presented typographically (or their color!).

1.10 Lemma An isomorphism maps a zero vector to a zero vector.

Proof Where $f\colon V \to W$ is an isomorphism, fix some $\vec{v} \in V$. Then $f(\vec{0}_V) = f(0 \cdot \vec{v}) = 0 \cdot f(\vec{v}) = \vec{0}_W$. QED

1.11 Lemma For any map $f\colon V \to W$ between vector spaces these statements are equivalent.

(1) f preserves structure

$$f(\vec{v}_1 + \vec{v}_2) = f(\vec{v}_1) + f(\vec{v}_2) \quad \text{and} \quad f(c\vec{v}) = c\,f(\vec{v})$$

(2) f preserves linear combinations of two vectors

$$f(c_1\vec{v}_1 + c_2\vec{v}_2) = c_1 f(\vec{v}_1) + c_2 f(\vec{v}_2)$$

(3) f preserves linear combinations of any finite number of vectors

$$f(c_1\vec{v}_1 + \cdots + c_n\vec{v}_n) = c_1 f(\vec{v}_1) + \cdots + c_n f(\vec{v}_n)$$

Proof Since the implications (3) $\Longrightarrow$ (2) and (2) $\Longrightarrow$ (1) are clear, we need only show that (1) $\Longrightarrow$ (3). So assume statement (1). We will prove (3) by induction on the number of summands n.

The one-summand base case, that $f(c\vec{v}_1) = c\,f(\vec{v}_1)$, is covered by the second clause of statement (1).

For the inductive step assume that statement (3) holds whenever there are k or fewer summands. Consider the $k+1$-summand case. Use the first half of (1) to break the sum along the final '+'.

$$f(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k + c_{k+1}\vec{v}_{k+1}) = f(c_1\vec{v}_1 + \cdots + c_k\vec{v}_k) + f(c_{k+1}\vec{v}_{k+1})$$

Use the inductive hypothesis to break up the k-term sum on the left.

$$= f(c_1\vec{v}_1) + \cdots + f(c_k\vec{v}_k) + f(c_{k+1}\vec{v}_{k+1})$$

Now the second half of (1) gives

$$= c_1\, f(\vec{v}_1) + \cdots + c_k\, f(\vec{v}_k) + c_{k+1}\, f(\vec{v}_{k+1})$$

when applied $k+1$ times. QED

We often use item (2) to simplify the verification that a map preserves structure.

Finally, a summary. In the prior chapter, after giving the definition of a vector space, we looked at examples and noted that some spaces seemed to be essentially the same as others. Here we have defined the relation '$\cong$' and have argued that it is the right way to precisely say what we mean by "the same" because it preserves the features of interest in a vector space — in particular, it preserves linear combinations. In the next section we will show that isomorphism is an equivalence relation and so partitions the collection of vector spaces.

Exercises

✓ **1.12** Verify, using Example 1.4 as a model, that the two correspondences given before the definition are isomorphisms.
  (a) Example 1.1 **(b)** Example 1.2

✓ **1.13** For the map $f\colon \mathcal{P}_1 \to \mathbb{R}^2$ given by

$$a + bx \stackrel{f}{\longmapsto} \begin{pmatrix} a-b \\ b \end{pmatrix}$$

Find the image of each of these elements of the domain.
  (a) $3-2x$ **(b)** $2+2x$ **(c)** x

Show that this map is an isomorphism.

1.14 Show that the natural map f_1 from Example 1.5 is an isomorphism.

✓ **1.15** Decide whether each map is an isomorphism (if it is an isomorphism then prove it and if it isn't then state a condition that it fails to satisfy).
  (a) $f\colon \mathcal{M}_{2\times 2} \to \mathbb{R}$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto ad - bc$$

  (b) $f\colon \mathcal{M}_{2\times 2} \to \mathbb{R}^4$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a+b+c+d \\ a+b+c \\ a+b \\ a \end{pmatrix}$$

(c) $f\colon \mathcal{M}_{2\times2} \to \mathcal{P}_3$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto c + (d+c)x + (b+a)x^2 + ax^3$$

(d) $f\colon \mathcal{M}_{2\times2} \to \mathcal{P}_3$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto c + (d+c)x + (b+a+1)x^2 + ax^3$$

1.16 Show that the map $f\colon \mathbb{R}^1 \to \mathbb{R}^1$ given by $f(x) = x^3$ is one-to-one and onto. Is it an isomorphism?

✓ 1.17 Refer to Example 1.1. Produce two more isomorphisms (of course, you must also verify that they satisfy the conditions in the definition of isomorphism).

1.18 Refer to Example 1.2. Produce two more isomorphisms (and verify that they satisfy the conditions).

✓ 1.19 Show that, although $\mathbb{R}^2$ is not itself a subspace of $\mathbb{R}^3$, it is isomorphic to the xy-plane subspace of $\mathbb{R}^3$.

1.20 Find two isomorphisms between $\mathbb{R}^{16}$ and $\mathcal{M}_{4\times4}$.

✓ 1.21 For what k is $\mathcal{M}_{m\times n}$ isomorphic to $\mathbb{R}^k$?

1.22 For what k is $\mathcal{P}_k$ isomorphic to $\mathbb{R}^n$?

1.23 Prove that the map in Example 1.9, from $\mathcal{P}_5$ to $\mathcal{P}_5$ given by $p(x) \mapsto p(x-1)$, is a vector space isomorphism.

1.24 Why, in Lemma 1.10, must there be a $\vec{v} \in V$? That is, why must V be nonempty?

1.25 Are any two trivial spaces isomorphic?

1.26 In the proof of Lemma 1.11, what about the zero-summands case (that is, if n is zero)?

1.27 Show that any isomorphism $f\colon \mathcal{P}_0 \to \mathbb{R}^1$ has the form $a \mapsto ka$ for some nonzero real number k.

✓ 1.28 These prove that isomorphism is an equivalence relation.

(a) Show that the identity map $\mathrm{id}\colon V \to V$ is an isomorphism. Thus, any vector space is isomorphic to itself.

(b) Show that if $f\colon V \to W$ is an isomorphism then so is its inverse $f^{-1}\colon W \to V$. Thus, if V is isomorphic to W then also W is isomorphic to V.

(c) Show that a composition of isomorphisms is an isomorphism: if $f\colon V \to W$ is an isomorphism and $g\colon W \to U$ is an isomorphism then so also is $g \circ f\colon V \to U$. Thus, if V is isomorphic to W and W is isomorphic to U, then also V is isomorphic to U.

1.29 Suppose that $f\colon V \to W$ preserves structure. Show that f is one-to-one if and only if the unique member of V mapped by f to $\vec{0}_W$ is $\vec{0}_V$.

1.30 Suppose that $f\colon V \to W$ is an isomorphism. Prove that the set $\{\vec{v}_1, \ldots, \vec{v}_k\} \subseteq V$ is linearly dependent if and only if the set of images $\{f(\vec{v}_1), \ldots, f(\vec{v}_k)\} \subseteq W$ is linearly dependent.

✓ 1.31 Show that each type of map from Example 1.8 is an automorphism.

(a) Dilation d_s by a nonzero scalar s.

(b) Rotation t_θ through an angle θ.

(c) Reflection f_ℓ over a line through the origin.

Hint. For the second and third items, polar coordinates are useful.

1.32 Produce an automorphism of $\mathcal{P}_2$ other than the identity map, and other than a shift map $p(x) \mapsto p(x-k)$.

1.33 **(a)** Show that a function $f\colon \mathbb{R}^1 \to \mathbb{R}^1$ is an automorphism if and only if it has the form $x \mapsto kx$ for some $k \neq 0$.

(b) Let f be an automorphism of $\mathbb{R}^1$ such that $f(3) = 7$. Find $f(-2)$.

(c) Show that a function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is an automorphism if and only if it has the form

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix}$$

for some $a, b, c, d \in \mathbb{R}$ with $ad - bc \neq 0$. *Hint.* Exercises in prior subsections have shown that

$$\begin{pmatrix} b \\ d \end{pmatrix} \text{ is not a multiple of } \begin{pmatrix} a \\ c \end{pmatrix}$$

if and only if $ad - bc \neq 0$.

(d) Let f be an automorphism of $\mathbb{R}^2$ with

$$f(\begin{pmatrix} 1 \\ 3 \end{pmatrix}) = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \quad\text{and}\quad f(\begin{pmatrix} 1 \\ 4 \end{pmatrix}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Find

$$f(\begin{pmatrix} 0 \\ -1 \end{pmatrix}).$$

1.34 Refer to Lemma 1.10 and Lemma 1.11. Find two more things preserved by isomorphism.

1.35 We show that isomorphisms can be tailored to fit in that, sometimes, given vectors in the domain and in the range we can produce an isomorphism associating those vectors.

(a) Let $B = \langle \vec\beta_1, \vec\beta_2, \vec\beta_3 \rangle$ be a basis for $\mathcal{P}_2$ so that any $\vec p \in \mathcal{P}_2$ has a unique representation as $\vec p = c_1\vec\beta_1 + c_2\vec\beta_2 + c_3\vec\beta_3$, which we denote in this way.

$$\mathrm{Rep}_B(\vec p) = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

Show that the $\mathrm{Rep}_B(\cdot)$ operation is a function from $\mathcal{P}_2$ to $\mathbb{R}^3$ (this entails showing that with every domain vector $\vec v \in \mathcal{P}_2$ there is an associated image vector in $\mathbb{R}^3$, and further, that with every domain vector $\vec v \in \mathcal{P}_2$ there is at most one associated image vector).

(b) Show that this $\mathrm{Rep}_B(\cdot)$ function is one-to-one and onto.

(c) Show that it preserves structure.

(d) Produce an isomorphism from $\mathcal{P}_2$ to $\mathbb{R}^3$ that fits these specifications.

$$x + x^2 \mapsto \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad\text{and}\quad 1 - x \mapsto \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

1.36 Prove that a space is n-dimensional if and only if it is isomorphic to $\mathbb{R}^n$. *Hint.* Fix a basis B for the space and consider the map sending a vector over to its representation with respect to B.

1.37 *(Requires the subsection on Combining Subspaces, which is optional.)* Let U and W be vector spaces. Define a new vector space, consisting of the set $U \times W = \{(\vec{u}, \vec{w}) \mid \vec{u} \in U \text{ and } \vec{w} \in W\}$ along with these operations.

$$(\vec{u}_1, \vec{w}_1) + (\vec{u}_2, \vec{w}_2) = (\vec{u}_1 + \vec{u}_2, \vec{w}_1 + \vec{w}_2) \quad \text{and} \quad r \cdot (\vec{u}, \vec{w}) = (r\vec{u}, r\vec{w})$$

This is a vector space, the *external direct sum* of U and W.

(a) Check that it is a vector space.

(b) Find a basis for, and the dimension of, the external direct sum $\mathcal{P}_2 \times \mathbb{R}^2$.

(c) What is the relationship among $\dim(U)$, $\dim(W)$, and $\dim(U \times W)$?

(d) Suppose that U and W are subspaces of a vector space V such that $V = U \oplus W$ (in this case we say that V is the *internal direct sum* of U and W). Show that the map $f\colon U \times W \to V$ given by

$$(\vec{u}, \vec{w}) \stackrel{f}{\longmapsto} \vec{u} + \vec{w}$$

is an isomorphism. Thus if the internal direct sum is defined then the internal and external direct sums are isomorphic.

I.2 Dimension Characterizes Isomorphism

In the prior subsection, after stating the definition of isomorphism, we gave some results supporting our sense that such a map describes spaces as "the same." Here we will develop this intuition. When two (unequal) spaces are isomorphic we think of them as almost equal, as equivalent. We shall make that precise by proving that the relationship 'is isomorphic to' is an equivalence relation.

2.1 Lemma The inverse of an isomorphism is also an isomorphism.

Proof Suppose that V is isomorphic to W via $f\colon V \to W$. An isomorphism is a correspondence between the sets so f has an inverse function $f^{-1}\colon W \to V$ that is also a correspondence.*

We will show that because f preserves linear combinations, so also does f^{-1}. Suppose that $\vec{w}_1, \vec{w}_2 \in W$. Because it is an isomorphism, f is onto and there are $\vec{v}_1, \vec{v}_2 \in V$ such that $\vec{w}_1 = f(\vec{v}_1)$ and $\vec{w}_2 = f(\vec{v}_2)$. Then

$$\begin{aligned} f^{-1}(c_1 \cdot \vec{w}_1 + c_2 \cdot \vec{w}_2) &= f^{-1}\big(c_1 \cdot f(\vec{v}_1) + c_2 \cdot f(\vec{v}_2)\big) \\ &= f^{-1}\big(f(c_1\vec{v}_1 + c_2\vec{v}_2)\big) = c_1\vec{v}_1 + c_2\vec{v}_2 = c_1 \cdot f^{-1}(\vec{w}_1) + c_2 \cdot f^{-1}(\vec{w}_2) \end{aligned}$$

since $f^{-1}(\vec{w}_1) = \vec{v}_1$ and $f^{-1}(\vec{w}_2) = \vec{v}_2$. With that, by Lemma 1.11's second statement, this map preserves structure. QED

* More information on inverse functions is in the appendix.

2.2 Theorem Isomorphism is an equivalence relation between vector spaces.

PROOF We must prove that the relation is symmetric, reflexive, and transitive.

To check reflexivity, that any space is isomorphic to itself, consider the identity map. It is clearly one-to-one and onto. This shows that it preserves linear combinations.

$$\mathrm{id}(c_1 \cdot \vec{v}_1 + c_2 \cdot \vec{v}_2) = c_1\vec{v}_1 + c_2\vec{v}_2 = c_1 \cdot \mathrm{id}(\vec{v}_1) + c_2 \cdot \mathrm{id}(\vec{v}_2)$$

Symmetry, that if V is isomorphic to W then also W is isomorphic to V, holds by Lemma 2.1 since each isomorphism map from V to W is paired with an isomorphism from W to V.

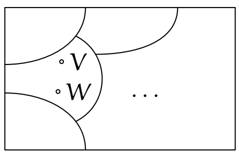
To finish we must check transitivity, that if V is isomorphic to W and W is isomorphic to U then V is isomorphic to U. Let $f\colon V \to W$ and $g\colon W \to U$ be isomorphisms. Consider their composition $g \circ f\colon V \to U$. Because the composition of correspondences is a correspondence, we need only check that the composition preserves linear combinations.

$$\begin{aligned}
g \circ f\,\big(c_1 \cdot \vec{v}_1 + c_2 \cdot \vec{v}_2\big) &= g\big(\, f(\, c_1 \cdot \vec{v}_1 + c_2 \cdot \vec{v}_2\,)\,\big) \\
&= g\big(\, c_1 \cdot f(\vec{v}_1) + c_2 \cdot f(\vec{v}_2)\,\big) \\
&= c_1 \cdot g\big(f(\vec{v}_1)\big) + c_2 \cdot g\big(f(\vec{v}_2)\big) \\
&= c_1 \cdot (g \circ f)\,(\vec{v}_1) + c_2 \cdot (g \circ f)\,(\vec{v}_2)
\end{aligned}$$

Thus the composition is an isomorphism. QED

Since it is an equivalence, isomorphism partitions the universe of vector spaces into classes: each space is in one and only one isomorphism class.

All finite dimensional vector spaces:



$V \cong W$

The next result characterizes these classes by dimension. That is, we can describe each class simply by giving the number that is the dimension of all of the spaces in that class.

2.3 Theorem Vector spaces are isomorphic if and only if they have the same dimension.

In this double implication statement the proof of each half involves a significant idea so we will do the two separately.

2.4 Lemma If spaces are isomorphic then they have the same dimension.

Proof We shall show that an isomorphism of two spaces gives a correspondence between their bases. That is, we shall show that if $f\colon V\to W$ is an isomorphism and a basis for the domain V is $B=\langle\vec{\beta}_1,\dots,\vec{\beta}_n\rangle$ then its image $D=\langle f(\vec{\beta}_1),\dots,f(\vec{\beta}_n)\rangle$ is a basis for the codomain W. (The other half of the correspondence, that for any basis of W the inverse image is a basis for V, follows from the fact that f^{-1} is also an isomorphism and so we can apply the prior sentence to f^{-1}.)

To see that D spans W, fix any $\vec{w}\in W$. Because f is an isomorphism it is onto and so there is a $\vec{v}\in V$ with $\vec{w}=f(\vec{v})$. Expand $\vec{v}$ as a combination of basis vectors.

$$\vec{w}=f(\vec{v})=f(v_1\vec{\beta}_1+\cdots+v_n\vec{\beta}_n)=v_1\cdot f(\vec{\beta}_1)+\cdots+v_n\cdot f(\vec{\beta}_n)$$

For linear independence of D, if

$$\vec{0}_W=c_1f(\vec{\beta}_1)+\cdots+c_nf(\vec{\beta}_n)=f(c_1\vec{\beta}_1+\cdots+c_n\vec{\beta}_n)$$

then, since f is one-to-one and so the only vector sent to $\vec{0}_W$ is $\vec{0}_V$, we have that $\vec{0}_V=c_1\vec{\beta}_1+\cdots+c_n\vec{\beta}_n$, which implies that all of the c's are zero. QED

2.5 Lemma If spaces have the same dimension then they are isomorphic.

Proof We will prove that any space of dimension n is isomorphic to $\mathbb{R}^n$. Then we will have that all such spaces are isomorphic to each other by transitivity, which was shown in Theorem 2.2.

Let V be n-dimensional. Fix a basis $B=\langle\vec{\beta}_1,\dots,\vec{\beta}_n\rangle$ for the domain V. Consider the operation of representing the members of V with respect to B as a function from V to $\mathbb{R}^n$.

$$\vec{v}=v_1\vec{\beta}_1+\cdots+v_n\vec{\beta}_n\;\overset{\text{Rep}_B}{\longmapsto}\;\begin{pmatrix}v_1\\ \vdots\\ v_n\end{pmatrix}$$

It is well-defined* since every $\vec{v}$ has one and only one such representation (see Remark 2.6 following this proof).

This function is one-to-one because if

$$\text{Rep}_B(u_1\vec{\beta}_1+\cdots+u_n\vec{\beta}_n)=\text{Rep}_B(v_1\vec{\beta}_1+\cdots+v_n\vec{\beta}_n)$$

* More information on well-defined is in the appendix.

then

$$\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

and so $u_1 = v_1$, ..., $u_n = v_n$, implying that the original arguments $u_1\vec{\beta}_1 + \cdots + u_n\vec{\beta}_n$ and $v_1\vec{\beta}_1 + \cdots + v_n\vec{\beta}_n$ are equal.

This function is onto; any member of $\mathbb{R}^n$

$$\vec{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$$

is the image of some $\vec{v} \in V$, namely $\vec{w} = \mathrm{Rep}_B(w_1\vec{\beta}_1 + \cdots + w_n\vec{\beta}_n)$.

Finally, this function preserves structure.

$$\begin{aligned}
\mathrm{Rep}_B(r\cdot\vec{u} + s\cdot\vec{v}) &= \mathrm{Rep}_B(\,(ru_1+sv_1)\vec{\beta}_1 + \cdots + (ru_n+sv_n)\vec{\beta}_n\,) \\
&= \begin{pmatrix} ru_1+sv_1 \\ \vdots \\ ru_n+sv_n \end{pmatrix} \\
&= r\cdot\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} + s\cdot\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \\
&= r\cdot\mathrm{Rep}_B(\vec{u}) + s\cdot\mathrm{Rep}_B(\vec{v})
\end{aligned}$$

Therefore Rep_B is an isomorphism. Consequently any n-dimensional space is isomorphic to $\mathbb{R}^n$. QED

2.6 Remark The proof has a sentence about 'well-defined.' Its point is that to be an isomorphism Rep_B must be a function, and the definition of function requires that for all inputs the associated output exists and is determined by the input. So we must check that every $\vec{v}$ is associated with at least one $\mathrm{Rep}_B(\vec{v})$, and no more than one.

In the proof we express elements $\vec{v}$ of the domain space as combinations of members of the basis B and then associate $\vec{v}$ with the column vector of coefficients. That there is at least one expansion of each $\vec{v}$ holds because B is a basis and so spans the space.

The worry that there is no more than one associated member of the codomain is subtler. A contrasting example, where an association fails this unique output requirement, illuminates the issue. Let the domain be $\mathcal{P}_2$ and consider a set that is not a basis (it is not linearly independent, although it does span the space).

$$A = \{1 + 0x + 0x^2, 0 + 1x + 0x^2, 0 + 0x + 1x^2, 1 + 1x + 2x^2\}$$

Call those polynomials $\vec{\alpha}_1, \ldots, \vec{\alpha}_4$. In contrast to the situation when the set is a basis, here there can be more than one expression of a domain vector in terms of members of the set. For instance, consider $\vec{v} = 1 + x + x^2$. Here are two different expansions.

$$\vec{v} = 1\vec{\alpha}_1 + 1\vec{\alpha}_2 + 1\vec{\alpha}_3 + 0\vec{\alpha}_4 \qquad \vec{v} = 0\vec{\alpha}_1 + 0\vec{\alpha}_2 - 1\vec{\alpha}_3 + 1\vec{\alpha}_4$$

So this input vector $\vec{v}$ is associated with more than one column.

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

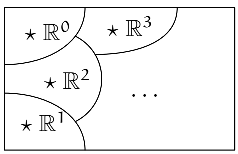
Thus, with A the association is not well-defined. (The issue is that A is not linearly independent; to show uniqueness Theorem Two.III.1.12's proof uses only linear independence.)

In general, any time that we define a function we must check that output values are well-defined. Most of the time that condition is perfectly obvious but in the above proof it needs verification. See Exercise 19.

2.7 Corollary A finite-dimensional vector space is isomorphic to one and only one of the $\mathbb{R}^n$.

This gives us a collection of representatives of the isomorphism classes.

All finite dimensional vector spaces:



One representative per class

The proofs above pack many ideas into a small space. Through the rest of this chapter we'll consider these ideas again, and fill them out. As a taste of this we will expand here on the proof of Lemma 2.5.

2.8 Example The space $\mathcal{M}_{2\times 2}$ of 2×2 matrices is isomorphic to $\mathbb{R}^4$. With this basis for the domain

$$B = \langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \rangle$$

the isomorphism given in the lemma, the representation map $f_1 = \mathrm{Rep}_B$, carries the entries over.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \stackrel{f_1}{\longmapsto} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

One way to think of the map f_1 is: fix the basis B for the domain, use the standard basis $\mathcal{E}_4$ for the codomain, and associate $\vec{\beta}_1$ with $\vec{e}_1$, $\vec{\beta}_2$ with $\vec{e}_2$, etc. Then extend this association to all of the members of two spaces.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\vec{\beta}_1 + b\vec{\beta}_2 + c\vec{\beta}_3 + d\vec{\beta}_4 \quad \stackrel{f_1}{\longmapsto} \quad a\vec{e}_1 + b\vec{e}_2 + c\vec{e}_3 + d\vec{e}_4 = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$$

We can do the same thing with different bases, for instance, taking this basis for the domain.

$$A = \langle \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} \rangle$$

Associating corresponding members of A and $\mathcal{E}_4$ gives this.

$$\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= (a/2)\vec{\alpha}_1 + (b/2)\vec{\alpha}_2 + (c/2)\vec{\alpha}_3 + (d/2)\vec{\alpha}_4 \\ &\stackrel{f_2}{\longmapsto} \quad (a/2)\vec{e}_1 + (b/2)\vec{e}_2 + (c/2)\vec{e}_3 + (d/2)\vec{e}_4 = \begin{pmatrix} a/2 \\ b/2 \\ c/2 \\ d/2 \end{pmatrix} \end{aligned}$$

gives rise to an isomorphism that is different than f_1.

The prior map arose by changing the basis for the domain. We can also change the basis for the codomain. Go back to the basis B above and use this basis for the codomain.

$$D = \langle \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \rangle$$

Associate $\vec{\beta}_1$ with $\vec{\delta}_1$, etc. Extending that gives another isomorphism.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\vec{\beta}_1 + b\vec{\beta}_2 + c\vec{\beta}_3 + d\vec{\beta}_4 \quad \stackrel{f_3}{\longmapsto} \quad a\vec{\delta}_1 + b\vec{\delta}_2 + c\vec{\delta}_3 + d\vec{\delta}_4 = \begin{pmatrix} a \\ b \\ d \\ c \end{pmatrix}$$

We close with a recap. Recall that the first chapter defines two matrices to be row equivalent if they can be derived from each other by row operations. There we showed that relation is an equivalence and so the collection of matrices is partitioned into classes, where all the matrices that are row equivalent together fall into a single class. Then for insight into which matrices are in each class we gave representatives for the classes, the reduced echelon form matrices.

In this section we have followed that pattern except that the notion here of "the same" is vector space isomorphism. We defined it and established some properties, including that it is an equivalence. Then, as before, we developed a list of class representatives to help us understand the partition — it classifies vector spaces by dimension.

In Chapter Two, with the definition of vector spaces, we seemed to have opened up our studies to many examples of new structures besides the familiar $\mathbb{R}^n$'s. We now know that isn't the case. Any finite-dimensional vector space is actually "the same" as a real space.

Exercises

✓ 2.9 Decide if the spaces are isomorphic.
(a) $\mathbb{R}^2$, $\mathbb{R}^4$ (b) $\mathcal{P}_5$, $\mathbb{R}^5$ (c) $\mathcal{M}_{2\times 3}$, $\mathbb{R}^6$ (d) $\mathcal{P}_5$, $\mathcal{M}_{2\times 3}$
(e) $\mathcal{M}_{2\times k}$, $\mathbb{C}^k$

✓ 2.10 Consider the isomorphism $\mathrm{Rep}_B(\cdot)\colon \mathcal{P}_1 \to \mathbb{R}^2$ where $B = \langle 1, 1+x\rangle$. Find the image of each of these elements of the domain.
(a) $3-2x$; (b) $2+2x$; (c) x

✓ 2.11 Show that if $m \neq n$ then $\mathbb{R}^m \ncong \mathbb{R}^n$.

✓ 2.12 Is $\mathcal{M}_{m\times n} \cong \mathcal{M}_{n\times m}$?

✓ 2.13 Are any two planes through the origin in $\mathbb{R}^3$ isomorphic?

2.14 Find a set of equivalence class representatives other than the set of $\mathbb{R}^n$'s.

2.15 True or false: between any n-dimensional space and $\mathbb{R}^n$ there is exactly one isomorphism.

2.16 Can a vector space be isomorphic to one of its (proper) subspaces?

✓ 2.17 This subsection shows that for any isomorphism, the inverse map is also an isomorphism. This subsection also shows that for a fixed basis B of an n-dimensional vector space V, the map $\mathrm{Rep}_B\colon V \to \mathbb{R}^n$ is an isomorphism. Find the inverse of this map.

✓ 2.18 Prove these facts about matrices.
(a) The row space of a matrix is isomorphic to the column space of its transpose.
(b) The row space of a matrix is isomorphic to its column space.

2.19 Show that the function from Theorem 2.3 is well-defined.

2.20 Is the proof of Theorem 2.3 valid when $n = 0$?

2.21 For each, decide if it is a set of isomorphism class representatives.
(a) $\{\mathbb{C}^k \mid k \in \mathbb{N}\}$
(b) $\{\mathcal{P}_k \mid k \in \{-1, 0, 1, \ldots\}\}$

(c) $\{\mathcal{M}_{m\times n} \mid m,n\in\mathbb{N}\}$

2.22 Let f be a correspondence between vector spaces V and W (that is, a map that is one-to-one and onto). Show that the spaces V and W are isomorphic via f if and only if there are bases $B\subset V$ and $D\subset W$ such that corresponding vectors have the same coordinates: $\operatorname{Rep}_B(\vec{v})=\operatorname{Rep}_D(f(\vec{v}))$.

2.23 Consider the isomorphism $\operatorname{Rep}_B\colon\mathcal{P}_3\to\mathbb{R}^4$.

(a) Vectors in a real space are orthogonal if and only if their dot product is zero. Give a definition of orthogonality for polynomials.

(b) The derivative of a member of $\mathcal{P}_3$ is in $\mathcal{P}_3$. Give a definition of the derivative of a vector in $\mathbb{R}^4$.

✓ 2.24 Does every correspondence between bases, when extended to the spaces, give an isomorphism? That is, suppose that V is a vector space with basis $B=\langle\vec{\beta}_1,\ldots,\vec{\beta}_n\rangle$ and that $f\colon B\to W$ is a correspondence such that $D=\langle f(\vec{\beta}_1),\ldots,f(\vec{\beta}_n)\rangle$ is basis for W. Must $\hat{f}\colon V\to W$ sending $\vec{v}=c_1\vec{\beta}_1+\cdots+c_n\vec{\beta}_n$ to $\hat{f}(\vec{v})=c_1\hat{f}(\vec{\beta}_1)+\cdots+c_n\hat{f}(\vec{\beta}_n)$ be an isomorphism?

2.25 *(Requires the subsection on Combining Subspaces, which is optional.)* Suppose that $V=V_1\oplus V_2$ and that V is isomorphic to the space U under the map f. Show that $U=f(V_1)\oplus f(U_2)$.

2.26 Show that this is not a well-defined function from the rational numbers to the integers: with each fraction, associate the value of its numerator.

II Homomorphisms

The definition of isomorphism has two conditions. In this section we will consider the second one. We will study maps that are required only to preserve structure, maps that are not also required to be correspondences.

Experience shows that these maps are tremendously useful. For one thing we shall see in the second subsection below that while isomorphisms describe how spaces are the same, we can think of these maps as describing how spaces are alike.

II.1 Definition

1.1 Definition A function between vector spaces $h\colon V \to W$ that preserves addition

$$\text{if } \vec{v}_1, \vec{v}_2 \in V \text{ then } h(\vec{v}_1 + \vec{v}_2) = h(\vec{v}_1) + h(\vec{v}_2)$$

and scalar multiplication

$$\text{if } \vec{v} \in V \text{ and } r \in \mathbb{R} \text{ then } h(r \cdot \vec{v}) = r \cdot h(\vec{v})$$

is a *homomorphism* or *linear map*.

1.2 Example The projection map $\pi\colon \mathbb{R}^3 \to \mathbb{R}^2$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \stackrel{\pi}{\longmapsto} \begin{pmatrix} x \\ y \end{pmatrix}$$

is a homomorphism. It preserves addition

$$\pi(\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}) = \pi(\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix}) = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} = \pi(\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}) + \pi(\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix})$$

and scalar multiplication.

$$\pi(r \cdot \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}) = \pi(\begin{pmatrix} rx_1 \\ ry_1 \\ rz_1 \end{pmatrix}) = \begin{pmatrix} rx_1 \\ ry_1 \end{pmatrix} = r \cdot \pi(\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix})$$

This is not an isomorphism since it is not one-to-one. For instance, both $\vec{0}$ and $\vec{e}_3$ in $\mathbb{R}^3$ map to the zero vector in $\mathbb{R}^2$.

1.3 Example The domain and codomain can be other than spaces of column vectors. Both of these are homomorphisms; the verifications are straightforward.

(1) $f_1 \colon \mathcal{P}_2 \to \mathcal{P}_3$ given by

$$a_0 + a_1 x + a_2 x^2 \;\mapsto\; a_0 x + (a_1/2)x^2 + (a_2/3)x^3$$

(2) $f_2 \colon \mathcal{M}_{2\times 2} \to \mathbb{R}$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto a + d$$

1.4 Example Between any two spaces there is a *zero homomorphism*, mapping every vector in the domain to the zero vector in the codomain.

1.5 Example These two suggest why we use the term 'linear map'.

(1) The map $g \colon \mathbb{R}^3 \to \mathbb{R}$ given by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \stackrel{g}{\longmapsto} 3x + 2y - 4.5z$$

is linear, that is, is a homomorphism. The check is easy. In contrast, the map $\hat{g} \colon \mathbb{R}^3 \to \mathbb{R}$ given by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \stackrel{\hat{g}}{\longmapsto} 3x + 2y - 4.5z + 1$$

is not linear. To show this we need only produce a single linear combination that the map does not preserve. Here is one.

$$\hat{g}(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}) = 4 \qquad \hat{g}(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}) + \hat{g}(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}) = 5$$

(2) The first of these two maps $t_1, t_2 \colon \mathbb{R}^3 \to \mathbb{R}^2$ is linear while the second is not.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \stackrel{t_1}{\longmapsto} \begin{pmatrix} 5x - 2y \\ x + y \end{pmatrix} \qquad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \stackrel{t_2}{\longmapsto} \begin{pmatrix} 5x - 2y \\ xy \end{pmatrix}$$

Finding a linear combination that the second map does not preserve is easy.

So one way to think of 'homomorphism' is that we are generalizing 'isomorphism' (by dropping the condition that the map is a correspondence), motivated by the observation that many of the properties of isomorphisms have only to do with the map's structure-preservation property. The next two results are examples of this motivation. In the prior section we saw a proof for each that only uses preservation of addition and preservation of scalar multiplication, and therefore applies to homomorphisms.

1.6 Lemma A homomorphism sends the zero vector to the zero vector.

1.7 Lemma The following are equivalent for any map $f\colon V \to W$ between vector spaces.

(1) f is a homomorphism
(2) $f(c_1 \cdot \vec{v}_1 + c_2 \cdot \vec{v}_2) = c_1 \cdot f(\vec{v}_1) + c_2 \cdot f(\vec{v}_2)$ for any $c_1, c_2 \in \mathbb{R}$ and $\vec{v}_1, \vec{v}_2 \in V$
(3) $f(c_1 \cdot \vec{v}_1 + \cdots + c_n \cdot \vec{v}_n) = c_1 \cdot f(\vec{v}_1) + \cdots + c_n \cdot f(\vec{v}_n)$ for any $c_1, \ldots, c_n \in \mathbb{R}$ and $\vec{v}_1, \ldots, \vec{v}_n \in V$

1.8 Example The function $f\colon \mathbb{R}^2 \to \mathbb{R}^4$ given by

$$\begin{pmatrix} x \\ y \end{pmatrix} \stackrel{f}{\longmapsto} \begin{pmatrix} x/2 \\ 0 \\ x+y \\ 3y \end{pmatrix}$$

is linear since it satisfies item (2).

$$\begin{pmatrix} r_1(x_1/2) + r_2(x_2/2) \\ 0 \\ r_1(x_1+y_1) + r_2(x_2+y_2) \\ r_1(3y_1) + r_2(3y_2) \end{pmatrix} = r_1 \begin{pmatrix} x_1/2 \\ 0 \\ x_1+y_1 \\ 3y_1 \end{pmatrix} + r_2 \begin{pmatrix} x_2/2 \\ 0 \\ x_2+y_2 \\ 3y_2 \end{pmatrix}$$

However, some things that hold for isomorphisms fail to hold for homomorphisms. One example is in the proof of Lemma I.2.4, which shows that an isomorphism between spaces gives a correspondence between their bases. Homomorphisms do not give any such correspondence; Example 1.2 shows this and another example is the zero map between two nontrivial spaces. Instead, for homomorphisms we have a weaker but still very useful result.

1.9 Theorem A homomorphism is determined by its action on a basis: if V is a vector space with basis $\langle \vec{\beta}_1, \ldots, \vec{\beta}_n \rangle$, if W is a vector space, and if $\vec{w}_1, \ldots, \vec{w}_n \in W$ (these codomain elements need not be distinct) then there exists a homomorphism from V to W sending each $\vec{\beta}_i$ to $\vec{w}_i$, and that homomorphism is unique.

Proof For any input $\vec{v} \in V$ let its expression with respect to the basis be $\vec{v} = c_1\vec{\beta}_1 + \cdots + c_n\vec{\beta}_n$. Define the associated output by using the same coordinates $h(\vec{v}) = c_1\vec{w}_1 + \cdots + c_n\vec{w}_n$. This is well defined because, with respect to the basis, the representation of each domain vector $\vec{v}$ is unique.

This map is a homomorphism because it preserves linear combinations: where $\vec{v_1} = c_1\vec{\beta}_1 + \cdots + c_n\vec{\beta}_n$ and $\vec{v_2} = d_1\vec{\beta}_1 + \cdots + d_n\vec{\beta}_n$, here is the calculation.

$$\begin{aligned} h(r_1\vec{v}_1 + r_2\vec{v}_2) &= h(\,(r_1c_1 + r_2d_1)\vec{\beta}_1 + \cdots + (r_1c_n + r_2d_n)\vec{\beta}_n\,) \\ &= (r_1c_1 + r_2d_1)\vec{w}_1 + \cdots + (r_1c_n + r_2d_n)\vec{w}_n \\ &= r_1h(\vec{v}_1) + r_2h(\vec{v}_2) \end{aligned}$$

This map is unique because if $\hat{h}\colon V \to W$ is another homomorphism satisfying that $\hat{h}(\vec{\beta}_i) = \vec{w}_i$ for each i then h and $\hat{h}$ have the same effect on all of the vectors in the domain.

$$\begin{aligned} \hat{h}(\vec{v}) = \hat{h}(c_1\vec{\beta}_1 + \cdots + c_n\vec{\beta}_n) &= c_1\hat{h}(\vec{\beta}_1) + \cdots + c_n\hat{h}(\vec{\beta}_n) \\ &= c_1\vec{w}_1 + \cdots + c_n\vec{w}_n = h(\vec{v}) \end{aligned}$$

They have the same action so they are the same function. QED

1.10 Definition Let V and W be vector spaces and let $B = \langle\vec{\beta}_1, \ldots, \vec{\beta}_n\rangle$ be a basis for V. A function defined on that basis $f\colon B \to W$ is *extended linearly* to a function $\hat{f}\colon V \to W$ if for all $\vec{v} \in V$ such that $\vec{v} = c_1\vec{\beta}_1 + \cdots + c_n\vec{\beta}_n$, the action of the map is $\hat{f}(\vec{v}) = c_1 \cdot f(\vec{\beta}_1) + \cdots + c_n \cdot f(\vec{\beta}_n)$.

1.11 Example If we specify a map $h\colon \mathbb{R}^2 \to \mathbb{R}^2$ that acts on the standard basis $\mathcal{E}_2$ in this way

$$h(\begin{pmatrix} 1 \\ 0 \end{pmatrix}) = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \qquad h(\begin{pmatrix} 0 \\ 1 \end{pmatrix}) = \begin{pmatrix} -4 \\ 4 \end{pmatrix}$$

then we have also specified the action of h on any other member of the domain. For instance, the value of h on this argument

$$h(\begin{pmatrix} 3 \\ -2 \end{pmatrix}) = h(3 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} - 2 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}) = 3 \cdot h(\begin{pmatrix} 1 \\ 0 \end{pmatrix}) - 2 \cdot h(\begin{pmatrix} 0 \\ 1 \end{pmatrix}) = \begin{pmatrix} 5 \\ -5 \end{pmatrix}$$

is a direct consequence of the value of h on the basis vectors.

Later in this chapter we shall develop a convenient scheme for computations like this one, using matrices.

1.12 Definition A linear map from a space into itself $t: V \to V$ is a *linear transformation*.

1.13 Remark In this book we use 'linear transformation' only in the case where the codomain equals the domain. However, be aware that other sources may instead use it as a synonym for 'homomorphism'.

1.14 Example The map on $\mathbb{R}^2$ that projects all vectors down to the x-axis is a linear transformation.

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ 0 \end{pmatrix}$$

1.15 Example The derivative map $d/dx: \mathcal{P}_n \to \mathcal{P}_n$

$$a_0 + a_1 x + \cdots + a_n x^n \xmapsto{d/dx} a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1}$$

is a linear transformation as this result from calculus shows: $d(c_1 f + c_2 g)/dx = c_1\,(df/dx) + c_2\,(dg/dx)$.

1.16 Example The matrix transpose operation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

is a linear transformation of $\mathcal{M}_{2\times 2}$. (Transpose is one-to-one and onto and so is in fact an automorphism.)

We finish this subsection about maps by recalling that we can linearly combine maps. For instance, for these maps from $\mathbb{R}^2$ to itself

$$\begin{pmatrix} x \\ y \end{pmatrix} \xmapsto{f} \begin{pmatrix} 2x \\ 3x - 2y \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x \\ y \end{pmatrix} \xmapsto{g} \begin{pmatrix} 0 \\ 5x \end{pmatrix}$$

the linear combination $5f - 2g$ is also a transformation of R^2.

$$\begin{pmatrix} x \\ y \end{pmatrix} \xmapsto{5f-2g} \begin{pmatrix} 10x \\ 5x - 10y \end{pmatrix}$$

1.17 Lemma For vector spaces V and W, the set of linear functions from V to W is itself a vector space, a subspace of the space of all functions from V to W.

We denote the space of linear maps from V to W by $\mathcal{L}(V, W)$.

Proof This set is non-empty because it contains the zero homomorphism. So to show that it is a subspace we need only check that it is closed under the operations. Let $f, g\colon V \to W$ be linear. Then the operation of function addition is preserved

$$\begin{aligned}(f+g)(c_1\vec{v}_1 + c_2\vec{v}_2) &= f(c_1\vec{v}_1 + c_2\vec{v}_2) + g(c_1\vec{v}_1 + c_2\vec{v}_2)\\ &= c_1 f(\vec{v}_1) + c_2 f(\vec{v}_2) + c_1 g(\vec{v}_1) + c_2 g(\vec{v}_2)\\ &= c_1\bigl(f+g\bigr)(\vec{v}_1) + c_2\bigl(f+g\bigr)(\vec{v}_2)\end{aligned}$$

as is the operation of scalar multiplication of a function.

$$\begin{aligned}(r\cdot f)(c_1\vec{v}_1 + c_2\vec{v}_2) &= r(c_1 f(\vec{v}_1) + c_2 f(\vec{v}_2))\\ &= c_1(r\cdot f)(\vec{v}_1) + c_2(r\cdot f)(\vec{v}_2)\end{aligned}$$

Hence $\mathcal{L}(V, W)$ is a subspace. QED

We started this section by defining 'homomorphism' as a generalization of 'isomorphism', by isolating the structure preservation property. Some of the points about isomorphisms carried over unchanged, while we adapted others.

Note, however, that the idea of 'homomorphism' is in no way somehow secondary to that of 'isomorphism'. In the rest of this chapter we shall work mostly with homomorphisms. This is partly because any statement made about homomorphisms is automatically true about isomorphisms but more because, while the isomorphism concept is more natural, our experience will show that the homomorphism concept is more fruitful and more central to progress.

Exercises

✓ **1.18** Decide if each $h\colon \mathbb{R}^3 \to \mathbb{R}^2$ is linear.

(a) $h(\begin{pmatrix}x\\y\\z\end{pmatrix}) = \begin{pmatrix}x\\x+y+z\end{pmatrix}$ **(b)** $h(\begin{pmatrix}x\\y\\z\end{pmatrix}) = \begin{pmatrix}0\\0\end{pmatrix}$ **(c)** $h(\begin{pmatrix}x\\y\\z\end{pmatrix}) = \begin{pmatrix}1\\1\end{pmatrix}$

(d) $h(\begin{pmatrix}x\\y\\z\end{pmatrix}) = \begin{pmatrix}2x+y\\3y-4z\end{pmatrix}$

✓ **1.19** Decide if each map $h\colon \mathcal{M}_{2\times 2} \to \mathbb{R}$ is linear.

(a) $h(\begin{pmatrix}a&b\\c&d\end{pmatrix}) = a+d$

(b) $h(\begin{pmatrix}a&b\\c&d\end{pmatrix}) = ad-bc$

(c) $h(\begin{pmatrix}a&b\\c&d\end{pmatrix}) = 2a+3b+c-d$

(d) $h(\begin{pmatrix}a&b\\c&d\end{pmatrix}) = a^2+b^2$

✓ **1.20** Show that these two maps are homomorphisms.

(a) $d/dx\colon \mathcal{P}_3 \to \mathcal{P}_2$ given by $a_0 + a_1x + a_2x^2 + a_3x^3$ maps to $a_1 + 2a_2x + 3a_3x^2$

(b) $\int\colon \mathcal{P}_2 \to \mathcal{P}_3$ given by $b_0 + b_1x + b_2x^2$ maps to $b_0x + (b_1/2)x^2 + (b_2/3)x^3$

Are these maps inverse to each other?

1.21 Is (perpendicular) projection from $\mathbb{R}^3$ to the xz-plane a homomorphism? Projection to the yz-plane? To the x-axis? The y-axis? The z-axis? Projection to the origin?

1.22 Show that, while the maps from Example 1.3 preserve linear operations, they are not isomorphisms.

1.23 Is an identity map a linear transformation?

✓ 1.24 Stating that a function is 'linear' is different than stating that its graph is a line.

(a) The function $f_1 \colon \mathbb{R} \to \mathbb{R}$ given by $f_1(x) = 2x - 1$ has a graph that is a line. Show that it is not a linear function.

(b) The function $f_2 \colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto x + 2y$$

does not have a graph that is a line. Show that it is a linear function.

✓ 1.25 Part of the definition of a linear function is that it respects addition. Does a linear function respect subtraction?

1.26 Assume that h is a linear transformation of V and that $\langle \vec{\beta}_1, \ldots, \vec{\beta}_n \rangle$ is a basis of V. Prove each statement.

(a) If $h(\vec{\beta}_i) = \vec{0}$ for each basis vector then h is the zero map.

(b) If $h(\vec{\beta}_i) = \vec{\beta}_i$ for each basis vector then h is the identity map.

(c) If there is a scalar r such that $h(\vec{\beta}_i) = r \cdot \vec{\beta}_i$ for each basis vector then $h(\vec{v}) = r \cdot \vec{v}$ for all vectors in V.

✓ 1.27 Consider the vector space $\mathbb{R}^+$ where vector addition and scalar multiplication are not the ones inherited from $\mathbb{R}$ but rather are these: $a + b$ is the product of a and b, and $r \cdot a$ is the r-th power of a. (This was shown to be a vector space in an earlier exercise.) Verify that the natural logarithm map $\ln \colon \mathbb{R}^+ \to \mathbb{R}$ is a homomorphism between these two spaces. Is it an isomorphism?

✓ 1.28 Consider this transformation of $\mathbb{R}^2$.

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x/2 \\ y/3 \end{pmatrix}$$

Find the image under this map of this ellipse.

$$\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid (x^2/4) + (y^2/9) = 1 \}$$

✓ 1.29 Imagine a rope wound around the earth's equator so that it fits snugly (suppose that the earth is a sphere). How much extra rope must we add to raise the circle to a constant six feet off the ground?

✓ 1.30 Verify that this map $h \colon \mathbb{R}^3 \to \mathbb{R}$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -1 \\ -1 \end{pmatrix} = 3x - y - z$$

is linear. Generalize.

1.31 Show that every homomorphism from $\mathbb{R}^1$ to $\mathbb{R}^1$ acts via multiplication by a scalar. Conclude that every nontrivial linear transformation of $\mathbb{R}^1$ is an isomorphism. Is that true for transformations of $\mathbb{R}^2$? $\mathbb{R}^n$?

1.32 (a) Show that for any scalars $a_{1,1},\ldots,a_{m,n}$ this map $h\colon\mathbb{R}^n\to\mathbb{R}^m$ is a homomorphism.

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} a_{1,1}x_1+\cdots+a_{1,n}x_n \\ \vdots \\ a_{m,1}x_1+\cdots+a_{m,n}x_n \end{pmatrix}$$

(b) Show that for each i, the i-th derivative operator d^i/dx^i is a linear transformation of $\mathcal{P}_n$. Conclude that for any scalars $c_k,\ldots,c_0$ this map is a linear transformation of that space.

$$f\mapsto \frac{d^k}{dx^k}f+c_{k-1}\frac{d^{k-1}}{dx^{k-1}}f+\cdots+c_1\frac{d}{dx}f+c_0f$$

1.33 Lemma 1.17 shows that a sum of linear functions is linear and that a scalar multiple of a linear function is linear. Show also that a composition of linear functions is linear.

✓ 1.34 Where $f\colon V\to W$ is linear, suppose that $f(\vec{v}_1)=\vec{w}_1,\ldots,f(\vec{v}_n)=\vec{w}_n$ for some vectors $\vec{w}_1,\ldots,\vec{w}_n$ from W.

(a) If the set of $\vec{w}\,$'s is independent, must the set of $\vec{v}\,$'s also be independent?

(b) If the set of $\vec{v}\,$'s is independent, must the set of $\vec{w}\,$'s also be independent?

(c) If the set of $\vec{w}\,$'s spans W, must the set of $\vec{v}\,$'s span V?

(d) If the set of $\vec{v}\,$'s spans V, must the set of $\vec{w}\,$'s span W?

1.35 Generalize Example 1.16 by proving that for every appropriate domain and codomain the matrix transpose map is linear. What are the appropriate domains and codomains?

1.36 (a) Where $\vec{u},\vec{v}\in\mathbb{R}^n$, by definition the line segment connecting them is the set $\ell=\{t\cdot\vec{u}+(1-t)\cdot\vec{v}\mid t\in[0..1]\}$. Show that the image, under a homomorphism h, of the segment between $\vec{u}$ and $\vec{v}$ is the segment between $h(\vec{u})$ and $h(\vec{v})$.

(b) A subset of $\mathbb{R}^n$ is *convex* if, for any two points in that set, the line segment joining them lies entirely in that set. (The inside of a sphere is convex while the skin of a sphere is not.) Prove that linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ preserve the property of set convexity.

✓ 1.37 Let $h\colon\mathbb{R}^n\to\mathbb{R}^m$ be a homomorphism.

(a) Show that the image under h of a line in $\mathbb{R}^n$ is a (possibly degenerate) line in $\mathbb{R}^m$.

(b) What happens to a k-dimensional linear surface?

1.38 Prove that the restriction of a homomorphism to a subspace of its domain is another homomorphism.

1.39 Assume that $h\colon V\to W$ is linear.

(a) Show that the *range space* of this map $\{h(\vec{v})\mid\vec{v}\in V\}$ is a subspace of the codomain W.

(b) Show that the *null space* of this map $\{\vec{v}\in V\mid h(\vec{v})=\vec{0}_W\}$ is a subspace of the domain V.

(c) Show that if U is a subspace of the domain V then its image $\{h(\vec{u}) \mid \vec{u} \in U\}$ is a subspace of the codomain W. This generalizes the first item.

(d) Generalize the second item.

1.40 Consider the set of isomorphisms from a vector space to itself. Is this a subspace of the space $\mathcal{L}(V, V)$ of homomorphisms from the space to itself?

1.41 Does Theorem 1.9 need that $\langle\vec{\beta}_1, \ldots, \vec{\beta}_n\rangle$ is a basis? That is, can we still get a well-defined and unique homomorphism if we drop either the condition that the set of $\vec{\beta}$'s be linearly independent, or the condition that it span the domain?

1.42 Let V be a vector space and assume that the maps $f_1, f_2 \colon V \to \mathbb{R}^1$ are linear.

(a) Define a map $F \colon V \to \mathbb{R}^2$ whose component functions are the given linear ones.

$$\vec{v} \mapsto \begin{pmatrix} f_1(\vec{v}) \\ f_2(\vec{v}) \end{pmatrix}$$

Show that F is linear.

(b) Does the converse hold — is any linear map from V to $\mathbb{R}^2$ made up of two linear component maps to $\mathbb{R}^1$?

(c) Generalize.

II.2 Range space and Null space

Isomorphisms and homomorphisms both preserve structure. The difference is that homomorphisms have fewer restrictions, since they needn't be onto and needn't be one-to-one. We will examine what can happen with homomorphisms that cannot happen with isomorphisms.

First consider the fact that homomorphisms need not be onto. Of course, each function is onto some set, namely its range. For example, the injection map $\iota \colon \mathbb{R}^2 \to \mathbb{R}^3$

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

is a homomorphism, and is not onto $\mathbb{R}^3$. But it is onto the xy-plane.

2.1 Lemma Under a homomorphism, the image of any subspace of the domain is a subspace of the codomain. In particular, the image of the entire space, the range of the homomorphism, is a subspace of the codomain.

Proof Let $h \colon V \to W$ be linear and let S be a subspace of the domain V. The image $h(S)$ is a subset of the codomain W, which is nonempty because S is nonempty. Thus, to show that $h(S)$ is a subspace of W we need only show that it is closed under linear combinations of two vectors. If $h(\vec{s}_1)$ and $h(\vec{s}_2)$ are members of $h(S)$ then $c_1 \cdot h(\vec{s}_1) + c_2 \cdot h(\vec{s}_2) = h(c_1 \cdot \vec{s}_1) + h(c_2 \cdot \vec{s}_2) = h(c_1 \cdot \vec{s}_1 + c_2 \cdot \vec{s}_2)$ is also a member of $h(S)$ because it is the image of $c_1 \cdot \vec{s}_1 + c_2 \cdot \vec{s}_2$ from S. QED

2.2 Definition The *range space* of a homomorphism $h\colon V \to W$ is

$$\mathscr{R}(h) = \{h(\vec{v}) \mid \vec{v} \in V\}$$

sometimes denoted $h(V)$. The dimension of the range space is the map's *rank*.

We shall soon see the connection between the rank of a map and the rank of a matrix.

2.3 Example For the derivative map $d/dx\colon \mathcal{P}_3 \to \mathcal{P}_3$ given by $a_0 + a_1x + a_2x^2 + a_3x^3 \mapsto a_1 + 2a_2x + 3a_3x^2$ the range space $\mathscr{R}(d/dx)$ is the set of quadratic polynomials $\{r + sx + tx^2 \mid r, s, t \in \mathbb{R}\}$. Thus, this map's rank is 3.

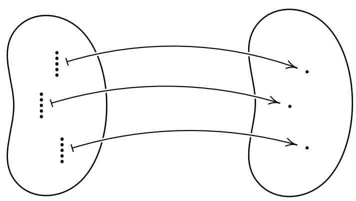
2.4 Example With this homomorphism $h\colon \mathcal{M}_{2\times 2} \to \mathcal{P}_3$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto (a + b + 2d) + cx^2 + cx^3$$

an image vector in the range can have any constant term, must have an x coefficient of zero, and must have the same coefficient of x^2 as of x^3. That is, the range space is $\mathscr{R}(h) = \{r + sx^2 + sx^3 \mid r, s \in \mathbb{R}\}$ and so the rank is 2.

The prior result shows that, in passing from the definition of isomorphism to the more general definition of homomorphism, omitting the onto requirement doesn't make an essential difference. Any homomorphism is onto some space, namely its range.

However, omitting the one-to-one condition does make a difference. A homomorphism may have many elements of the domain that map to one element of the codomain. Below is a bean sketch of a many-to-one map between sets.* It shows three elements of the codomain that are each the image of many members of the domain. (Rather than picture lots of individual $\mapsto$ arrows, each association of many inputs with one output shows only one such arrow.)



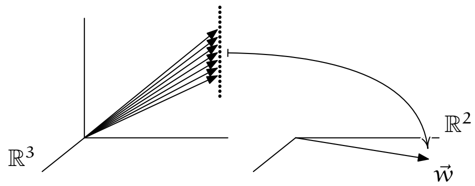
* More information on many-to-one maps is in the appendix.

Recall that for any function $h\colon V \to W$, the set of elements of V that map to $\vec{w} \in W$ is the *inverse image* $h^{-1}(\vec{w}) = \{\vec{v} \in V \mid h(\vec{v}) = \vec{w}\}$. Above, the left side shows three inverse image sets.

2.5 Example Consider the projection $\pi\colon \mathbb{R}^3 \to \mathbb{R}^2$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \overset{\pi}{\longmapsto} \begin{pmatrix} x \\ y \end{pmatrix}$$

which is a homomorphism that is many-to-one. An inverse image set is a vertical line of vectors in the domain.



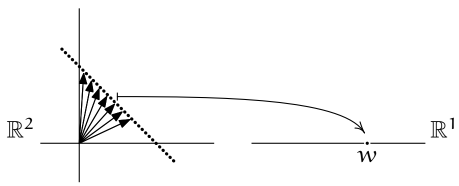
One example is this.

$$\pi^{-1}(\begin{pmatrix} 1 \\ 3 \end{pmatrix}) = \{ \begin{pmatrix} 1 \\ 3 \\ z \end{pmatrix} \mid z \in \mathbb{R}\}$$

2.6 Example This homomorphism $h\colon \mathbb{R}^2 \to \mathbb{R}^1$

$$\begin{pmatrix} x \\ y \end{pmatrix} \overset{h}{\longmapsto} x + y$$

is also many-to-one. For a fixed $w \in \mathbb{R}^1$ the inverse image $h^{-1}(w)$

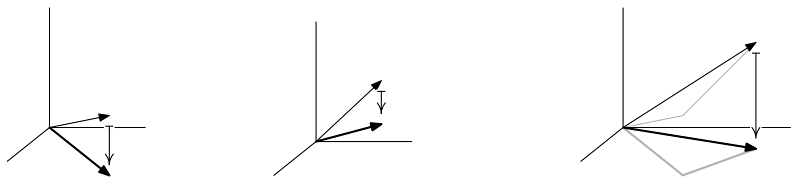


is the set of plane vectors whose components add to w.

In generalizing from isomorphisms to homomorphisms by dropping the one-to-one condition we lose the property that, intuitively, the domain is "the same" as the range. We lose, that is, that the domain corresponds perfectly to the range. The examples below illustrate that what we retain is that a homomorphism describes how the domain is "analogous to" or "like" the range.

2.7 Example We think of $\mathbb{R}^3$ as like $\mathbb{R}^2$ except that vectors have an extra component. That is, we think of the vector with components x, y, and z as like the vector with components x and y. Defining the projection map π makes precise which members of the domain we are thinking of as related to which members of the codomain.

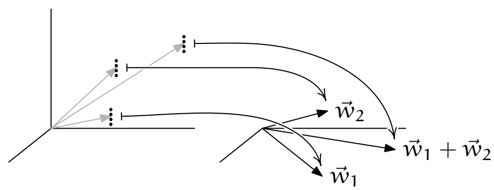
To understanding how the preservation conditions in the definition of homomorphism show that the domain elements are like the codomain elements, start by picturing $\mathbb{R}^2$ as the xy-plane inside of $\mathbb{R}^3$ (the xy plane inside of $\mathbb{R}^3$ is a set of three-tall vectors with a third component of zero and so does not precisely equal the set of two-tall vectors $\mathbb{R}^2$, but this embedding makes the picture much clearer). The preservation of addition property says that vectors in $\mathbb{R}^3$ act like their shadows in the plane.



$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} \text{ above } \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \text{ plus } \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} \text{ above } \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \text{ equals } \begin{pmatrix} x_1 + y_1 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix} \text{ above } \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}$$

Thinking of $\pi(\vec{v})$ as the "shadow" of $\vec{v}$ in the plane gives this restatement: the sum of the shadows $\pi(\vec{v}_1) + \pi(\vec{v}_2)$ equals the shadow of the sum $\pi(\vec{v}_1 + \vec{v}_2)$. Preservation of scalar multiplication is similar.

Drawing the codomain $\mathbb{R}^2$ on the right gives a picture that is uglier but is more faithful to the bean sketch above.



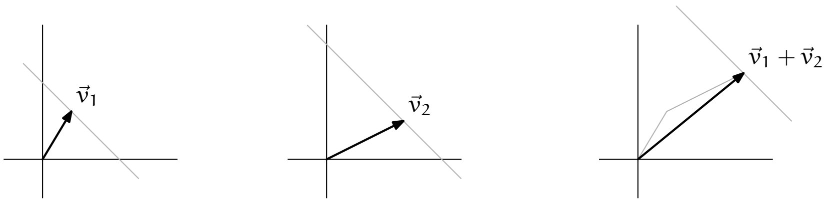
Again, the domain vectors that map to $\vec{w}_1$ lie in a vertical line; one is drawn, in gray. Call any member of this inverse image $\pi^{-1}(\vec{w}_1)$ a "$\vec{w}_1$ vector." Similarly, there is a vertical line of "$\vec{w}_2$ vectors" and a vertical line of "$\vec{w}_1 + \vec{w}_2$ vectors." Now, saying that π is a homomorphism is recognizing that if $\pi(\vec{v}_1) = \vec{w}_1$ and $\pi(\vec{v}_2) = \vec{w}_2$ then $\pi(\vec{v}_1 + \vec{v}_2) = \pi(\vec{v}_1) + \pi(\vec{v}_2) = \vec{w}_1 + \vec{w}_2$. That is, the classes add: any $\vec{w}_1$ vector plus any $\vec{w}_2$ vector equals a $\vec{w}_1 + \vec{w}_2$ vector. Scalar multiplication is similar.

So although $\mathbb{R}^3$ and $\mathbb{R}^2$ are not isomorphic π describes a way in which they are alike: vectors in $\mathbb{R}^3$ add as do the associated vectors in $\mathbb{R}^2$ — vectors add as their shadows add.

2.8 Example A homomorphism can express an analogy between spaces that is more subtle than the prior one. For the map from Example 2.6

$$\begin{pmatrix} x \\ y \end{pmatrix} \stackrel{h}{\longmapsto} x + y$$

fix two numbers in the range $w_1, w_2 \in \mathbb{R}$. A $\vec{v}_1$ that maps to w_1 has components that add to w_1, so the inverse image $h^{-1}(w_1)$ is the set of vectors with endpoint on the diagonal line $x + y = w_1$. Think of these as "w_1 vectors." Similarly we have "w_2 vectors" and "$w_1 + w_2$ vectors." The addition preservation property says this.



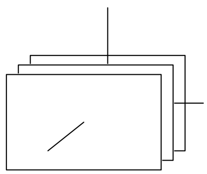
a "w_1 vector" plus a "w_2 vector" equals a "$w_1 + w_2$ vector"

Restated, if we add a w_1 vector to a w_2 vector then h maps the result to a $w_1 + w_2$ vector. Briefly, the sum of the images is the image of the sum. Even more briefly, $h(\vec{v}_1) + h(\vec{v}_2) = h(\vec{v}_1 + \vec{v}_2)$.

2.9 Example The inverse images can be structures other than lines. For the linear map $h\colon \mathbb{R}^3 \to \mathbb{R}^2$

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x \\ x \end{pmatrix}$$

the inverse image sets are planes $x = 0$, $x = 1$, etc., perpendicular to the x-axis.



We won't describe how every homomorphism that we will use is an analogy because the formal sense that we make of "alike in that ..." is 'a homomorphism exists such that ...'. Nonetheless, the idea that a homomorphism between two spaces expresses how the domain's vectors fall into classes that act like the range's vectors is a good way to view homomorphisms.

Another reason that we won't treat all of the homomorphisms that we see as above is that many vector spaces are hard to draw, e.g., a space of polynomials. But there is nothing wrong with leveraging spaces that we can draw: from the three examples 2.7, 2.8, and 2.9 we draw two insights.

The first insight is that in all three examples the inverse image of the range's zero vector is a line or plane through the origin. It is therefore a subspace of the domain.

2.10 Lemma For any homomorphism the inverse image of a subspace of the range is a subspace of the domain. In particular, the inverse image of the trivial subspace of the range is a subspace of the domain.

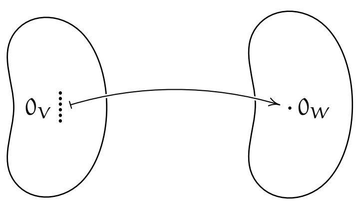
(The examples above consider inverse images of single vectors but this result is about inverse images of sets $h^{-1}(S) = \{\vec{v} \in V \mid h(\vec{v}) \in S\}$. We use the same term for both by taking the inverse image of a single element $h^{-1}(\vec{w})$ to be the inverse image of the one-element set $h^{-1}(\{\vec{w}\})$.)

Proof Let $h\colon V \to W$ be a homomorphism and let S be a subspace of the range space of h. Consider the inverse image of S. It is nonempty because it contains $\vec{0}_V$, since $h(\vec{0}_V) = \vec{0}_W$ and $\vec{0}_W$ is an element of S as S is a subspace. To finish we show that $h^{-1}(S)$ is closed under linear combinations. Let $\vec{v}_1$ and $\vec{v}_2$ be two of its elements, so that $h(\vec{v}_1)$ and $h(\vec{v}_2)$ are elements of S. Then $c_1\vec{v}_1 + c_2\vec{v}_2$ is an element of the inverse image $h^{-1}(S)$ because $h(c_1\vec{v}_1 + c_2\vec{v}_2) = c_1h(\vec{v}_1) + c_2h(\vec{v}_2)$ is a member of S. QED

2.11 Definition The *null space* or *kernel* of a linear map $h\colon V \to W$ is the inverse image of $\vec{0}_W$.

$$\mathscr{N}(h) = h^{-1}(\vec{0}_W) = \{\vec{v} \in V \mid h(\vec{v}) = \vec{0}_W\}$$

The dimension of the null space is the map's *nullity*.



2.12 Example The map from Example 2.3 has this null space $\mathscr{N}(d/dx) = \{a_0 + 0x + 0x^2 + 0x^3 \mid a_0 \in \mathbb{R}\}$ so its nullity is 1.

2.13 Example The map from Example 2.4 has this null space, and nullity 2.

$$\mathscr{N}(h)=\{\begin{pmatrix}a&b\\0&-(a+b)/2\end{pmatrix}\mid a,b\in\mathbb{R}\}$$

Now for the second insight from the above examples. In Example 2.7 each of the vertical lines squashes down to a single point — in passing from the domain to the range, π takes all of these one-dimensional vertical lines and maps them to a point, leaving the range smaller than the domain by one dimension. Similarly, in Example 2.8 the two-dimensional domain compresses to a one-dimensional range by breaking the domain into the diagonal lines and maps each of those to a single member of the range. Finally, in Example 2.9 the domain breaks into planes which get squashed to a point and so the map starts with a three-dimensional domain but ends two smaller, with a one-dimensional range. (The codomain is two-dimensional but the range is one-dimensional and the dimension of the range is what matters.)

2.14 Theorem A linear map's rank plus its nullity equals the dimension of its domain.

Proof Let $h\colon V\to W$ be linear and let $B_N=\langle\vec{\beta}_1,\ldots,\vec{\beta}_k\rangle$ be a basis for the null space. Expand that to a basis $B_V=\langle\vec{\beta}_1,\ldots,\vec{\beta}_k,\vec{\beta}_{k+1},\ldots,\vec{\beta}_n\rangle$ for the entire domain, using Corollary Two.III.2.13. We shall show that $B_R=\langle h(\vec{\beta}_{k+1}),\ldots,h(\vec{\beta}_n)\rangle$ is a basis for the range space. Then counting the size of the bases gives the result.

To see that B_R is linearly independent, consider $\vec{0}_W=c_{k+1}h(\vec{\beta}_{k+1})+\cdots+c_nh(\vec{\beta}_n)$. We have $\vec{0}_W=h(c_{k+1}\vec{\beta}_{k+1}+\cdots+c_n\vec{\beta}_n)$ and so $c_{k+1}\vec{\beta}_{k+1}+\cdots+c_n\vec{\beta}_n$ is in the null space of h. As B_N is a basis for the null space there are scalars $c_1,\ldots,c_k$ satisfying this relationship.

$$c_1\vec{\beta}_1+\cdots+c_k\vec{\beta}_k=c_{k+1}\vec{\beta}_{k+1}+\cdots+c_n\vec{\beta}_n$$

But this is an equation among members of B_V, which is a basis for V, so each c_i equals 0. Therefore B_R is linearly independent.

To show that B_R spans the range space consider a member of the range space $h(\vec{v})$. Express $\vec{v}$ as a linear combination $\vec{v}=c_1\vec{\beta}_1+\cdots+c_n\vec{\beta}_n$ of members of B_V. This gives $h(\vec{v})=h(c_1\vec{\beta}_1+\cdots+c_n\vec{\beta}_n)=c_1h(\vec{\beta}_1)+\cdots+c_kh(\vec{\beta}_k)+c_{k+1}h(\vec{\beta}_{k+1})+\cdots+c_nh(\vec{\beta}_n)$ and since $\vec{\beta}_1$, ..., $\vec{\beta}_k$ are in the null space, we have that $h(\vec{v})=\vec{0}+\cdots+\vec{0}+c_{k+1}h(\vec{\beta}_{k+1})+\cdots+c_nh(\vec{\beta}_n)$. Thus, $h(\vec{v})$ is a linear combination of members of B_R, and so B_R spans the range space. QED

2.15 Example Where $h\colon \mathbb{R}^3 \to \mathbb{R}^4$ is

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \stackrel{h}{\longmapsto} \begin{pmatrix} x \\ 0 \\ y \\ 0 \end{pmatrix}$$

the range space and null space are

$$\mathscr{R}(h) = \{\begin{pmatrix} a \\ 0 \\ b \\ 0 \end{pmatrix} \mid a, b \in \mathbb{R}\} \quad\text{and}\quad \mathscr{N}(h) = \{\begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix} \mid z \in \mathbb{R}\}$$

and so the rank of h is 2 while the nullity is 1.

2.16 Example If $t\colon \mathbb{R} \to \mathbb{R}$ is the linear transformation $x \mapsto -4x$, then the range is $\mathscr{R}(t) = \mathbb{R}$. The rank is 1 and the nullity is 0.

2.17 Corollary The rank of a linear map is less than or equal to the dimension of the domain. Equality holds if and only if the nullity of the map is 0.

We know that an isomorphism exists between two spaces if and only if the dimension of the range equals the dimension of the domain. We have now seen that for a homomorphism to exist a necessary condition is that the dimension of the range must be less than or equal to the dimension of the domain. For instance, there is no homomorphism from $\mathbb{R}^2$ onto $\mathbb{R}^3$. There are many homomorphisms from $\mathbb{R}^2$ into $\mathbb{R}^3$, but none onto.

The range space of a linear map can be of dimension strictly less than the dimension of the domain and so linearly independent sets in the domain may map to linearly dependent sets in the range. (Example 2.3's derivative transformation on $\mathcal{P}_3$ has a domain of dimension 4 but a range of dimension 3 and the derivative sends $\{1, x, x^2, x^3\}$ to $\{0, 1, 2x, 3x^2\}$). That is, under a homomorphism independence may be lost. In contrast, dependence stays.

2.18 Lemma Under a linear map, the image of a linearly dependent set is linearly dependent.

Proof Suppose that $c_1\vec{v}_1 + \cdots + c_n\vec{v}_n = \vec{0}_V$ with some c_i nonzero. Apply h to both sides: $h(c_1\vec{v}_1 + \cdots + c_n\vec{v}_n) = c_1h(\vec{v}_1) + \cdots + c_nh(\vec{v}_n)$ and $h(\vec{0}_V) = \vec{0}_W$. Thus we have $c_1h(\vec{v}_1) + \cdots + c_nh(\vec{v}_n) = \vec{0}_W$ with some c_i nonzero. QED

When is independence not lost? The obvious sufficient condition is when the homomorphism is an isomorphism. This condition is also necessary; see

Exercise 34. We will finish this subsection comparing homomorphisms with isomorphisms by observing that a one-to-one homomorphism is an isomorphism from its domain onto its range.

2.19 Example This one-to-one homomorphism $\iota\colon \mathbb{R}^2 \to \mathbb{R}^3$

$$\begin{pmatrix} x \\ y \end{pmatrix} \overset{\iota}{\longmapsto} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

gives a correspondence between $\mathbb{R}^2$ and the xy-plane subset of $\mathbb{R}^3$.

2.20 Theorem Where V is an n-dimensional vector space, these are equivalent statements about a linear map $h\colon V \to W$.

(1) h is one-to-one
(2) h has an inverse from its range to its domain that is a linear map
(3) $\mathscr{N}(h) = \{\vec{0}\,\}$, that is, $\text{nullity}(h) = 0$
(4) $\text{rank}(h) = n$
(5) if $\langle \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$ is a basis for V then $\langle h(\vec{\beta}_1), \dots, h(\vec{\beta}_n) \rangle$ is a basis for $\mathscr{R}(h)$

Proof We will first show that (1) $\iff$ (2). We will then show that (1) $\implies$ (3) $\implies$ (4) $\implies$ (5) $\implies$ (2).

For (1) $\implies$ (2), suppose that the linear map h is one-to-one, and therefore has an inverse $h^{-1}\colon \mathscr{R}(h) \to V$. The domain of that inverse is the range of h and thus a linear combination of two members of it has the form $c_1h(\vec{v}_1) + c_2h(\vec{v}_2)$. On that combination, the inverse h^{-1} gives this.

$$\begin{aligned} h^{-1}(c_1h(\vec{v}_1) + c_2h(\vec{v}_2)) &= h^{-1}(h(c_1\vec{v}_1 + c_2\vec{v}_2)) \\ &= h^{-1} \circ h\ (c_1\vec{v}_1 + c_2\vec{v}_2) \\ &= c_1\vec{v}_1 + c_2\vec{v}_2 \\ &= c_1 \cdot h^{-1}(h(\vec{v}_1)) + c_2 \cdot h^{-1}(h(\vec{v}_2)) \end{aligned}$$

Thus if a linear map has an inverse then the inverse must be linear. But this also gives the (2) $\implies$ (1) implication, because the inverse itself must be one-to-one.

Of the remaining implications, (1) $\implies$ (3) holds because any homomorphism maps $\vec{0}_V$ to $\vec{0}_W$, but a one-to-one map sends at most one member of V to $\vec{0}_W$.

Next, (3) $\implies$ (4) is true since rank plus nullity equals the dimension of the domain.

For (4) $\implies$ (5), to show that $\langle h(\vec{\beta}_1), \dots, h(\vec{\beta}_n) \rangle$ is a basis for the range space we need only show that it is a spanning set, because by assumption the range has dimension n. Consider $h(\vec{v}) \in \mathscr{R}(h)$. Expressing $\vec{v}$ as a linear combination of basis elements produces $h(\vec{v}) = h(c_1\vec{\beta}_1 + c_2\vec{\beta}_2 + \cdots + c_n\vec{\beta}_n)$, which gives that $h(\vec{v}) = c_1h(\vec{\beta}_1) + \cdots + c_nh(\vec{\beta}_n)$, as desired.

Finally, for the (5) $\implies$ (2) implication, assume that $\langle\vec{\beta}_1, \ldots, \vec{\beta}_n\rangle$ is a basis for V so that $\langle h(\vec{\beta}_1), \ldots, h(\vec{\beta}_n)\rangle$ is a basis for $\mathscr{R}(h)$. Then every $\vec{w} \in \mathscr{R}(h)$ has the unique representation $\vec{w} = c_1h(\vec{\beta}_1) + \cdots + c_nh(\vec{\beta}_n)$. Define a map from $\mathscr{R}(h)$ to V by

$$\vec{w} \;\mapsto\; c_1\vec{\beta}_1 + c_2\vec{\beta}_2 + \cdots + c_n\vec{\beta}_n$$

(uniqueness of the representation makes this well-defined). Checking that it is linear and that it is the inverse of h are easy. QED

We have seen that a linear map expresses how the structure of the domain is like that of the range. We can think of such a map as organizing the domain space into inverse images of points in the range. In the special case that the map is one-to-one, each inverse image is a single point and the map is an isomorphism between the domain and the range.

Exercises

✓ 2.21 Let $h\colon \mathcal{P}_3 \to \mathcal{P}_4$ be given by $p(x) \mapsto x \cdot p(x)$. Which of these are in the null space? Which are in the range space?

(a) x^3 **(b)** 0 **(c)** 7 **(d)** $12x - 0.5x^3$ **(e)** $1 + 3x^2 - x^3$

✓ 2.22 Find the null space, nullity, range space, and rank of each map.

(a) $h\colon \mathbb{R}^2 \to \mathcal{P}_3$ given by

$$\begin{pmatrix} a \\ b \end{pmatrix} \mapsto a + ax + ax^2$$

(b) $h\colon \mathcal{M}_{2\times2} \to \mathbb{R}$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto a + d$$

(c) $h\colon \mathcal{M}_{2\times2} \to \mathcal{P}_2$ given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto a + b + c + dx^2$$

(d) the zero map $Z\colon \mathbb{R}^3 \to \mathbb{R}^4$

✓ 2.23 Find the nullity of each map.

(a) $h\colon \mathbb{R}^5 \to \mathbb{R}^8$ of rank five **(b)** $h\colon \mathcal{P}_3 \to \mathcal{P}_3$ of rank one
(c) $h\colon \mathbb{R}^6 \to \mathbb{R}^3$, an onto map **(d)** $h\colon \mathcal{M}_{3\times3} \to \mathcal{M}_{3\times3}$, onto

✓ 2.24 What is the null space of the differentiation transformation $d/dx\colon \mathcal{P}_n \to \mathcal{P}_n$? What is the null space of the second derivative, as a transformation of $\mathcal{P}_n$? The k-th derivative?

2.25 Example 2.7 restates the first condition in the definition of homomorphism as ‘the shadow of a sum is the sum of the shadows’. Restate the second condition in the same style.

2.26 For the homomorphism $h\colon \mathcal{P}_3 \to \mathcal{P}_3$ given by $h(a_0 + a_1x + a_2x^2 + a_3x^3) = a_0 + (a_0 + a_1)x + (a_2 + a_3)x^3$ find these.

(a) $\mathscr{N}(h)$ **(b)** $h^{-1}(2-x^3)$ **(c)** $h^{-1}(1+x^2)$

✓ 2.27 For the map $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(\begin{pmatrix} x \\ y \end{pmatrix}) = 2x + y$$

sketch these inverse image sets: $f^{-1}(-3)$, $f^{-1}(0)$, and $f^{-1}(1)$.

✓ 2.28 Each of these transformations of $\mathcal{P}_3$ is one-to-one. For each, find the inverse.

(a) $a_0 + a_1x + a_2x^2 + a_3x^3 \mapsto a_0 + a_1x + 2a_2x^2 + 3a_3x^3$
(b) $a_0 + a_1x + a_2x^2 + a_3x^3 \mapsto a_0 + a_2x + a_1x^2 + a_3x^3$
(c) $a_0 + a_1x + a_2x^2 + a_3x^3 \mapsto a_1 + a_2x + a_3x^2 + a_0x^3$
(d) $a_0+a_1x+a_2x^2+a_3x^3 \mapsto a_0+(a_0+a_1)x+(a_0+a_1+a_2)x^2+(a_0+a_1+a_2+a_3)x^3$

2.29 Describe the null space and range space of a transformation given by $\vec{v} \mapsto 2\vec{v}$.

2.30 List all pairs $(\text{rank}(h), \text{nullity}(h))$ that are possible for linear maps from $\mathbb{R}^5$ to $\mathbb{R}^3$.

2.31 Does the differentiation map $d/dx\colon \mathcal{P}_n \to \mathcal{P}_n$ have an inverse?

✓ 2.32 Find the nullity of this map $h\colon \mathcal{P}_n \to \mathbb{R}$.

$$a_0 + a_1x + \cdots + a_nx^n \mapsto \int_{x=0}^{x=1} a_0 + a_1x + \cdots + a_nx^n \, dx$$

2.33 **(a)** Prove that a homomorphism is onto if and only if its rank equals the dimension of its codomain.
(b) Conclude that a homomorphism between vector spaces with the same dimension is one-to-one if and only if it is onto.

2.34 Show that a linear map is one-to-one if and only if it preserves linear independence.

2.35 Corollary 2.17 says that for there to be an onto homomorphism from a vector space V to a vector space W, it is necessary that the dimension of W be less than or equal to the dimension of V. Prove that this condition is also sufficient; use Theorem 1.9 to show that if the dimension of W is less than or equal to the dimension of V, then there is a homomorphism from V to W that is onto.

✓ 2.36 Recall that the null space is a subset of the domain and the range space is a subset of the codomain. Are they necessarily distinct? Is there a homomorphism that has a nontrivial intersection of its null space and its range space?

2.37 Prove that the image of a span equals the span of the images. That is, where $h\colon V \to W$ is linear, prove that if S is a subset of V then $h([S])$ equals $[h(S)]$. This generalizes Lemma 2.1 since it shows that if U is any subspace of V then its image $\{h(\vec{u}) \mid \vec{u} \in U\}$ is a subspace of W, because the span of the set U is U.

✓ 2.38 **(a)** Prove that for any linear map $h\colon V \to W$ and any $\vec{w} \in W$, the set $h^{-1}(\vec{w})$ has the form

$$\{\vec{v} + \vec{n} \mid \vec{n} \in \mathscr{N}(h)\}$$

for $\vec{v} \in V$ with $h(\vec{v}) = \vec{w}$ (if h is not onto then this set may be empty). Such a set is a *coset* of $\mathscr{N}(h)$ and we denote it as $\vec{v} + \mathscr{N}(h)$.
(b) Consider the map $t\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\begin{pmatrix} x \\ y \end{pmatrix} \overset{t}{\longmapsto} \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}$$

for some scalars a, b, c, and d. Prove that t is linear.

(c) Conclude from the prior two items that for any linear system of the form

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned}$$

we can write the solution set (the vectors are members of $\mathbb{R}^2$)

$$\{\vec{p} + \vec{h} \mid \vec{h} \text{ satisfies the associated homogeneous system}\}$$

where $\vec{p}$ is a particular solution of that linear system (if there is no particular solution then the above set is empty).

(d) Show that this map $h\colon \mathbb{R}^n \to \mathbb{R}^m$ is linear

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} a_{1,1}x_1 + \cdots + a_{1,n}x_n \\ \vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n \end{pmatrix}$$

for any scalars $a_{1,1}$, ..., $a_{m,n}$. Extend the conclusion made in the prior item.

(e) Show that the k-th derivative map is a linear transformation of $\mathcal{P}_n$ for each k. Prove that this map is a linear transformation of the space

$$f \mapsto \frac{d^k}{dx^k}f + c_{k-1}\frac{d^{k-1}}{dx^{k-1}}f + \cdots + c_1\frac{d}{dx}f + c_0 f$$

for any scalars c_k, ..., c_0. Draw a conclusion as above.

2.39 Prove that for any transformation $t\colon V \to V$ that is rank one, the map given by composing the operator with itself $t \circ t\colon V \to V$ satisfies $t \circ t = r \cdot t$ for some real number r.

2.40 Let $h\colon V \to \mathbb{R}$ be a homomorphism, but not the zero homomorphism. Prove that if $\langle \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$ is a basis for the null space and if $\vec{v} \in V$ is not in the null space then $\langle \vec{v}, \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$ is a basis for the entire domain V.

2.41 Show that for any space V of dimension n, the *dual space*

$$\mathcal{L}(V, \mathbb{R}) = \{h\colon V \to \mathbb{R} \mid h \text{ is linear}\}$$

is isomorphic to $\mathbb{R}^n$. It is often denoted V^*. Conclude that $V^* \cong V$.

2.42 Show that any linear map is the sum of maps of rank one.

2.43 Is 'is homomorphic to' an equivalence relation? (*Hint:* the difficulty is to decide on an appropriate meaning for the quoted phrase.)

2.44 Show that the range spaces and null spaces of powers of linear maps $t\colon V \to V$ form descending

$$V \supseteq \mathscr{R}(t) \supseteq \mathscr{R}(t^2) \supseteq \dots$$

and ascending

$$\{\vec{0}\} \subseteq \mathscr{N}(t) \subseteq \mathscr{N}(t^2) \subseteq \dots$$

chains. Also show that if k is such that $\mathscr{R}(t^k) = \mathscr{R}(t^{k+1})$ then all following range spaces are equal: $\mathscr{R}(t^k) = \mathscr{R}(t^{k+1}) = \mathscr{R}(t^{k+2}) \dots$. Similarly, if $\mathscr{N}(t^k) = \mathscr{N}(t^{k+1})$ then $\mathscr{N}(t^k) = \mathscr{N}(t^{k+1}) = \mathscr{N}(t^{k+2}) = \dots$.

III Computing Linear Maps

The prior section shows that a linear map is determined by its action on a basis. The equation

$$h(\vec{v}) = h(c_1 \cdot \vec{\beta}_1 + \cdots + c_n \cdot \vec{\beta}_n) = c_1 \cdot h(\vec{\beta}_1) + \cdots + c_n \cdot h(\vec{\beta}_n)$$

describes how we get the value of the map on any vector $\vec{v}$ by starting from the value of the map on the vectors $\vec{\beta}_i$ in a basis and extending linearly.

This section gives a convenient scheme based on matrices to use the representations of $h(\vec{\beta}_1), \ldots, h(\vec{\beta}_n)$ to compute, from the representation of a vector in the domain $\mathrm{Rep}_B(\vec{v})$, the representation of that vector's image in the codomain $\mathrm{Rep}_D(h(\vec{v}))$.

III.1 Representing Linear Maps with Matrices

1.1 Example For the spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ fix these bases.

$$B = \langle \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \end{pmatrix} \rangle \qquad D = \langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \rangle$$

Consider the map $h\colon \mathbb{R}^2 \to \mathbb{R}^3$ that is determined by this association.

$$\begin{pmatrix} 2 \\ 0 \end{pmatrix} \stackrel{h}{\longmapsto} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ 4 \end{pmatrix} \stackrel{h}{\longmapsto} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

To compute the action of this map on any vector at all from the domain we first represent the vector $h(\vec{\beta}_1)$

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \qquad \mathrm{Rep}_D(h(\vec{\beta}_1)) = \begin{pmatrix} 0 \\ -1/2 \\ 1 \end{pmatrix}_D$$

and the vector $h(\vec{\beta}_2)$.

$$\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - 1 \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \qquad \mathrm{Rep}_D(h(\vec{\beta}_2)) = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}_D$$

With these, for any member $\vec{v}$ of the domain we can compute $h(\vec{v})$.

$$\begin{aligned}
h(\vec{v}) &= h(c_1 \cdot \begin{pmatrix} 2 \\ 0 \end{pmatrix} + c_2 \cdot \begin{pmatrix} 1 \\ 4 \end{pmatrix}) \\
&= c_1 \cdot h(\begin{pmatrix} 2 \\ 0 \end{pmatrix}) + c_2 \cdot h(\begin{pmatrix} 1 \\ 4 \end{pmatrix}) \\
&= c_1 \cdot (0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}) + c_2 \cdot (1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - 1 \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}) \\
&= (0c_1 + 1c_2) \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (-\frac{1}{2}c_1 - 1c_2) \cdot \begin{pmatrix} 0 \\ -2 \\ 0 \end{pmatrix} + (1c_1 + 0c_2) \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}
\end{aligned}$$

Thus,

$$\text{if } \mathrm{Rep}_B(\vec{v}) = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \text{ then } \mathrm{Rep}_D(\, h(\vec{v})\,) = \begin{pmatrix} 0c_1 + 1c_2 \\ -(1/2)c_1 - 1c_2 \\ 1c_1 + 0c_2 \end{pmatrix}.$$

For instance,

$$\text{since } \mathrm{Rep}_B(\begin{pmatrix} 4 \\ 8 \end{pmatrix}) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}_B \text{ we have } \mathrm{Rep}_D(\, h(\begin{pmatrix} 4 \\ 8 \end{pmatrix})\,) = \begin{pmatrix} 2 \\ -5/2 \\ 1 \end{pmatrix}.$$

We express computations like the one above with a matrix notation.

$$\begin{pmatrix} 0 & 1 \\ -1/2 & -1 \\ 1 & 0 \end{pmatrix}_{B,D} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}_B = \begin{pmatrix} 0c_1 + 1c_2 \\ (-1/2)c_1 - 1c_2 \\ 1c_1 + 0c_2 \end{pmatrix}_D$$

In the middle is the argument $\vec{v}$ to the map, represented with respect to the domain's basis B by the column vector with components c_1 and c_2. On the right is the value of the map on that argument $h(\vec{v})$, represented with respect to the codomain's basis D. The matrix on the left is the new thing. We will use it to represent the map and we will think of the above equation as representing an application of the map to the matrix.

That matrix consists of the coefficients from the vector on the right, 0 and 1 from the first row, $-1/2$ and -1 from the second row, and 1 and 0 from the third row. That is, we make it by adjoining the vectors representing the $h(\vec{\beta}_i)$'s.

$$\left(\begin{array}{c|c} \vdots & \vdots \\ \mathrm{Rep}_D(\, h(\vec{\beta}_1)\,) & \mathrm{Rep}_D(\, h(\vec{\beta}_2)\,) \\ \vdots & \vdots \end{array}\right)$$

1.2 Definition Suppose that V and W are vector spaces of dimensions n and m with bases B and D, and that $h\colon V \to W$ is a linear map. If

$$\mathrm{Rep}_D(h(\vec{\beta}_1)) = \begin{pmatrix} h_{1,1} \\ h_{2,1} \\ \vdots \\ h_{m,1} \end{pmatrix}_D \quad \ldots \quad \mathrm{Rep}_D(h(\vec{\beta}_n)) = \begin{pmatrix} h_{1,n} \\ h_{2,n} \\ \vdots \\ h_{m,n} \end{pmatrix}_D$$

then

$$\mathrm{Rep}_{B,D}(h) = \begin{pmatrix} h_{1,1} & h_{1,2} & \ldots & h_{1,n} \\ h_{2,1} & h_{2,2} & \ldots & h_{2,n} \\ & \vdots & & \\ h_{m,1} & h_{m,2} & \ldots & h_{m,n} \end{pmatrix}_{B,D}$$

is the *matrix representation of* h *with respect to* B, D.

In that matrix the number of columns n is the dimension of the map's domain while the number of rows m is the dimension of the codomain.

We use lower case letters for a map, upper case for the matrix, and lower case again for the entries of the matrix. Thus for the map h, the matrix representing it is H, with entries $h_{i,j}$.

1.3 Example If $h\colon \mathbb{R}^3 \to \mathcal{P}_1$ is

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \stackrel{h}{\longmapsto} (2a_1 + a_2) + (-a_3)x$$

then where

$$B = \langle \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \rangle \qquad D = \langle 1 + x, -1 + x \rangle$$

the action of h on B is this.

$$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \stackrel{h}{\longmapsto} -x \qquad \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} \stackrel{h}{\longmapsto} 2 \qquad \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \stackrel{h}{\longmapsto} 4$$

A simple calculation

$$\mathrm{Rep}_D(-x) = \begin{pmatrix} -1/2 \\ -1/2 \end{pmatrix}_D \quad \mathrm{Rep}_D(2) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}_D \quad \mathrm{Rep}_D(4) = \begin{pmatrix} 2 \\ -2 \end{pmatrix}_D$$

shows that this is the matrix representing h with respect to the bases.

$$\mathrm{Rep}_{B,D}(h) = \begin{pmatrix} -1/2 & 1 & 2 \\ -1/2 & -1 & -2 \end{pmatrix}_{B,D}$$

1.4 Theorem Assume that V and W are vector spaces of dimensions n and m with bases B and D, and that $h\colon V \to W$ is a linear map. If h is represented by

$$\mathrm{Rep}_{B,D}(h) = \begin{pmatrix} h_{1,1} & h_{1,2} & \ldots & h_{1,n} \\ h_{2,1} & h_{2,2} & \ldots & h_{2,n} \\ & \vdots & & \\ h_{m,1} & h_{m,2} & \ldots & h_{m,n} \end{pmatrix}_{B,D}$$

and $\vec{v} \in V$ is represented by

$$\mathrm{Rep}_{B}(\vec{v}) = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}_{B}$$

then the representation of the image of $\vec{v}$ is this.

$$\mathrm{Rep}_{D}(\,h(\vec{v})\,) = \begin{pmatrix} h_{1,1}c_1 + h_{1,2}c_2 + \cdots + h_{1,n}c_n \\ h_{2,1}c_1 + h_{2,2}c_2 + \cdots + h_{2,n}c_n \\ \vdots \\ h_{m,1}c_1 + h_{m,2}c_2 + \cdots + h_{m,n}c_n \end{pmatrix}_{D}$$

Proof This formalizes Example 1.1. See Exercise 29. QED

1.5 Definition The *matrix-vector product* of a $m\times n$ matrix and a $n\times 1$ vector is this.

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ & \vdots & & \\ a_{m,1} & a_{m,2} & \ldots & a_{m,n} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} a_{1,1}c_1 + \cdots + a_{1,n}c_n \\ a_{2,1}c_1 + \cdots + a_{2,n}c_n \\ \vdots \\ a_{m,1}c_1 + \cdots + a_{m,n}c_n \end{pmatrix}$$

Briefly, application of a linear map is represented by the matrix-vector product of the map's representative and the vector's representative.

1.6 Remark Theorem 1.4 is not surprising, because we chose the matrix representative in Definition 1.2 precisely to make the theorem true — if the theorem were not true then we would adjust the definition to make it so. Nonetheless, we need the verification.

1.7 Example For the matrix from Example 1.3 we can calculate where that map sends this vector.

$$\vec{v}=\begin{pmatrix}4\\1\\0\end{pmatrix}$$

With respect to the domain basis B the representation of this vector is

$$\operatorname{Rep}_B(\vec{v})=\begin{pmatrix}0\\1/2\\2\end{pmatrix}_B$$

and so the matrix-vector product gives the representation of the value $h(\vec{v})$ with respect to the codomain basis D.

$$\begin{aligned}\operatorname{Rep}_D(h(\vec{v}))&=\begin{pmatrix}-1/2&1&2\\-1/2&-1&-2\end{pmatrix}_{B,D}\begin{pmatrix}0\\1/2\\2\end{pmatrix}_B\\&=\begin{pmatrix}(-1/2)\cdot 0+1\cdot(1/2)+2\cdot 2\\(-1/2)\cdot 0-1\cdot(1/2)-2\cdot 2\end{pmatrix}_D=\begin{pmatrix}9/2\\-9/2\end{pmatrix}_D\end{aligned}$$

To find $h(\vec{v})$ itself, not its representation, take $(9/2)(1+x)-(9/2)(-1+x)=9$.

1.8 Example Let $\pi\colon\mathbb{R}^3\to\mathbb{R}^2$ be projection onto the xy-plane. To give a matrix representing this map, we first fix some bases.

$$B=\langle\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}1\\1\\0\end{pmatrix},\begin{pmatrix}-1\\0\\1\end{pmatrix}\rangle\qquad D=\langle\begin{pmatrix}2\\1\end{pmatrix},\begin{pmatrix}1\\1\end{pmatrix}\rangle$$

For each vector in the domain's basis, find its image under the map.

$$\begin{pmatrix}1\\0\\0\end{pmatrix}\stackrel{\pi}{\longmapsto}\begin{pmatrix}1\\0\end{pmatrix}\quad\begin{pmatrix}1\\1\\0\end{pmatrix}\stackrel{\pi}{\longmapsto}\begin{pmatrix}1\\1\end{pmatrix}\quad\begin{pmatrix}-1\\0\\1\end{pmatrix}\stackrel{\pi}{\longmapsto}\begin{pmatrix}-1\\0\end{pmatrix}$$

Then find the representation of each image with respect to the codomain's basis.

$$\operatorname{Rep}_D(\begin{pmatrix}1\\0\end{pmatrix})=\begin{pmatrix}1\\-1\end{pmatrix}\quad\operatorname{Rep}_D(\begin{pmatrix}1\\1\end{pmatrix})=\begin{pmatrix}0\\1\end{pmatrix}\quad\operatorname{Rep}_D(\begin{pmatrix}-1\\0\end{pmatrix})=\begin{pmatrix}-1\\1\end{pmatrix}$$

Finally, adjoining these representations gives the matrix representing π with respect to B,D.

$$\operatorname{Rep}_{B,D}(\pi)=\begin{pmatrix}1&0&-1\\-1&1&1\end{pmatrix}_{B,D}$$

We can illustrate Theorem 1.4 by computing the matrix-vector product representing this action by the projection map.

$$\pi(\begin{pmatrix}2\\2\\1\end{pmatrix}) = \begin{pmatrix}2\\2\end{pmatrix}$$

Represent the domain vector with respect to the domain's basis

$$\mathrm{Rep}_B(\begin{pmatrix}2\\2\\1\end{pmatrix}) = \begin{pmatrix}1\\2\\1\end{pmatrix}_B$$

to get this matrix-vector product.

$$\mathrm{Rep}_D(\,\pi(\begin{pmatrix}2\\1\\1\end{pmatrix})\,) = \begin{pmatrix}1&0&-1\\-1&1&1\end{pmatrix}_{B,D}\begin{pmatrix}1\\2\\1\end{pmatrix}_B = \begin{pmatrix}0\\2\end{pmatrix}_D$$

Expanding this into a linear combination of vectors from D

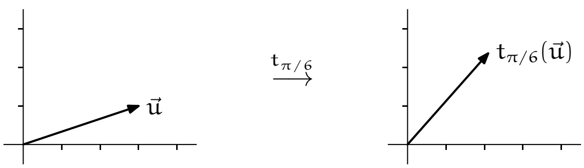
$$0\cdot\begin{pmatrix}2\\1\end{pmatrix} + 2\cdot\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix}$$

checks that the map's action is indeed reflected in the operation of the matrix. We will sometimes compress these three displayed equations into one.

$$\begin{pmatrix}2\\2\\1\end{pmatrix} = \begin{pmatrix}1\\2\\1\end{pmatrix}_B \overset{h}{\underset{H}{\longmapsto}} \begin{pmatrix}0\\2\end{pmatrix}_D = \begin{pmatrix}2\\2\end{pmatrix}$$

We now have two ways to compute the effect of projection, the straightforward formula that drops each three-tall vector's third component to make a two-tall vector, and the above formula that uses representations and matrix-vector multiplication. The second way may seem complicated compared to the first, but it has advantages. The next example shows that for some maps this new scheme simplifies the formula.

1.9 Example To represent a rotation map $t_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ that turns all vectors in the plane counterclockwise through an angle θ



we start by fixing the standard bases $\mathcal{E}_2$ for both the domain and codomain basis, Now find the image under the map of each vector in the domain's basis.

$$\begin{pmatrix}1\\0\end{pmatrix}\overset{t_\theta}{\longmapsto}\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}\qquad\begin{pmatrix}0\\1\end{pmatrix}\overset{t_\theta}{\longmapsto}\begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}\tag{$*$}$$

Represent these images with respect to the codomain's basis. Because this basis is $\mathcal{E}_2$, vectors represent themselves. Adjoin the representations to get the matrix representing the map.

$$\mathrm{Rep}_{\mathcal{E}_2,\mathcal{E}_2}(t_\theta)=\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}$$

The advantage of this scheme is that we get a formula for the image of any vector at all just by knowing in $(*)$ how to represent the image of the two basis vectors. For instance, here we rotate a vector by $\theta=\pi/6$.

$$\begin{pmatrix}3\\-2\end{pmatrix}=\begin{pmatrix}3\\-2\end{pmatrix}_{\mathcal{E}_2}\overset{t_{\pi/6}}{\longmapsto}\begin{pmatrix}\sqrt{3}/2 & -1/2\\ 1/2 & \sqrt{3}/2\end{pmatrix}\begin{pmatrix}3\\-2\end{pmatrix}\approx\begin{pmatrix}3.598\\-0.232\end{pmatrix}_{\mathcal{E}_2}=\begin{pmatrix}3.598\\-0.232\end{pmatrix}$$

More generally, we have a formula for rotation by $\theta=\pi/6$.

$$\begin{pmatrix}x\\y\end{pmatrix}\overset{t_{\pi/6}}{\longmapsto}\begin{pmatrix}\sqrt{3}/2 & -1/2\\ 1/2 & \sqrt{3}/2\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}(\sqrt{3}/2)x-(1/2)y\\(1/2)x+(\sqrt{3}/2)y\end{pmatrix}$$

1.10 Example In the definition of matrix-vector product the width of the matrix equals the height of the vector. Hence, this product is not defined.

$$\begin{pmatrix}1&0&0\\4&3&1\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}$$

It is undefined for a reason: the three-wide matrix represents a map with a three-dimensional domain while the two-tall vector represents a member of a two-dimensional space. So the vector cannot be in the domain of the map.

Nothing in Definition 1.5 forces us to view matrix-vector product in terms of representations. We can get some insights by focusing on how the entries combine.

A good way to view matrix-vector product is that it is formed from the dot products of the rows of the matrix with the column vector.

$$\begin{pmatrix} & \vdots & & \\ a_{i,1} & a_{i,2} & \ldots & a_{i,n}\\ & \vdots & & \end{pmatrix}\begin{pmatrix}c_1\\c_2\\ \vdots\\ c_n\end{pmatrix}=\begin{pmatrix}\vdots\\ a_{i,1}c_1+a_{i,2}c_2+\cdots+a_{i,n}c_n\\ \vdots\end{pmatrix}$$

Looked at in this row-by-row way, this new operation generalizes dot product.

We can also view the operation column-by-column.

$$\begin{pmatrix} h_{1,1} & h_{1,2} & \ldots & h_{1,n} \\ h_{2,1} & h_{2,2} & \ldots & h_{2,n} \\ & \vdots & & \\ h_{m,1} & h_{m,2} & \ldots & h_{m,n} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} h_{1,1}c_1 + h_{1,2}c_2 + \cdots + h_{1,n}c_n \\ h_{2,1}c_1 + h_{2,2}c_2 + \cdots + h_{2,n}c_n \\ \vdots \\ h_{m,1}c_1 + h_{m,2}c_2 + \cdots + h_{m,n}c_n \end{pmatrix}$$
$$= c_1 \begin{pmatrix} h_{1,1} \\ h_{2,1} \\ \vdots \\ h_{m,1} \end{pmatrix} + \cdots + c_n \begin{pmatrix} h_{1,n} \\ h_{2,n} \\ \vdots \\ h_{m,n} \end{pmatrix}$$

The result is the columns of the matrix weighted by the entries of the vector.

1.11 Example

$$\begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} - 1 \begin{pmatrix} 0 \\ 0 \end{pmatrix} + 1 \begin{pmatrix} -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 7 \end{pmatrix}$$

This way of looking at matrix-vector product brings us back to the objective stated at the start of this section, to compute $h(c_1\vec{\beta}_1 + \cdots + c_n\vec{\beta}_n)$ as $c_1h(\vec{\beta}_1) + \cdots + c_nh(\vec{\beta}_n)$.

We began this section by noting that the equality of these two enables us to compute the action of h on any argument knowing only $h(\vec{\beta}_1)$, ..., $h(\vec{\beta}_n)$. We have developed this into a scheme to compute the action of the map by taking the matrix-vector product of the matrix representing the map with the vector representing the argument. In this way, with respect to any bases, for any linear map there is a matrix representation. The next subsection will show the converse, that if we fix bases then for any matrix there is an associated linear map.

Exercises

✓ 1.12 Multiply the matrix

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

by each vector, or state "not defined."

(a) $\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$ **(b)** $\begin{pmatrix} -2 \\ -2 \end{pmatrix}$ **(c)** $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

1.13 Perform, if possible, each matrix-vector multiplication.

(a) $\begin{pmatrix}2&1\\3&-1/2\end{pmatrix}\begin{pmatrix}4\\2\end{pmatrix}$ **(b)** $\begin{pmatrix}1&1&0\\-2&1&0\end{pmatrix}\begin{pmatrix}1\\3\\1\end{pmatrix}$ **(c)** $\begin{pmatrix}1&1\\-2&1\end{pmatrix}\begin{pmatrix}1\\3\\1\end{pmatrix}$

✓ **1.14** Solve this matrix equation.

$$\begin{pmatrix}2&1&1\\0&1&3\\1&-1&2\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}8\\4\\4\end{pmatrix}$$

✓ **1.15** For a homomorphism from $\mathcal{P}_2$ to $\mathcal{P}_3$ that sends

$$1\mapsto 1+x,\quad x\mapsto 1+2x,\quad\text{and}\quad x^2\mapsto x-x^3$$

where does $1-3x+2x^2$ go?

✓ **1.16** Assume that $h\colon\mathbb{R}^2\to\mathbb{R}^3$ is determined by this action.

$$\begin{pmatrix}1\\0\end{pmatrix}\mapsto\begin{pmatrix}2\\2\\0\end{pmatrix}\qquad\begin{pmatrix}0\\1\end{pmatrix}\mapsto\begin{pmatrix}0\\1\\-1\end{pmatrix}$$

Using the standard bases, find

(a) the matrix representing this map;

(b) a general formula for $h(\vec{v})$.

✓ **1.17** Let $d/dx\colon\mathcal{P}_3\to\mathcal{P}_3$ be the derivative transformation.

(a) Represent d/dx with respect to B,B where $B=\langle 1,x,x^2,x^3\rangle$.

(b) Represent d/dx with respect to B,D where $D=\langle 1,2x,3x^2,4x^3\rangle$.

✓ **1.18** Represent each linear map with respect to each pair of bases.

(a) $d/dx\colon\mathcal{P}_n\to\mathcal{P}_n$ with respect to B,B where $B=\langle 1,x,\ldots,x^n\rangle$, given by

$$a_0+a_1x+a_2x^2+\cdots+a_nx^n\mapsto a_1+2a_2x+\cdots+na_nx^{n-1}$$

(b) $\int\colon\mathcal{P}_n\to\mathcal{P}_{n+1}$ with respect to B_n,B_{n+1} where $B_i=\langle 1,x,\ldots,x^i\rangle$, given by

$$a_0+a_1x+a_2x^2+\cdots+a_nx^n\mapsto a_0x+\frac{a_1}{2}x^2+\cdots+\frac{a_n}{n+1}x^{n+1}$$

(c) $\int_0^1\colon\mathcal{P}_n\to\mathbb{R}$ with respect to $B,\mathcal{E}_1$ where $B=\langle 1,x,\ldots,x^n\rangle$ and $\mathcal{E}_1=\langle 1\rangle$, given by

$$a_0+a_1x+a_2x^2+\cdots+a_nx^n\mapsto a_0+\frac{a_1}{2}+\cdots+\frac{a_n}{n+1}$$

(d) $\mathrm{eval}_3\colon\mathcal{P}_n\to\mathbb{R}$ with respect to $B,\mathcal{E}_1$ where $B=\langle 1,x,\ldots,x^n\rangle$ and $\mathcal{E}_1=\langle 1\rangle$, given by

$$a_0+a_1x+a_2x^2+\cdots+a_nx^n\mapsto a_0+a_1\cdot 3+a_2\cdot 3^2+\cdots+a_n\cdot 3^n$$

(e) $\mathrm{slide}_{-1}\colon\mathcal{P}_n\to\mathcal{P}_n$ with respect to B,B where $B=\langle 1,x,\ldots,x^n\rangle$, given by

$$a_0+a_1x+a_2x^2+\cdots+a_nx^n\mapsto a_0+a_1\cdot(x+1)+\cdots+a_n\cdot(x+1)^n$$

1.19 Represent the identity map on any nontrivial space with respect to B,B, where B is any basis.

1.20 Represent, with respect to the natural basis, the transpose transformation on the space $\mathcal{M}_{2\times 2}$ of 2×2 matrices.

1.21 Assume that $B=\langle\vec{\beta}_1,\vec{\beta}_2,\vec{\beta}_3,\vec{\beta}_4\rangle$ is a basis for a vector space. Represent with respect to B,B the transformation that is determined by each.

(a) $\vec{\beta}_1\mapsto\vec{\beta}_2$, $\vec{\beta}_2\mapsto\vec{\beta}_3$, $\vec{\beta}_3\mapsto\vec{\beta}_4$, $\vec{\beta}_4\mapsto\vec{0}$

(b) $\vec{\beta}_1\mapsto\vec{\beta}_2$, $\vec{\beta}_2\mapsto\vec{0}$, $\vec{\beta}_3\mapsto\vec{\beta}_4$, $\vec{\beta}_4\mapsto\vec{0}$

(c) $\vec{\beta}_1 \mapsto \vec{\beta}_2$, $\vec{\beta}_2 \mapsto \vec{\beta}_3$, $\vec{\beta}_3 \mapsto \vec{0}$, $\vec{\beta}_4 \mapsto \vec{0}$

1.22 Example 1.9 shows how to represent the rotation transformation of the plane with respect to the standard basis. Express these other transformations also with respect to the standard basis.

(a) the *dilation* map d_s, which multiplies all vectors by the same scalar s

(b) the *reflection* map f_ℓ, which reflects all all vectors across a line ℓ through the origin

✓ 1.23 Consider a linear transformation of $\mathbb{R}^2$ determined by these two.

$$\begin{pmatrix}1\\1\end{pmatrix} \mapsto \begin{pmatrix}2\\0\end{pmatrix} \qquad \begin{pmatrix}1\\0\end{pmatrix} \mapsto \begin{pmatrix}-1\\0\end{pmatrix}$$

(a) Represent this transformation with respect to the standard bases.

(b) Where does the transformation send this vector?

$$\begin{pmatrix}0\\5\end{pmatrix}$$

(c) Represent this transformation with respect to these bases.

$$B = \langle \begin{pmatrix}1\\-1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix} \rangle \qquad D = \langle \begin{pmatrix}2\\2\end{pmatrix}, \begin{pmatrix}-1\\1\end{pmatrix} \rangle$$

(d) Using B from the prior item, represent the transformation with respect to B, B.

1.24 Suppose that $h\colon V \to W$ is one-to-one so that by Theorem 2.20, for any basis $B = \langle \vec{\beta}_1, \ldots, \vec{\beta}_n \rangle \subset V$ the image $h(B) = \langle h(\vec{\beta}_1), \ldots, h(\vec{\beta}_n) \rangle$ is a basis for W.

(a) Represent the map h with respect to $B, h(B)$.

(b) For a member $\vec{v}$ of the domain, where the representation of $\vec{v}$ has components $c_1, \ldots, c_n$, represent the image vector $h(\vec{v})$ with respect to the image basis $h(B)$.

1.25 Give a formula for the product of a matrix and $\vec{e}_i$, the column vector that is all zeroes except for a single one in the i-th position.

✓ 1.26 For each vector space of functions of one real variable, represent the derivative transformation with respect to B, B.

(a) $\{a\cos x + b\sin x \mid a, b \in \mathbb{R}\}$, $B = \langle \cos x, \sin x \rangle$

(b) $\{ae^x + be^{2x} \mid a, b \in \mathbb{R}\}$, $B = \langle e^x, e^{2x} \rangle$

(c) $\{a + bx + ce^x + dxe^x \mid a, b, c, d \in \mathbb{R}\}$, $B = \langle 1, x, e^x, xe^x \rangle$

1.27 Find the range of the linear transformation of $\mathbb{R}^2$ represented with respect to the standard bases by each matrix.

(a) $\begin{pmatrix}1&0\\0&0\end{pmatrix}$ (b) $\begin{pmatrix}0&0\\3&2\end{pmatrix}$ (c) a matrix of the form $\begin{pmatrix}a&b\\2a&2b\end{pmatrix}$

✓ 1.28 Can one matrix represent two different linear maps? That is, can $\text{Rep}_{B,D}(h) = \text{Rep}_{\hat{B},\hat{D}}(\hat{h})$?

1.29 Prove Theorem 1.4.

✓ 1.30 Example 1.9 shows how to represent rotation of all vectors in the plane through an angle θ about the origin, with respect to the standard bases.

(a) Rotation of all vectors in three-space through an angle θ about the x-axis is a transformation of $\mathbb{R}^3$. Represent it with respect to the standard bases. Arrange the rotation so that to someone whose feet are at the origin and whose head is at $(1, 0, 0)$, the movement appears clockwise.

(b) Repeat the prior item, only rotate about the y-axis instead. (Put the person's head at $\vec{e}_2$.)

(c) Repeat, about the z-axis.

(d) Extend the prior item to $\mathbb{R}^4$. (*Hint:* we can restate 'rotate about the z-axis' as 'rotate parallel to the xy-plane'.)

1.31 (Schur's Triangularization Lemma)

(a) Let U be a subspace of V and fix bases $B_U \subseteq B_V$. What is the relationship between the representation of a vector from U with respect to B_U and the representation of that vector (viewed as a member of V) with respect to B_V?

(b) What about maps?

(c) Fix a basis $B = \langle\vec{\beta}_1, \ldots, \vec{\beta}_n\rangle$ for V and observe that the spans

$$[\varnothing] = \{\vec{0}\} \subset [\{\vec{\beta}_1\}] \subset [\{\vec{\beta}_1, \vec{\beta}_2\}] \subset \quad \cdots \quad \subset [B] = V$$

form a strictly increasing chain of subspaces. Show that for any linear map $h\colon V \to W$ there is a chain $W_0 = \{\vec{0}\} \subseteq W_1 \subseteq \cdots \subseteq W_m = W$ of subspaces of W such that

$$h([\{\vec{\beta}_1, \ldots, \vec{\beta}_i\}]) \subset W_i$$

for each i.

(d) Conclude that for every linear map $h\colon V \to W$ there are bases B, D so the matrix representing h with respect to B, D is upper-triangular (that is, each entry $h_{i,j}$ with $i > j$ is zero).

(e) Is an upper-triangular representation unique?

III.2 Any Matrix Represents a Linear Map

The prior subsection shows that the action of a linear map h is described by a matrix H, with respect to appropriate bases, in this way.

$$\vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}_B \quad \overset{h}{\underset{H}{\longmapsto}} \quad h(\vec{v}) = \begin{pmatrix} h_{1,1}v_1 + \cdots + h_{1,n}v_n \\ \vdots \\ h_{m,1}v_1 + \cdots + h_{m,n}v_n \end{pmatrix}_D \qquad (*)$$

Here we will show the converse, that each matrix represents a linear map.

So we start with a matrix

$$H = \begin{pmatrix} h_{1,1} & h_{1,2} & \ldots & h_{1,n} \\ h_{2,1} & h_{2,2} & \ldots & h_{2,n} \\ & \vdots & & \\ h_{m,1} & h_{m,2} & \ldots & h_{m,n} \end{pmatrix}$$

and we will describe how it defines a map h. We require that the map be represented by the matrix so first note that in $(*)$ the dimension of the map's domain is the number of columns n of the matrix and the dimension of the codomain is the number of rows m. Thus, for h's domain fix an n-dimensional vector space V and for the codomain fix an m-dimensional space W. Also fix bases $B = \langle \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$ and $D = \langle \vec{\delta}_1, \dots, \vec{\delta}_m \rangle$ for those spaces.

Now let $h\colon V \to W$ be: where $\vec{v}$ in the domain has the representation

$$\mathrm{Rep}_B(\vec{v}) = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}_B$$

then its image $h(\vec{v})$ is the member the codomain with this representation.

$$\mathrm{Rep}_D(\, h(\vec{v})\,) = \begin{pmatrix} h_{1,1}v_1 + \cdots + h_{1,n}v_n \\ \vdots \\ h_{m,1}v_1 + \cdots + h_{m,n}v_n \end{pmatrix}_D$$

That is, to compute the action of h on any $\vec{v} \in V$, first express $\vec{v}$ with respect to the basis $\vec{v} = v_1\vec{\beta}_1 + \cdots + v_n\vec{\beta}_n$ and then $h(\vec{v}) = (h_{1,1}v_1 + \cdots + h_{1,n}v_n) \cdot \vec{\delta}_1 + \cdots + (h_{m,1}v_1 + \cdots + h_{m,n}v_n) \cdot \vec{\delta}_m$.

Above we have made some choices; for instance V can be any n-dimensional space and B could be any basis for V, so H does not define a unique function. However, note once we have fixed V, B, W, and D then h is well-defined since $\vec{v}$ has a unique representation with respect to the basis B and the calculation of $\vec{w}$ from its representation is also uniquely determined.

2.1 Example Consider this matrix.

$$H = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

It is 3×2 so any map that it defines must carry a dimension 2 domain to a dimension 3 codomain. We can choose the domain and codomain to be $\mathbb{R}^2$ and $\mathcal{P}_2$, with these bases.

$$B = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rangle \qquad D = \langle x^2, x^2 + x, x^2 + x + 1 \rangle$$

Then let $h\colon \mathbb{R}^2 \to \mathcal{P}_2$ be the function defined by H. We will compute the image under h of this member of the domain.

$$\vec{v} = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$$

The computation is straightforward.

$$\mathrm{Rep}_D(h(\vec{v})) = H \cdot \mathrm{Rep}_B(\vec{v}) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} -1/2 \\ -5/2 \end{pmatrix} = \begin{pmatrix} -11/2 \\ -23/2 \\ -35/2 \end{pmatrix}$$

From its representation, computation of $\vec{w}$ is routine $(-11/2)(x^2) - (23/2)(x^2 + x) - (35/2)(x^2 + x + 1) = (-69/2)x^2 - (58/2)x - (35/2)$.

2.2 Theorem Any matrix represents a homomorphism between vector spaces of appropriate dimensions, with respect to any pair of bases.

Proof We must check that for any matrix H and any domain and codomain bases B, D, the defined map h is linear. If $\vec{v}, \vec{u} \in V$ are such that

$$\mathrm{Rep}_B(\vec{v}) = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \qquad \mathrm{Rep}_B(\vec{u}) = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

and $c, d \in \mathbb{R}$ then the calculation

$$\begin{aligned} h(c\vec{v} + d\vec{u}) &= \bigl(h_{1,1}(cv_1 + du_1) + \cdots + h_{1,n}(cv_n + du_n)\bigr) \cdot \vec{\delta}_1 + \\ &\qquad \cdots + \bigl(h_{m,1}(cv_1 + du_1) + \cdots + h_{m,n}(cv_n + du_n)\bigr) \cdot \vec{\delta}_m \\ &= c \cdot h(\vec{v}) + d \cdot h(\vec{u}) \end{aligned}$$

supplies that check. QED

2.3 Example Even if the domain and codomain are the same, the map that the matrix represents depends on the bases that we choose. If

$$H = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad B_1 = D_1 = \langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rangle, \quad \text{and} \quad B_2 = D_2 = \langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle,$$

then $h_1 \colon \mathbb{R}^2 \to \mathbb{R}^2$ represented by H with respect to B_1, D_1 maps

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}_{B_1} \quad \mapsto \quad \begin{pmatrix} c_1 \\ 0 \end{pmatrix}_{D_1} = \begin{pmatrix} c_1 \\ 0 \end{pmatrix}$$

while $h_2 \colon \mathbb{R}^2 \to \mathbb{R}^2$ represented by H with respect to B_2, D_2 is this map.

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} c_2 \\ c_1 \end{pmatrix}_{B_2} \quad \mapsto \quad \begin{pmatrix} c_2 \\ 0 \end{pmatrix}_{D_2} = \begin{pmatrix} 0 \\ c_2 \end{pmatrix}$$

These are different functions. The first is projection onto the x-axis while the second is projection onto the y-axis.

This result means that when convenient we can work solely with matrices, just doing the computations without having to worry whether a matrix of interest represents a linear map on some pair of spaces.

When we are working with a matrix but we do not have particular spaces or bases in mind then we can take the domain and codomain to be $\mathbb{R}^n$ and $\mathbb{R}^m$, with the standard bases. This is convenient because with the standard bases vector representation is transparent — the representation of $\vec{v}$ is $\vec{v}$. (In this case the column space of the matrix equals the range of the map and consequently the column space of H is often denoted by $\mathscr{R}(H)$.)

Given a matrix, to come up with an associated map we can choose among many domain and codomain spaces, and many bases for those. So a matrix can represent many maps. We finish this section by illustrating how the matrix can give us information about the associated maps.

2.4 Theorem The rank of a matrix equals the rank of any map that it represents.

Proof Suppose that the matrix H is $m \times n$. Fix domain and codomain spaces V and W of dimension n and m with bases $B = \langle \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$ and D. Then H represents some linear map h between those spaces with respect to these bases whose range space

$$\begin{aligned}\{h(\vec{v}) \mid \vec{v} \in V\} &= \{h(c_1\vec{\beta}_1 + \cdots + c_n\vec{\beta}_n) \mid c_1, \dots, c_n \in \mathbb{R}\} \\ &= \{c_1 h(\vec{\beta}_1) + \cdots + c_n h(\vec{\beta}_n) \mid c_1, \dots, c_n \in \mathbb{R}\}\end{aligned}$$

is the span $[\{h(\vec{\beta}_1), \dots, h(\vec{\beta}_n)\}]$. The rank of the map h is the dimension of this range space.

The rank of the matrix is the dimension of its column space, the span of the set of its columns $[\{\mathrm{Rep}_D(h(\vec{\beta}_1)), \dots, \mathrm{Rep}_D(h(\vec{\beta}_n))\}]$.

To see that the two spans have the same dimension, recall from the proof of Lemma I.2.5 that if we fix a basis then representation with respect to that basis gives an isomorphism $\mathrm{Rep}_D \colon W \to \mathbb{R}^m$. Under this isomorphism there is a linear relationship among members of the range space if and only if the same relationship holds in the column space, e.g, $\vec{0} = c_1 \cdot h(\vec{\beta}_1) + \cdots + c_n \cdot h(\vec{\beta}_n)$ if and only if $\vec{0} = c_1 \cdot \mathrm{Rep}_D(h(\vec{\beta}_1)) + \cdots + c_n \cdot \mathrm{Rep}_D(h(\vec{\beta}_n))$. Hence, a subset of the range space is linearly independent if and only if the corresponding subset of the column space is linearly independent. Therefore the size of the largest linearly independent subset of the range space equals the size of the largest linearly independent subset of the column space, and so the two spaces have the same dimension. QED

That settles the apparent ambiguity in our use of the same word 'rank' to apply both to matrices and to maps.

2.5 Example Any map represented by

$$\begin{pmatrix} 1 & 2 & 2 \\ 1 & 2 & 1 \\ 0 & 0 & 3 \\ 0 & 0 & 2 \end{pmatrix}$$

must have three-dimensional domain and a four-dimensional codomain. In addition, because the rank of this matrix is two (we can spot this by eye or get it with Gauss's Method), any map represented by this matrix has a two-dimensional range space.

2.6 Corollary Let h be a linear map represented by a matrix H. Then h is onto if and only if the rank of H equals the number of its rows, and h is one-to-one if and only if the rank of H equals the number of its columns.

PROOF For the onto half, the dimension of the range space of h is the rank of h, which equals the rank of H by the theorem. Since the dimension of the codomain of h equals the number of rows in H, if the rank of H equals the number of rows then the dimension of the range space equals the dimension of the codomain. But a subspace with the same dimension as its superspace must equal that superspace (because any basis for the range space is a linearly independent subset of the codomain whose size is equal to the dimension of the codomain, and thus so this basis for the range space must also be a basis for the codomain).

For the other half, a linear map is one-to-one if and only if it is an isomorphism between its domain and its range, that is, if and only if its domain has the same dimension as its range. The number of columns in h is the dimension of h's domain and by the theorem the rank of H equals the dimension of h's range. QED

2.7 Definition A linear map that is one-to-one and onto is *nonsingular*, otherwise it is *singular*. That is, a linear map is nonsingular if and only if it is an isomorphism.

2.8 Remark Some authors use 'nonsingular' as a synonym for one-to-one while others use it the way that we have here. The difference is slight because any map is onto its range space, so a one-to-one map is an isomorphism with its range.

In the first chapter we defined a matrix to be nonsingular if it is square and is the matrix of coefficients of a linear system with a unique solution. The next result justifies our dual use of the term.

2.9 Lemma A nonsingular linear map is represented by a square matrix. A square matrix represents nonsingular maps if and only if it is a nonsingular matrix. Thus, a matrix represents isomorphisms if and only if it is square and nonsingular.

Proof Assume that the map $h: V \to W$ is nonsingular. Corollary 2.6 says that for any matrix H representing that map, because h is onto the number of rows of H equals the rank of H, and because h is one-to-one the number of columns of H is also equal to the rank of H. Hence H is square.

Next assume that H is square, $n \times n$. The matrix H is nonsingular if and only if its row rank is n, which is true if and only if H's rank is n by Theorem Two.III.3.11, which is true if and only if h's rank is n by Theorem 2.4, which is true if and only if h is an isomorphism by Theorem I.2.3. (This last holds because the domain of h is n-dimensional as it is the number of columns in H.) QED

2.10 Example Any map from $\mathbb{R}^2$ to $\mathcal{P}_1$ represented with respect to any pair of bases by

$$\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$$

is nonsingular because this matrix has rank two.

2.11 Example Any map $g: V \to W$ represented by

$$\begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$$

is singular because this matrix is singular.

We've now seen that the relationship between maps and matrices goes both ways: for a particular pair of bases, any linear map is represented by a matrix and any matrix describes a linear map. That is, by fixing spaces and bases we get a correspondence between maps and matrices. In the rest of this chapter we will explore this correspondence. For instance, we've defined for linear maps the operations of addition and scalar multiplication and we shall see what the corresponding matrix operations are. We shall also see the matrix operation that represent the map operation of composition. And, we shall see how to find the matrix that represents a map's inverse.

Exercises

✓ 2.12 Let h be the linear map defined by this matrix on the domain $\mathcal{P}_1$ and codomain $\mathbb{R}^2$ with respect to the given bases.

$$H = \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix} \qquad B = \langle 1 + x, x \rangle,\ D = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle$$

What is the image under h of the vector $\vec{v} = 2x - 1$?

✓ 2.13 Decide if each vector lies in the range of the map from $\mathbb{R}^3$ to $\mathbb{R}^2$ represented with respect to the standard bases by the matrix.

(a) $\begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ **(b)** $\begin{pmatrix} 2 & 0 & 3 \\ 4 & 0 & 6 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

✓ 2.14 Consider this matrix, representing a transformation of $\mathbb{R}^2$, and these bases for that space.

$$\frac{1}{2} \cdot \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \qquad B = \langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle \quad D = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rangle$$

(a) To what vector in the codomain is the first member of B mapped?

(b) The second member?

(c) Where is a general vector from the domain (a vector with components x and y) mapped? That is, what transformation of $\mathbb{R}^2$ is represented with respect to B, D by this matrix?

2.15 What transformation of $F = \{a\cos\theta + b\sin\theta \mid a, b \in \mathbb{R}\}$ is represented with respect to $B = \langle \cos\theta - \sin\theta, \sin\theta \rangle$ and $D = \langle \cos\theta + \sin\theta, \cos\theta \rangle$ by this matrix?

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

✓ 2.16 Decide whether $1 + 2x$ is in the range of the map from $\mathbb{R}^3$ to $\mathcal{P}_2$ represented with respect to $\mathcal{E}_3$ and $\langle 1, 1 + x^2, x \rangle$ by this matrix.

$$\begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

2.17 Example 2.11 gives a matrix that is nonsingular and is therefore associated with maps that are nonsingular.

(a) Find the set of column vectors representing the members of the null space of any map represented by this matrix.

(b) Find the nullity of any such map.

(c) Find the set of column vectors representing the members of the range space of any map represented by this matrix.

(d) Find the rank of any such map.

(e) Check that rank plus nullity equals the dimension of the domain.

2.18 This is an alternative proof of Lemma 2.9. Given an $n \times n$ matrix H, fix a domain V and codomain W of appropriate dimension n, and bases B, D for those spaces, and consider the map h represented by the matrix.

(a) Show that h is onto if and only if there is at least one $\text{Rep}_B(\vec{v})$ associated by H with each $\text{Rep}_D(\vec{w})$.

(b) Show that h is one-to-one if and only if there is at most one $\text{Rep}_B(\vec{v})$ associated by H with each $\text{Rep}_D(\vec{w})$.

(c) Consider the linear system $H \cdot \text{Rep}_B(\vec{v}) = \text{Rep}_D(\vec{w})$. Show that H is nonsingular if and only if there is exactly one solution $\text{Rep}_B(\vec{v})$ for each $\text{Rep}_D(\vec{w})$.

✓ 2.19 Because the rank of a matrix equals the rank of any map it represents, if one matrix represents two different maps $H = \text{Rep}_{B,D}(h) = \text{Rep}_{\hat{B},\hat{D}}(\hat{h})$ (where

$h, \hat{h}\colon V \to W$) then the dimension of the range space of h equals the dimension of the range space of $\hat{h}$. Must these equal-dimensioned range spaces actually be the same?

✓ 2.20 Let V be an n-dimensional space with bases B and D. Consider a map that sends, for $\vec{v} \in V$, the column vector representing $\vec{v}$ with respect to B to the column vector representing $\vec{v}$ with respect to D. Show that map is a linear transformation of $\mathbb{R}^n$.

2.21 Example 2.3 shows that changing the pair of bases can change the map that a matrix represents, even though the domain and codomain remain the same. Could the map ever not change? Is there a matrix H, vector spaces V and W, and associated pairs of bases B_1, D_1 and B_2, D_2 (with $B_1 \neq B_2$ or $D_1 \neq D_2$ or both) such that the map represented by H with respect to B_1, D_1 equals the map represented by H with respect to B_2, D_2?

✓ 2.22 A square matrix is a *diagonal* matrix if it is all zeroes except possibly for the entries on its upper-left to lower-right diagonal—its $1,1$ entry, its $2,2$ entry, etc. Show that a linear map is an isomorphism if there are bases such that, with respect to those bases, the map is represented by a diagonal matrix with no zeroes on the diagonal.

2.23 Describe geometrically the action on $\mathbb{R}^2$ of the map represented with respect to the standard bases $\mathcal{E}_2, \mathcal{E}_2$ by this matrix.

$$\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

Do the same for these.

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$$

2.24 The fact that for any linear map the rank plus the nullity equals the dimension of the domain shows that a necessary condition for the existence of a homomorphism between two spaces, onto the second space, is that there be no gain in dimension. That is, where $h\colon V \to W$ is onto, the dimension of W must be less than or equal to the dimension of V.

(a) Show that this (strong) converse holds: no gain in dimension implies that there is a homomorphism and, further, any matrix with the correct size and correct rank represents such a map.

(b) Are there bases for $\mathbb{R}^3$ such that this matrix

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

represents a map from $\mathbb{R}^3$ to $\mathbb{R}^3$ whose range is the xy plane subspace of $\mathbb{R}^3$?

2.25 Let V be an n-dimensional space and suppose that $\vec{x} \in \mathbb{R}^n$. Fix a basis B for V and consider the map $h_{\vec{x}}\colon V \to \mathbb{R}$ given $\vec{v} \mapsto \vec{x} \cdot \mathrm{Rep}_B(\vec{v})$ by the dot product.

(a) Show that this map is linear.

(b) Show that for any linear map $g\colon V \to \mathbb{R}$ there is an $\vec{x} \in \mathbb{R}^n$ such that $g = h_{\vec{x}}$.

(c) In the prior item we fixed the basis and varied the $\vec{x}$ to get all possible linear maps. Can we get all possible linear maps by fixing an $\vec{x}$ and varying the basis?

2.26 Let V, W, X be vector spaces with bases B, C, D.

(a) Suppose that $h: V \to W$ is represented with respect to B, C by the matrix H. Give the matrix representing the scalar multiple rh (where $r \in \mathbb{R}$) with respect to B, C by expressing it in terms of H.

(b) Suppose that $h, g: V \to W$ are represented with respect to B, C by H and G. Give the matrix representing $h + g$ with respect to B, C by expressing it in terms of H and G.

(c) Suppose that $h: V \to W$ is represented with respect to B, C by H and $g: W \to X$ is represented with respect to C, D by G. Give the matrix representing $g \circ h$ with respect to B, D by expressing it in terms of H and G.

IV Matrix Operations

The prior section shows how matrices represent linear maps. We now explore how this representation interacts with things that we already know. First we will see how the representation of a scalar product $r \cdot f$ of a linear map relates to the representation of f, and also how the representation of a sum $f + g$ relates to the representations of the two summands. Later we will do the same comparison for the map operations of composition and inverse.

IV.1 Sums and Scalar Products

1.1 Example Let $f\colon V \to W$ be a linear function represented with respect to some bases by this matrix.

$$\mathrm{Rep}_{B,D}(f) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

Consider the map that is the scalar multiple $5f\colon V \to W$. We want to relate $\mathrm{Rep}_{B,D}(5f)$ to $\mathrm{Rep}_{B,D}(f)$.

Suppose that this function f takes $\vec{v} \mapsto \vec{w}$. Consider the representations.

$$\mathrm{Rep}_B(\vec{v}) = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \qquad \mathrm{Rep}_D(\vec{w}) = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$$

Where the codomain space basis is $D = \langle \vec{\delta}_1, \vec{\delta}_2 \rangle$ that representation gives that that the output vector is $\vec{w} = w_1\vec{\delta}_1 + w_2\vec{\delta}_2$.

The action of the map $5f$ is $\vec{v} \mapsto 5\vec{w}$. Since $5\vec{w} = 5 \cdot (w_1\vec{\delta}_1 + w_2\vec{\delta}_2) = (5w_1)\vec{\delta}_1 + (5w_2)\vec{\delta}_2$ we have that $5f$ associates $\vec{v}$ with the vector having this representation.

$$\mathrm{Rep}_D(5\vec{w}) = \begin{pmatrix} 5w_1 \\ 5w_2 \end{pmatrix}$$

Thus going from the map from f to the map $5f$ changes the representation of the output by multiplying each entry by 5.

Because of that, $\mathrm{Rep}_{B,D}(5f)$ is this matrix.

$$\mathrm{Rep}_{B,D}(5f) \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 5v_1 \\ 5v_1 + 5v_2 \end{pmatrix} \qquad \mathrm{Rep}_{B,D}(5f) = \begin{pmatrix} 5 & 0 \\ 5 & 5 \end{pmatrix}$$

Therefore, going from the matrix representing f to the one representing $5f$ just means multiplying all the matrix entries by 5.

1.2 Example We can do a similar exploration for the sum of two maps. Suppose that two linear maps with the same domain and codomain $f,g\colon \mathbb{R}^2 \to \mathbb{R}^2$ are represented with respect to bases B and D by these matrices.

$$\mathrm{Rep}_{B,D}(f)=\begin{pmatrix}1&3\\2&0\end{pmatrix}\qquad \mathrm{Rep}_{B,D}(g)=\begin{pmatrix}-2&-1\\2&4\end{pmatrix}$$

Recall the definition of sum: if f does $\vec{v}\mapsto\vec{u}$ and g does $\vec{v}\mapsto\vec{w}$ then $f+g$ is the function whose action is $\vec{v}\mapsto\vec{u}+\vec{w}$. Let these be the representations of the one input and two output vectors.

$$\mathrm{Rep}_B(\vec{v})=\begin{pmatrix}v_1\\v_2\end{pmatrix}\qquad \mathrm{Rep}_D(\vec{u})=\begin{pmatrix}u_1\\u_2\end{pmatrix}\quad \mathrm{Rep}_D(\vec{w})=\begin{pmatrix}w_1\\w_2\end{pmatrix}$$

Where $D=\langle\vec{\delta}_1,\vec{\delta}_2\rangle$ we have $\vec{u}+\vec{w}=(u_1\vec{\delta}_1+u_2\vec{\delta}_2)+(w_1\vec{\delta}_1+w_2\vec{\delta}_2)=(u_1+w_1)\vec{\delta}_1+(u_2+w_2)\vec{\delta}_2$ and so this is the representation of the vector sum.

$$\mathrm{Rep}_D(\vec{u}+\vec{w})=\begin{pmatrix}u_1+w_1\\u_2+w_2\end{pmatrix}$$

Since these represent the actions of of the maps f and g on the input $\vec{v}$

$$\begin{pmatrix}1&3\\2&0\end{pmatrix}\begin{pmatrix}v_1\\v_2\end{pmatrix}=\begin{pmatrix}v_1+3v_2\\2v_1\end{pmatrix}\qquad \begin{pmatrix}-2&-1\\2&4\end{pmatrix}\begin{pmatrix}v_1\\v_2\end{pmatrix}=\begin{pmatrix}-2v_1-v_2\\2v_1+4v_2\end{pmatrix}$$

this represents the action of the map $f+g$.

$$\mathrm{Rep}_{B,D}(f+g)\cdot\begin{pmatrix}v_1\\v_2\end{pmatrix}=\begin{pmatrix}-v_1+2v_2\\4v_1+4v_2\end{pmatrix}$$

Therefore, we compute the matrix representing the function sum by adding the entries of the matrices representing the functions.

$$\mathrm{Rep}_{B,D}(f+g)=\begin{pmatrix}-1&2\\4&4\end{pmatrix}$$

1.3 Definition The *scalar multiple* of a matrix is the result of entry-by-entry scalar multiplication. The *sum* of two same-sized matrices is their entry-by-entry sum.

These operations extend the first chapter's operations of addition and scalar multiplication of vectors.

We need a result that proves these matrix operations do what the examples suggest that they do.

1.4 Theorem Let $h, g\colon V \to W$ be linear maps represented with respect to bases B, D by the matrices H and G and let r be a scalar. Then with respect to B, D the map $r \cdot h\colon V \to W$ is represented by rH and the map $h + g\colon V \to W$ is represented by $H + G$.

Proof This is Exercise 9. Generalize the examples. QED

1.5 Remark These two matrix operations are simple. But, we did not define them this way because they are simple. We defined them this way because they represent function addition and function scalar multiplication. Simplicity is only a pleasant bonus.

In the next subsection we will define another operation, matrix multiplication. A first thought may be to define it to be the entry-by-entry product of the two matrices. While in theory we could do whatever we please, we will instead be practical and combine the entries to represent function composition. That is, we define matrix operations by referencing function operations.

We can express this program in another way. Recall Theorem III.1.4, which says that matrix-vector multiplication represents the application of a linear map. Following it, Remark III.1.6 notes that the theorem justifies the definition of matrix-vector multiplication and so in some sense the theorem must hold — if the theorem didn't hold then we would adjust the definition until the theorem did hold. The above Theorem 1.4 is another example of such a result. It justifies the given definition of the matrix operations.

A special case of scalar multiplication is multiplication by zero. For any map $0 \cdot h$ is the zero homomorphism and for any matrix $0 \cdot H$ is the matrix with all entries zero.

1.6 Definition A *zero matrix* has all entries 0. We write $Z_{n\times m}$ or simply Z (another common notation is $0_{n\times m}$ or just 0).

1.7 Example The zero map from any three-dimensional space to any two-dimensional space is represented by the 2×3 zero matrix

$$Z = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

no matter what domain and codomain bases we use.

Exercises

✓ 1.8 Perform the indicated operations, if defined.

(a) $\begin{pmatrix} 5 & -1 & 2 \\ 6 & 1 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 1 & 4 \\ 3 & 0 & 5 \end{pmatrix}$

(b) $6 \cdot \begin{pmatrix} 2 & -1 & -1 \\ 1 & 2 & 3 \end{pmatrix}$

(c) $\begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}$

(d) $4\begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix} + 5\begin{pmatrix} -1 & 4 \\ -2 & 1 \end{pmatrix}$

(e) $3\begin{pmatrix} 2 & 1 \\ 3 & 0 \end{pmatrix} + 2\begin{pmatrix} 1 & 1 & 4 \\ 3 & 0 & 5 \end{pmatrix}$

1.9 Prove Theorem 1.4.

(a) Prove that matrix addition represents addition of linear maps.

(b) Prove that matrix scalar multiplication represents scalar multiplication of linear maps.

✓ 1.10 Prove each, assuming that the operations are defined, where G, H, and J are matrices, where Z is the zero matrix, and where r and s are scalars.

(a) Matrix addition is commutative $G + H = H + G$.

(b) Matrix addition is associative $G + (H + J) = (G + H) + J$.

(c) The zero matrix is an additive identity $G + Z = G$.

(d) $0 \cdot G = Z$

(e) $(r + s)G = rG + sG$

(f) Matrices have an additive inverse $G + (-1) \cdot G = Z$.

(g) $r(G + H) = rG + rH$

(h) $(rs)G = r(sG)$

1.11 Fix domain and codomain spaces. In general, one matrix can represent many different maps with respect to different bases. However, prove that a zero matrix represents only a zero map. Are there other such matrices?

✓ 1.12 Let V and W be vector spaces of dimensions n and m. Show that the space $\mathcal{L}(V, W)$ of linear maps from V to W is isomorphic to $\mathcal{M}_{m \times n}$.

✓ 1.13 Show that it follows from the prior questions that for any six transformations $t_1, \ldots, t_6 \colon \mathbb{R}^2 \to \mathbb{R}^2$ there are scalars $c_1, \ldots, c_6 \in \mathbb{R}$ such that $c_1 t_1 + \cdots + c_6 t_6$ is the zero map. (*Hint:* the six is slightly misleading.)

1.14 The *trace* of a square matrix is the sum of the entries on the main diagonal (the $1,1$ entry plus the $2,2$ entry, etc.; we will see the significance of the trace in Chapter Five). Show that $\text{trace}(H + G) = \text{trace}(H) + \text{trace}(G)$. Is there a similar result for scalar multiplication?

1.15 Recall that the *transpose* of a matrix M is another matrix, whose i, j entry is the j, i entry of M. Verify these identities.

(a) $(G + H)^{\mathsf{T}} = G^{\mathsf{T}} + H^{\mathsf{T}}$

(b) $(r \cdot H)^{\mathsf{T}} = r \cdot H^{\mathsf{T}}$

✓ 1.16 A square matrix is *symmetric* if each i, j entry equals the j, i entry, that is, if the matrix equals its transpose.

(a) Prove that for any square H, the matrix $H + H^{\mathsf{T}}$ is symmetric. Does every symmetric matrix have this form?

(b) Prove that the set of $n \times n$ symmetric matrices is a subspace of $\mathcal{M}_{n \times n}$.

✓ 1.17 (a) How does matrix rank interact with scalar multiplication — can a scalar product of a rank n matrix have rank less than n? Greater?
(b) How does matrix rank interact with matrix addition — can a sum of rank n matrices have rank less than n? Greater?

IV.2 Matrix Multiplication

After representing addition and scalar multiplication of linear maps in the prior subsection, the natural next operation to consider is function composition.

2.1 Lemma The composition of linear maps is linear.

Proof *(Note: this argument has already appeared, as part of the proof of Theorem I.2.2.)* Let $h\colon V \to W$ and $g\colon W \to U$ be linear. The calculation

$$\begin{aligned} g \circ h\,(c_1 \cdot \vec{v}_1 + c_2 \cdot \vec{v}_2) &= g\big(\,h(c_1 \cdot \vec{v}_1 + c_2 \cdot \vec{v}_2)\,\big) = g\big(\,c_1 \cdot h(\vec{v}_1) + c_2 \cdot h(\vec{v}_2)\,\big) \\ &= c_1 \cdot g\big(h(\vec{v}_1)\big) + c_2 \cdot g\big(h(\vec{v}_2)\big) = c_1 \cdot (g \circ h)(\vec{v}_1) + c_2 \cdot (g \circ h)(\vec{v}_2) \end{aligned}$$

shows that $g \circ h\colon V \to U$ preserves linear combinations, and so is linear. QED

As we did with the operation of matrix addition and scalar multiplication, we will see how the representation of the composite relates to the representations of the compositors by first considering an example.

2.2 Example Let $h\colon \mathbb{R}^4 \to \mathbb{R}^2$ and $g\colon \mathbb{R}^2 \to \mathbb{R}^3$, fix bases $B \subset \mathbb{R}^4$, $C \subset \mathbb{R}^2$, $D \subset \mathbb{R}^3$, and let these be the representations.

$$H = \mathrm{Rep}_{B,C}(h) = \begin{pmatrix} 4 & 6 & 8 & 2 \\ 5 & 7 & 9 & 3 \end{pmatrix}_{B,C} \qquad G = \mathrm{Rep}_{C,D}(g) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}_{C,D}$$

To represent the composition $g \circ h\colon \mathbb{R}^4 \to \mathbb{R}^3$ we start with a $\vec{v}$, represent h of $\vec{v}$, and then represent g of that. The representation of $h(\vec{v})$ is the product of h's matrix and $\vec{v}$'s vector.

$$\mathrm{Rep}_C(\,h(\vec{v})\,) = \begin{pmatrix} 4 & 6 & 8 & 2 \\ 5 & 7 & 9 & 3 \end{pmatrix}_{B,C} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix}_B = \begin{pmatrix} 4v_1 + 6v_2 + 8v_3 + 2v_4 \\ 5v_1 + 7v_2 + 9v_3 + 3v_4 \end{pmatrix}_C$$

The representation of $g(\,h(\vec{v})\,)$ is the product of g's matrix and $h(\vec{v})$'s vector.

$$\mathrm{Rep}_D(\,g(h(\vec{v}))\,)=\begin{pmatrix}1&1\\0&1\\1&0\end{pmatrix}_{C,D}\begin{pmatrix}4v_1+6v_2+8v_3+2v_4\\5v_1+7v_2+9v_3+3v_4\end{pmatrix}_C$$

$$=\begin{pmatrix}1\cdot(4v_1+6v_2+8v_3+2v_4)+1\cdot(5v_1+7v_2+9v_3+3v_4)\\0\cdot(4v_1+6v_2+8v_3+2v_4)+1\cdot(5v_1+7v_2+9v_3+3v_4)\\1\cdot(4v_1+6v_2+8v_3+2v_4)+0\cdot(5v_1+7v_2+9v_3+3v_4)\end{pmatrix}_D$$

Distributing and regrouping on the v's gives

$$=\begin{pmatrix}(1\cdot4+1\cdot5)v_1+(1\cdot6+1\cdot7)v_2+(1\cdot8+1\cdot9)v_3+(1\cdot2+1\cdot3)v_4\\(0\cdot4+1\cdot5)v_1+(0\cdot6+1\cdot7)v_2+(0\cdot8+1\cdot9)v_3+(0\cdot2+1\cdot3)v_4\\(1\cdot4+0\cdot5)v_1+(1\cdot6+0\cdot7)v_2+(1\cdot8+0\cdot9)v_3+(1\cdot2+0\cdot3)v_4\end{pmatrix}_D$$

which is this matrix-vector product.

$$=\begin{pmatrix}1\cdot4+1\cdot5&1\cdot6+1\cdot7&1\cdot8+1\cdot9&1\cdot2+1\cdot3\\0\cdot4+1\cdot5&0\cdot6+1\cdot7&0\cdot8+1\cdot9&0\cdot2+1\cdot3\\1\cdot4+0\cdot5&1\cdot6+0\cdot7&1\cdot8+0\cdot9&1\cdot2+0\cdot3\end{pmatrix}_{B,D}\begin{pmatrix}v_1\\v_2\\v_3\\v_4\end{pmatrix}_D$$

The matrix representing $g\circ h$ has the rows of G combined with the columns of H.

2.3 Definition The *matrix-multiplicative product* of the $m\times r$ matrix G and the $r\times n$ matrix H is the $m\times n$ matrix P, where

$$p_{i,j}=g_{i,1}h_{1,j}+g_{i,2}h_{2,j}+\cdots+g_{i,r}h_{r,j}$$

so that the i,j-th entry of the product is the dot product of the i-th row of the first matrix with the j-th column of the second.

$$GH=\begin{pmatrix}&\vdots&&\\g_{i,1}&g_{i,2}&\cdots&g_{i,r}\\&\vdots&&\end{pmatrix}\begin{pmatrix}&h_{1,j}&\\\cdots&h_{2,j}&\cdots\\&\vdots&\\&h_{r,j}&\end{pmatrix}=\begin{pmatrix}&\vdots&\\\cdots&p_{i,j}&\cdots\\&\vdots&\end{pmatrix}$$

2.4 Example

$$\begin{pmatrix}2&0\\4&6\\8&2\end{pmatrix}\begin{pmatrix}1&3\\5&7\end{pmatrix}=\begin{pmatrix}2\cdot1+0\cdot5&2\cdot3+0\cdot7\\4\cdot1+6\cdot5&4\cdot3+6\cdot7\\8\cdot1+2\cdot5&8\cdot3+2\cdot7\end{pmatrix}=\begin{pmatrix}2&6\\34&54\\18&38\end{pmatrix}$$

2.5 Example The matrices from Example 2.2 combine in this way.

$$\begin{aligned}
\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}
\begin{pmatrix} 4 & 6 & 8 & 2 \\ 5 & 7 & 9 & 3 \end{pmatrix}
&= \begin{pmatrix} 1\cdot 4+1\cdot 5 & 1\cdot 6+1\cdot 7 & 1\cdot 8+1\cdot 9 & 1\cdot 2+1\cdot 3 \\ 0\cdot 4+1\cdot 5 & 0\cdot 6+1\cdot 7 & 0\cdot 8+1\cdot 9 & 0\cdot 2+1\cdot 3 \\ 1\cdot 4+0\cdot 5 & 1\cdot 6+0\cdot 7 & 1\cdot 8+0\cdot 9 & 1\cdot 2+0\cdot 3 \end{pmatrix} \\
&= \begin{pmatrix} 9 & 13 & 17 & 5 \\ 5 & 7 & 9 & 3 \\ 4 & 6 & 8 & 2 \end{pmatrix}
\end{aligned}$$

2.6 Theorem A composition of linear maps is represented by the matrix product of the representatives.

Proof This argument generalizes Example 2.2. Let $h\colon V\to W$ and $g\colon W\to X$ be represented by H and G with respect to bases $B\subset V$, $C\subset W$, and $D\subset X$, of sizes n, r, and m. For any $\vec{v}\in V$ the k-th component of $\operatorname{Rep}_C(\,h(\vec{v})\,)$ is

$$h_{k,1}v_1+\cdots+h_{k,n}v_n$$

and so the i-th component of $\operatorname{Rep}_D(\,g\circ h\,(\vec{v})\,)$ is this.

$$\begin{aligned}
g_{i,1}\cdot(h_{1,1}v_1+\cdots+h_{1,n}v_n)+g_{i,2}\cdot(h_{2,1}v_1+\cdots+h_{2,n}v_n)& \\
+\cdots+g_{i,r}\cdot(h_{r,1}v_1+\cdots+h_{r,n}v_n)&
\end{aligned}$$

Distribute and regroup on the v's.

$$\begin{aligned}
=(g_{i,1}h_{1,1}+g_{i,2}h_{2,1}+\cdots+g_{i,r}h_{r,1})\cdot v_1& \\
+\cdots+(g_{i,1}h_{1,n}+g_{i,2}h_{2,n}+\cdots+g_{i,r}h_{r,n})\cdot v_n&
\end{aligned}$$

Finish by recognizing that the coefficient of each v_j

$$g_{i,1}h_{1,j}+g_{i,2}h_{2,j}+\cdots+g_{i,r}h_{r,j}$$

matches the definition of the i,j entry of the product GH. QED

This *arrow diagram* pictures the relationship between maps and matrices ('wrt' abbreviates 'with respect to').

$$\begin{tikzcd}
& W_{\textit{wrt}\ C} \arrow[dr, "g", "G"'] & \\
V_{\textit{wrt}\ B} \arrow[ur, "h", "H"'] \arrow[rr, "g\circ h", "GH"'] & & X_{\textit{wrt}\ D}
\end{tikzcd}$$

Above the arrows, the maps show that the two ways of going from V to X, straight over via the composition or else in two steps by way of W, have the same effect

$$\vec{v} \xmapsto{g\circ h} g(h(\vec{v})) \qquad \vec{v} \xmapsto{h} h(\vec{v}) \xmapsto{g} g(h(\vec{v}))$$

(this is just the definition of composition). Below the arrows, the matrices indicate that multiplying GH into the column vector $\operatorname{Rep}_B(\vec{v})$ has the same effect as multiplying the column vector first by H and then multiplying the result by G.

$$\operatorname{Rep}_{B,D}(g\circ h) = GH \qquad \operatorname{Rep}_{C,D}(g)\operatorname{Rep}_{B,C}(h) = GH$$

2.7 Example Because the number of columns on the left does not equal the number of rows on the right, this product is not defined.

$$\begin{pmatrix} -1 & 2 & 0 \\ 0 & 10 & 1.1 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$$

The combination in the prior example is undefined because of the underlying maps. We require that the sizes match because we want that the underlying function composition is possible.

$$\text{dimension } n \text{ space} \xrightarrow{h} \text{dimension } r \text{ space} \xrightarrow{g} \text{dimension } m \text{ space} \qquad (*)$$

Thus, matrix product combines an $m\times r$ matrix G with an $r\times n$ matrix F to yield the $m\times n$ result GF. Briefly: $m\times r$ times $r\times n$ equals $m\times n$.

2.8 Remark The order in which we write things can be confusing. In '$m\times r$ times $r\times n$ equals $m\times n$' the number written first m is the dimension of g's codomain and is thus the number that appears last in the map dimension description $(*)$. The explanation is that while h is done first and is followed by g, we write the composition as $g\circ h$, with g on the left, from the notation $g(h(\vec{v}))$. (Some people try to lessen confusion by reading '$g\circ h$' aloud as "g following h.") That carries over to matrices, so that $g\circ h$ is represented by GH.

We can get insight into matrix-matrix product operation by studying how the entries combine. For instance, an alternative way to understand why we require above that the sizes match is that the row of the left-hand matrix must have the same number of entries as the column of the right-hand matrix, or else some entry will be left without a matching entry from the other matrix.

Another aspect of the combinatorics of matrix multiplication, in the sum defining the i,j entry, is brought out here by the boxing the equal subscripts.

$$p_{i,j} = g_{i,\boxed{1}}h_{\boxed{1},j} + g_{i,\boxed{2}}h_{\boxed{2},j} + \cdots + g_{i,\boxed{r}}h_{\boxed{r},j}$$

The highlighted subscripts on the g's are column indices while those on the h's are for rows. That is, the summation takes place over the columns of G but over the rows of H — the definition treats left differently than right. So we may reasonably suspect that GH can be unequal to HG.

2.9 Example Matrix multiplication is not commutative.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix} \qquad \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 23 & 34 \\ 31 & 46 \end{pmatrix}$$

2.10 Example Commutativity can fail more dramatically:

$$\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \end{pmatrix} = \begin{pmatrix} 23 & 34 & 0 \\ 31 & 46 & 0 \end{pmatrix}$$

while

$$\begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & 0 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$$

isn't even defined.

2.11 Remark The fact that matrix multiplication is not commutative can seem odd at first, perhaps because most mathematical operations in prior courses are commutative. But matrix multiplication represents function composition and function composition is not commutative: if $f(x) = 2x$ and $g(x) = x + 1$ then $g \circ f(x) = 2x + 1$ while $f \circ g(x) = 2(x + 1) = 2x + 2$.

Except for the lack of commutativity, matrix multiplication is algebraically well-behaved. The next result gives some nice properties and more are in Exercise 24 and Exercise 25.

2.12 Theorem If F, G, and H are matrices, and the matrix products are defined, then the product is associative $(FG)H = F(GH)$ and distributes over matrix addition $F(G + H) = FG + FH$ and $(G + H)F = GF + HF$.

Proof Associativity holds because matrix multiplication represents function composition, which is associative: the maps $(f \circ g) \circ h$ and $f \circ (g \circ h)$ are equal as both send $\vec{v}$ to $f(g(h(\vec{v})))$.

Distributivity is similar. For instance, the first one goes $f \circ (g + h)(\vec{v}) = f\big((g + h)(\vec{v})\big) = f\big(g(\vec{v}) + h(\vec{v})\big) = f(g(\vec{v})) + f(h(\vec{v})) = f \circ g(\vec{v}) + f \circ h(\vec{v})$ (the third equality uses the linearity of f). Right-distributivity goes the same way. QED

2.13 Remark We could instead prove that result by slogging through indices. For example, for associativity the i,j entry of $(FG)H$ is

$$\begin{aligned}&(f_{i,1}g_{1,1}+f_{i,2}g_{2,1}+\cdots+f_{i,r}g_{r,1})h_{1,j}\\ &\quad+(f_{i,1}g_{1,2}+f_{i,2}g_{2,2}+\cdots+f_{i,r}g_{r,2})h_{2,j}\\ &\quad\ \vdots\\ &\quad+(f_{i,1}g_{1,s}+f_{i,2}g_{2,s}+\cdots+f_{i,r}g_{r,s})h_{s,j}\end{aligned}$$

where F, G, and H are $m\times r$, $r\times s$, and $s\times n$ matrices. Distribute

$$\begin{aligned}&f_{i,1}g_{1,1}h_{1,j}+f_{i,2}g_{2,1}h_{1,j}+\cdots+f_{i,r}g_{r,1}h_{1,j}\\ &\quad+f_{i,1}g_{1,2}h_{2,j}+f_{i,2}g_{2,2}h_{2,j}+\cdots+f_{i,r}g_{r,2}h_{2,j}\\ &\quad\ \vdots\\ &\quad+f_{i,1}g_{1,s}h_{s,j}+f_{i,2}g_{2,s}h_{s,j}+\cdots+f_{i,r}g_{r,s}h_{s,j}\end{aligned}$$

and regroup around the f's

$$\begin{aligned}&f_{i,1}(g_{1,1}h_{1,j}+g_{1,2}h_{2,j}+\cdots+g_{1,s}h_{s,j})\\ &\quad+f_{i,2}(g_{2,1}h_{1,j}+g_{2,2}h_{2,j}+\cdots+g_{2,s}h_{s,j})\\ &\quad\ \vdots\\ &\quad+f_{i,r}(g_{r,1}h_{1,j}+g_{r,2}h_{2,j}+\cdots+g_{r,s}h_{s,j})\end{aligned}$$

to get the i,j entry of $F(GH)$.

Contrast the two proofs. The index-heavy argument is hard to understand in that while the calculations are easy to check, the arithmetic seems unconnected to any idea. The argument in the proof is shorter and also says why this property "really" holds. This illustrates the comments made at the start of the chapter on vector spaces — at least sometimes an argument from higher-level constructs is clearer.

We have now seen how to represent the composition of linear maps. The next subsection will continue to explore this operation.

Exercises

✓ 2.14 Compute, or state "not defined".

(a) $\begin{pmatrix}3&1\\-4&2\end{pmatrix}\begin{pmatrix}0&5\\0&0.5\end{pmatrix}$ **(b)** $\begin{pmatrix}1&1&-1\\4&0&3\end{pmatrix}\begin{pmatrix}2&-1&-1\\3&1&1\\3&1&1\end{pmatrix}$

(c) $\begin{pmatrix}2&-7\\7&4\end{pmatrix}\begin{pmatrix}1&0&5\\-1&1&1\\3&8&4\end{pmatrix}$ **(d)** $\begin{pmatrix}5&2\\3&1\end{pmatrix}\begin{pmatrix}-1&2\\3&-5\end{pmatrix}$

✓ 2.15 Where

$$A = \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 5 & 2 \\ 4 & 4 \end{pmatrix} \quad C = \begin{pmatrix} -2 & 3 \\ -4 & 1 \end{pmatrix}$$

compute or state 'not defined'.

(a) AB **(b)** $(AB)C$ **(c)** BC **(d)** $A(BC)$

2.16 Which products are defined?

(a) 3×2 times 2×3 **(b)** 2×3 times 3×2 **(c)** 2×2 times 3×3
(d) 3×3 times 2×2

✓ 2.17 Give the size of the product or state "not defined".

(a) a 2×3 matrix times a 3×1 matrix
(b) a 1×12 matrix times a 12×1 matrix
(c) a 2×3 matrix times a 2×1 matrix
(d) a 2×2 matrix times a 2×2 matrix

✓ 2.18 Find the system of equations resulting from starting with

$$\begin{aligned} h_{1,1}x_1 + h_{1,2}x_2 + h_{1,3}x_3 &= d_1 \\ h_{2,1}x_1 + h_{2,2}x_2 + h_{2,3}x_3 &= d_2 \end{aligned}$$

and making this change of variable (i.e., substitution).

$$\begin{aligned} x_1 &= g_{1,1}y_1 + g_{1,2}y_2 \\ x_2 &= g_{2,1}y_1 + g_{2,2}y_2 \\ x_3 &= g_{3,1}y_1 + g_{3,2}y_2 \end{aligned}$$

2.19 As Definition 2.3 points out, the matrix product operation generalizes the dot product. Is the dot product of a $1\times n$ row vector and a $n\times 1$ column vector the same as their matrix-multiplicative product?

✓ 2.20 Represent the derivative map on $\mathcal{P}_n$ with respect to B, B where B is the natural basis $\langle 1, x, \ldots, x^n\rangle$. Show that the product of this matrix with itself is defined; what map does it represent?

2.21 [Cleary] Match each type of matrix with all these descriptions that could fit: (i) can be multiplied by its transpose to make a 1×1 matrix, (ii) is similar to the 3×3 matrix of all zeros, (iii) can represent a linear map from $\mathbb{R}^3$ to $\mathbb{R}^2$ that is not onto, (iv) can represent an isomorphism from $\mathbb{R}^3$ to $\mathcal{P}^2$.

(a) a 2×3 matrix whose rank is 1
(b) a 3×3 matrix that is nonsingular
(c) a 2×2 matrix that is singular
(d) an $n\times 1$ column vector

2.22 Show that composition of linear transformations on $\mathbb{R}^1$ is commutative. Is this true for any one-dimensional space?

2.23 Why is matrix multiplication not defined as entry-wise multiplication? That would be easier, and commutative too.

✓ 2.24 **(a)** Prove that $H^pH^q = H^{p+q}$ and $(H^p)^q = H^{pq}$ for positive integers p, q.
(b) Prove that $(rH)^p = r^p \cdot H^p$ for any positive integer p and scalar $r \in \mathbb{R}$.

✓ 2.25 **(a)** How does matrix multiplication interact with scalar multiplication: is $r(GH) = (rG)H$? Is $G(rH) = r(GH)$?
(b) How does matrix multiplication interact with linear combinations: is $F(rG + sH) = r(FG) + s(FH)$? Is $(rF + sG)H = rFH + sGH$?

2.26 We can ask how the matrix product operation interacts with the transpose operation.

(a) Show that $(GH)^{\mathsf{T}} = H^{\mathsf{T}}G^{\mathsf{T}}$.

(b) A square matrix is *symmetric* if each i,j entry equals the j,i entry, that is, if the matrix equals its own transpose. Show that the matrices HH^{T} and $H^{\mathsf{T}}H$ are symmetric.

✓ 2.27 Rotation of vectors in $\mathbb{R}^3$ about an axis is a linear map. Show that linear maps do not commute by showing geometrically that rotations do not commute.

2.28 In the proof of Theorem 2.12 we used some maps. What are the domains and codomains?

2.29 How does matrix rank interact with matrix multiplication?

(a) Can the product of rank n matrices have rank less than n? Greater?

(b) Show that the rank of the product of two matrices is less than or equal to the minimum of the rank of each factor.

2.30 Is 'commutes with' an equivalence relation among $n \times n$ matrices?

✓ 2.31 *(We will use this exercise in the Matrix Inverses exercises.)* Here is another property of matrix multiplication that might be puzzling at first sight.

(a) Prove that the composition of the projections $\pi_x, \pi_y \colon \mathbb{R}^3 \to \mathbb{R}^3$ onto the x and y axes is the zero map despite that neither one is itself the zero map.

(b) Prove that the composition of the derivatives d^2/dx^2, $d^3/dx^3 \colon \mathcal{P}_4 \to \mathcal{P}_4$ is the zero map despite that neither is the zero map.

(c) Give a matrix equation representing the first fact.

(d) Give a matrix equation representing the second.

When two things multiply to give zero despite that neither is zero we say that each is a *zero divisor*.

2.32 Show that, for square matrices, $(S+T)(S-T)$ need not equal $S^2 - T^2$.

✓ 2.33 Represent the identity transformation $\mathrm{id} \colon V \to V$ with respect to B, B for any basis B. This is the *identity matrix* I. Show that this matrix plays the role in matrix multiplication that the number 1 plays in real number multiplication: $HI = IH = H$ (for all matrices H for which the product is defined).

2.34 In real number algebra, quadratic equations have at most two solutions. That is not so with matrix algebra. Show that the 2×2 matrix equation $T^2 = I$ has more than two solutions, where I is the identity matrix (this matrix has ones in its $1,1$ and $2,2$ entries and zeroes elsewhere; see Exercise 33).

2.35 **(a)** Prove that for any 2×2 matrix T there are scalars $c_0, \ldots, c_4$ that are not all 0 such that the combination $c_4T^4 + c_3T^3 + c_2T^2 + c_1T + c_0I$ is the zero matrix (where I is the 2×2 identity matrix, with 1's in its $1,1$ and $2,2$ entries and zeroes elsewhere; see Exercise 33).

(b) Let $p(x)$ be a polynomial $p(x) = c_nx^n + \cdots + c_1x + c_0$. If T is a square matrix we define $p(T)$ to be the matrix $c_nT^n + \cdots + c_1T + c_0I$ (where I is the appropriately-sized identity matrix). Prove that for any square matrix there is a polynomial such that $p(T)$ is the zero matrix.

(c) The *minimal polynomial* $m(x)$ of a square matrix is the polynomial of least degree, and with leading coefficient 1, such that $m(T)$ is the zero matrix. Find the minimal polynomial of this matrix.

$$\begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix}$$

(This is the representation with respect to $\mathcal{E}_2, \mathcal{E}_2$, the standard basis, of a rotation through $\pi/6$ radians counterclockwise.)

2.36 The infinite-dimensional space $\mathcal{P}$ of all finite-degree polynomials gives a memorable example of the non-commutativity of linear maps. Let $d/dx\colon \mathcal{P} \to \mathcal{P}$ be the usual derivative and let $s\colon \mathcal{P} \to \mathcal{P}$ be the *shift* map.

$$a_0 + a_1x + \cdots + a_nx^n \;\stackrel{s}{\longmapsto}\; 0 + a_0x + a_1x^2 + \cdots + a_nx^{n+1}$$

Show that the two maps don't commute $d/dx \circ s \neq s \circ d/dx$; in fact, not only is $(d/dx \circ s) - (s \circ d/dx)$ not the zero map, it is the identity map.

2.37 Recall the notation for the sum of the sequence of numbers $a_1, a_2, \ldots, a_n$.

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n$$

In this notation, the i, j entry of the product of G and H is this.

$$p_{i,j} = \sum_{k=1}^{r} g_{i,k}h_{k,j}$$

Using this notation,

(a) reprove that matrix multiplication is associative;

(b) reprove Theorem 2.6.

IV.3 Mechanics of Matrix Multiplication

We can consider matrix multiplication as a mechanical process, putting aside for the moment any implications about the underlying maps.

The striking thing about this operation is the way that rows and columns combine. The i, j entry of the matrix product is the dot product of row i of the left matrix with column j of the right one. For instance, here a second row and a third column combine to make a $2, 3$ entry.

$$\begin{pmatrix} 1 & 1 \\ \boxed{0 \quad 1} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 4 & 6 & \boxed{\begin{matrix} 8 \\ 9 \end{matrix}} & 2 \\ 5 & 7 & & 3 \end{pmatrix} = \begin{pmatrix} 9 & 13 & 17 & 5 \\ 5 & 7 & \boxed{9} & 3 \\ 4 & 6 & 8 & 2 \end{pmatrix}$$

We can view this as the left matrix acting by multiplying its rows into the columns of the right matrix. Or, it is the right matrix using its columns to act on the rows of the left matrix. Below, we will examine actions from the left and from the right for some simple matrices.

Simplest is the zero matrix.

3.1 Example Multiplying by a zero matrix from the left or from the right results in a zero matrix.

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 3 & 2 \\ -1 & 1 & -1 \end{pmatrix}=\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 2 & 3 \\ 1 & 4 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}=\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

The next easiest matrices are the ones with a single nonzero entry.

3.2 Definition A matrix with all 0's except for a 1 in the i, j entry is an i, j *unit* matrix (or *matrix unit*).

3.3 Example This is the $1, 2$ unit matrix with three rows and two columns, multiplying from the left.

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}=\begin{pmatrix} 7 & 8 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Acting from the left, an i, j unit matrix copies row j of the multiplicand into row i of the result. From the right an i, j unit matrix picks out column i of the multiplicand and copies it into column j of the result.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}=\begin{pmatrix} 0 & 1 \\ 0 & 4 \\ 0 & 7 \end{pmatrix}$$

3.4 Example Rescaling unit matrices simply rescales the result. This is the action from the left of the matrix that is twice the one in the prior example.

$$\begin{pmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}=\begin{pmatrix} 14 & 16 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Next in complication are matrices with two nonzero entries.

3.5 Example There are two cases. If a left-multiplier has entries in different rows then their actions don't interact.

$$\begin{aligned}
\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} &= (\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}+\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix})\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}+\begin{pmatrix} 0 & 0 & 0 \\ 14 & 16 & 18 \\ 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 2 & 3 \\ 14 & 16 & 18 \\ 0 & 0 & 0 \end{pmatrix}
\end{aligned}$$

But if the left-multiplier's nonzero entries are in the same row then that row of the result is a combination.

$$\begin{pmatrix}1&0&2\\0&0&0\\0&0&0\end{pmatrix}\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}=(\begin{pmatrix}1&0&0\\0&0&0\\0&0&0\end{pmatrix}+\begin{pmatrix}0&0&2\\0&0&0\\0&0&0\end{pmatrix})\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}$$
$$=\begin{pmatrix}1&2&3\\0&0&0\\0&0&0\end{pmatrix}+\begin{pmatrix}14&16&18\\0&0&0\\0&0&0\end{pmatrix}$$
$$=\begin{pmatrix}15&18&21\\0&0&0\\0&0&0\end{pmatrix}$$

Right-multiplication acts in the same way, but with columns.

3.6 Example Consider the columns of the product of two 2×2 matrices.

$$\begin{pmatrix}g_{1,1}&g_{1,2}\\g_{2,1}&g_{2,2}\end{pmatrix}\begin{pmatrix}h_{1,1}&h_{1,2}\\h_{2,1}&h_{2,2}\end{pmatrix}=\begin{pmatrix}g_{1,1}h_{1,1}+g_{1,2}h_{2,1}&g_{1,1}h_{1,2}+g_{1,2}h_{2,2}\\g_{2,1}h_{1,1}+g_{2,2}h_{2,1}&g_{2,1}h_{1,2}+g_{2,2}h_{2,2}\end{pmatrix}$$

Each column is the result of multiplying G by the corresponding column of H.

$$G\begin{pmatrix}h_{1,1}\\h_{2,1}\end{pmatrix}=\begin{pmatrix}g_{1,1}h_{1,1}+g_{1,2}h_{2,1}\\g_{2,1}h_{1,1}+g_{2,2}h_{2,1}\end{pmatrix}\qquad G\begin{pmatrix}h_{1,2}\\h_{2,2}\end{pmatrix}=\begin{pmatrix}g_{1,1}h_{1,2}+g_{1,2}h_{2,2}\\g_{2,1}h_{1,2}+g_{2,2}h_{2,2}\end{pmatrix}$$

3.7 Lemma In a product of two matrices G and H, the columns of GH are formed by taking G times the columns of H

$$G\cdot\begin{pmatrix}\vdots&&\vdots\\\vec{h}_1&\cdots&\vec{h}_n\\\vdots&&\vdots\end{pmatrix}=\begin{pmatrix}\vdots&&\vdots\\G\cdot\vec{h}_1&\cdots&G\cdot\vec{h}_n\\\vdots&&\vdots\end{pmatrix}$$

and the rows of GH are formed by taking the rows of G times H

$$\begin{pmatrix}\cdots&\vec{g}_1&\cdots\\&\vdots&\\\cdots&\vec{g}_r&\cdots\end{pmatrix}\cdot H=\begin{pmatrix}\cdots&\vec{g}_1\cdot H&\cdots\\&\vdots&\\\cdots&\vec{g}_r\cdot H&\cdots\end{pmatrix}$$

(ignoring the extra parentheses).

Proof We will check that in a product of 2×2 matrices, the rows of the product equal the product of the rows of G with the entire matrix H.

$$\begin{pmatrix} g_{1,1} & g_{1,2} \\ g_{2,1} & g_{2,2} \end{pmatrix} \begin{pmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{pmatrix} = \begin{pmatrix} (g_{1,1} \;\; g_{1,2})H \\ (g_{2,1} \;\; g_{2,2})H \end{pmatrix}$$

$$= \begin{pmatrix} (g_{1,1}h_{1,1} + g_{1,2}h_{2,1} & g_{1,1}h_{1,2} + g_{1,2}h_{2,2}) \\ (g_{2,1}h_{1,1} + g_{2,2}h_{2,1} & g_{2,1}h_{1,2} + g_{2,2}h_{2,2}) \end{pmatrix}$$

We leave the more general check as an exercise. QED

An application of those observations is that there is a matrix that just copies out the rows and columns.

3.8 Definition The *main diagonal* (or *principle diagonal* or *diagonal*) of a square matrix goes from the upper left to the lower right.

3.9 Definition An *identity matrix* is square and every entry is 0 except for 1's in the main diagonal.

$$I_{n \times n} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & \vdots & & \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

3.10 Example Here is the 2×2 identity matrix leaving its multiplicand unchanged when it acts from the right.

$$\begin{pmatrix} 1 & -2 \\ 0 & -2 \\ 1 & -1 \\ 4 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 0 & -2 \\ 1 & -1 \\ 4 & 3 \end{pmatrix}$$

3.11 Example Here the 3×3 identity leaves its multiplicand unchanged both from the left

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 6 \\ 1 & 3 & 8 \\ -7 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 6 \\ 1 & 3 & 8 \\ -7 & 1 & 0 \end{pmatrix}$$

and from the right.

$$\begin{pmatrix} 2 & 3 & 6 \\ 1 & 3 & 8 \\ -7 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 6 \\ 1 & 3 & 8 \\ -7 & 1 & 0 \end{pmatrix}$$

In short, an identity matrix is the identity element of the set of $n\times n$ matrices with respect to the operation of matrix multiplication.

We can generalize the identity matrix by relaxing the ones to arbitrary reals. The resulting matrix rescales whole rows or columns.

3.12 Definition A *diagonal matrix* is square and has 0's off the main diagonal.

$$\begin{pmatrix} a_{1,1} & 0 & \ldots & 0 \\ 0 & a_{2,2} & \ldots & 0 \\ & \vdots & & \\ 0 & 0 & \ldots & a_{n,n} \end{pmatrix}$$

3.13 Example From the left, the action of multiplication by a diagonal matrix is to rescales the rows.

$$\begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 2 & 1 & 4 & -1 \\ -1 & 3 & 4 & 4 \end{pmatrix}=\begin{pmatrix} 4 & 2 & 8 & -2 \\ 1 & -3 & -4 & -4 \end{pmatrix}$$

From the right such a matrix rescales the columns.

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 2 & 2 \end{pmatrix}\begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{pmatrix}=\begin{pmatrix} 3 & 4 & -2 \\ 6 & 4 & -4 \end{pmatrix}$$

We can also generalize identity matrices by putting a single one in each row and column in ways other than putting them down the diagonal.

3.14 Definition A *permutation matrix* is square and is all 0's except for a single 1 in each row and column.

3.15 Example From the left these matrices permute rows.

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}=\begin{pmatrix} 7 & 8 & 9 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

From the right they permute columns.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}=\begin{pmatrix} 2 & 3 & 1 \\ 5 & 6 & 4 \\ 8 & 9 & 7 \end{pmatrix}$$

We finish this subsection by applying these observations to get matrices that perform Gauss's Method and Gauss-Jordan reduction. We have already seen how to produce a matrix that rescales rows, and a row swapper.

3.16 Example Multiplying by this matrix rescales the second row by three.

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

3.17 Example This multiplication swaps the first and third rows.

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

To see how to perform a row combination, we observe something about those two examples. The matrix that rescales the second row by a factor of three arises in this way from the identity.

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\xrightarrow{3\rho_2}\begin{pmatrix}1&0&0\\0&3&0\\0&0&1\end{pmatrix}$$

Similarly, the matrix that swaps first and third rows arises in this way.

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\xrightarrow{\rho_1\leftrightarrow\rho_3}\begin{pmatrix}0&0&1\\0&1&0\\1&0&0\end{pmatrix}$$

3.18 Example The 3×3 matrix that arises as

$$\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\xrightarrow{-2\rho_2+\rho_3}\begin{pmatrix}1&0&0\\0&1&0\\0&-2&1\end{pmatrix}$$

will, when it acts from the left, perform the combination operation $-2\rho_2+\rho_3$.

$$\begin{pmatrix}1&0&0\\0&1&0\\0&-2&1\end{pmatrix}\begin{pmatrix}1&0&2&0\\0&1&3&-3\\0&2&1&1\end{pmatrix}=\begin{pmatrix}1&0&2&0\\0&1&3&-3\\0&0&-5&7\end{pmatrix}$$

3.19 Definition The *elementary reduction matrices* (or just *elementary matrices*) result from applying a one Gaussian operation to an identity matrix.

(1) $I\xrightarrow{k\rho_i} M_i(k)$ for $k\neq 0$

(2) $I\xrightarrow{\rho_i\leftrightarrow\rho_j} P_{i,j}$ for $i\neq j$

(3) $I\xrightarrow{k\rho_i+\rho_j} C_{i,j}(k)$ for $i\neq j$

3.20 Lemma Matrix multiplication can do Gaussian reduction.

(1) If $H \xrightarrow{k\rho_i} G$ then $M_i(k)H = G$.

(2) If $H \xrightarrow{\rho_i \leftrightarrow \rho_j} G$ then $P_{i,j}H = G$.

(3) If $H \xrightarrow{k\rho_i + \rho_j} G$ then $C_{i,j}(k)H = G$.

Proof Clear. QED

3.21 Example This is the first system, from the first chapter, on which we performed Gauss's Method.

$$\begin{aligned} 3x_3 &= 9 \\ x_1 + 5x_2 - 2x_3 &= 2 \\ (1/3)x_1 + 2x_2 \qquad &= 3 \end{aligned}$$

We can reduce it with matrix multiplication. Swap the first and third rows,

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \left(\begin{array}{ccc|c} 0 & 0 & 3 & 9 \\ 1 & 5 & -2 & 2 \\ 1/3 & 2 & 0 & 3 \end{array}\right) = \left(\begin{array}{ccc|c} 1/3 & 2 & 0 & 3 \\ 1 & 5 & -2 & 2 \\ 0 & 0 & 3 & 9 \end{array}\right)$$

triple the first row,

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \left(\begin{array}{ccc|c} 1/3 & 2 & 0 & 3 \\ 1 & 5 & -2 & 2 \\ 0 & 0 & 3 & 9 \end{array}\right) = \left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 1 & 5 & -2 & 2 \\ 0 & 0 & 3 & 9 \end{array}\right)$$

and then add -1 times the first row to the second.

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 1 & 5 & -2 & 2 \\ 0 & 0 & 3 & 9 \end{array}\right) = \left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 0 & -1 & -2 & -7 \\ 0 & 0 & 3 & 9 \end{array}\right)$$

Now back substitution will give the solution.

3.22 Example Gauss-Jordan reduction works the same way. For the matrix ending the prior example, first turn the leading entries to ones,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1/3 \end{pmatrix} \left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 0 & -1 & -2 & -7 \\ 0 & 0 & 3 & 9 \end{array}\right) = \left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

then clear the third column, and then the second column.

$$\begin{pmatrix} 1 & -6 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix} \left(\begin{array}{ccc|c} 1 & 6 & 0 & 9 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 1 & 3 \end{array}\right) = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

3.23 Corollary For any matrix H there are elementary reduction matrices $R_1, \ldots, R_r$ such that $R_r \cdot R_{r-1} \cdots R_1 \cdot H$ is in reduced echelon form.

Until now we have taken the point of view that our primary objects of study are vector spaces and the maps between them, and we seemed to have adopted matrices only for computational convenience. This subsection show that this isn't the entire story.

Understanding matrices operations by understanding the mechanics of how the entries combine is also useful. In the rest of this book we shall continue to focus on maps as the primary objects but we will be pragmatic — if the matrix point of view gives some clearer idea then we will go with it.

Exercises

✓ 3.24 Predict the result of each multiplication by an elementary reduction matrix, and then check by multiplying it out.

(a) $\begin{pmatrix}3&0\\0&1\end{pmatrix}\begin{pmatrix}1&2\\3&4\end{pmatrix}$ (b) $\begin{pmatrix}1&0\\0&2\end{pmatrix}\begin{pmatrix}1&2\\3&4\end{pmatrix}$ (c) $\begin{pmatrix}1&0\\-2&1\end{pmatrix}\begin{pmatrix}1&2\\3&4\end{pmatrix}$

(d) $\begin{pmatrix}1&2\\3&4\end{pmatrix}\begin{pmatrix}1&-1\\0&1\end{pmatrix}$ (e) $\begin{pmatrix}1&2\\3&4\end{pmatrix}\begin{pmatrix}0&1\\1&0\end{pmatrix}$

3.25 Predict the result of each multiplication by a diagonal matrix, and then check by multiplying it out.

(a) $\begin{pmatrix}3&0\\0&0\end{pmatrix}\begin{pmatrix}1&2\\3&4\end{pmatrix}$ (b) $\begin{pmatrix}4&0\\0&2\end{pmatrix}\begin{pmatrix}1&2\\3&4\end{pmatrix}$

3.26 Produce each.

(a) a 3×3 matrix that, acting from the left, swaps rows one and two

(b) a 2×2 matrix that, acting from the right, swaps column one and two

✓ 3.27 This table gives the number of hours of each type done by each worker, and the associated pay rates. Use matrices to compute the wages due.

	regular	*overtime*
Alan	40	12
Betty	35	6
Catherine	40	18
Donald	28	0

	wage
regular	\$25.00
overtime	\$45.00

Remark. This illustrates that in practice we often want to compute linear combinations of rows and columns in a context where we really aren't interested in any associated linear maps.

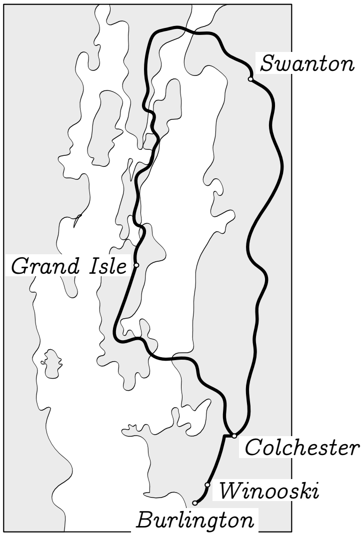
3.28 Find the product of this matrix with its transpose.

$$\begin{pmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{pmatrix}$$

✓ 3.29 The need to take linear combinations of rows and columns in tables of numbers arises often in practice. For instance, this is a map of part of Vermont and New

York.

In part because of Lake Champlain, there are no roads directly connecting some pairs of towns. For instance, there is no way to go from Winooski to Grand Isle without going through Colchester. (To simplify the graph many other roads and towns have been omitted. From top to bottom of this map is about forty miles.)



(a) The *adjacency matrix* of a map is the square matrix whose i, j entry is the number of roads from city i to city j. Produce the incidence matrix of this map (take the cities in alphabetical order).

(b) A matrix is *symmetric* if it equals its transpose. Show that an adjacency matrix is symmetric. (These are all two-way streets. Vermont doesn't have many one-way streets.)

(c) What is the significance of the square of the incidence matrix? The cube?

✓ 3.30 Prove that the diagonal matrices form a subspace of $\mathcal{M}_{n\times n}$. What is its dimension?

3.31 Does the identity matrix represent the identity map if the bases are unequal?

3.32 Show that every multiple of the identity commutes with every square matrix. Are there other matrices that commute with all square matrices?

3.33 Prove or disprove: nonsingular matrices commute.

✓ 3.34 Show that the product of a permutation matrix and its transpose is an identity matrix.

3.35 Show that if the first and second rows of G are equal then so are the first and second rows of GH. Generalize.

3.36 Describe the product of two diagonal matrices.

3.37 Write

$$\begin{pmatrix} 1 & 0 \\ -3 & 3 \end{pmatrix}$$

as the product of two elementary reduction matrices.

✓ 3.38 Show that if G has a row of zeros then GH (if defined) has a row of zeros. Does that work for columns?

3.39 Show that the set of unit matrices forms a basis for $\mathcal{M}_{n\times m}$.

3.40 Find the formula for the n-th power of this matrix.

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$$

✓ 3.41 The *trace* of a square matrix is the sum of the entries on its diagonal (its significance appears in Chapter Five). Show that $\operatorname{Tr}(GH) = \operatorname{Tr}(HG)$.

✓ 3.42 A square matrix is *upper triangular* if its only nonzero entries lie above, or on, the diagonal. Show that the product of two upper triangular matrices is upper triangular. Does this hold for lower triangular also?

3.43 A square matrix is a *Markov matrix* if each entry is between zero and one and the sum along each row is one. Prove that a product of Markov matrices is Markov.

✓ 3.44 Give an example of two matrices of the same rank and size with squares of differing rank.

3.45 Combine the two generalizations of the identity matrix, the one allowing entries to be other than ones, and the one allowing the single one in each row and column to be off the diagonal. What is the action of this type of matrix?

3.46 On a computer multiplications have traditionally been more costly than additions, so people have tried to in reduce the number of multiplications used to compute a matrix product.

(a) How many real number multiplications do we need in the formula we gave for the product of a $m \times r$ matrix and a $r \times n$ matrix?

(b) Matrix multiplication is associative, so all associations yield the same result. The cost in number of multiplications, however, varies. Find the association requiring the fewest real number multiplications to compute the matrix product of a 5×10 matrix, a 10×20 matrix, a 20×5 matrix, and a 5×1 matrix.

(c) *(Very hard.)* Find a way to multiply two 2×2 matrices using only seven multiplications instead of the eight suggested by the naive approach.

? 3.47 [Putnam, 1990, A-5] If A and B are square matrices of the same size such that $ABAB = 0$, does it follow that $BABA = 0$?

3.48 [Am. Math. Mon., Dec. 1966] Demonstrate these four assertions to get an alternate proof that column rank equals row rank.

(a) $\vec{y} \cdot \vec{y} = 0$ iff $\vec{y} = \vec{0}$.

(b) $A\vec{x} = \vec{0}$ iff $A^{\mathsf{T}}A\vec{x} = \vec{0}$.

(c) $\dim(\mathscr{R}(A)) = \dim(\mathscr{R}(A^{\mathsf{T}}A))$.

(d) $\text{col rank}(A) = \text{col rank}(A^{\mathsf{T}}) = \text{row rank}(A)$.

3.49 [Ackerson] Prove (where A is an $n \times n$ matrix and so defines a transformation of any n-dimensional space V with respect to B, B where B is a basis) that $\dim(\mathscr{R}(A) \cap \mathscr{N}(A)) = \dim(\mathscr{R}(A)) - \dim(\mathscr{R}(A^2))$. Conclude

(a) $\mathscr{N}(A) \subset \mathscr{R}(A)$ iff $\dim(\mathscr{N}(A)) = \dim(\mathscr{R}(A)) - \dim(\mathscr{R}(A^2))$;

(b) $\mathscr{R}(A) \subseteq \mathscr{N}(A)$ iff $A^2 = 0$;

(c) $\mathscr{R}(A) = \mathscr{N}(A)$ iff $A^2 = 0$ and $\dim(\mathscr{N}(A)) = \dim(\mathscr{R}(A))$;

(d) $\dim(\mathscr{R}(A) \cap \mathscr{N}(A)) = 0$ iff $\dim(\mathscr{R}(A)) = \dim(\mathscr{R}(A^2))$;

(e) *(Requires the Direct Sum subsection, which is optional.)* $V = \mathscr{R}(A) \oplus \mathscr{N}(A)$ iff $\dim(\mathscr{R}(A)) = \dim(\mathscr{R}(A^2))$.

IV.4 Inverses

We finish this section by considering how to represent the inverse of a linear map.

We first recall some things about inverses. Where $\pi\colon \mathbb{R}^3 \to \mathbb{R}^2$ is the projection map and $\iota\colon \mathbb{R}^2 \to \mathbb{R}^3$ is the embedding

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \xmapsto{\pi} \begin{pmatrix} x \\ y \end{pmatrix} \qquad \begin{pmatrix} x \\ y \end{pmatrix} \xmapsto{\iota} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

then the composition $\pi \circ \iota$ is the identity map $\pi \circ \iota = \mathrm{id}$ on $\mathbb{R}^2$.

$$\begin{pmatrix} x \\ y \end{pmatrix} \xmapsto{\iota} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \xmapsto{\pi} \begin{pmatrix} x \\ y \end{pmatrix}$$

We say that ι is a *right inverse* of π or, what is the same thing, that π is a *left inverse* of ι. However, composition in the other order $\iota \circ \pi$ doesn't give the identity map — here is a vector that is not sent to itself under $\iota \circ \pi$.

$$\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \xmapsto{\pi} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \xmapsto{\iota} \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

In fact, π has no left inverse at all. For, if f were to be a left inverse of π then we would have

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \xmapsto{\pi} \begin{pmatrix} x \\ y \end{pmatrix} \xmapsto{f} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

for all of the infinitely many z's. But a function f cannot send a single argument $\binom{x}{y}$ to more than one value.

So a function can have a right inverse but no left inverse, or a left inverse but no right inverse. A function can also fail to have an inverse on either side; one example is the zero transformation on $\mathbb{R}^2$.

Some functions have a *two-sided inverse*, another function that is the inverse both from the left and from the right. For instance, the transformation given by $\vec{v} \mapsto 2 \cdot \vec{v}$ has the two-sided inverse $\vec{v} \mapsto (1/2) \cdot \vec{v}$. The appendix shows that a function has a two-sided inverse if and only if it is both one-to-one and onto. The appendix also shows that if a function f has a two-sided inverse then it is unique, so we call it 'the' inverse and write f^{-1}.

In addition, recall that we have shown in Theorem II.2.20 that if a linear map has a two-sided inverse then that inverse is also linear.

Thus, our goal in this subsection is, where a linear h has an inverse, to find the relationship between $\text{Rep}_{B,D}(h)$ and $\text{Rep}_{D,B}(h^{-1})$.

4.1 Definition A matrix G is a *left inverse matrix* of the matrix H if GH is the identity matrix. It is a *right inverse* if HG is the identity. A matrix H with a two-sided inverse is an *invertible matrix*. That two-sided inverse is denoted H^{-1}.

Because of the correspondence between linear maps and matrices, statements about map inverses translate into statements about matrix inverses.

4.2 Lemma If a matrix has both a left inverse and a right inverse then the two are equal.

4.3 Theorem A matrix is invertible if and only if it is nonsingular.

Proof *(For both results.)* Given a matrix H, fix spaces of appropriate dimension for the domain and codomain and fix bases for these spaces. With respect to these bases, H represents a map h. The statements are true about the map and therefore they are true about the matrix. QED

4.4 Lemma A product of invertible matrices is invertible: if G and H are invertible and GH is defined then GH is invertible and $(GH)^{-1} = H^{-1}G^{-1}$.

Proof Because the two matrices are invertible they are square, and because their product is defined they must both be $n \times n$. Fix spaces and bases — say, $\mathbb{R}^n$ with the standard bases — to get maps $g, h\colon \mathbb{R}^n \to \mathbb{R}^n$ that are associated with the matrices, $G = \text{Rep}_{\mathcal{E}_n,\mathcal{E}_n}(g)$ and $H = \text{Rep}_{\mathcal{E}_n,\mathcal{E}_n}(h)$.

Consider $h^{-1}g^{-1}$. By the prior paragraph this composition is defined. This map is a two-sided inverse of gh since $(h^{-1}g^{-1})(gh) = h^{-1}(\text{id})h = h^{-1}h = \text{id}$ and $(gh)(h^{-1}g^{-1}) = g(\text{id})g^{-1} = gg^{-1} = \text{id}$. The matrices representing the maps reflect this equality. QED

This is the arrow diagram giving the relationship between map inverses and matrix inverses. It is a special case of the diagram relating function composition to matrix multiplication.

$$\begin{tikzcd} & W_{\text{wrt } C} \arrow[dr, "h^{-1}", "H^{-1}"'] & \\ V_{\text{wrt } B} \arrow[ur, "h", "H"'] \arrow[rr, "\text{id}", "I"'] & & V_{\text{wrt } B} \end{tikzcd}$$

Beyond its place in our program of seeing how to represent map operations, another reason for our interest in inverses comes from linear systems. A linear system is equivalent to a matrix equation, as here.

$$\begin{aligned} x_1 + x_2 &= 3 \\ 2x_1 - x_2 &= 2 \end{aligned} \quad\Longleftrightarrow\quad \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

By fixing spaces and bases (for instance, $\mathbb{R}^2, \mathbb{R}^2$ with the standard bases), we take the matrix H to represent a map h. The matrix equation then becomes this linear map equation.

$$h(\vec{x}) = \vec{d}$$

If we had a left inverse map g then we could apply it to both sides $g \circ h(\vec{x}) = g(\vec{d})$ to get $\vec{x} = g(\vec{d})$. Restating in terms of the matrices, we want to multiply by the inverse matrix $\mathrm{Rep}_{C,B}(g) \cdot \mathrm{Rep}_C(\vec{d})$ to get $\mathrm{Rep}_B(\vec{x})$.

4.5 Example We can find a left inverse for the matrix just given

$$\begin{pmatrix} m & n \\ p & q \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

by using Gauss's Method to solve the resulting linear system.

$$\begin{array}{rcl} m + 2n & & = 1 \\ m - n & & = 0 \\ & p + 2q & = 0 \\ & p - q & = 1 \end{array}$$

Answer: $m = 1/3$, $n = 1/3$, $p = 2/3$, and $q = -1/3$. (This matrix is actually the two-sided inverse of H; the check is easy.) With it, we can solve the system from the prior example.

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 1/3 \\ 2/3 & -1/3 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 5/3 \\ 4/3 \end{pmatrix}$$

4.6 Remark Why do inverse matrices when we have Gauss's Method? Beyond the conceptual appeal of representing the map inverse operation, solving linear systems this way has two advantages.

First, once we have done the work of finding an inverse then solving a system with the same coefficients but different constants is fast: if we change the constants on the right of the system above then we get a related problem

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 5 \\ 1 \end{pmatrix}$$

that our inverse method solves quickly.

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1/3 & 1/3 \\ 2/3 & -1/3 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Another advantage of inverses is that we can explore a system's sensitivity to changes in the constants. For example, tweaking the 3 on the right of the prior example's system to

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3.01 \\ 2 \end{pmatrix}$$

and solving with the inverse

$$\begin{pmatrix} 1/3 & 1/3 \\ 2/3 & -1/3 \end{pmatrix} \begin{pmatrix} 3.01 \\ 2 \end{pmatrix} = \begin{pmatrix} (1/3)(3.01) + (1/3)(2) \\ (2/3)(3.01) - (1/3)(2) \end{pmatrix}$$

shows that the first component of the solution changes by $1/3$ of the tweak, while the second component moves by $2/3$ of the tweak. This is *sensitivity analysis*. We could use it to decide how accurately we must specify the data in a linear model to ensure that the solution has a desired accuracy.

4.7 Lemma A matrix H is invertible if and only if it can be written as the product of elementary reduction matrices. We can compute the inverse by applying to the identity matrix the same row steps, in the same order, that Gauss-Jordan reduce H.

Proof The matrix H is invertible if and only if it is nonsingular and thus Gauss-Jordan reduces to the identity. By Corollary 3.23 we can do this reduction with elementary matrices.

$$R_r \cdot R_{r-1} \ldots R_1 \cdot H = I \qquad (*)$$

For the first sentence of the result, note that elementary matrices are invertible because elementary row operations are reversible, and that their inverses are also elementary. Apply R_r^{-1} from the left to both sides of $(*)$. Then apply R_{r-1}^{-1}, etc. The result gives H as the product of elementary matrices $H = R_1^{-1} \cdots R_r^{-1} \cdot I$. (The I there covers the case $r = 0$.)

For the second sentence, group $(*)$ as $(R_r \cdot R_{r-1} \ldots R_1) \cdot H = I$ and recognize what's in the parentheses as the inverse $H^{-1} = R_r \cdot R_{r-1} \ldots R_1 \cdot I$. Restated: applying R_1 to the identity, followed by R_2, etc., yields the inverse of H. QED

4.8 Example To find the inverse of

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}$$

do Gauss-Jordan reduction, meanwhile performing the same operations on the identity. For clerical convenience we write the matrix and the identity side-by-side and do the reduction steps together.

$$\begin{aligned}
\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right)
&\xrightarrow{-2\rho_1+\rho_2} \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & -3 & -2 & 1 \end{array}\right) \\
&\xrightarrow{-1/3\rho_2} \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & 2/3 & -1/3 \end{array}\right) \\
&\xrightarrow{-\rho_2+\rho_1} \left(\begin{array}{cc|cc} 1 & 0 & 1/3 & 1/3 \\ 0 & 1 & 2/3 & -1/3 \end{array}\right)
\end{aligned}$$

This calculation has found the inverse.

$$\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 1/3 & 1/3 \\ 2/3 & -1/3 \end{pmatrix}$$

4.9 Example This one happens to start with a row swap.

$$\begin{aligned}
\left(\begin{array}{ccc|ccc} 0 & 3 & -1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \end{array}\right)
&\xrightarrow{\rho_1\leftrightarrow\rho_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 3 & -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 1 \end{array}\right) \\
&\xrightarrow{-\rho_1+\rho_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 3 & -1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & -1 & 1 \end{array}\right) \\
&\quad\vdots \\
&\longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/4 & 1/4 & 3/4 \\ 0 & 1 & 0 & 1/4 & 1/4 & -1/4 \\ 0 & 0 & 1 & -1/4 & 3/4 & -3/4 \end{array}\right)
\end{aligned}$$

4.10 Example This algorithm detects a non-invertible matrix when the left half won't reduce to the identity.

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 2 & 2 & 0 & 1 \end{array}\right) \xrightarrow{-2\rho_1+\rho_2} \left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{array}\right)$$

With this procedure we can give a formula for the inverse of a general 2×2 matrix, which is worth memorizing.

4.11 Corollary The inverse for a 2×2 matrix exists and equals

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

if and only if $ad - bc \neq 0$.

Proof This computation is Exercise 21. QED

We have seen in this subsection, as in the subsection on Mechanics of Matrix Multiplication, how to exploit the correspondence between linear maps and matrices. We can fruitfully study both maps and matrices, translating back and forth to use whichever is handiest.

Over the course of this entire section we have developed an algebra system for matrices. We can compare it with the familiar algebra of real numbers. Matrix addition and subtraction work in much the same way as the real number operations except that they only combine same-sized matrices. Scalar multiplication is in some ways an extension of real number multiplication. We also have a matrix multiplication operation and its inverse that are somewhat like the familiar real number operations (associativity, and distributivity over addition, for example), but there are differences (failure of commutativity). This section provides an example that algebra systems other than the usual real number one can be interesting and useful.

Exercises

4.12 Supply the intermediate steps in Example 4.9.

✓ 4.13 Use Corollary 4.11 to decide if each matrix has an inverse.

(a) $\begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$ **(b)** $\begin{pmatrix} 0 & 4 \\ 1 & -3 \end{pmatrix}$ **(c)** $\begin{pmatrix} 2 & -3 \\ -4 & 6 \end{pmatrix}$

✓ 4.14 For each invertible matrix in the prior problem, use Corollary 4.11 to find its inverse.

✓ 4.15 Find the inverse, if it exists, by using the Gauss-Jordan Method. Check the answers for the 2×2 matrices with Corollary 4.11.

(a) $\begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix}$ **(b)** $\begin{pmatrix} 2 & 1/2 \\ 3 & 1 \end{pmatrix}$ **(c)** $\begin{pmatrix} 2 & -4 \\ -1 & 2 \end{pmatrix}$ **(d)** $\begin{pmatrix} 1 & 1 & 3 \\ 0 & 2 & 4 \\ -1 & 1 & 0 \end{pmatrix}$

(e) $\begin{pmatrix} 0 & 1 & 5 \\ 0 & -2 & 4 \\ 2 & 3 & -2 \end{pmatrix}$ **(f)** $\begin{pmatrix} 2 & 2 & 3 \\ 1 & -2 & -3 \\ 4 & -2 & -3 \end{pmatrix}$

✓ 4.16 What matrix has this one for its inverse?

$$\begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix}$$

4.17 How does the inverse operation interact with scalar multiplication and addition of matrices?

(a) What is the inverse of rH?

(b) Is $(H+G)^{-1} = H^{-1} + G^{-1}$?

✓ 4.18 Is $(T^k)^{-1} = (T^{-1})^k$?

4.19 Is H^{-1} invertible?

4.20 For each real number θ let $t_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ be represented with respect to the standard bases by this matrix.

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Show that $t_{\theta_1+\theta_2} = t_{\theta_1} \cdot t_{\theta_2}$. Show also that ${t_\theta}^{-1} = t_{-\theta}$.

4.21 Do the calculations for the proof of Corollary 4.11.

4.22 Show that this matrix

$$H = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

has infinitely many right inverses. Show also that it has no left inverse.

4.23 In the review of inverses example, starting this subsection, how many left inverses has ι?

4.24 If a matrix has infinitely many right-inverses, can it have infinitely many left-inverses? Must it have?

4.25 Assume that $g\colon V \to W$ is linear. One of these is true, the other is false. Which is which?

(a) If $f\colon W \to V$ is a left inverse of g then f must be linear.

(b) If $f\colon W \to V$ is a right inverse of g then f must be linear.

✓ 4.26 Assume that H is invertible and that HG is the zero matrix. Show that G is a zero matrix.

4.27 Prove that if H is invertible then the inverse commutes with a matrix $GH^{-1} = H^{-1}G$ if and only if H itself commutes with that matrix $GH = HG$.

✓ 4.28 Show that if T is square and if T^4 is the zero matrix then $(I-T)^{-1} = I+T+T^2+T^3$. Generalize.

✓ 4.29 Let D be diagonal. Describe D^2, D^3, ..., etc. Describe D^{-1}, D^{-2}, ..., etc. Define D^0 appropriately.

4.30 Prove that any matrix row-equivalent to an invertible matrix is also invertible.

4.31 *The first question below appeared as Exercise 29.*

(a) Show that the rank of the product of two matrices is less than or equal to the minimum of the rank of each.

(b) Show that if T and S are square then $TS = I$ if and only if $ST = I$.

4.32 Show that the inverse of a permutation matrix is its transpose.

4.33 *The first two parts of this question appeared as Exercise 26.*

(a) Show that $(GH)^{\mathsf{T}} = H^{\mathsf{T}}G^{\mathsf{T}}$.

(b) A square matrix is *symmetric* if each i,j entry equals the j,i entry (that is, if the matrix equals its transpose). Show that the matrices HH^{T} and $H^{\mathsf{T}}H$ are symmetric.

(c) Show that the inverse of the transpose is the transpose of the inverse.

(d) Show that the inverse of a symmetric matrix is symmetric.

✓ **4.34** *The items starting this question appeared as Exercise 31.*

(a) Prove that the composition of the projections $\pi_x, \pi_y \colon \mathbb{R}^3 \to \mathbb{R}^3$ is the zero map despite that neither is the zero map.

(b) Prove that the composition of the derivatives d^2/dx^2, $d^3/dx^3 \colon \mathcal{P}_4 \to \mathcal{P}_4$ is the zero map despite that neither map is the zero map.

(c) Give matrix equations representing each of the prior two items.

When two things multiply to give zero despite that neither is zero, each is said to be a *zero divisor*. Prove that no zero divisor is invertible.

4.35 In real number algebra, there are exactly two numbers, 1 and -1, that are their own multiplicative inverse. Does $H^2 = I$ have exactly two solutions for $2{\times}2$ matrices?

4.36 Is the relation 'is a two-sided inverse of' transitive? Reflexive? Symmetric?

4.37 [Am. Math. Mon., Nov. 1951] Prove: if the sum of the elements of each row of a square matrix is k, then the sum of the elements in each row of the inverse matrix is $1/k$.

V Change of Basis

Representations vary with the bases. For instance, with respect to the bases $\mathcal{E}_2$ and

$$B = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rangle$$

$\vec{e}_1 \in \mathbb{R}^2$ has these different representations.

$$\mathrm{Rep}_{\mathcal{E}_2}(\vec{e}_1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \mathrm{Rep}_B(\vec{e}_1) = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

The same holds for maps: with respect to the basis pairs $\mathcal{E}_2, \mathcal{E}_2$ and $\mathcal{E}_2, B$, the identity map has these representations.

$$\mathrm{Rep}_{\mathcal{E}_2,\mathcal{E}_2}(\mathrm{id}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \mathrm{Rep}_{\mathcal{E}_2,B}(\mathrm{id}) = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}$$

This section shows how to translate among the representations. That is, we will compute how the representations vary as the bases vary.

V.1 Changing Representations of Vectors

In converting $\mathrm{Rep}_B(\vec{v})$ to $\mathrm{Rep}_D(\vec{v})$ the underlying vector $\vec{v}$ doesn't change. Thus, the translation between these two ways of expressing the vector is accomplished by the identity map on the space, described so that the domain space vectors are represented with respect to B and the codomain space vectors are represented with respect to D.

$$\begin{array}{c} V_{\mathit{wrt}\ B} \\ \Big\downarrow{\scriptstyle \mathrm{id}} \\ V_{\mathit{wrt}\ D} \end{array}$$

(This diagram is vertical to fit with the ones in the next subsection.)

1.1 Definition The *change of basis matrix* for bases $B, D \subset V$ is the representation of the identity map $\mathrm{id}\colon V \to V$ with respect to those bases.

$$\mathrm{Rep}_{B,D}(\mathrm{id}) = \begin{pmatrix} \vdots & & \vdots \\ \mathrm{Rep}_D(\vec{\beta}_1) & \cdots & \mathrm{Rep}_D(\vec{\beta}_n) \\ \vdots & & \vdots \end{pmatrix}$$

1.2 Remark A better name would be 'change of representation matrix' but the above name is standard.

The next result supports the definition.

1.3 Lemma Left-multiplication by the change of basis matrix for B, D converts a representation with respect to B to one with respect to D. Conversely, if left-multiplication by a matrix changes bases $M \cdot \mathrm{Rep}_B(\vec{v}) = \mathrm{Rep}_D(\vec{v})$ then M is a change of basis matrix.

Proof The first sentence holds because matrix-vector multiplication represents a map application and so $\mathrm{Rep}_{B,D}(\mathrm{id}) \cdot \mathrm{Rep}_B(\vec{v}) = \mathrm{Rep}_D(\,\mathrm{id}(\vec{v})\,) = \mathrm{Rep}_D(\vec{v})$ for each $\vec{v}$. For the second sentence, with respect to B, D the matrix M represents a linear map whose action is to map each vector to itself, and is therefore the identity map. QED

1.4 Example With these bases for $\mathbb{R}^2$,

$$B = \langle \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle \qquad D = \langle \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle$$

because

$$\mathrm{Rep}_D(\,\mathrm{id}(\begin{pmatrix} 2 \\ 1 \end{pmatrix})) = \begin{pmatrix} -1/2 \\ 3/2 \end{pmatrix}_D \qquad \mathrm{Rep}_D(\,\mathrm{id}(\begin{pmatrix} 1 \\ 0 \end{pmatrix})) = \begin{pmatrix} -1/2 \\ 1/2 \end{pmatrix}_D$$

the change of basis matrix is this.

$$\mathrm{Rep}_{B,D}(\mathrm{id}) = \begin{pmatrix} -1/2 & -1/2 \\ 3/2 & 1/2 \end{pmatrix}$$

For instance, this is the representation of $\vec{e}_2$

$$\mathrm{Rep}_B(\begin{pmatrix} 0 \\ 1 \end{pmatrix}) = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

and the matrix does the conversion.

$$\begin{pmatrix} -1/2 & -1/2 \\ 3/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$$

Checking that vector on the right is $\mathrm{Rep}_D(\vec{e}_2)$ is easy.

We finish this subsection by recognizing the change of basis matrices as a familiar set.

1.5 Lemma A matrix changes bases if and only if it is nonsingular.

Proof For the 'only if' direction, if left-multiplication by a matrix changes bases then the matrix represents an invertible function, simply because we can invert the function by changing the bases back. Because it represents a function that is invertible, the matrix itself is invertible, and so is nonsingular.

For 'if' we will show that any nonsingular matrix M performs a change of basis operation from any given starting basis B (having n vectors, where the matrix is $n\times n$) to some ending basis.

If the matrix is the identity I then the statement is obvious. Otherwise because the matrix is nonsingular Corollary IV.3.23 says there are elementary reduction matrices such that $R_r\cdots R_1\cdot M=I$ with $r\geqslant 1$. Elementary matrices are invertible and their inverses are also elementary so multiplying both sides of that equation from the left by ${R_r}^{-1}$, then by ${R_{r-1}}^{-1}$, etc., gives M as a product of elementary matrices $M={R_1}^{-1}\cdots {R_r}^{-1}$.

We will be done if we show that elementary matrices change a given basis to another basis, since then ${R_r}^{-1}$ changes B to some other basis B_r and ${R_{r-1}}^{-1}$ changes B_r to some B_{r-1}, etc. We will cover the three types of elementary matrices separately; recall the notation for the three.

$$M_i(k)\begin{pmatrix}c_1\\ \vdots\\ c_i\\ \vdots\\ c_n\end{pmatrix}=\begin{pmatrix}c_1\\ \vdots\\ kc_i\\ \vdots\\ c_n\end{pmatrix}\qquad P_{i,j}\begin{pmatrix}c_1\\ \vdots\\ c_i\\ \vdots\\ c_j\\ \vdots\\ c_n\end{pmatrix}=\begin{pmatrix}c_1\\ \vdots\\ c_j\\ \vdots\\ c_i\\ \vdots\\ c_n\end{pmatrix}\qquad C_{i,j}(k)\begin{pmatrix}c_1\\ \vdots\\ c_i\\ \vdots\\ c_j\\ \vdots\\ c_n\end{pmatrix}=\begin{pmatrix}c_1\\ \vdots\\ c_i\\ \vdots\\ kc_i+c_j\\ \vdots\\ c_n\end{pmatrix}$$

Applying a row-multiplication matrix $M_i(k)$ changes a representation with respect to $\langle\vec{\beta}_1,\ldots,\vec{\beta}_i,\ldots,\vec{\beta}_n\rangle$ to one with respect to $\langle\vec{\beta}_1,\ldots,(1/k)\vec{\beta}_i,\ldots,\vec{\beta}_n\rangle$.

$$\begin{aligned}\vec{v}&=c_1\cdot\vec{\beta}_1+\cdots+c_i\cdot\vec{\beta}_i+\cdots+c_n\cdot\vec{\beta}_n\\ &\qquad\mapsto\ c_1\cdot\vec{\beta}_1+\cdots+kc_i\cdot(1/k)\vec{\beta}_i+\cdots+c_n\cdot\vec{\beta}_n=\vec{v}\end{aligned}$$

The second one is a basis because the first is a basis and because of the $k\neq 0$ restriction in the definition of a row-multiplication matrix. Similarly, left-multiplication by a row-swap matrix $P_{i,j}$ changes a representation with respect to the basis $\langle\vec{\beta}_1,\ldots,\vec{\beta}_i,\ldots,\vec{\beta}_j,\ldots,\vec{\beta}_n\rangle$ into one with respect to this basis

$\langle\vec{\beta}_1,\ldots,\vec{\beta}_j,\ldots,\vec{\beta}_i,\ldots,\vec{\beta}_n\rangle$.

$$\begin{aligned}\vec{v}=c_1\cdot\vec{\beta}_1+\cdots+c_i\cdot\vec{\beta}_i+\cdots+c_j\vec{\beta}_j+\cdots+c_n\cdot\vec{\beta}_n\\ \mapsto\; c_1\cdot\vec{\beta}_1+\cdots+c_j\cdot\vec{\beta}_j+\cdots+c_i\cdot\vec{\beta}_i+\cdots+c_n\cdot\vec{\beta}_n=\vec{v}\end{aligned}$$

And, a representation with respect to $\langle\vec{\beta}_1,\ldots,\vec{\beta}_i,\ldots,\vec{\beta}_j,\ldots,\vec{\beta}_n\rangle$ changes via left-multiplication by a row-combination matrix $C_{i,j}(k)$ into a representation with respect to $\langle\vec{\beta}_1,\ldots,\vec{\beta}_i-k\vec{\beta}_j,\ldots,\vec{\beta}_j,\ldots,\vec{\beta}_n\rangle$

$$\begin{aligned}\vec{v}=c_1\cdot\vec{\beta}_1+\cdots+c_i\cdot\vec{\beta}_i+c_j\vec{\beta}_j+\cdots+c_n\cdot\vec{\beta}_n\\ \mapsto\; c_1\cdot\vec{\beta}_1+\cdots+c_i\cdot(\vec{\beta}_i-k\vec{\beta}_j)+\cdots+(kc_i+c_j)\cdot\vec{\beta}_j+\cdots+c_n\cdot\vec{\beta}_n=\vec{v}\end{aligned}$$

(the definition of $C_{i,j}(k)$ specifies that $i\neq j$ and $k\neq 0$). QED

1.6 Corollary A matrix is nonsingular if and only if it represents the identity map with respect to some pair of bases.

Exercises

✓ 1.7 In $\mathbb{R}^2$, where

$$D=\langle\begin{pmatrix}2\\1\end{pmatrix},\begin{pmatrix}-2\\4\end{pmatrix}\rangle$$

find the change of basis matrices from D to $\mathcal{E}_2$ and from $\mathcal{E}_2$ to D. Multiply the two.

✓ 1.8 Find the change of basis matrix for $B,D\subseteq\mathbb{R}^2$.

(a) $B=\mathcal{E}_2$, $D=\langle\vec{e}_2,\vec{e}_1\rangle$ **(b)** $B=\mathcal{E}_2$, $D=\langle\begin{pmatrix}1\\2\end{pmatrix},\begin{pmatrix}1\\4\end{pmatrix}\rangle$

(c) $B=\langle\begin{pmatrix}1\\2\end{pmatrix},\begin{pmatrix}1\\4\end{pmatrix}\rangle$, $D=\mathcal{E}_2$ **(d)** $B=\langle\begin{pmatrix}-1\\1\end{pmatrix},\begin{pmatrix}2\\2\end{pmatrix}\rangle$, $D=\langle\begin{pmatrix}0\\4\end{pmatrix},\begin{pmatrix}1\\3\end{pmatrix}\rangle$

✓ 1.9 Find the change of basis matrix for each $B,D\subseteq\mathcal{P}_2$.

(a) $B=\langle 1,x,x^2\rangle, D=\langle x^2,1,x\rangle$ **(b)** $B=\langle 1,x,x^2\rangle, D=\langle 1,1+x,1+x+x^2\rangle$

(c) $B=\langle 2,2x,x^2\rangle, D=\langle 1+x^2,1-x^2,x+x^2\rangle$

1.10 For the bases in Exercise 8, find the change of basis matrix in the other direction, from D to B.

✓ 1.11 Decide if each changes bases on $\mathbb{R}^2$. To what basis is $\mathcal{E}_2$ changed?

(a) $\begin{pmatrix}5&0\\0&4\end{pmatrix}$ **(b)** $\begin{pmatrix}2&1\\3&1\end{pmatrix}$ **(c)** $\begin{pmatrix}-1&4\\2&-8\end{pmatrix}$ **(d)** $\begin{pmatrix}1&-1\\1&1\end{pmatrix}$

1.12 Find bases such that this matrix represents the identity map with respect to those bases.

$$\begin{pmatrix}3&1&4\\2&-1&1\\0&0&4\end{pmatrix}$$

1.13 Consider the vector space of real-valued functions with basis $\langle\sin(x),\cos(x)\rangle$. Show that $\langle 2\sin(x)+\cos(x),3\cos(x)\rangle$ is also a basis for this space. Find the change of basis matrix in each direction.

1.14 Where does this matrix

$$\begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}$$

send the standard basis for $\mathbb{R}^2$? Any other bases? *Hint.* Consider the inverse.

✓ 1.15 What is the change of basis matrix with respect to B, B?

1.16 Prove that a matrix changes bases if and only if it is invertible.

1.17 Finish the proof of Lemma 1.5.

✓ 1.18 Let H be an $n{\times}n$ nonsingular matrix. What basis of $\mathbb{R}^n$ does H change to the standard basis?

✓ 1.19 **(a)** In $\mathcal{P}_3$ with basis $B = \langle 1+x, 1-x, x^2+x^3, x^2-x^3\rangle$ we have this representation.

$$\mathrm{Rep}_B(1-x+3x^2-x^3) = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix}_B$$

Find a basis D giving this different representation for the same polynomial.

$$\mathrm{Rep}_D(1-x+3x^2-x^3) = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 0 \end{pmatrix}_D$$

(b) State and prove that we can change any nonzero vector representation to any other.

Hint. The proof of Lemma 1.5 is constructive—it not only says the bases change, it shows how they change.

1.20 Let V, W be vector spaces, and let $B, \hat{B}$ be bases for V and $D, \hat{D}$ be bases for W. Where $h\colon V \to W$ is linear, find a formula relating $\mathrm{Rep}_{B,D}(h)$ to $\mathrm{Rep}_{\hat{B},\hat{D}}(h)$.

✓ 1.21 Show that the columns of an $n{\times}n$ change of basis matrix form a basis for $\mathbb{R}^n$. Do all bases appear in that way: can the vectors from any $\mathbb{R}^n$ basis make the columns of a change of basis matrix?

✓ 1.22 Find a matrix having this effect.

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} \mapsto \begin{pmatrix} 4 \\ -1 \end{pmatrix}$$

That is, find a M that left-multiplies the starting vector to yield the ending vector. Is there a matrix having these two effects?

(a) $\begin{pmatrix} 1 \\ 3 \end{pmatrix} \mapsto \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\begin{pmatrix} 2 \\ -1 \end{pmatrix} \mapsto \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ **(b)** $\begin{pmatrix} 1 \\ 3 \end{pmatrix} \mapsto \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\begin{pmatrix} 2 \\ 6 \end{pmatrix} \mapsto \begin{pmatrix} -1 \\ -1 \end{pmatrix}$

Give a necessary and sufficient condition for there to be a matrix such that $\vec{v}_1 \mapsto \vec{w}_1$ and $\vec{v}_2 \mapsto \vec{w}_2$.

V.2 Changing Map Representations

The first subsection shows how to convert the representation of a vector with respect to one basis to the representation of that same vector with respect to another basis. We next convert the representation of a map with respect to one pair of bases to the representation with respect to a different pair — we convert from $\mathrm{Rep}_{B,D}(h)$ to $\mathrm{Rep}_{\hat{B},\hat{D}}(h)$. Here is the arrow diagram.

$$\begin{array}{ccc} V_{\mathit{wrt}\ B} & \xrightarrow[H]{h} & W_{\mathit{wrt}\ D} \\ {\scriptstyle \mathrm{id}}\big\downarrow & & {\scriptstyle \mathrm{id}}\big\downarrow \\ V_{\mathit{wrt}\ \hat{B}} & \xrightarrow[\hat{H}]{h} & W_{\mathit{wrt}\ \hat{D}} \end{array}$$

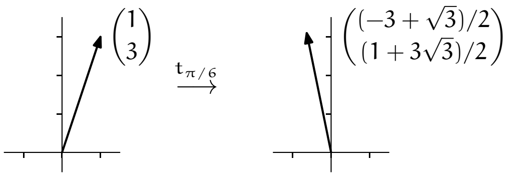
To move from the lower-left to the lower-right we can either go straight over, or else up to V_B then over to W_D and then down. So we can calculate $\hat{H} = \mathrm{Rep}_{\hat{B},\hat{D}}(h)$ either by directly using $\hat{B}$ and $\hat{D}$, or else by first changing bases with $\mathrm{Rep}_{\hat{B},B}(\mathrm{id})$ then multiplying by $H = \mathrm{Rep}_{B,D}(h)$ and then changing bases with $\mathrm{Rep}_{D,\hat{D}}(\mathrm{id})$.

$$\hat{H} = \mathrm{Rep}_{D,\hat{D}}(\mathrm{id}) \cdot H \cdot \mathrm{Rep}_{\hat{B},B}(\mathrm{id}) \qquad (*)$$

2.1 Example The matrix

$$T = \begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{pmatrix} = \begin{pmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{pmatrix}$$

represents, with respect to $\mathcal{E}_2, \mathcal{E}_2$, the transformation $t\colon \mathbb{R}^2 \to \mathbb{R}^2$ that rotates vectors through the counterclockwise angle of $\pi/6$ radians.



Translate T to a representation with respect to these

$$\hat{B} = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix} \rangle \qquad \hat{D} = \langle \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \rangle$$

by using the arrow diagram and formula $(*)$ above.

$$\begin{array}{ccc} \mathbb{R}^2_{\mathit{wrt}\ \mathcal{E}_2} & \xrightarrow[T]{t} & \mathbb{R}^2_{\mathit{wrt}\ \mathcal{E}_2} \\ {\scriptstyle \mathrm{id}}\big\downarrow & & {\scriptstyle \mathrm{id}}\big\downarrow \\ \mathbb{R}^2_{\mathit{wrt}\ \hat{B}} & \xrightarrow[\hat{T}]{t} & \mathbb{R}^2_{\mathit{wrt}\ \hat{D}} \end{array} \qquad \hat{T} = \mathrm{Rep}_{\mathcal{E}_2,\hat{D}}(\mathrm{id}) \cdot T \cdot \mathrm{Rep}_{\hat{B},\mathcal{E}_2}(\mathrm{id})$$

Note that we can compute $\text{Rep}_{\mathcal{E}_2,\hat{D}}(\text{id})$ as the matrix inverse of $\text{Rep}_{\hat{D},\mathcal{E}_2}(\text{id})$.

$$\begin{aligned}\text{Rep}_{\hat{B},\hat{D}}(t) &= \begin{pmatrix}-1 & 2\\ 0 & 3\end{pmatrix}^{-1}\begin{pmatrix}\sqrt{3}/2 & -1/2\\ 1/2 & \sqrt{3}/2\end{pmatrix}\begin{pmatrix}1 & 0\\ 1 & 2\end{pmatrix}\\ &= \begin{pmatrix}(5-\sqrt{3})/6 & (3+2\sqrt{3})/3\\ (1+\sqrt{3})/6 & \sqrt{3}/3\end{pmatrix}\end{aligned}$$

The matrix is messier but the map that it represents is the same. For instance, to replicate the effect of t in the picture, start with $\hat{B}$,

$$\text{Rep}_{\hat{B}}(\begin{pmatrix}1\\ 3\end{pmatrix}) = \begin{pmatrix}1\\ 1\end{pmatrix}_{\hat{B}}$$

apply $\hat{T}$,

$$\begin{pmatrix}(5-\sqrt{3})/6 & (3+2\sqrt{3})/3\\ (1+\sqrt{3})/6 & \sqrt{3}/3\end{pmatrix}_{\hat{B},\hat{D}}\begin{pmatrix}1\\ 1\end{pmatrix}_{\hat{B}} = \begin{pmatrix}(11+3\sqrt{3})/6\\ (1+3\sqrt{3})/6\end{pmatrix}_{\hat{D}}$$

and check it against $\hat{D}$.

$$\frac{11+3\sqrt{3}}{6}\cdot\begin{pmatrix}-1\\ 0\end{pmatrix} + \frac{1+3\sqrt{3}}{6}\cdot\begin{pmatrix}2\\ 3\end{pmatrix} = \begin{pmatrix}(-3+\sqrt{3})/2\\ (1+3\sqrt{3})/2\end{pmatrix}$$

2.2 Example Changing bases can make the matrix simpler. On $\mathbb{R}^3$ the map

$$\begin{pmatrix}x\\ y\\ z\end{pmatrix} \stackrel{t}{\longmapsto} \begin{pmatrix}y+z\\ x+z\\ x+y\end{pmatrix}$$

is represented with respect to the standard basis in this way.

$$\text{Rep}_{\mathcal{E}_3,\mathcal{E}_3}(t) = \begin{pmatrix}0 & 1 & 1\\ 1 & 0 & 1\\ 1 & 1 & 0\end{pmatrix}$$

Representing it with respect to

$$B = \langle \begin{pmatrix}1\\ -1\\ 0\end{pmatrix}, \begin{pmatrix}1\\ 1\\ -2\end{pmatrix}, \begin{pmatrix}1\\ 1\\ 1\end{pmatrix} \rangle$$

gives a matrix that is diagonal.

$$\text{Rep}_{B,B}(t) = \begin{pmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 2\end{pmatrix}$$

Naturally we usually prefer representations that are easier to understand. We say that a map or matrix has been *diagonalized* when we find a basis B such that the representation is diagonal with respect to B, B, that is, with respect to the same starting basis as ending basis. Chapter Five finds which maps and matrices are diagonalizable.

The rest of this subsection develops the easier case of finding two bases B, D such that a representation is simple. Recall that the prior subsection shows that a matrix is a change of basis matrix if and only if it is nonsingular.

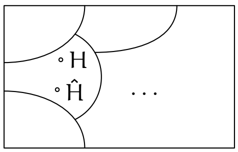
2.3 Definition Same-sized matrices H and $\hat{H}$ are *matrix equivalent* if there are nonsingular matrices P and Q such that $\hat{H} = PHQ$.

2.4 Corollary Matrix equivalent matrices represent the same map, with respect to appropriate pairs of bases.

Proof This is immediate from equation $(*)$ above. QED

Exercise 19 checks that matrix equivalence is an equivalence relation. Thus it partitions the set of matrices into matrix equivalence classes.

All matrices:



H matrix equivalent to $\hat{H}$

We can get insight into the classes by comparing matrix equivalence with row equivalence (remember that matrices are row equivalent when they can be reduced to each other by row operations). In $\hat{H} = PHQ$, the matrices P and Q are nonsingular and thus each is a product of elementary reduction matrices by Lemma IV.4.7. Left-multiplication by the reduction matrices making up P performs row operations. Right-multiplication by the reduction matrices making up Q performs column operations. Hence, matrix equivalence is a generalization of row equivalence — two matrices are row equivalent if one can be converted to the other by a sequence of row reduction steps, while two matrices are matrix equivalent if one can be converted to the other by a sequence of row reduction steps followed by a sequence of column reduction steps.

Consequently, if matrices are row equivalent then they are also matrix equivalent since we can take Q to be the identity matrix. The converse, however, does not hold: two matrices can be matrix equivalent but not row equivalent.

2.5 Example These two are matrix equivalent

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

because the second reduces to the first by the column operation of taking -1 times the first column and adding to the second. They are not row equivalent because they have different reduced echelon forms (both are already in reduced form).

We close this section by giving a set of representatives for the matrix equivalence classes.

2.6 Theorem Any $m\times n$ matrix of rank k is matrix equivalent to the $m\times n$ matrix that is all zeros except that the first k diagonal entries are ones.

$$\begin{pmatrix} 1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & 0 & \ldots & 0 \\ & \vdots & & & & & \\ 0 & 0 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ & \vdots & & & & & \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{pmatrix}$$

This is a *block partial-identity* form.

$$\left(\begin{array}{c|c} I & Z \\ \hline Z & Z \end{array}\right)$$

Proof Gauss-Jordan reduce the given matrix and combine all the row reduction matrices to make P. Then use the leading entries to do column reduction and finish by swapping the columns to put the leading ones on the diagonal. Combine the column reduction matrices into Q. QED

2.7 Example We illustrate the proof by finding P and Q for this matrix.

$$\begin{pmatrix} 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & -1 \\ 2 & 4 & 2 & -2 \end{pmatrix}$$

First Gauss-Jordan row-reduce.

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & -1 \\ 0 & 0 & 1 & -1 \\ 2 & 4 & 2 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Then column-reduce, which involves right-multiplication.

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

Finish by swapping columns.

$$\begin{pmatrix}1&0&0&0\\0&0&1&0\\0&0&0&0\end{pmatrix}\begin{pmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{pmatrix}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&0\end{pmatrix}$$

Finally, combine the left-multipliers together as P and the right-multipliers together as Q to get PHQ.

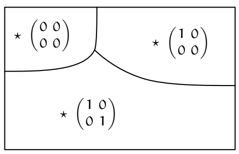
$$\begin{pmatrix}1&-1&0\\0&1&0\\-2&0&1\end{pmatrix}\begin{pmatrix}1&2&1&-1\\0&0&1&-1\\2&4&2&-2\end{pmatrix}\begin{pmatrix}1&0&-2&0\\0&0&1&0\\0&1&0&1\\0&0&0&1\end{pmatrix}=\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&0&0\end{pmatrix}$$

2.8 Corollary Matrix equivalence classes are characterized by rank: two same-sized matrices are matrix equivalent if and only if they have the same rank.

Proof Two same-sized matrices with the same rank are equivalent to the same block partial-identity matrix. QED

2.9 Example The 2×2 matrices have only three possible ranks: zero, one, or two. Thus there are three matrix equivalence classes.

All 2×2 matrices:



Three equivalence classes

Each class consists of all of the 2×2 matrices with the same rank. There is only one rank zero matrix. The other two classes have infinitely many members; we've shown only the canonical representative.

One nice thing about the representative in Theorem 2.6 is that we can completely understand the linear map when it is expressed in this way: where the bases are $B=\langle\vec{\beta}_1,\dots,\vec{\beta}_n\rangle$ and $D=\langle\vec{\delta}_1,\dots,\vec{\delta}_m\rangle$ then the map's action is

$$c_1\vec{\beta}_1+\cdots+c_k\vec{\beta}_k+c_{k+1}\vec{\beta}_{k+1}+\cdots+c_n\vec{\beta}_n\;\mapsto\;c_1\vec{\delta}_1+\cdots+c_k\vec{\delta}_k+\vec{0}+\cdots+\vec{0}$$

where k is the rank. Thus we can view any linear map as a projection.

$$\begin{pmatrix}c_1\\\vdots\\c_k\\c_{k+1}\\\vdots\\c_n\end{pmatrix}_B\longmapsto\begin{pmatrix}c_1\\\vdots\\c_k\\0\\\vdots\\0\end{pmatrix}_D$$

Exercises

✓ 2.10 Decide if these matrices are matrix equivalent.

(a) $\begin{pmatrix} 1 & 3 & 0 \\ 2 & 3 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 2 & 1 \\ 0 & 5 & -1 \end{pmatrix}$

(b) $\begin{pmatrix} 0 & 3 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 5 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 2 & -6 \end{pmatrix}$

✓ 2.11 Find the canonical representative of the matrix equivalence class of each matrix.

(a) $\begin{pmatrix} 2 & 1 & 0 \\ 4 & 2 & 0 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & 1 & 0 & 2 \\ 1 & 1 & 0 & 4 \\ 3 & 3 & 3 & -1 \end{pmatrix}$

2.12 Suppose that, with respect to

$$B = \mathcal{E}_2 \qquad D = \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \rangle$$

the transformation $t\colon \mathbb{R}^2 \to \mathbb{R}^2$ is represented by this matrix.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

Use change of basis matrices to represent t with respect to each pair.

(a) $\hat{B} = \langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle$, $\hat{D} = \langle \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \rangle$

(b) $\hat{B} = \langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle$, $\hat{D} = \langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \rangle$

✓ 2.13 What sizes are P and Q in the equation $\hat{H} = PHQ$?

✓ 2.14 Use Theorem 2.6 to show that a square matrix is nonsingular if and only if it is equivalent to an identity matrix.

✓ 2.15 Show that, where A is a nonsingular square matrix, if P and Q are nonsingular square matrices such that $PAQ = I$ then $QP = A^{-1}$.

✓ 2.16 Why does Theorem 2.6 not show that every matrix is diagonalizable (see Example 2.2)?

2.17 Must matrix equivalent matrices have matrix equivalent transposes?

2.18 What happens in Theorem 2.6 if $k = 0$?

✓ 2.19 Show that matrix equivalence is an equivalence relation.

✓ 2.20 Show that a zero matrix is alone in its matrix equivalence class. Are there other matrices like that?

2.21 What are the matrix equivalence classes of matrices of transformations on $\mathbb{R}^1$? $\mathbb{R}^3$?

2.22 How many matrix equivalence classes are there?

2.23 Are matrix equivalence classes closed under scalar multiplication? Addition?

2.24 Let $t\colon \mathbb{R}^n \to \mathbb{R}^n$ represented by T with respect to $\mathcal{E}_n, \mathcal{E}_n$.

(a) Find $\text{Rep}_{B,B}(t)$ in this specific case.

$$T = \begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix} \qquad B = \langle \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \rangle$$

(b) Describe $\mathrm{Rep}_{B,B}(t)$ in the general case where $B = \langle \vec{\beta}_1, \dots, \vec{\beta}_n \rangle$.

2.25 **(a)** Let V have bases B_1 and B_2 and suppose that W has the basis D. Where $h\colon V \to W$, find the formula that computes $\mathrm{Rep}_{B_2,D}(h)$ from $\mathrm{Rep}_{B_1,D}(h)$.

(b) Repeat the prior question with one basis for V and two bases for W.

2.26 **(a)** If two matrices are matrix equivalent and invertible, must their inverses be matrix equivalent?

(b) If two matrices have matrix equivalent inverses, must the two be matrix equivalent?

(c) If two matrices are square and matrix equivalent, must their squares be matrix equivalent?

(d) If two matrices are square and have matrix equivalent squares, must they be matrix equivalent?

✓ 2.27 Square matrices are *similar* if they represent the same transformation, but each with respect to the same ending as starting basis. That is, $\mathrm{Rep}_{B_1,B_1}(t)$ is similar to $\mathrm{Rep}_{B_2,B_2}(t)$.

(a) Give a definition of matrix similarity like that of Definition 2.3.

(b) Prove that similar matrices are matrix equivalent.

(c) Show that similarity is an equivalence relation.

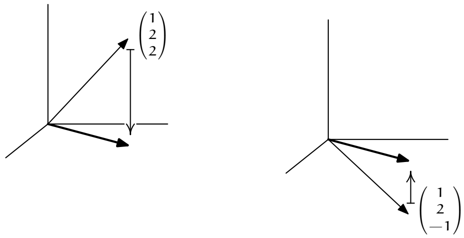
(d) Show that if T is similar to $\hat{T}$ then T^2 is similar to $\hat{T}^2$, the cubes are similar, etc. *Contrast with the prior exercise.*

(e) Prove that there are matrix equivalent matrices that are not similar.

VI Projection

This section is optional. It is a prerequisite only for the final two sections of Chapter Five, and some Topics.

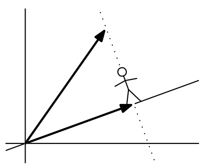
We have described projection from $\mathbb{R}^3$ into its xy-plane subspace as a shadow map. This shows why but it also shows that some shadows fall upward.



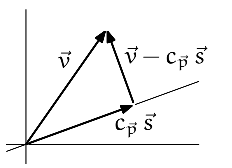
So perhaps a better description is: the projection of $\vec{v}$ is the vector $\vec{p}$ in the plane with the property that someone standing on $\vec{p}$ and looking straight up or down — that is, looking orthogonally to the plane — sees the tip of $\vec{v}$. In this section we will generalize this to other projections, orthogonal and non-orthogonal.

VI.1 Orthogonal Projection Into a Line

We first consider orthogonal projection of a vector $\vec{v}$ into a line ℓ. This shows a figure walking out on the line to a point $\vec{p}$ such that the tip of $\vec{v}$ is directly above them, where "above" does not mean parallel to the y-axis but instead means orthogonal to the line.



Since the line is the span of some vector $\ell = \{c \cdot \vec{s} \mid c \in \mathbb{R}\}$, we have a coefficient $c_{\vec{p}}$ with the property that $\vec{v} - c_{\vec{p}}\vec{s}$ is orthogonal to $c_{\vec{p}}\vec{s}$.



To solve for this coefficient, observe that because $\vec{v} - c_{\vec{p}}\vec{s}$ is orthogonal to a scalar multiple of $\vec{s}$, it must be orthogonal to $\vec{s}$ itself. Then $(\vec{v} - c_{\vec{p}}\vec{s}) \cdot \vec{s} = 0$ gives that $c_{\vec{p}} = \vec{v} \cdot \vec{s} / \vec{s} \cdot \vec{s}$.

1.1 Definition The *orthogonal projection of $\vec{v}$ into the line spanned by a nonzero $\vec{s}$* is this vector.

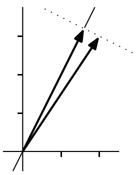
$$\mathrm{proj}_{[\vec{s}\,]}(\vec{v}) = \frac{\vec{v} \cdot \vec{s}}{\vec{s} \cdot \vec{s}} \cdot \vec{s}$$

(That says 'spanned by $\vec{s}$' instead the more formal 'span of the set $\{\vec{s}\,\}$'. This more casual phrase is common.)

1.2 Example To orthogonally project the vector $\binom{2}{3}$ into the line $y = 2x$, first pick a direction vector for the line.

$$\vec{s} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

The calculation is easy.



$$\frac{\begin{pmatrix} 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}}{\begin{pmatrix} 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{8}{5} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 8/5 \\ 16/5 \end{pmatrix}$$

1.3 Example In $\mathbb{R}^3$, the orthogonal projection of a general vector

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

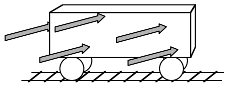
into the y-axis is

$$\frac{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}}{\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix}$$

which matches our intuitive expectation.

The picture above showing the figure walking out on the line until $\vec{v}$'s tip is overhead is one way to think of the orthogonal projection of a vector into a line. We finish this subsection with two other ways.

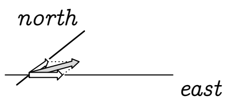
1.4 Example A railroad car left on an east-west track without its brake is pushed by a wind blowing toward the northeast at fifteen miles per hour; what speed will the car reach?



For the wind we use a vector of length 15 that points toward the northeast.

$$\vec{v} = \begin{pmatrix} 15\sqrt{1/2} \\ 15\sqrt{1/2} \end{pmatrix}$$

The car is only affected by the part of the wind blowing in the east-west direction — the part of $\vec{v}$ in the direction of the x-axis is this (the picture has the same perspective as the railroad car picture above).

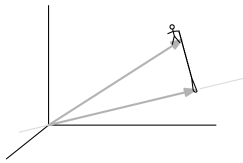


$$\vec{p} = \begin{pmatrix} 15\sqrt{1/2} \\ 0 \end{pmatrix}$$

So the car will reach a velocity of $15\sqrt{1/2}$ miles per hour toward the east.

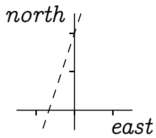
Thus, another way to think of the picture that precedes the definition is that it shows $\vec{v}$ as decomposed into two parts, the part $\vec{p}$ with the line, and the part that is orthogonal to the line (shown above on the north-south axis). These two are non-interacting in the sense that the east-west car is not at all affected by the north-south part of the wind (see Exercise 10). So we can think of the orthogonal projection of $\vec{v}$ into the line spanned by $\vec{s}$ as the part of $\vec{v}$ that lies in the direction of $\vec{s}$.

Still another useful way to think of orthogonal projection into a line is to have the person stand on the vector, not the line. This person holds a rope looped over the line. As they pull, the loop slides on the line.



When it is tight, the rope is orthogonal to the line. That is, we can think of the projection $\vec{p}$ as being the vector in the line that is closest to $\vec{v}$ (see Exercise 16).

1.5 Example A submarine is tracking a ship moving along the line $y = 3x + 2$. Torpedo range is one-half mile. If the sub stays where it is, at the origin on the chart below, will the ship pass within range?



The formula for projection into a line does not immediately apply because the line doesn't pass through the origin, and so isn't the span of any $\vec{s}$. To adjust for this, we start by shifting the entire map down two units. Now the line is $y = 3x$, a subspace. We project to get the point $\vec{p}$ on the line closest to

$$\vec{v} = \begin{pmatrix} 0 \\ -2 \end{pmatrix}$$

the sub's shifted position.

$$\vec{p} = \frac{\begin{pmatrix} 0 \\ -2 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 3 \end{pmatrix}}{\begin{pmatrix} 1 \\ 3 \end{pmatrix} \bullet \begin{pmatrix} 1 \\ 3 \end{pmatrix}} \cdot \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -3/5 \\ -9/5 \end{pmatrix}$$

The distance between $\vec{v}$ and $\vec{p}$ is about 0.63 miles. The ship will never be in range.

Exercises

✓ **1.6** Project the first vector orthogonally into the line spanned by the second vector.

(a) $\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -2 \end{pmatrix}$ **(b)** $\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \end{pmatrix}$ **(c)** $\begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ **(d)** $\begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \\ 12 \end{pmatrix}$

✓ **1.7** Project the vector orthogonally into the line.

(a) $\begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix}, \{c \begin{pmatrix} -3 \\ 1 \\ -3 \end{pmatrix} \mid c \in \mathbb{R}\}$ **(b)** $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$, the line $y = 3x$

1.8 Although pictures guided our development of Definition 1.1, we are not restricted to spaces that we can draw. In $\mathbb{R}^4$ project this vector into this line.

$$\vec{v} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 3 \end{pmatrix} \qquad \ell = \{c \cdot \begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix} \mid c \in \mathbb{R}\}$$

✓ **1.9** Definition 1.1 uses two vectors $\vec{s}$ and $\vec{v}$. Consider the transformation of $\mathbb{R}^2$ resulting from fixing

$$\vec{s} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

and projecting $\vec{v}$ into the line that is the span of $\vec{s}$. Apply it to these vectors.

(a) $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ **(b)** $\begin{pmatrix} 0 \\ 4 \end{pmatrix}$

Show that in general the projection transformation is this.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} (x_1 + 3x_2)/10 \\ (3x_1 + 9x_2)/10 \end{pmatrix}$$

Express the action of this transformation with a matrix.

1.10 Example 1.4 suggests that projection breaks $\vec{v}$ into two parts, $\mathrm{proj}_{[\vec{s}\,]}(\vec{v}\,)$ and $\vec{v} - \mathrm{proj}_{[\vec{s}\,]}(\vec{v}\,)$, that are non-interacting. Recall that the two are orthogonal. Show that any two nonzero orthogonal vectors make up a linearly independent set.

1.11 **(a)** What is the orthogonal projection of $\vec{v}$ into a line if $\vec{v}$ is a member of that line?

(b) Show that if $\vec{v}$ is not a member of the line then the set $\{\vec{v}, \vec{v} - \mathrm{proj}_{[\vec{s}\,]}(\vec{v}\,)\}$ is linearly independent.

1.12 Definition 1.1 requires that $\vec{s}$ be nonzero. Why? What is the right definition of the orthogonal projection of a vector into the (degenerate) line spanned by the zero vector?

1.13 Are all vectors the projection of some other vector into some line?

✓ 1.14 Show that the projection of $\vec{v}$ into the line spanned by $\vec{s}$ has length equal to the absolute value of the number $\vec{v} \bullet \vec{s}$ divided by the length of the vector $\vec{s}$.

1.15 Find the formula for the distance from a point to a line.

1.16 Find the scalar c such that the point (cs_1, cs_2) is a minimum distance from the point (v_1, v_2) by using calculus (i.e., consider the distance function, set the first derivative equal to zero, and solve). Generalize to $\mathbb{R}^n$.

✓ 1.17 Prove that the orthogonal projection of a vector into a line is shorter than the vector.

✓ 1.18 Show that the definition of orthogonal projection into a line does not depend on the spanning vector: if $\vec{s}$ is a nonzero multiple of $\vec{q}$ then $(\vec{v} \bullet \vec{s}/\vec{s} \bullet \vec{s}\,) \cdot \vec{s}$ equals $(\vec{v} \bullet \vec{q}/\vec{q} \bullet \vec{q}\,) \cdot \vec{q}$.

✓ 1.19 Consider the function mapping the plane to itself that takes a vector to its projection into the line $y = x$. These two each show that the map is linear, the first one in a way that is coordinate-bound (that is, it fixes a basis and then computes) and the second in a way that is more conceptual.

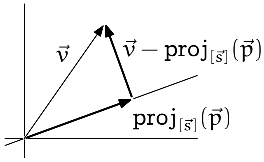
(a) Produce a matrix that describes the function's action.

(b) Show that we can obtain this map by first rotating everything in the plane $\pi/4$ radians clockwise, then projecting into the x-axis, and then rotating $\pi/4$ radians counterclockwise.

1.20 For $\vec{a}, \vec{b} \in \mathbb{R}^n$ let $\vec{v}_1$ be the projection of $\vec{a}$ into the line spanned by $\vec{b}$, let $\vec{v}_2$ be the projection of $\vec{v}_1$ into the line spanned by $\vec{a}$, let $\vec{v}_3$ be the projection of $\vec{v}_2$ into the line spanned by $\vec{b}$, etc., back and forth between the spans of $\vec{a}$ and $\vec{b}$. That is, $\vec{v}_{i+1}$ is the projection of $\vec{v}_i$ into the span of $\vec{a}$ if $i + 1$ is even, and into the span of $\vec{b}$ if $i + 1$ is odd. Must that sequence of vectors eventually settle down — must there be a sufficiently large i such that $\vec{v}_{i+2}$ equals $\vec{v}_i$ and $\vec{v}_{i+3}$ equals $\vec{v}_{i+1}$? If so, what is the earliest such i?

VI.2 Gram-Schmidt Orthogonalization

The prior subsection suggests that projecting $\vec{v}$ into the line spanned by $\vec{s}$ decomposes that vector into two parts



$$\vec{v} = \mathrm{proj}_{[\vec{s}\,]}(\vec{v}) + \left(\vec{v} - \mathrm{proj}_{[\vec{s}\,]}(\vec{v})\right)$$

that are orthogonal and so are "non-interacting." We now develop that suggestion.

2.1 Definition Vectors $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{R}^n$ are *mutually orthogonal* when any two are orthogonal: if $i \neq j$ then the dot product $\vec{v}_i \cdot \vec{v}_j$ is zero.

2.2 Theorem If the vectors in a set $\{\vec{v}_1, \ldots, \vec{v}_k\} \subset \mathbb{R}^n$ are mutually orthogonal and nonzero then that set is linearly independent.

Proof Consider $\vec{0} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k$. For $i \in \{1, .., k\}$, taking the dot product of $\vec{v}_i$ with both sides of the equation $\vec{v}_i \cdot (c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k) = \vec{v}_i \cdot \vec{0}$, which gives $c_i \cdot (\vec{v}_i \cdot \vec{v}_i) = 0$, shows that $c_i = 0$ since $\vec{v}_i \neq \vec{0}$. QED

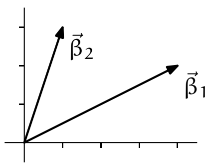
2.3 Corollary In a k dimensional vector space, if the vectors in a size k set are mutually orthogonal and nonzero then that set is a basis for the space.

Proof Any linearly independent size k subset of a k dimensional space is a basis. QED

Of course, the converse of Corollary 2.3 does not hold — not every basis of every subspace of $\mathbb{R}^n$ has mutually orthogonal vectors. However, we can get the partial converse that for every subspace of $\mathbb{R}^n$ there is at least one basis consisting of mutually orthogonal vectors.

2.4 Example The members $\vec{\beta}_1$ and $\vec{\beta}_2$ of this basis for $\mathbb{R}^2$ are not orthogonal.

$$B = \langle \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix} \rangle$$

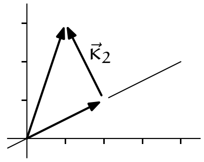


We will derive from B a new basis for the space $\langle \vec{\kappa}_1, \vec{\kappa}_2 \rangle$ consisting of mutually orthogonal vectors. The first member of the new basis is just $\vec{\beta}_1$.

$$\vec{\kappa}_1 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$$

For the second member of the new basis, we subtract from $\vec{\beta}_2$ the part in the direction of $\vec{\kappa}_1$. This leaves the part of $\vec{\beta}_2$ that is orthogonal to $\vec{\kappa}_1$.

$$\vec{\kappa}_2=\begin{pmatrix}1\\3\end{pmatrix}-\mathrm{proj}_{[\vec{\kappa}_1]}(\begin{pmatrix}1\\3\end{pmatrix})=\begin{pmatrix}1\\3\end{pmatrix}-\begin{pmatrix}2\\1\end{pmatrix}=\begin{pmatrix}-1\\2\end{pmatrix}$$



By the corollary $\langle\vec{\kappa}_1,\vec{\kappa}_2\rangle$ is a basis for $\mathbb{R}^2$.

2.5 Definition An *orthogonal basis* for a vector space is a basis of mutually orthogonal vectors.

2.6 Example To produce from this basis for $\mathbb{R}^3$

$$B=\langle\begin{pmatrix}1\\1\\1\end{pmatrix},\begin{pmatrix}0\\2\\0\end{pmatrix},\begin{pmatrix}1\\0\\3\end{pmatrix}\rangle$$

an orthogonal basis, start by taking the first vector unchanged.

$$\vec{\kappa}_1=\begin{pmatrix}1\\1\\1\end{pmatrix}$$

Get $\vec{\kappa}_2$ by subtracting from $\vec{\beta}_2$ its part in the direction of $\vec{\kappa}_1$.

$$\vec{\kappa}_2=\begin{pmatrix}0\\2\\0\end{pmatrix}-\mathrm{proj}_{[\vec{\kappa}_1]}(\begin{pmatrix}0\\2\\0\end{pmatrix})=\begin{pmatrix}0\\2\\0\end{pmatrix}-\begin{pmatrix}2/3\\2/3\\2/3\end{pmatrix}=\begin{pmatrix}-2/3\\4/3\\-2/3\end{pmatrix}$$

Find $\vec{\kappa}_3$ by subtracting from $\vec{\beta}_3$ the part in the direction of $\vec{\kappa}_1$ and also the part in the direction of $\vec{\kappa}_2$.

$$\vec{\kappa}_3=\begin{pmatrix}1\\0\\3\end{pmatrix}-\mathrm{proj}_{[\vec{\kappa}_1]}(\begin{pmatrix}1\\0\\3\end{pmatrix})-\mathrm{proj}_{[\vec{\kappa}_2]}(\begin{pmatrix}1\\0\\3\end{pmatrix})=\begin{pmatrix}-1\\0\\1\end{pmatrix}$$

As above, the corollary gives that the result is a basis for $\mathbb{R}^3$.

$$\langle\begin{pmatrix}1\\1\\1\end{pmatrix},\begin{pmatrix}-2/3\\4/3\\-2/3\end{pmatrix},\begin{pmatrix}-1\\0\\1\end{pmatrix}\rangle$$

2.7 Theorem (Gram-Schmidt orthogonalization) If $\langle\vec{\beta}_1, \ldots \vec{\beta}_k\rangle$ is a basis for a subspace of $\mathbb{R}^n$ then the vectors

$$\begin{aligned}
\vec{\kappa}_1 &= \vec{\beta}_1 \\
\vec{\kappa}_2 &= \vec{\beta}_2 - \mathrm{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_2) \\
\vec{\kappa}_3 &= \vec{\beta}_3 - \mathrm{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_3) - \mathrm{proj}_{[\vec{\kappa}_2]}(\vec{\beta}_3) \\
&\vdots \\
\vec{\kappa}_k &= \vec{\beta}_k - \mathrm{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_k) - \cdots - \mathrm{proj}_{[\vec{\kappa}_{k-1}]}(\vec{\beta}_k)
\end{aligned}$$

form an orthogonal basis for the same subspace.

2.8 Remark This is restricted to $\mathbb{R}^n$ only because we have not given a definition of orthogonality for other spaces.

PROOF We will use induction to check that each $\vec{\kappa}_i$ is nonzero, is in the span of $\langle\vec{\beta}_1, \ldots \vec{\beta}_i\rangle$, and is orthogonal to all preceding vectors $\vec{\kappa}_1 \cdot \vec{\kappa}_i = \cdots = \vec{\kappa}_{i-1} \cdot \vec{\kappa}_i = 0$. Then Corollary 2.3 gives that $\langle\vec{\kappa}_1, \ldots \vec{\kappa}_k\rangle$ is a basis for the same space as is the starting basis.

We shall only cover the cases up to $i = 3$, to give the sense of the argument. The full argument is Exercise 25.

The $i = 1$ case is trivial; taking $\vec{\kappa}_1$ to be $\vec{\beta}_1$ makes it a nonzero vector since $\vec{\beta}_1$ is a member of a basis, it is obviously in the span of $\langle\vec{\beta}_1\rangle$, and the 'orthogonal to all preceding vectors' condition is satisfied vacuously.

In the $i = 2$ case the expansion

$$\vec{\kappa}_2 = \vec{\beta}_2 - \mathrm{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_2) = \vec{\beta}_2 - \frac{\vec{\beta}_2 \cdot \vec{\kappa}_1}{\vec{\kappa}_1 \cdot \vec{\kappa}_1} \cdot \vec{\kappa}_1 = \vec{\beta}_2 - \frac{\vec{\beta}_2 \cdot \vec{\kappa}_1}{\vec{\kappa}_1 \cdot \vec{\kappa}_1} \cdot \vec{\beta}_1$$

shows that $\vec{\kappa}_2 \neq \vec{0}$ or else this would be a non-trivial linear dependence among the $\vec{\beta}$'s (it is nontrivial because the coefficient of $\vec{\beta}_2$ is 1). It also shows that $\vec{\kappa}_2$ is in the span of $\langle\vec{\beta}_1, \vec{\beta}_2\rangle$. And, $\vec{\kappa}_2$ is orthogonal to the only preceding vector

$$\vec{\kappa}_1 \cdot \vec{\kappa}_2 = \vec{\kappa}_1 \cdot (\vec{\beta}_2 - \mathrm{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_2)) = 0$$

because this projection is orthogonal.

The $i = 3$ case is the same as the $i = 2$ case except for one detail. As in the $i = 2$ case, expand the definition.

$$\begin{aligned}
\vec{\kappa}_3 &= \vec{\beta}_3 - \frac{\vec{\beta}_3 \cdot \vec{\kappa}_1}{\vec{\kappa}_1 \cdot \vec{\kappa}_1} \cdot \vec{\kappa}_1 - \frac{\vec{\beta}_3 \cdot \vec{\kappa}_2}{\vec{\kappa}_2 \cdot \vec{\kappa}_2} \cdot \vec{\kappa}_2 \\
&= \vec{\beta}_3 - \frac{\vec{\beta}_3 \cdot \vec{\kappa}_1}{\vec{\kappa}_1 \cdot \vec{\kappa}_1} \cdot \vec{\beta}_1 - \frac{\vec{\beta}_3 \cdot \vec{\kappa}_2}{\vec{\kappa}_2 \cdot \vec{\kappa}_2} \cdot \left(\vec{\beta}_2 - \frac{\vec{\beta}_2 \cdot \vec{\kappa}_1}{\vec{\kappa}_1 \cdot \vec{\kappa}_1} \cdot \vec{\beta}_1\right)
\end{aligned}$$

By the first line $\vec{\kappa}_3 \neq \vec{0}$, since $\vec{\beta}_3$ isn't in the span $[\vec{\beta}_1, \vec{\beta}_2]$ and therefore by the inductive hypothesis it isn't in the span $[\vec{\kappa}_1, \vec{\kappa}_2]$. By the second line $\vec{\kappa}_3$ is in the span of the first three $\vec{\beta}$'s. Finally, the calculation below shows that $\vec{\kappa}_3$ is orthogonal to $\vec{\kappa}_1$.

$$\begin{aligned}
\vec{\kappa}_1 \cdot \vec{\kappa}_3 &= \vec{\kappa}_1 \cdot \big(\, \vec{\beta}_3 - \mathrm{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_3) - \mathrm{proj}_{[\vec{\kappa}_2]}(\vec{\beta}_3) \,\big) \\
&= \vec{\kappa}_1 \cdot \big(\vec{\beta}_3 - \mathrm{proj}_{[\vec{\kappa}_1]}(\vec{\beta}_3)\big) - \vec{\kappa}_1 \cdot \mathrm{proj}_{[\vec{\kappa}_2]}(\vec{\beta}_3) \\
&= 0
\end{aligned}$$

(Here is the difference with the $i = 2$ case: as happened for $i = 2$ the first term is 0 because this projection is orthogonal, but here the second term in the second line is 0 because $\vec{\kappa}_1$ is orthogonal to $\vec{\kappa}_2$ and so is orthogonal to any vector in the line spanned by $\vec{\kappa}_2$.) A similar check shows that $\vec{\kappa}_3$ is also orthogonal to $\vec{\kappa}_2$. QED

In addition to having the vectors in the basis be orthogonal, we can also *normalize* each vector by dividing by its length, to end with an *orthonormal basis.*.

2.9 Example From the orthogonal basis of Example 2.6, normalizing produces this orthonormal basis.

$$\langle \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, \begin{pmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ -1/\sqrt{6} \end{pmatrix}, \begin{pmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix} \rangle$$

Besides its intuitive appeal, and its analogy with the standard basis $\mathcal{E}_n$ for $\mathbb{R}^n$, an orthonormal basis also simplifies some computations. Exercise 19 is an example.

Exercises

2.10 Perform the Gram-Schmidt process on each of these bases for $\mathbb{R}^2$.

(a) $\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \rangle$ **(b)** $\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix} \rangle$ **(c)** $\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix} \rangle$

Then turn those orthogonal bases into orthonormal bases.

✓ 2.11 Perform the Gram-Schmidt process on each of these bases for $\mathbb{R}^3$.

(a) $\langle \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix} \rangle$ **(b)** $\langle \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} \rangle$

Then turn those orthogonal bases into orthonormal bases.

✓ 2.12 Find an orthonormal basis for this subspace of $\mathbb{R}^3$: the plane $x - y + z = 0$.

2.13 Find an orthonormal basis for this subspace of $\mathbb{R}^4$.

$$\{ \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} \mid x - y - z + w = 0 \text{ and } x + z = 0 \}$$

2.14 Show that any linearly independent subset of $\mathbb{R}^n$ can be orthogonalized without changing its span.

2.15 What happens if we try to apply the Gram-Schmidt process to a finite set that is not a basis?

✓ 2.16 What happens if we apply the Gram-Schmidt process to a basis that is already orthogonal?

2.17 Let $\langle\vec{\kappa}_1,\ldots,\vec{\kappa}_k\rangle$ be a set of mutually orthogonal vectors in $\mathbb{R}^n$.

(a) Prove that for any $\vec{v}$ in the space, the vector $\vec{v}-(\text{proj}_{[\vec{\kappa}_1]}(\vec{v})+\cdots+\text{proj}_{[\vec{v}_k]}(\vec{v}))$ is orthogonal to each of $\vec{\kappa}_1, \ldots, \vec{\kappa}_k$.

(b) Illustrate the prior item in $\mathbb{R}^3$ by using $\vec{e}_1$ as $\vec{\kappa}_1$, using $\vec{e}_2$ as $\vec{\kappa}_2$, and taking $\vec{v}$ to have components 1, 2, and 3.

(c) Show that $\text{proj}_{[\vec{\kappa}_1]}(\vec{v})+\cdots+\text{proj}_{[\vec{v}_k]}(\vec{v})$ is the vector in the span of the set of $\vec{\kappa}$'s that is closest to $\vec{v}$. *Hint.* To the illustration done for the prior part, add a vector $d_1\vec{\kappa}_1+d_2\vec{\kappa}_2$ and apply the Pythagorean Theorem to the resulting triangle.

2.18 Find a nonzero vector in $\mathbb{R}^3$ that is orthogonal to both of these.

$$\begin{pmatrix}1\\5\\-1\end{pmatrix}\qquad\begin{pmatrix}2\\2\\0\end{pmatrix}$$

✓ 2.19 One advantage of orthogonal bases is that they simplify finding the representation of a vector with respect to that basis.

(a) For this vector and this non-orthogonal basis for $\mathbb{R}^2$

$$\vec{v}=\begin{pmatrix}2\\3\end{pmatrix}\qquad B=\langle\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\0\end{pmatrix}\rangle$$

first represent the vector with respect to the basis. Then project the vector into the span of each basis vector $[\vec{\beta}_1]$ and $[\vec{\beta}_2]$.

(b) With this orthogonal basis for $\mathbb{R}^2$

$$K=\langle\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\-1\end{pmatrix}\rangle$$

represent the same vector $\vec{v}$ with respect to the basis. Then project the vector into the span of each basis vector. Note that the coefficients in the representation and the projection are the same.

(c) Let $K=\langle\vec{\kappa}_1,\ldots,\vec{\kappa}_k\rangle$ be an orthogonal basis for some subspace of $\mathbb{R}^n$. Prove that for any $\vec{v}$ in the subspace, the i-th component of the representation $\text{Rep}_K(\vec{v})$ is the scalar coefficient $(\vec{v}\cdot\vec{\kappa}_i)/(\vec{\kappa}_i\cdot\vec{\kappa}_i)$ from $\text{proj}_{[\vec{\kappa}_i]}(\vec{v})$.

(d) Prove that $\vec{v}=\text{proj}_{[\vec{\kappa}_1]}(\vec{v})+\cdots+\text{proj}_{[\vec{\kappa}_k]}(\vec{v})$.

2.20 *Bessel's Inequality.* Consider these orthonormal sets

$$B_1=\{\vec{e}_1\}\quad B_2=\{\vec{e}_1,\vec{e}_2\}\quad B_3=\{\vec{e}_1,\vec{e}_2,\vec{e}_3\}\quad B_4=\{\vec{e}_1,\vec{e}_2,\vec{e}_3,\vec{e}_4\}$$

along with the vector $\vec{v}\in\mathbb{R}^4$ whose components are 4, 3, 2, and 1.

(a) Find the coefficient c_1 for the projection of $\vec{v}$ into the span of the vector in B_1. Check that $\|\vec{v}\|^2\geqslant|c_1|^2$.

(b) Find the coefficients c_1 and c_2 for the projection of $\vec{v}$ into the spans of the two vectors in B_2. Check that $\|\vec{v}\|^2\geqslant|c_1|^2+|c_2|^2$.

(c) Find c_1, c_2, and c_3 associated with the vectors in B_3, and c_1, c_2, c_3, and c_4 for the vectors in B_4. Check that $\|\vec{v}\|^2 \geqslant |c_1|^2 + \cdots + |c_3|^2$ and that $\|\vec{v}\|^2 \geqslant |c_1|^2 + \cdots + |c_4|^2$.

Show that this holds in general: where $\{\vec{\kappa}_1, \ldots, \vec{\kappa}_k\}$ is an orthonormal set and c_i is coefficient of the projection of a vector $\vec{v}$ from the space then $\|\vec{v}\|^2 \geqslant |c_1|^2 + \cdots + |c_k|^2$. *Hint.* One way is to look at the inequality $0 \leqslant \|\vec{v} - (c_1\vec{\kappa}_1 + \cdots + c_k\vec{\kappa}_k)\|^2$ and expand the c's.

2.21 Prove or disprove: every vector in $\mathbb{R}^n$ is in some orthogonal basis.

2.22 Show that the columns of an $n \times n$ matrix form an orthonormal set if and only if the inverse of the matrix is its transpose. Produce such a matrix.

2.23 Does the proof of Theorem 2.2 fail to consider the possibility that the set of vectors is empty (i.e., that $k = 0$)?

2.24 Theorem 2.7 describes a change of basis from any basis $B = \langle\vec{\beta}_1, \ldots, \vec{\beta}_k\rangle$ to one that is orthogonal $K = \langle\vec{\kappa}_1, \ldots, \vec{\kappa}_k\rangle$. Consider the change of basis matrix $\mathrm{Rep}_{B,K}(\mathrm{id})$.

(a) Prove that the matrix $\mathrm{Rep}_{K,B}(\mathrm{id})$ changing bases in the direction opposite to that of the theorem has an upper triangular shape — all of its entries below the main diagonal are zeros.

(b) Prove that the inverse of an upper triangular matrix is also upper triangular (if the matrix is invertible, that is). This shows that the matrix $\mathrm{Rep}_{B,K}(\mathrm{id})$ changing bases in the direction described in the theorem is upper triangular.

2.25 Complete the induction argument in the proof of Theorem 2.7.

VI.3 Projection Into a Subspace

This subsection uses material from the optional earlier subsection on Combining Subspaces.

The prior subsections project a vector into a line by decomposing it into two parts: the part in the line $\mathrm{proj}_{[\vec{s}\,]}(\vec{v}\,)$ and the rest $\vec{v} - \mathrm{proj}_{[\vec{s}\,]}(\vec{v}\,)$. To generalize projection to arbitrary subspaces we will follow this decomposition idea.

3.1 Definition Let a vector space be a direct sum $V = M \oplus N$. Then for any $\vec{v} \in V$ with $\vec{v} = \vec{m} + \vec{n}$ where $\vec{m} \in M$, $\vec{n} \in N$, the *projection of $\vec{v}$ into M along N* is $\mathrm{proj}_{M,N}(\vec{v}\,) = \vec{m}$.

This definition applies in spaces where we don't have a ready definition of orthogonal. (Definitions of orthogonality for spaces other than the $\mathbb{R}^n$ are perfectly possible but we haven't seen any in this book.)

3.2 Example The space $\mathcal{M}_{2\times 2}$ of 2×2 matrices is the direct sum of these two.

$$M=\{\begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \mid a,b\in\mathbb{R}\} \qquad N=\{\begin{pmatrix} 0 & 0 \\ c & d \end{pmatrix} \mid c,d\in\mathbb{R}\}$$

To project

$$A=\begin{pmatrix} 3 & 1 \\ 0 & 4 \end{pmatrix}$$

into M along N, we first fix bases for the two subspaces.

$$B_M=\langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \rangle \qquad B_N=\langle \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \rangle$$

Their concatenation

$$B=B_M\frown B_N=\langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \rangle$$

is a basis for the entire space because $\mathcal{M}_{2\times 2}$ is the direct sum. So we can use it to represent A.

$$\begin{pmatrix} 3 & 1 \\ 0 & 4 \end{pmatrix}=3\cdot\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}+1\cdot\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}+0\cdot\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}+4\cdot\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

The projection of A into M along N keeps the M part and drops the N part.

$$\mbox{proj}_{M,N}(\begin{pmatrix} 3 & 1 \\ 0 & 4 \end{pmatrix})=3\cdot\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}+1\cdot\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}=\begin{pmatrix} 3 & 1 \\ 0 & 0 \end{pmatrix}$$

3.3 Example Both subscripts on $\mbox{proj}_{M,N}(\vec{v}\,)$ are significant. The first subscript M matters because the result of the projection is a member of M. For an example showing that the second one matters, fix this plane subspace of $\mathbb{R}^3$ and its basis.

$$M=\{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid y-2z=0\} \qquad B_M=\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \rangle$$

We will compare the projections of this element of $\mathbb{R}^3$

$$\vec{v}=\begin{pmatrix} 2 \\ 2 \\ 5 \end{pmatrix}$$

into M along these two subspaces (verification that $\mathbb{R}^3 = M \oplus N$ and $\mathbb{R}^3 = M \oplus \hat{N}$ is routine).

$$N = \{k\begin{pmatrix}0\\0\\1\end{pmatrix} \mid k \in \mathbb{R}\} \qquad \hat{N} = \{k\begin{pmatrix}0\\1\\-2\end{pmatrix} \mid k \in \mathbb{R}\}$$

Here are natural bases for N and $\hat{N}$.

$$B_N = \langle \begin{pmatrix}0\\0\\1\end{pmatrix} \rangle \qquad B_{\hat{N}} = \langle \begin{pmatrix}0\\1\\-2\end{pmatrix} \rangle$$

To project into M along N, represent $\vec{v}$ with respect to the concatenation $B_M \frown B_N$

$$\begin{pmatrix}2\\2\\5\end{pmatrix} = 2\cdot\begin{pmatrix}1\\0\\0\end{pmatrix} + 1\cdot\begin{pmatrix}0\\2\\1\end{pmatrix} + 4\cdot\begin{pmatrix}0\\0\\1\end{pmatrix}$$

and drop the N term.

$$\mathrm{proj}_{M,N}(\vec{v}\,) = 2\cdot\begin{pmatrix}1\\0\\0\end{pmatrix} + 1\cdot\begin{pmatrix}0\\2\\1\end{pmatrix} = \begin{pmatrix}2\\2\\1\end{pmatrix}$$

To project into M along $\hat{N}$ represent $\vec{v}$ with respect to $B_M \frown B_{\hat{N}}$

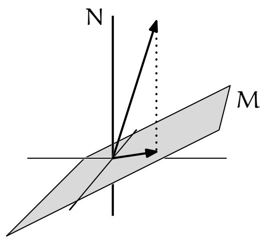
$$\begin{pmatrix}2\\2\\5\end{pmatrix} = 2\cdot\begin{pmatrix}1\\0\\0\end{pmatrix} + (9/5)\cdot\begin{pmatrix}0\\2\\1\end{pmatrix} - (8/5)\cdot\begin{pmatrix}0\\1\\-2\end{pmatrix}$$

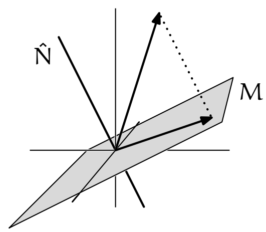
and omit the $\hat{N}$ part.

$$\mathrm{proj}_{M,\hat{N}}(\vec{v}\,) = 2\cdot\begin{pmatrix}1\\0\\0\end{pmatrix} + (9/5)\cdot\begin{pmatrix}0\\2\\1\end{pmatrix} = \begin{pmatrix}2\\18/5\\9/5\end{pmatrix}$$

So projecting along different subspaces can give different results.

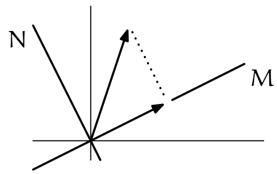
These pictures compare the two maps. Both show that the projection is indeed ‘into’ the plane and ‘along’ the line.





Notice that the projection along N is not orthogonal since there are members of the plane M that are not orthogonal to the dotted line. But the projection along $\hat{N}$ is orthogonal.

We have seen two projection operations, orthogonal projection into a line as well as this subsections's projection into an M and along an N, and we naturally ask whether they are related. The right-hand picture above suggests the answer — orthogonal projection into a line is a special case of this subsection's projection; it is projection along a subspace perpendicular to the line.



3.4 Definition The *orthogonal complement* of a subspace M of $\mathbb{R}^n$ is

$$M^{\perp} = \{\vec{v} \in \mathbb{R}^n \mid \vec{v} \text{ is perpendicular to all vectors in } M\}$$

(read "M perp"). The *orthogonal projection* $\mathrm{proj}_M(\vec{v}\,)$ of a vector is its projection into M along $M^{\perp}$.

3.5 Example In $\mathbb{R}^3$, to find the orthogonal complement of the plane

$$P = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid 3x + 2y - z = 0\}$$

we start with a basis for P.

$$B = \langle \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \rangle$$

Any $\vec{v}$ perpendicular to every vector in B is perpendicular to every vector in the span of B (the proof of this is Exercise 19). Therefore, the subspace $P^{\perp}$ consists of the vectors that satisfy these two conditions.

$$\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0 \qquad \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0$$

Those conditions give a linear system.

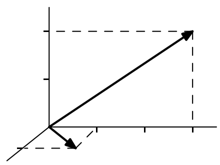
$$P^{\perp} = \{\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \mid \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\}$$

We are thus left with finding the null space of the map represented by the matrix, that is, with calculating the solution set of the homogeneous linear system.

$$\begin{aligned} v_1 + 3v_3 &= 0 \\ v_2 + 2v_3 &= 0 \end{aligned} \quad\Longrightarrow\quad P^{\perp} = \{k\begin{pmatrix} -3 \\ -2 \\ 1 \end{pmatrix} \mid k \in \mathbb{R}\}$$

3.6 Example Where M is the xy-plane subspace of $\mathbb{R}^3$, what is $M^{\perp}$? A common first reaction is that $M^{\perp}$ is the yz-plane but that's not right because some vectors from the yz-plane are not perpendicular to every vector in the xy-plane.

$$\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \not\perp \begin{pmatrix} 0 \\ 3 \\ 2 \end{pmatrix}$$



$$\theta = \arccos(\frac{1 \cdot 0 + 1 \cdot 3 + 0 \cdot 2}{\sqrt{2} \cdot \sqrt{13}}) \approx 0.94 \text{ rad}$$

Instead $M^{\perp}$ is the z-axis, since proceeding as in the prior example and taking the natural basis for the xy-plane gives this.

$$M^{\perp} = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}\} = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x = 0 \text{ and } y = 0\}$$

3.7 Lemma If M is a subspace of $\mathbb{R}^n$ then its orthogonal complement $M^{\perp}$ is also a subspace. The space is the direct sum of the two $\mathbb{R}^n = M \oplus M^{\perp}$. And, for any $\vec{v} \in \mathbb{R}^n$ the vector $\vec{v} - \mathrm{proj}_M(\vec{v})$ is perpendicular to every vector in M.

Proof First, the orthogonal complement $M^{\perp}$ is a subspace of $\mathbb{R}^n$ because, as noted in the prior two examples, it is a null space.

Next, to show that the space is the direct sum of the two, start with any basis $B_M = \langle \vec{\mu}_1, \ldots, \vec{\mu}_k \rangle$ for M and expand it to a basis for the entire space. Apply the Gram-Schmidt process to get an orthogonal basis $K = \langle \vec{\kappa}_1, \ldots, \vec{\kappa}_n \rangle$ for $\mathbb{R}^n$. This K is the concatenation of two bases: $\langle \vec{\kappa}_1, \ldots, \vec{\kappa}_k \rangle$ with the same number of members k as B_M, and $\langle \vec{\kappa}_{k+1}, \ldots, \vec{\kappa}_n \rangle$. The first is a basis for M so if we show that the second is a basis for $M^{\perp}$ then we will have that the entire space is the direct sum.

Exercise 19 from the prior subsection proves this about any orthogonal basis: each vector $\vec{v}$ in the space is the sum of its orthogonal projections into the lines spanned by the basis vectors.

$$\vec{v} = \mathrm{proj}_{[\vec{\kappa}_1]}(\vec{v}) + \cdots + \mathrm{proj}_{[\vec{\kappa}_n]}(\vec{v}) \qquad (*)$$

To check this, represent the vector as $\vec{v} = r_1\vec{\kappa}_1 + \cdots + r_n\vec{\kappa}_n$, apply $\vec{\kappa}_i$ to both sides $\vec{v} \cdot \vec{\kappa}_i = (r_1\vec{\kappa}_1 + \cdots + r_n\vec{\kappa}_n) \cdot \vec{\kappa}_i = r_1 \cdot 0 + \cdots + r_i \cdot (\vec{\kappa}_i \cdot \vec{\kappa}_i) + \cdots + r_n \cdot 0$, and solve to get $r_i = (\vec{v} \cdot \vec{\kappa}_i)/(\vec{\kappa}_i \cdot \vec{\kappa}_i)$, as desired.

Since obviously any member of the span of $\langle\vec{\kappa}_{k+1}, \ldots, \vec{\kappa}_n\rangle$ is orthogonal to any vector in M, to show that this is a basis for $M^\perp$ we need only show the other containment — that any $\vec{w} \in M^\perp$ is in the span of this basis. The prior paragraph does this. Any $\vec{w} \in M^\perp$ gives this on projections into basis vectors from M: $\mathrm{proj}_{[\vec{\kappa}_1]}(\vec{w}\,) = \vec{0}, \ldots, \mathrm{proj}_{[\vec{\kappa}_k]}(\vec{w}\,) = \vec{0}$. Therefore equation $(*)$ gives that $\vec{w}$ is a linear combination of $\vec{\kappa}_{k+1}, \ldots, \vec{\kappa}_n$. Thus this is a basis for $M^\perp$ and $\mathbb{R}^n$ is the direct sum of the two.

The final sentence of the lemma is proved in much the same way. Write $\vec{v} = \mathrm{proj}_{[\vec{\kappa}_1]}(\vec{v}\,) + \cdots + \mathrm{proj}_{[\vec{\kappa}_n]}(\vec{v}\,)$. Then $\mathrm{proj}_M(\vec{v}\,)$ keeps only the M part and drops the $M^\perp$ part $\mathrm{proj}_M(\vec{v}\,) = \mathrm{proj}_{[\vec{\kappa}_{k+1}]}(\vec{v}\,) + \cdots + \mathrm{proj}_{[\vec{\kappa}_k]}(\vec{v}\,)$. Therefore $\vec{v} - \mathrm{proj}_M(\vec{v}\,)$ consists of a linear combination of elements of $M^\perp$ and so is perpendicular to every vector in M. QED

Given a subspace, we could compute the orthogonal projection into that subspace by following the steps of that proof: finding a basis, expanding it to a basis for the entire space, applying Gram-Schmidt to get an orthogonal basis, and projecting into each linear subspace. However we will instead use a convenient formula.

3.8 Theorem Let M be a subspace of $\mathbb{R}^n$ with basis $\langle\vec{\beta}_1, \ldots, \vec{\beta}_k\rangle$ and let A be the matrix whose columns are the $\vec{\beta}$'s. Then for any $\vec{v} \in \mathbb{R}^n$ the orthogonal projection is $\mathrm{proj}_M(\vec{v}\,) = c_1\vec{\beta}_1 + \cdots + c_k\vec{\beta}_k$, where the coefficients c_i are the entries of the vector $(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} \cdot \vec{v}$. That is, $\mathrm{proj}_M(\vec{v}\,) = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} \cdot \vec{v}$.

Proof The vector $\mathrm{proj}_M(\vec{v})$ is a member of M and so is a linear combination of basis vectors $c_1 \cdot \vec{\beta}_1 + \cdots + c_k \cdot \vec{\beta}_k$. Since A's columns are the $\vec{\beta}$'s, there is a $\vec{c} \in \mathbb{R}^k$ such that $\mathrm{proj}_M(\vec{v}\,) = A\vec{c}$. To find $\vec{c}$ note that the vector $\vec{v} - \mathrm{proj}_M(\vec{v}\,)$ is perpendicular to each member of the basis so

$$\vec{0} = A^{\mathsf{T}}\bigl(\vec{v} - A\vec{c}\bigr) = A^{\mathsf{T}}\vec{v} - A^{\mathsf{T}}A\vec{c}$$

and solving gives this (showing that $A^{\mathsf{T}}A$ is invertible is an exercise).

$$\vec{c} = \bigl(A^{\mathsf{T}}A\bigr)^{-1}A^{\mathsf{T}} \cdot \vec{v}$$

Therefore $\mathrm{proj}_M(\vec{v}\,) = A \cdot \vec{c} = A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} \cdot \vec{v}$, as required. QED

3.9 Example To orthogonally project this vector into this subspace

$$\vec{v} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \qquad P = \{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x + z = 0 \}$$

first make a matrix whose columns are a basis for the subspace

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

and then compute.

$$\begin{aligned} A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & 0 & -1/2 \\ 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \end{pmatrix} \end{aligned}$$

With the matrix, calculating the orthogonal projection of any vector into P is easy.

$$\operatorname{proj}_P(\vec{v}) = \begin{pmatrix} 1/2 & 0 & -1/2 \\ 0 & 1 & 0 \\ -1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}$$

Note, as a check, that this result is indeed in P.

Exercises

✓ 3.10 Project the vectors into M along N.

(a) $\begin{pmatrix} 3 \\ -2 \end{pmatrix}$, $M = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid x + y = 0\}$, $N = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid -x - 2y = 0\}$

(b) $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $M = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid x - y = 0\}$, $N = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid 2x + y = 0\}$

(c) $\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}$, $M = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x + y = 0\}$, $N = \{c \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \mid c \in \mathbb{R}\}$

✓ 3.11 Find $M^{\perp}$.

(a) $M = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid x + y = 0\}$ **(b)** $M = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid -2x + 3y = 0\}$

(c) $M = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid x - y = 0\}$ **(d)** $M = \{\vec{0}\}$ **(e)** $M = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid x = 0\}$

(f) $M = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid -x + 3y + z = 0\}$ **(g)** $M = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x = 0 \text{ and } y + z = 0\}$

3.12 This subsection shows how to project orthogonally in two ways, the method of Example 3.2 and 3.3, and the method of Theorem 3.8. To compare them, consider the plane P specified by $3x + 2y - z = 0$ in $\mathbb{R}^3$.

(a) Find a basis for P.

(b) Find $P^{\perp}$ and a basis for $P^{\perp}$.

(c) Represent this vector with respect to the concatenation of the two bases from the prior item.

$$\vec{v} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

(d) Find the orthogonal projection of $\vec{v}$ into P by keeping only the P part from the prior item.

(e) Check that against the result from applying Theorem 3.8.

✓ **3.13** We have three ways to find the orthogonal projection of a vector into a line, the Definition 1.1 way from the first subsection of this section, the Example 3.2 and 3.3 way of representing the vector with respect to a basis for the space and then keeping the M part, and the way of Theorem 3.8. For these cases, do all three ways.

(a) $\vec{v} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}, \quad M = \{\begin{pmatrix} x \\ y \end{pmatrix} \mid x + y = 0\}$

(b) $\vec{v} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \quad M = \{\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x + z = 0 \text{ and } y = 0\}$

3.14 Check that the operation of Definition 3.1 is well-defined. That is, in Example 3.2 and 3.3, doesn't the answer depend on the choice of bases?

3.15 What is the orthogonal projection into the trivial subspace?

3.16 What is the projection of $\vec{v}$ into M along N if $\vec{v} \in M$?

3.17 Show that if $M \subseteq \mathbb{R}^n$ is a subspace with orthonormal basis $\langle \vec{\kappa}_1, \ldots, \vec{\kappa}_n \rangle$ then the orthogonal projection of $\vec{v}$ into M is this.

$$(\vec{v} \cdot \vec{\kappa}_1) \cdot \vec{\kappa}_1 + \cdots + (\vec{v} \cdot \vec{\kappa}_n) \cdot \vec{\kappa}_n$$

✓ **3.18** Prove that the map $p\colon V \to V$ is the projection into M along N if and only if the map $\mathrm{id} - p$ is the projection into N along M. (Recall the definition of the difference of two maps: $(\mathrm{id} - p)(\vec{v}) = \mathrm{id}(\vec{v}) - p(\vec{v}) = \vec{v} - p(\vec{v})$.)

✓ **3.19** Show that if a vector is perpendicular to every vector in a set then it is perpendicular to every vector in the span of that set.

3.20 True or false: the intersection of a subspace and its orthogonal complement is trivial.

3.21 Show that the dimensions of orthogonal complements add to the dimension of the entire space.

✓ **3.22** Suppose that $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^n$ are such that for all complements $M, N \subseteq \mathbb{R}^n$, the projections of $\vec{v}_1$ and $\vec{v}_2$ into M along N are equal. Must $\vec{v}_1$ equal $\vec{v}_2$? (If so, what if we relax the condition to: all orthogonal projections of the two are equal?)

✓ **3.23** Let M, N be subspaces of $\mathbb{R}^n$. The perp operator acts on subspaces; we can ask how it interacts with other such operations.

(a) Show that two perps cancel: $(M^{\perp})^{\perp} = M$.

(b) Prove that $M \subseteq N$ implies that $N^{\perp} \subseteq M^{\perp}$.

(c) Show that $(M + N)^{\perp} = M^{\perp} \cap N^{\perp}$.

✓ 3.24 The material in this subsection allows us to express a geometric relationship that we have not yet seen between the range space and the null space of a linear map.

(a) Represent $f\colon \mathbb{R}^3 \to \mathbb{R}$ given by
$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \mapsto 1v_1 + 2v_2 + 3v_3$$
with respect to the standard bases and show that
$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$
is a member of the perp of the null space. Prove that $\mathscr{N}(f)^\perp$ is equal to the span of this vector.

(b) Generalize that to apply to any $f\colon \mathbb{R}^n \to \mathbb{R}$.

(c) Represent $f\colon \mathbb{R}^3 \to \mathbb{R}^2$
$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \mapsto \begin{pmatrix} 1v_1 + 2v_2 + 3v_3 \\ 4v_1 + 5v_2 + 6v_3 \end{pmatrix}$$
with respect to the standard bases and show that
$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$$
are both members of the perp of the null space. Prove that $\mathscr{N}(f)^\perp$ is the span of these two. (*Hint.* See the third item of Exercise 23.)

(d) Generalize that to apply to any $f\colon \mathbb{R}^n \to \mathbb{R}^m$.

In [Strang 93] this is called the *Fundamental Theorem of Linear Algebra*

3.25 Define a *projection* to be a linear transformation $t\colon V \to V$ with the property that repeating the projection does nothing more than does the projection alone: $(t \circ t)(\vec{v}) = t(\vec{v})$ for all $\vec{v} \in V$.

(a) Show that orthogonal projection into a line has that property.

(b) Show that projection along a subspace has that property.

(c) Show that for any such t there is a basis $B = \langle \vec{\beta}_1, \ldots, \vec{\beta}_n \rangle$ for V such that
$$t(\vec{\beta}_i) = \begin{cases} \vec{\beta}_i & i = 1, 2, \ldots, r \\ \vec{0} & i = r+1, r+2, \ldots, n \end{cases}$$
where r is the rank of t.

(d) Conclude that every projection is a projection along a subspace.

(e) Also conclude that every projection has a representation
$$\mathrm{Rep}_{B,B}(t) = \left(\begin{array}{c|c} I & Z \\ \hline Z & Z \end{array}\right)$$
in block partial-identity form.

3.26 A square matrix is *symmetric* if each i, j entry equals the j, i entry (i.e., if the matrix equals its transpose). Show that the projection matrix $A(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$ is symmetric. [Strang 80] *Hint.* Find properties of transposes by looking in the index under 'transpose'.

Topic

Line of Best Fit

This Topic requires the formulas from the subsections on Orthogonal Projection Into a Line and Projection Into a Subspace.

Scientists are often presented with a system that has no solution and they must find an answer anyway. More precisely, they must find a best answer. For instance, this is the result of flipping a penny, including some intermediate numbers.

number of flips	30	60	90
number of heads	16	34	51

Because of the randomness in this experiment we expect that the ratio of heads to flips will fluctuate around a penny's long-term ratio of 50-50. So the system for such an experiment likely has no solution, and that's what happened here.

$$\begin{aligned} 30m &= 16 \\ 60m &= 34 \\ 90m &= 51 \end{aligned}$$

That is, the vector of data that we collected is not in the subspace where theory has it.

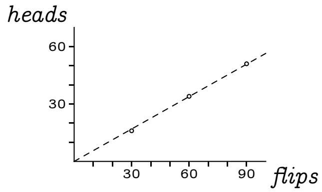
$$\begin{pmatrix} 16 \\ 34 \\ 51 \end{pmatrix} \notin \{ m \begin{pmatrix} 30 \\ 60 \\ 90 \end{pmatrix} \mid m \in \mathbb{R} \}$$

However, we have to do something so we look for the m that most nearly works. An orthogonal projection of the data vector into the line subspace gives this best guess.

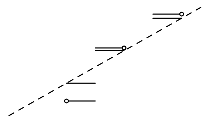
$$\frac{\begin{pmatrix} 16 \\ 34 \\ 51 \end{pmatrix} \bullet \begin{pmatrix} 30 \\ 60 \\ 90 \end{pmatrix}}{\begin{pmatrix} 30 \\ 60 \\ 90 \end{pmatrix} \bullet \begin{pmatrix} 30 \\ 60 \\ 90 \end{pmatrix}} \cdot \begin{pmatrix} 30 \\ 60 \\ 90 \end{pmatrix} = \frac{7110}{12600} \cdot \begin{pmatrix} 30 \\ 60 \\ 90 \end{pmatrix}$$

The estimate ($m = 7110/12600 \approx 0.56$) is a bit more than one half, but not much more than half, so probably the penny is fair enough.

The line with the slope $m \approx 0.56$ is the *line of best fit* for this data.



Minimizing the distance between the given vector and the vector used as the right-hand side minimizes the total of these vertical lengths, and consequently we say that the line comes from *fitting by least-squares*.



This diagram exaggerates the vertical scale by a factor of ten to make the lengths more visible.

In the above equation the line must pass through $(0, 0)$, because we take it to be the line whose slope is this coin's true proportion of heads to flips. We can also handle cases where the line need not pass through the origin.

Here is the progression of world record times for the men's mile race [Oakley & Baker]. In the early 1900's many people wondered when, or if, this record would fall below the four minute mark. Here are the times that were in force on January first of each decade through the first half of that century.

year	1870	1880	1890	1900	1910	1920	1930	1940	1950
secs	268.8	264.5	258.4	255.6	255.6	252.6	250.4	246.4	241.4

We can use this to give a circa 1950 prediction of the date for 240 seconds, and then compare that to the actual date. As with the penny data, these numbers do not line in a perfect line. That is, this system does not have an exact solution for the slope and intercept.

$$\begin{aligned} b + 1870m &= 268.8 \\ b + 1880m &= 264.5 \\ &\vdots \\ b + 1950m &= 241.4 \end{aligned}$$

We find a best approximation by using orthogonal projection.

(*Comments on the data.* Restricting to the times at the start of each decade reduces the data entry burden, smooths the data to some extent, and gives much the same result as entering all of the dates and records. There are different sequences of times from competing standards bodies but the ones here are from [Wikipedia, Mens Mile]. We've started the plot at 1870 because at one point there were two classes of records, called 'professional' and 'amateur', and after a while the first class stopped being active so we've followed the second class.)

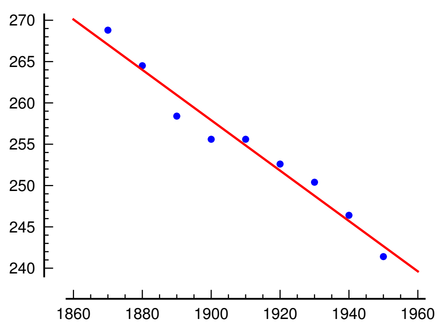
Write the linear system's matrix of coefficients and also its vector of constants, the world record times.

$$A = \begin{pmatrix} 1 & 1870 \\ 1 & 1880 \\ \vdots & \vdots \\ 1 & 1950 \end{pmatrix} \qquad \vec{v} = \begin{pmatrix} 268.8 \\ 264.5 \\ \vdots \\ 241.4 \end{pmatrix}$$

The ending result in the subsection on Projection into a Subspace gives the formula for the the coefficients b and m that make the linear combination of A's columns as close as possible to $\vec{v}$. Those coefficients are the entries of the vector $(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}} \cdot \vec{v}$.

Sage can do the computation for us.

```
sage: year = [1870, 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1950]
sage: secs = [268.8, 264.5, 258.4, 255.6, 255.6, 252.6, 250.4, 246.4, 241.4]
sage: var('a, b, t')
(a, b, t)
sage: model(t) = a*t+b
sage: data = zip(year, secs)
sage: fit = find_fit(data, model, solution_dict=True)
sage: model.subs(fit)
t |--> -0.3048333333333295*t + 837.0872222222147
sage: g=points(data)+plot(model.subs(fit),(t,1860,1960),color='red',
....:                    figsize=3,fontsize=7,typeset='latex')
sage: g.save("four_minute_mile.pdf")
sage: g
```



The progression makes a surprisingly good line. From the slope and intercept we predict 1958.73; the actual date of Roger Bannister's record was 1954-May-06.

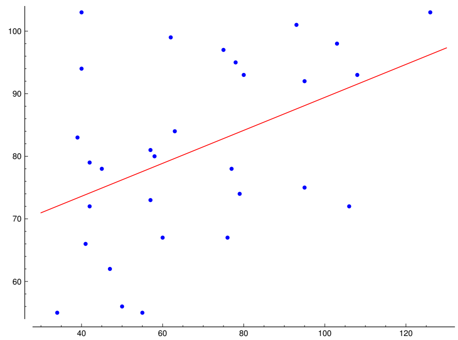
The final example compares team salaries from US major league baseball against the number of wins the team had, for the year 2002. In this year the

Oakland Athletics used mathematical techniques to optimize the players that they fielded for the money that they could spend, as told in the film *Moneyball*. (Salaries are in millions of dollars and the number of wins is out of 162 games).

To do the computations we again use *Sage*.

```
sage: sal = [40, 40, 39, 42, 45, 42, 62, 34, 41, 57, 58, 63, 47, 75, 57, 78, 80, 50, 60, 93,
....:        77, 55, 95, 103, 79, 76, 108, 126, 95, 106]
sage: wins = [103, 94, 83, 79, 78, 72, 99, 55, 66, 81, 80, 84, 62, 97, 73, 95, 93, 56, 67,
....:        101, 78, 55, 92, 98, 74, 67, 93, 103, 75, 72]
sage: var('a, b, t')
(a, b, t)
sage: model(t) = a*t+b
sage: data = zip(sal,wins)
sage: fit = find_fit(data, model, solution_dict=True)
sage: model.subs(fit)
t |--> 0.2634981251436269*t + 63.06477642781477
sage: p = points(data,size=25)+plot(model.subs(fit),(t,30,130),color='red',typeset='latex')
sage: p.save('moneyball.pdf')
```

The graph is below. The team in the upper left, who paid little for many wins, is the Oakland A's.



Judging this line by eye would be error-prone. So the equations give us a certainty about the 'best' in best fit. In addition, the model's equation tells us roughly that by spending an additional million dollars a team owner can expect to buy $1/4$ of a win (and that expectation is not very sure, thank goodness).

Exercises

The calculations here are best done on a computer. Some of the problems require data from the Internet.

1 Use least-squares to judge if the coin in this experiment is fair.

flips	8	16	24	32	40
heads	4	9	13	17	20

2 For the men's mile record, rather than give each of the many records and its exact date, we've "smoothed" the data somewhat by taking a periodic sample. Do the longer calculation and compare the conclusions.

3 Find the line of best fit for the men's 1500 meter run. How does the slope compare with that for the men's mile? (The distances are close; a mile is about 1609 meters.)

4 Find the line of best fit for the records for women's mile.

5 Do the lines of best fit for the men's and women's miles cross?

6 *(This illustrates that there are data sets for which a linear model is not right, and that the line of best fit doesn't in that case have any predictive value.)* In a highway restaurant a trucker told me that his boss often sends him by a roundabout route, using more gas but paying lower bridge tolls. He said that New York State calibrates the toll for each bridge across the Hudson, playing off the extra gas to get there from New York City against a lower crossing cost, to encourage people to go upstate. This table, from [Cost Of Tolls] and [Google Maps], lists for each toll crossing of the Hudson River, the distance to drive from Times Square in miles and the cost in US dollars for a passenger car (if a crossings has a one-way toll then it shows half that number).

Crossing	*Distance*	*Toll*
Lincoln Tunnel	2	6.00
Holland Tunnel	7	6.00
George Washington Bridge	8	6.00
Verrazano-Narrows Bridge	16	6.50
Tappan Zee Bridge	27	2.50
Bear Mountain Bridge	47	1.00
Newburgh-Beacon Bridge	67	1.00
Mid-Hudson Bridge	82	1.00
Kingston-Rhinecliff Bridge	102	1.00
Rip Van Winkle Bridge	120	1.00

Find the line of best fit and graph the data to show that the driver was practicing on my credulity.

7 When the space shuttle Challenger exploded in 1986, one of the criticisms made of NASA's decision to launch was in the way they did the analysis of number of O-ring failures versus temperature (O-ring failure caused the explosion). Four O-ring failures would be fatal. NASA had data from 24 previous flights.

temp $^\circ F$	53	75	57	58	63	70	70	66	67	67	67
failures	3	2	1	1	1	1	1	0	0	0	0

temp $^\circ F$	68	69	70	70	72	73	75	76	76	78	79	80	81
failures	0	0	0	0	0	0	0	0	0	0	0	0	0

The temperature that day was forecast to be 31°F.

(a) NASA based the decision to launch partially on a chart showing only the flights that had at least one O-ring failure. Find the line that best fits these seven flights. On the basis of this data, predict the number of O-ring failures when the temperature is 31, and when the number of failures will exceed four.

(b) Find the line that best fits all 24 flights. On the basis of this extra data, predict the number of O-ring failures when the temperature is 31, and when the number of failures will exceed four.

Which do you think is the more accurate method of predicting? (An excellent discussion is in [Dalal, et. al.].)

8 This table lists the average distance from the sun to each of the first seven planets, using Earth's average as a unit.

Mercury	Venus	Earth	Mars	Jupiter	Saturn	Uranus
0.39	0.72	1.00	1.52	5.20	9.54	19.2

(a) Plot the number of the planet (Mercury is 1, etc.) versus the distance. Note that it does not look like a line, and so finding the line of best fit is not fruitful.

(b) It does, however look like an exponential curve. Therefore, plot the number of the planet versus the logarithm of the distance. Does this look like a line?

(c) The asteroid belt between Mars and Jupiter is what is left of a planet that broke apart. Renumber so that Jupiter is 6, Saturn is 7, and Uranus is 8, and plot against the log again. Does this look better?

(d) Use least squares on that data to predict the location of Neptune.

(e) Repeat to predict where Pluto is.

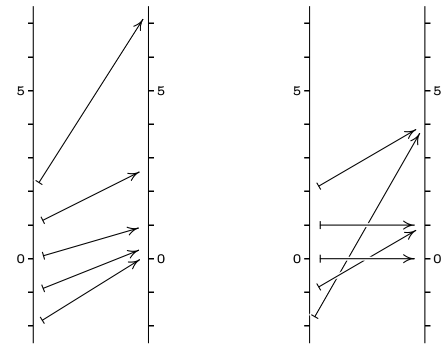
(f) Is the formula accurate for Neptune and Pluto?

This method was used to help discover Neptune (although the second item is misleading about the history; actually, the discovery of Neptune in position 9 prompted people to look for the "missing planet" in position 5). See [Gardner, 1970]

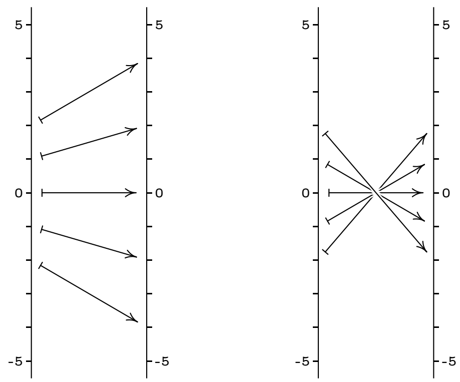
Topic

Geometry of Linear Maps

These pairs of pictures contrast the geometric action of the nonlinear maps $f_1(x) = e^x$ and $f_2(x) = x^2$



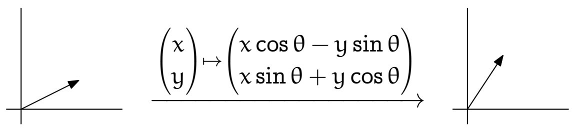
with the linear maps $h_1(x) = 2x$ and $h_2(x) = -x$.



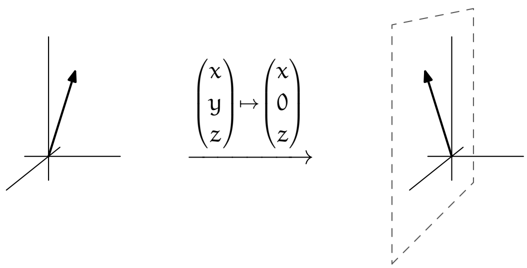
Each of the four pictures shows the domain $\mathbb{R}$ on the left mapped to the codomain $\mathbb{R}$ on the right. Arrows trace where each map sends $x = 0$, $x = 1$, $x = 2$, $x = -1$, and $x = -2$.

The nonlinear maps distort the domain in transforming it into the range. For instance, $f_1(1)$ is further from $f_1(2)$ than it is from $f_1(0)$ — this map spreads the domain out unevenly so that a domain interval near $x = 2$ is spread apart more than is a domain interval near $x = 0$. The linear maps are nicer, more regular, in that for each map all of the domain spreads by the same factor. The map h_1 on the left spreads all intervals apart to be twice as wide while on the right h_2 keeps intervals the same length but reverses their orientation, as with the rising interval from 1 to 2 being transformed to the falling interval from -1 to -2.

The only linear maps from $\mathbb{R}$ to $\mathbb{R}$ are multiplications by a scalar but in higher dimensions more can happen. For instance, this linear transformation of $\mathbb{R}^2$ rotates vectors counterclockwise.



The transformation of $\mathbb{R}^3$ that projects vectors into the xz-plane is also not simply a rescaling.



Despite this additional variety, even in higher dimensions linear maps behave nicely. Consider a linear $h\colon \mathbb{R}^n \to \mathbb{R}^m$ and use the standard bases to represent it by a matrix H. Recall that H factors into $H = PBQ$ where P and Q are nonsingular and B is a partial-identity matrix. Recall also that nonsingular matrices factor into elementary matrices $PBQ = T_nT_{n-1}\cdots T_sBT_{s-1}\cdots T_1$, which are matrices that come from the identity I after one Gaussian row operation, so each T matrix is one of these three kinds

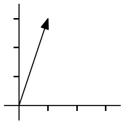
$$I \xrightarrow{k\rho_i} M_i(k) \qquad I \xrightarrow{\rho_i \leftrightarrow \rho_j} P_{i,j} \qquad I \xrightarrow{k\rho_i + \rho_j} C_{i,j}(k)$$

with $i \neq j$, $k \neq 0$. So if we understand the geometric effect of a linear map described by a partial-identity matrix and the effect of the linear maps described by the elementary matrices then we will in some sense completely understand the effect of any linear map. (The pictures below stick to transformations of $\mathbb{R}^2$ for ease of drawing but the principles extend for maps from any $\mathbb{R}^n$ to any $\mathbb{R}^m$.)

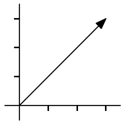
The geometric effect of the linear transformation represented by a partial-identity matrix is projection.

$$\begin{pmatrix}x\\y\\z\end{pmatrix}\xrightarrow{\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}}\begin{pmatrix}x\\y\\0\end{pmatrix}$$

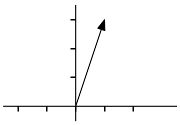
The geometric effect of the $M_i(k)$ matrices is to stretch vectors by a factor of k along the i-th axis. This map stretches by a factor of 3 along the x-axis.



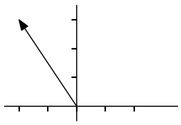
$$\xrightarrow{\begin{pmatrix}x\\y\end{pmatrix}\mapsto\begin{pmatrix}3x\\y\end{pmatrix}}$$



If $0 \leqslant k < 1$ or if $k < 0$ then the i-th component goes the other way, here to the left.

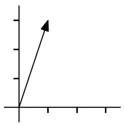


$$\xrightarrow{\begin{pmatrix}x\\y\end{pmatrix}\mapsto\begin{pmatrix}-2x\\y\end{pmatrix}}$$

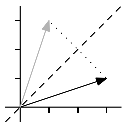


Either of these stretches is a *dilation*.

A transformation represented by a $P_{i,j}$ matrix interchanges the i-th and j-th axes. This is *reflection* about the line $x_i = x_j$.



$$\xrightarrow{\begin{pmatrix}x\\y\end{pmatrix}\mapsto\begin{pmatrix}y\\x\end{pmatrix}}$$

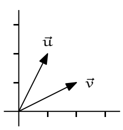


Permutations involving more than two axes decompose into a combination of swaps of pairs of axes; see Exercise 5.

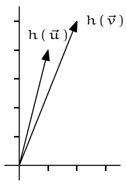
The remaining matrices have the form $C_{i,j}(k)$. For instance $C_{1,2}(2)$ performs $2\rho_1 + \rho_2$.

$$\begin{pmatrix}x\\y\end{pmatrix}\xrightarrow{\begin{pmatrix}1&0\\2&1\end{pmatrix}}\begin{pmatrix}x\\2x+y\end{pmatrix}$$

In the picture below, the vector $\vec{u}$ with the first component of 1 is affected less than the vector $\vec{v}$ with the first component of 2. The vector $\vec{u}$ is mapped to a $h(\vec{u})$ that is only 2 higher than $\vec{u}$ while $h(\vec{v})$ is 4 higher than $\vec{v}$.

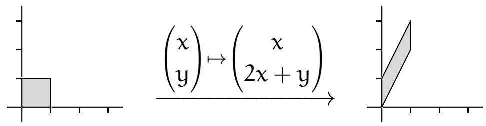


$$\xrightarrow{\begin{pmatrix}x\\y\end{pmatrix}\mapsto\begin{pmatrix}x\\2x+y\end{pmatrix}}$$

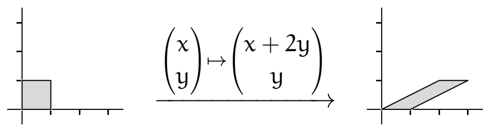


Any vector with a first component of 1 would be affected in the same way as $\vec{u}$: it would slide up by 2. And any vector with a first component of 2 would slide up 4, as was $\vec{v}$. That is, the transformation represented by $C_{i,j}(k)$ affects vectors depending on their i-th component.

Another way to see this point is to consider the action of this map on the unit square. In the next picture, vectors with a first component of 0, such as the origin, are not pushed vertically at all but vectors with a positive first component slide up. Here, all vectors with a first component of 1, the entire right side of the square, slide to the same extent. In general, vectors on the same vertical line slide by the same amount, by twice their first component. The resulting shape has the same base and height as the square (and thus the same area) but the right angle corners are gone.



For contrast, the next picture shows the effect of the map represented by $C_{2,1}(2)$. Here vectors are affected according to their second component: $\binom{x}{y}$ slides horizontally by twice y.

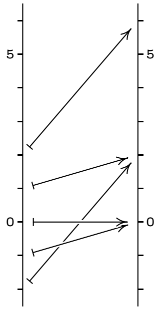


In general, for any $C_{i,j}(k)$, the sliding happens so that vectors with the same i-th component are slid by the same amount. This kind of map is a *shear*.

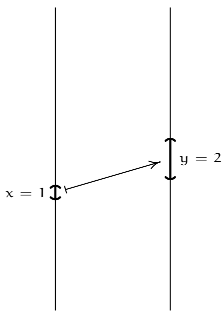
With that we understand the geometric effect of the four types of matrices on the right-hand side of $H = T_n T_{n-1} \cdots T_j B T_{j-1} \cdots T_1$ and so in some sense we understand the action of any matrix H. Thus, even in higher dimensions the geometry of linear maps is easy: it is built by putting together a number of components, each of which acts in a simple way.

We will apply this understanding in two ways. The first way is to prove something general about the geometry of linear maps. Recall that under a linear map, the image of a subspace is a subspace and thus the linear transformation h represented by H maps lines through the origin to lines through the origin. (The dimension of the image space cannot be greater than the dimension of the domain space, so a line can't map onto, say, a plane.) We will show that h maps any line — not just one through the origin — to a line. The proof is simple: the partial-identity projection B and the elementary T_i's each turn a line input into a line output; verifying the four cases is Exercise 6. Therefore their composition also preserves lines.

The second way that we will apply the geometric understanding of linear maps is to elucidate a point from Calculus. Below is a picture of the action of the one-variable real function $y(x) = x^2 + x$. As with the nonlinear functions pictured earlier, the geometric effect of this map is irregular in that at different domain points it has different effects; for example as the input x goes from 2 to -2, the associated output $f(x)$ at first decreases, then pauses for an instant, and then increases.



But in Calculus we focus less on the map overall and more on the local effect of the map. Below we look closely at what this map does near $x = 1$. The derivative is $dy/dx = 2x + 1$ so that near $x = 1$ we have $\Delta y \approx 3 \cdot \Delta x$. That is, in a neighborhood of $x = 1$, in carrying the domain over this map causes it to grow by a factor of 3—it is, locally, approximately, a dilation. The picture below shows this as a small interval in the domain $(1 - \Delta x \,..\, 1 + \Delta x)$ carried over to an interval in the codomain $(2 - \Delta y \,..\, 2 + \Delta y)$ that is three times as wide.



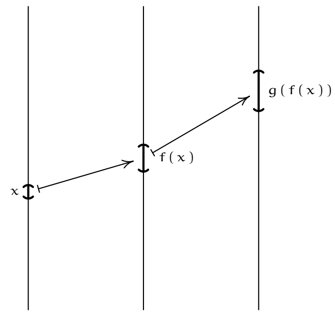
In higher dimensions the core idea is the same but more can happen. For a function $y\colon \mathbb{R}^n \to \mathbb{R}^m$ and a point $\vec{x} \in \mathbb{R}^n$, the derivative is defined to be the linear map $h\colon \mathbb{R}^n \to \mathbb{R}^m$ that best approximates how y changes near $y(\vec{x})$. So the geometry described above directly applies to the derivative.

We close by remarking how this point of view makes clear an often misunderstood result about derivatives, the Chain Rule. Recall that, under suitable conditions on the two functions, the derivative of the composition is this.

$$\frac{d\,(g \circ f)}{dx}(x) = \frac{dg}{dx}(f(x)) \cdot \frac{df}{dx}(x)$$

For instance the derivative of $\sin(x^2 + 3x)$ is $\cos(x^2 + 3x) \cdot (2x + 3)$.

Where does this come from? Consider $f, g \colon \mathbb{R} \to \mathbb{R}$.



The first map f dilates the neighborhood of x by a factor of

$$\frac{df}{dx}(x)$$

and the second map g follows that by dilating a neighborhood of $f(x)$ by a factor of

$$\frac{dg}{dx}(\,f(x)\,)$$

and when combined, the composition dilates by the product of the two. In higher dimensions the map expressing how a function changes near a point is a linear map, and is represented by a matrix. The Chain Rule multiplies the matrices.

Exercises

1 Let $h \colon \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates vectors clockwise by $\pi/4$ radians.

(a) Find the matrix H representing h with respect to the standard bases. Use Gauss's Method to reduce H to the identity.

(b) Translate the row reduction to a matrix equation $T_j T_{j-1} \cdots T_1 H = I$ (the prior item shows both that H is similar to I, and that we need no column operations to derive I from H).

(c) Solve this matrix equation for H.

(d) Sketch how H is a combination of dilations, flips, skews, and projections (the identity is a trivial projection).

2 What combination of dilations, flips, skews, and projections produces a rotation counterclockwise by $2\pi/3$ radians?

3 What combination of dilations, flips, skews, and projections produces the map $h\colon \mathbb{R}^3 \to \mathbb{R}^3$ represented with respect to the standard bases by this matrix?

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & 6 & 0 \\ 1 & 2 & 2 \end{pmatrix}$$

4 Show that any linear transformation of $\mathbb{R}^1$ is the map that multiplies by a scalar $x \mapsto kx$.

5 Show that for any permutation (that is, reordering) p of the numbers $1, \ldots, n$, the map

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} x_{p(1)} \\ x_{p(2)} \\ \vdots \\ x_{p(n)} \end{pmatrix}$$

can be done with a composition of maps, each of which only swaps a single pair of coordinates. *Hint:* you can use induction on n. (*Remark:* in the fourth chapter we will show this and we will also show that the parity of the number of swaps used is determined by p. That is, although a particular permutation could be expressed in two different ways with two different numbers of swaps, either both ways use an even number of swaps, or both use an odd number.)

6 Show that linear maps preserve the linear structures of a space.

(a) Show that for any linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$, the image of any line is a line. The image may be a degenerate line, that is, a single point.

(b) Show that the image of any linear surface is a linear surface. This generalizes the result that under a linear map the image of a subspace is a subspace.

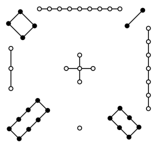
(c) Linear maps preserve other linear ideas. Show that linear maps preserve "betweeness": if the point B is between A and C then the image of B is between the image of A and the image of C.

7 Use a picture like the one that appears in the discussion of the Chain Rule to answer: if a function $f\colon \mathbb{R} \to \mathbb{R}$ has an inverse, what's the relationship between how the function — locally, approximately — dilates space, and how its inverse dilates space (assuming, of course, that it has an inverse)?

Topic

Magic Squares

A Chinese legend tells the story of a flood by the Lo river. People offered sacrifices to appease the river. Each time a turtle emerged, walked around the sacrifice, and returned to the water. Fuh-Hi, the founder of Chinese civilization, interpreted this to mean that the river was still cranky. Fortunately, a child noticed that on its shell the turtle had the pattern on the left below, which is today called Lo Shu ("river scroll").



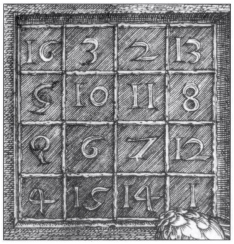
4	9	2
3	5	7
8	1	6

The dots make the matrix on the right where the rows, columns, and diagonals add to 15. Now that the people knew how much to sacrifice, the river's anger cooled.

A square matrix is *magic* if each row, column, and diagonal add to the same number, the matrix's *magic number*.

Another magic square appears in the engraving *Melencolia I* by Dürer.

One interpretation is that it depicts melancholy, a depressed state. The figure, genius, has a wealth of fascinating things to explore including the compass, the geometrical solid, the scale, and the hourglass. But the figure is unmoved; all of the things lie unused. One of the potential delights, in the upper right, is a 4×4 matrix whose rows, columns, and diagonals add to 34.



16	3	2	13
5	10	11	8
9	6	7	12
4	15	14	1

The middle entries on the bottom row give 1514, the date of the engraving.

The above two squares are arrangements of $1 \ldots n^2$. They are *normal*. The 1×1 square whose sole entry is 1 is normal, Exercise 2 shows that there is no normal 2×2 magic square, and there are normal magic squares of every other size; see [Wikipedia, Magic Square]. Finding how many normal magic squares there are of each size is an unsolved problem; see [Online Encyclopedia of Integer Sequences].

If we don't require that the squares be normal then we can say much more. Every 1×1 square is magic, trivially. If the rows, columns, and diagonals of a 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

add to s then $a+b=s$, $c+d=s$, $a+c=s$, $b+d=s$, $a+d=s$, and $b+c=s$. Exercise 2 shows that this system has the unique solution $a=b=c=d=s/2$. So the set of 2×2 magic squares is a one-dimensional subspace of $\mathcal{M}_{2\times2}$.

A sum of two same-sized magic squares is magic and a scalar multiple of a magic square is magic so the set of $n\times n$ magic squares $\mathscr{M}_n$ is a vector space, a subspace of $\mathcal{M}_{n\times n}$. This Topic shows that for $n\geqslant3$ the dimension of $\mathscr{M}_n$ is n^2-n. The set $\mathscr{M}_{n,0}$ of $n\times n$ magic squares with magic number 0 is another subspace and we will verify the formula for its dimension also: n^2-2n-1 when $n\geqslant3$.

We will first prove that $\dim\mathscr{M}_n = \dim\mathscr{M}_{n,0}+1$. Define the *trace* of a matrix to be the sum down its upper-left to lower-right diagonal $\operatorname{Tr}(M) = m_{1,1}+\cdots+m_{n,n}$. Consider the restriction of the trace to the magic squares $\operatorname{Tr}\colon \mathscr{M}_n \to \mathbb{R}$. The null space $\mathscr{N}(\operatorname{Tr})$ is the set of magic squares with magic number zero $\mathscr{M}_{n,0}$. Observe that the trace is onto because for any r in the codomain $\mathbb{R}$ the $n\times n$ matrix whose entries are all r/n is a magic square with magic number r. Theorem Two.II.2.14 says that for any linear map the dimension of the domain equals the dimension of the range space plus the dimension of the null space, the map's rank plus its nullity. Here the domain is $\mathscr{M}_n$, the range space is $\mathbb{R}$ and the null space is $\mathscr{M}_{n,0}$, so we have that $\dim\mathscr{M}_n = 1+\dim\mathscr{M}_{n,0}$.

We will finish by finding the dimension of the vector space $\mathscr{M}_{n,0}$. For $n=1$ the dimension is clearly 0. Exercise 3 shows that $\dim\mathscr{M}_{n,0}$ is also 0 for $n=2$.

That leaves showing that $\dim\mathscr{M}_{n,0}=n^2-2n-1$ for $n\geqslant 3$. The fact that the squares in this vector space are magic gives us a linear system of restrictions, and the fact that they have magic number zero makes this system homogeneous: for instance consider the 3×3 case. The restriction that the rows, columns, and diagonals of

$$\begin{pmatrix} a&b&c\\ d&e&f\\ g&h&i \end{pmatrix}$$

add to zero gives this $(2n+2)\times n^2$ linear system.

$$\begin{array}{rrrrrrrrrl}
a+b+c & & & & & & & & &=0\\
 & & & d+e+f & & & & & &=0\\
 & & & & & & g+h+i & & &=0\\
a & & +d & & & +g & & & &=0\\
 & b & & +e & & & +h & & &=0\\
 & & c & & +f & & & +i & &=0\\
a & & & +e & & & & +i & &=0\\
 & & c & +e & & +g & & & &=0
\end{array}$$

We will find the dimension of the space by finding the number of free variables in the linear system.

The matrix of coefficients for the particular cases of $n=3$ and $n=4$ are below, with the rows and columns numbered to help in reading the proof. With respect to the standard basis, each represents a linear map $h\colon\mathbb{R}^{n^2}\to\mathbb{R}^{2n+2}$. The domain has dimension n^2 so if we show that the rank of the matrix is $2n+1$ then we will have what we want, that the dimension of the null space $\mathscr{M}_{n,0}$ is $n^2-(2n+1)$.

$$\begin{array}{c|ccc ccc ccc}
 & 1&2&3&4&5&6&7&8&9\\
\hline
\vec\rho_1 & 1&1&1&0&0&0&0&0&0\\
\vec\rho_2 & 0&0&0&1&1&1&0&0&0\\
\vec\rho_3 & 0&0&0&0&0&0&1&1&1\\
\vec\rho_4 & 1&0&0&1&0&0&1&0&0\\
\vec\rho_5 & 0&1&0&0&1&0&0&1&0\\
\vec\rho_6 & 0&0&1&0&0&1&0&0&1\\
\vec\rho_7 & 1&0&0&0&1&0&0&0&1\\
\vec\rho_8 & 0&0&1&0&1&0&1&0&0
\end{array}$$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\vec{\rho}_1$	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
$\vec{\rho}_2$	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
$\vec{\rho}_3$	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
$\vec{\rho}_4$	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
$\vec{\rho}_5$	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
$\vec{\rho}_6$	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0
$\vec{\rho}_7$	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
$\vec{\rho}_8$	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
$\vec{\rho}_9$	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
$\vec{\rho}_{10}$	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0

We want to show that the rank of the matrix of coefficients, the number of rows in a maximal linearly independent set, is $2n+1$. The first n rows of the matrix of coefficients add to the same vector as the second n rows, the vector of all ones. So a maximal linearly independent must omit at least one row. We will show that the set of all rows but the first $\{\vec{\rho}_2\ \ldots\ \vec{\rho}_{2n+2}\}$ is linearly independent. So consider this linear relationship.

$$c_2\vec{\rho}_2 + \cdots + c_{2n}\vec{\rho}_{2n} + c_{2n+1}\vec{\rho}_{2n+1} + c_{2n+2}\vec{\rho}_{2n+2} = \vec{0} \qquad (*)$$

Now it gets messy. Focus on the lower left of the tables. Observe that in the final two rows, in the first n columns, is a subrow that is all zeros except that it starts with a one in column 1 and a subrow that is all zeros except that it ends with a one in column n.

First, with $\vec{\rho}_1$ omitted, both column 1 and column n contain only two ones. Since the only rows in $(*)$ with nonzero column 1 entries are rows $\vec{\rho}_{n+1}$ and $\vec{\rho}_{2n+1}$, which have ones, we must have $c_{2n+1} = -c_{n+1}$. Likewise considering the n-th entries of the vectors in $(*)$ gives that $c_{2n+2} = -c_{2n}$.

Next consider the columns between those two — in the $n = 3$ table this includes only column 2 while in the $n = 4$ table it includes both columns 2 and 3. Each such column has a single one. That is, for each column index $j \in \{2\ \ldots\ n-2\}$ the column consists of only zeros except for a one in row $n+j$, and hence $c_{n+j} = 0$.

On to the next block of columns, from $n+1$ through $2n$. Column $n+1$ has only two ones (because $n \geqslant 3$ the ones in the last two rows do not fall in the first column of this block). Thus $c_2 = -c_{n+1}$ and therefore $c_2 = c_{2n+1}$. Likewise, from column $2n$ we conclude that $c_2 = -c_{2n}$ and so $c_2 = c_{2n+2}$.

Because $n \geqslant 3$ there is at least one column between column $n+1$ and column $2n-1$. In at least one of those columns a one appears in $\vec{\rho}_{2n+1}$. If a one also appears in that column in $\vec{\rho}_{2n+2}$ then we have $c_2 = -(c_{2n+1} + c_{2n+2})$ since

$c_{n+j} = 0$ for $j \in \{2 \ldots n-2\}$. If a one does not appear in that column in $\vec{\rho}_{2n+2}$ then we have $c_2 = -c_{2n+1}$. In either case $c_2 = 0$, and thus $c_{2n+1} = c_{2n+2} = 0$ and $c_{n+1} = c_{2n} = 0$.

If the next block of n-many columns is not the last then similarly conclude from its first column that $c_3 = c_{n+1} = 0$.

Keep this up until we reach the last block of columns, those numbered $(n-1)n+1$ through n^2. Because $c_{n+1} = \cdots = c_{2n} = 0$ column n^2 gives that $c_n = -c_{2n+1} = 0$.

Therefore the rank of the matrix is $2n+1$, as required.

The classic source on normal magic squares is [Ball & Coxeter]. More on the Lo Shu square is at [Wikipedia, Lo Shu Square]. The proof given here began with [Ward].

Exercises

1 Let M be a 3×3 magic square with magic number s.
  (a) Prove that the sum of M's entries is $3s$.
  (b) Prove that $s = 3 \cdot m_{2,2}$.
  (c) Prove that $m_{2,2}$ is the average of the entries in its row, its column, and in each diagonal.
  (d) Prove that $m_{2,2}$ is the median of M's entries.

2 Solve the system $a+b=s$, $c+d=s$, $a+c=s$, $b+d=s$, $a+d=s$, and $b+c=s$.

3 Show that $\dim \mathscr{M}_{2,0} = 0$.

4 Let the *trace* function be $\operatorname{Tr}(M) = m_{1,1} + \cdots + m_{n,n}$. Define also the sum down the other diagonal $\operatorname{Tr}^*(M) = m_{1,n} + \cdots + m_{n,1}$.
  (a) Show that the two functions $\operatorname{Tr}, \operatorname{Tr}^*\colon \mathcal{M}_{n\times n} \to \mathbb{R}$ are linear.
  (b) Show that the function $\theta\colon \mathcal{M}_{n\times n} \to \mathbb{R}^2$ given by $\theta(M) = (\operatorname{Tr}(M), \operatorname{Tr}^*(m))$ is linear.
  (c) Generalize the prior item.

5 A square matrix is *semimagic* if the rows and columns add to the same value, that is, if we drop the condition on the diagonals.
  (a) Show that the set of semimagic squares $\mathscr{H}_n$ is a subspace of $\mathcal{M}_{n\times n}$.
  (b) Show that the set $\mathscr{H}_{n,0}$ of $n\times n$ semimagic squares with magic number 0 is also a subspace of $\mathcal{M}_{n\times n}$.

6 [Beardon] Here is a slicker proof of the result of this Topic, when $n \geqslant 3$. See the prior two exercises for some definitions and needed results.
  (a) First show that $\dim \mathscr{M}_{n,0} = \dim \mathscr{H}_{n,0} + 2$. To do this, consider the function $\theta\colon \mathcal{M}_n \to \mathbb{R}^2$ sending a matrix M to the ordered pair $(\operatorname{Tr}(M), \operatorname{Tr}^*(M))$. Specifically, consider the restriction of that map $\theta\colon \mathscr{H}_n \to \mathbb{R}^2$ to the semimagic squares. Clearly its null space is $\mathscr{M}_{n,0}$. Show that when $n \geqslant 3$ this restriction θ is onto. (*Hint:* we need only find a basis for $\mathbb{R}^2$ that is the image of two members of $\mathcal{H}_n$)
  (b) Let the function $\phi\colon \mathcal{M}_{n\times n} \to \mathcal{M}_{(n-1)\times(n-1)}$ be the identity map except that it drops the final row and column: $\phi(M) = \hat{M}$ where $\hat{m}_{i,j} = m_{i,j}$ for all $i,j \in \{1 \ldots n-1\}$. The check that ϕ is linear is easy. Consider ϕ's restriction to

the semimagic squares with magic number zero $\phi\colon \mathscr{H}_{n,0} \to \mathcal{M}_{(n-1)\times(n-1)}$. Show that ϕ is one-to-one

(c) Show that ϕ is onto.

(d) Conclude that $\mathscr{H}_{n,0}$ has dimension $(n-1)^2$.

(e) Conclude that $\dim(\mathscr{M}_n) = n^2 - n$

Topic

Markov Chains

Here is a simple game: a player bets on coin tosses, a dollar each time, and the game ends either when the player has no money or is up to five dollars. If the player starts with three dollars, what is the chance that the game takes at least five flips? Twenty-five flips?

At any point, this player has either \$0, or \$1, ..., or \$5. We say that the player is in the *state* s_0, s_1, ..., or s_5. In the game the player moves from state to state. For instance, a player now in state s_3 has on the next flip a 0.5 chance of moving to state s_2 and a 0.5 chance of moving to s_4. The boundary states are different; a player never leaves state s_0 or state s_5.

Let $p_i(n)$ be the probability that the player is in state s_i after n flips. Then for instance the probability of being in state s_0 after flip $n+1$ is $p_0(n+1) = p_0(n) + 0.5 \cdot p_1(n)$. This equation summarizes.

$$\begin{pmatrix} 1.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 1.0 \end{pmatrix} \begin{pmatrix} p_0(n) \\ p_1(n) \\ p_2(n) \\ p_3(n) \\ p_4(n) \\ p_5(n) \end{pmatrix} = \begin{pmatrix} p_0(n+1) \\ p_1(n+1) \\ p_2(n+1) \\ p_3(n+1) \\ p_4(n+1) \\ p_5(n+1) \end{pmatrix}$$

Sage will compute the evolution of this game.

```
sage: M = matrix(RDF, [[1.0, 0.5, 0.0, 0.0, 0.0, 0.0],
....:                  [0.5, 0.0, 0.5, 0.0, 0.0, 0.0],
....:                  [0.0, 0.5, 0.0, 0.5, 0.0, 0.0],
....:                  [0.0, 0.0, 0.5, 0.0, 0.5, 0.0],
....:                  [0.0, 0.0, 0.0, 0.5, 0.0, 0.5],
....:                  [0.0, 0.0, 0.0, 0.0, 0.5, 1.0]])
sage: M = M.transpose()
sage: v0 = vector(RDF, [0.0, 0.0, 0.0, 1.0, 0.0, 0.0])
sage: v1 = v0*M
sage: v1
(0.0, 0.0, 0.5, 0.0, 0.5, 0.0)
sage: v2 = v1*M
sage: v2
(0.0, 0.25, 0.0, 0.5, 0.0, 0.25)
```

(Two notes: (1) *Sage* can use various number systems to make the matrix entries and here we have used Real Double Float, and (2) *Sage* likes to do matrix multiplication from the right, as $\vec{v}M$ instead of our usual $M\vec{v}$, so we needed to take the matrix's transpose.)

These are components of the resulting vectors.

$n=0$	$n=1$	$n=2$	$n=3$	$n=4$	$\cdots$	$n=24$
0	0	0	0.125	0.125		0.39600
0	0	0.25	0	0.1875		0.00276
0	0.5	0	0.375	0		0
1	0	0.5	0	0.3125		0.00447
0	0.5	0	0.25	0		0
0	0	0.25	0.25	0.375		0.59676

This game is not likely to go on for long since the player quickly moves to an ending state. For instance, after the fourth flip there is already a 0.50 probability that the game is over.

This is a *Markov chain*. Each vector is a *probability vector*, whose entries are nonnegative real numbers that sum to 1. The matrix is a *transition matrix* or *stochastic matrix*, whose entries are nonnegative reals and whose columns sum to 1.

A characteristic feature of a Markov chain model is that it is *historyless* in that the next state depends only on the current state, not on any prior ones. Thus, a player who arrives at s_2 by starting in state s_3 and then going to state s_2 has exactly the same chance of moving next to s_3 as does a player whose history was to start in s_3 then go to s_4 then to s_3 and then to s_2.

Here is a Markov chain from sociology. A study ([Macdonald & Ridge], p. 202) divided occupations in the United Kingdom into three levels: executives and professionals, supervisors and skilled manual workers, and unskilled workers. They asked about two thousand men, "At what level are you, and at what level was your father when you were fourteen years old?" Here the Markov model assumption about history may seem reasonable — we may guess that while a parent's occupation has a direct influence on the occupation of the child, the grandparent's occupation likely has no such direct influence. This summarizes the study's conclusions.

$$\begin{pmatrix} .60 & .29 & .16 \\ .26 & .37 & .27 \\ .14 & .34 & .57 \end{pmatrix} \begin{pmatrix} p_U(n) \\ p_M(n) \\ p_L(n) \end{pmatrix} = \begin{pmatrix} p_U(n+1) \\ p_M(n+1) \\ p_L(n+1) \end{pmatrix}$$

For instance, looking at the middle class for the next generation, a child of an upper class worker has a 0.26 probability of becoming middle class, a child of a middle class worker has a 0.37 chance of being middle class, and a child of a lower class worker has a 0.27 probability of becoming middle class.

Sage will compute the successive stages of this system (the current class distribution is $\vec{v}_0$).

```
sage: M = matrix(RDF, [[0.60, 0.29, 0.16],
....:                  [0.26, 0.37, 0.27],
....:                  [0.14, 0.34, 0.57]])
sage: M = M.transpose()
sage: v0 = vector(RDF, [0.12, 0.32, 0.56])
sage: v0*M
(0.2544, 0.3008, 0.4448)
sage: v0*M^2
(0.31104, 0.297536, 0.391424)
sage: v0*M^3
(0.33553728, 0.2966432, 0.36781952)
```

Here are the next five generations. They show upward mobility, especially in the first generation. In particular, lower class shrinks a good bit.

$n=0$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
.12	.25	.31	.34	.35	.35
.32	.30	.30	.30	.30	.30
.56	.44	.39	.37	.36	.35

One more example. In professional American baseball there are two leagues, the American League and the National League. At the end of the annual season the team winning the American League and the team winning the National League play the World Series. The winner is the first team to take four games. That means that a series is in one of twenty-four states: 0-0 (no games won yet by either team), 1-0 (one game won for the American League team and no games for the National League team), etc.

Consider a series with a probability p that the American League team wins each game. We have this.

$$\begin{pmatrix} 0 & 0 & 0 & 0 & \ldots \\ p & 0 & 0 & 0 & \ldots \\ 1-p & 0 & 0 & 0 & \ldots \\ 0 & p & 0 & 0 & \ldots \\ 0 & 1-p & p & 0 & \ldots \\ 0 & 0 & 1-p & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \begin{pmatrix} p_{\text{0-0}}(n) \\ p_{\text{1-0}}(n) \\ p_{\text{0-1}}(n) \\ p_{\text{2-0}}(n) \\ p_{\text{1-1}}(n) \\ p_{\text{0-2}}(n) \\ \vdots \end{pmatrix} = \begin{pmatrix} p_{\text{0-0}}(n+1) \\ p_{\text{1-0}}(n+1) \\ p_{\text{0-1}}(n+1) \\ p_{\text{2-0}}(n+1) \\ p_{\text{1-1}}(n+1) \\ p_{\text{0-2}}(n+1) \\ \vdots \end{pmatrix}$$

An especially interesting special case is when the teams are evenly matched, $p = 0.50$. This table below lists the resulting components of the $n = 0$ through $n = 7$ vectors.

Note that evenly-matched teams are likely to have a long series — there is a probability of 0.625 that the series goes at least six games.

	$n=0$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
$0-0$	1	0	0	0	0	0	0	0
$1-0$	0	0.5	0	0	0	0	0	0
$0-1$	0	0.5	0	0	0	0	0	0
$2-0$	0	0	0.25	0	0	0	0	0
$1-1$	0	0	0.5	0	0	0	0	0
$0-2$	0	0	0.25	0	0	0	0	0
$3-0$	0	0	0	0.125	0	0	0	0
$2-1$	0	0	0	0.375	0	0	0	0
$1-2$	0	0	0	0.375	0	0	0	0
$0-3$	0	0	0	0.125	0	0	0	0
$4-0$	0	0	0	0	0.0625	0.0625	0.0625	0.0625
$3-1$	0	0	0	0	0.25	0	0	0
$2-2$	0	0	0	0	0.375	0	0	0
$1-3$	0	0	0	0	0.25	0	0	0
$0-4$	0	0	0	0	0.0625	0.0625	0.0625	0.0625
$4-1$	0	0	0	0	0	0.125	0.125	0.125
$3-2$	0	0	0	0	0	0.3125	0	0
$2-3$	0	0	0	0	0	0.3125	0	0
$1-4$	0	0	0	0	0	0.125	0.125	0.125
$4-2$	0	0	0	0	0	0	0.15625	0.15625
$3-3$	0	0	0	0	0	0	0.3125	0
$2-4$	0	0	0	0	0	0	0.15625	0.15625
$4-3$	0	0	0	0	0	0	0	0.15625
$3-4$	0	0	0	0	0	0	0	0.15625

Markov chains are a widely used application of matrix operations. They also give us an example of the use of matrices where we do not consider the significance of the maps represented by the matrices. For more on Markov chains, there are many sources such as [Kemeny & Snell] and [Iosifescu].

Exercises

1 These questions refer to the coin-flipping game.

(a) Check the computations in the table at the end of the first paragraph.

(b) Consider the second row of the vector table. Note that this row has alternating 0's. Must $p_1(j)$ be 0 when j is odd? Prove that it must be, or produce a counterexample.

(c) Perform a computational experiment to estimate the chance that the player ends at five dollars, starting with one dollar, two dollars, and four dollars.

2 [Feller] We consider throws of a die, and say the system is in state s_i if the largest number yet appearing on the die was i.

(a) Give the transition matrix.

(b) Start the system in state s_1, and run it for five throws. What is the vector at the end?

3 [Kelton] There has been much interest in whether industries in the United States are moving from the Northeast and North Central regions to the South and West, motivated by the warmer climate, by lower wages, and by less unionization. Here is the transition matrix for large firms in Electric and Electronic Equipment.

	NE	*NC*	*S*	*W*	*Z*
NE	0.787	0	0	0.111	0.102
NC	0	0.966	0.034	0	0
S	0	0.063	0.937	0	0
W	0	0	0.074	0.612	0.314
Z	0.021	0.009	0.005	0.010	0.954

For example, a firm in the Northeast region will be in the West region next year with probability 0.111. (The *Z* entry is a "birth-death" state. For instance, with probability 0.102 a large Electric and Electronic Equipment firm from the Northeast will move out of this system next year: go out of business, move abroad, or move to another category of firm. There is a 0.021 probability that a firm in the *National Census of Manufacturers* will move into Electronics, or be created, or move in from abroad, into the Northeast. Finally, with probability 0.954 a firm out of the categories will stay out, according to this research.)

(a) Does the Markov model assumption of lack of history seem justified?

(b) Assume that the initial distribution is even, except that the value at Z is 0.9. Compute the vectors for $n = 1$ through $n = 4$.

(c) Suppose that the initial distribution is this.

NE	*NC*	*S*	*W*	*Z*
0.0000	0.6522	0.3478	0.0000	0.0000

Calculate the distributions for $n = 1$ through $n = 4$.

(d) Find the distribution for $n = 50$ and $n = 51$. Has the system settled down to an equilibrium?

4 [Wickens] Here is a model of some kinds of learning The learner starts in an undecided state s_U. Eventually the learner has to decide to do either response A (that is, end in state s_A) or response B (ending in s_B). However, the learner doesn't jump right from undecided to sure that A is the correct thing to do (or B). Instead, the learner spends some time in a "tentative-A" state, or a "tentative-B" state, trying the response out (denoted here t_A and t_B). Imagine that once the learner has decided, it is final, so once in s_A or s_B, the learner stays there. For the other state changes, we can posit transitions with probability p in either direction.

(a) Construct the transition matrix.

(b) Take $p = 0.25$ and take the initial vector to be 1 at s_U. Run this for five steps. What is the chance of ending up at s_A?

(c) Do the same for $p = 0.20$.

(d) Graph p versus the chance of ending at s_A. Is there a threshold value for p, above which the learner is almost sure not to take longer than five steps?

5 A certain town is in a certain country (this is a hypothetical problem). Each year ten percent of the town dwellers move to other parts of the country. Each year one percent of the people from elsewhere move to the town. Assume that there are two

states s_T, living in town, and s_C, living elsewhere.

(a) Construct the transition matrix.

(b) Starting with an initial distribution $s_T = 0.3$ and $s_C = 0.7$, get the results for the first ten years.

(c) Do the same for $s_T = 0.2$.

(d) Are the two outcomes alike or different?

6 For the World Series application, use a computer to generate the seven vectors for $p = 0.55$ and $p = 0.6$.

(a) What is the chance of the National League team winning it all, even though they have only a probability of 0.45 or 0.40 of winning any one game?

(b) Graph the probability p against the chance that the American League team wins it all. Is there a threshold value—a p above which the better team is essentially ensured of winning?

7 Above we define a transition matrix to have each entry nonnegative and each column sum to 1.

(a) Check that the three transition matrices shown in this Topic meet these two conditions. Must any transition matrix do so?

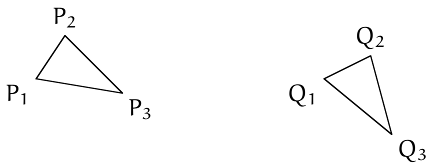
(b) Observe that if $A\vec{v}_0 = \vec{v}_1$ and $A\vec{v}_1 = \vec{v}_2$ then A^2 is a transition matrix from $\vec{v}_0$ to $\vec{v}_2$. Show that a power of a transition matrix is also a transition matrix.

(c) Generalize the prior item by proving that the product of two appropriately-sized transition matrices is a transition matrix.

Topic

Orthonormal Matrices

In *The Elements*, Euclid considers two figures to be the same if they have the same size and shape. That is, while the triangles below are not equal because they are not the same set of points, they are, for Euclid's purposes, essentially indistinguishable because we can imagine picking the plane up, sliding it over and rotating it a bit, although not warping or stretching it, and then putting it back down, to superimpose the first figure on the second. (Euclid never explicitly states this principle but he uses it often [Casey].)



In modern terms "picking the plane up ..." is taking a map from the plane to itself. Euclid considers only transformations that may slide or turn the plane but not bend or stretch it. Accordingly, define a map $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ to be *distance-preserving* or a *rigid motion* or an *isometry* if for all points $P_1, P_2 \in \mathbb{R}^2$, the distance from $f(P_1)$ to $f(P_2)$ equals the distance from P_1 to P_2. We also define a plane *figure* to be a set of points in the plane and we say that two figures are *congruent* if there is a distance-preserving map from the plane to itself that carries one figure onto the other.

Many statements from Euclidean geometry follow easily from these definitions. Some are: (i) collinearity is invariant under any distance-preserving map (that is, if P_1, P_2, and P_3 are collinear then so are $f(P_1)$, $f(P_2)$, and $f(P_3)$), (ii) betweeness is invariant under any distance-preserving map (if P_2 is between P_1 and P_3 then so is $f(P_2)$ between $f(P_1)$ and $f(P_3)$), (iii) the property of being a triangle is invariant under any distance-preserving map (if a figure is a triangle then the image of that figure is also a triangle), (iv) and the property of being a circle is invariant under any distance-preserving map. In 1872, F. Klein suggested that we can define Euclidean geometry as the study of properties that are invariant under these maps. (This forms part of Klein's Erlanger Program, which proposes the organizing principle that we can describe each kind of geometry — Euclidean, projective, etc. — as the study of the properties that are invariant under some group of transformations. The word 'group' here means more than just 'collection' but that lies outside of our scope.)

We can use linear algebra to characterize the distance-preserving maps of the plane.

To begin, observe that there are distance-preserving transformations of the plane that are not linear. The obvious example is this *translation*.

$$\begin{pmatrix} x \\ y \end{pmatrix} \quad \mapsto \quad \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} x+1 \\ y \end{pmatrix}$$

However, this example turns out to be the only one, in that if f is distance-preserving and sends $\vec{0}$ to $\vec{v}_0$ then the map $\vec{v} \mapsto f(\vec{v}) - \vec{v}_0$ is linear. That will follow immediately from this statement: a map t that is distance-preserving and sends $\vec{0}$ to itself is linear. To prove this equivalent statement, consider the standard basis and suppose that

$$t(\vec{e}_1) = \begin{pmatrix} a \\ b \end{pmatrix} \qquad t(\vec{e}_2) = \begin{pmatrix} c \\ d \end{pmatrix}$$

for some $a, b, c, d \in \mathbb{R}$. To show that t is linear we can show that it can be represented by a matrix, that is, that t acts in this way for all $x, y \in \mathbb{R}$.

$$\vec{v} = \begin{pmatrix} x \\ y \end{pmatrix} \stackrel{t}{\longmapsto} \begin{pmatrix} ax+cy \\ bx+dy \end{pmatrix} \tag{$*$}$$

Recall that if we fix three non-collinear points then we can determine any point by giving its distance from those three. So we can determine any point $\vec{v}$ in the domain by its distance from $\vec{0}$, $\vec{e}_1$, and $\vec{e}_2$. Similarly, we can determine any point $t(\vec{v})$ in the codomain by its distance from the three fixed points $t(\vec{0})$, $t(\vec{e}_1)$, and $t(\vec{e}_2)$ (these three are not collinear because, as mentioned above, collinearity is invariant and $\vec{0}$, $\vec{e}_1$, and $\vec{e}_2$ are not collinear). Because t is distance-preserving we can say more: for the point $\vec{v}$ in the plane that is determined by being the distance d_0 from $\vec{0}$, the distance d_1 from $\vec{e}_1$, and the distance d_2 from $\vec{e}_2$, its image $t(\vec{v})$ must be the unique point in the codomain that is determined by being d_0 from $t(\vec{0})$, d_1 from $t(\vec{e}_1)$, and d_2 from $t(\vec{e}_2)$. Because of the uniqueness, checking that the action in $(*)$ works in the d_0, d_1, and d_2 cases

$$\mathrm{dist}(\begin{pmatrix} x \\ y \end{pmatrix}, \vec{0}) = \mathrm{dist}(t(\begin{pmatrix} x \\ y \end{pmatrix}), t(\vec{0})) = \mathrm{dist}(\begin{pmatrix} ax+cy \\ bx+dy \end{pmatrix}, \vec{0})$$

(we assumed that t maps $\vec{0}$ to itself)

$$\operatorname{dist}(\begin{pmatrix} x \\ y \end{pmatrix}, \vec{e}_1) = \operatorname{dist}(t(\begin{pmatrix} x \\ y \end{pmatrix}), t(\vec{e}_1)) = \operatorname{dist}(\begin{pmatrix} ax+cy \\ bx+dy \end{pmatrix}, \begin{pmatrix} a \\ b \end{pmatrix})$$

and

$$\operatorname{dist}(\begin{pmatrix} x \\ y \end{pmatrix}, \vec{e}_2) = \operatorname{dist}(t(\begin{pmatrix} x \\ y \end{pmatrix}), t(\vec{e}_2)) = \operatorname{dist}(\begin{pmatrix} ax+cy \\ bx+dy \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix})$$

suffices to show that $(*)$ describes t. Those checks are routine.

Thus any distance-preserving $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map plus a translation, $f(\vec{v}) = t(\vec{v}) + \vec{v}_0$ for some constant vector $\vec{v}_0$ and linear map t that is distance-preserving. So in order to understand distance-preserving maps what remains is to understand distance-preserving linear maps.

Not every linear map is distance-preserving. For example $\vec{v} \mapsto 2\vec{v}$ does not preserve distances.

But there is a neat characterization: a linear transformation t of the plane is distance-preserving if and only if both $\|t(\vec{e}_1)\| = \|t(\vec{e}_2)\| = 1$, and $t(\vec{e}_1)$ is orthogonal to $t(\vec{e}_2)$. The 'only if' half of that statement is easy — because t is distance-preserving it must preserve the lengths of vectors and because t is distance-preserving the Pythagorean theorem shows that it must preserve orthogonality. To show the 'if' half we can check that the map preserves lengths of vectors because then for all $\vec{p}$ and $\vec{q}$ the distance between the two is preserved $\|t(\vec{p}-\vec{q}\,)\| = \|t(\vec{p}) - t(\vec{q}\,)\| = \|\vec{p}-\vec{q}\,\|$. For that check let

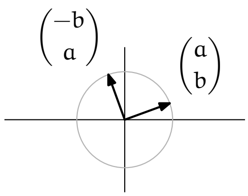
$$\vec{v} = \begin{pmatrix} x \\ y \end{pmatrix} \quad t(\vec{e}_1) = \begin{pmatrix} a \\ b \end{pmatrix} \quad t(\vec{e}_2) = \begin{pmatrix} c \\ d \end{pmatrix}$$

and with the 'if' assumptions that $a^2+b^2 = c^2+d^2 = 1$ and $ac+bd = 0$ we have this.

$$\begin{aligned}
\|t(\vec{v}\,)\|^2 &= (ax+cy)^2 + (bx+dy)^2 \\
&= a^2x^2 + 2acxy + c^2y^2 + b^2x^2 + 2bdxy + d^2y^2 \\
&= x^2(a^2+b^2) + y^2(c^2+d^2) + 2xy(ac+bd) \\
&= x^2 + y^2 \\
&= \|\vec{v}\,\|^2
\end{aligned}$$

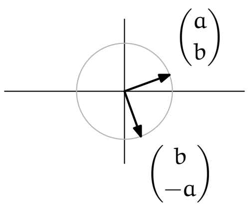
One thing that is neat about this characterization is that we can easily recognize matrices that represent such a map with respect to the standard bases: the columns are of length one and are mutually orthogonal. This is an *orthonormal matrix* (or, more informally, *orthogonal matrix* since people often use this term to mean not just that the columns are orthogonal but also that they have length one).

We can leverage this characterization to understand the geometric actions of distance-preserving maps. Because $\|t(\vec{v}\,)\| = \|\vec{v}\,\|$, the map t sends any $\vec{v}$ somewhere on the circle about the origin that has radius equal to the length of $\vec{v}$. In particular, $\vec{e}_1$ and $\vec{e}_2$ map to the unit circle. What's more, once we fix the unit vector $\vec{e}_1$ as mapped to the vector with components a and b then there are only two places where $\vec{e}_2$ can go if its image is to be perpendicular to the first vector's image: it can map either to one where $\vec{e}_2$ maintains its position a quarter circle clockwise from $\vec{e}_1$



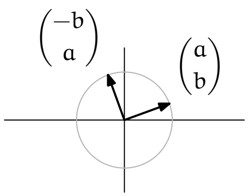
$$\text{Rep}_{\mathcal{E}_2,\mathcal{E}_2}(t) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

or to one where it goes a quarter circle counterclockwise.



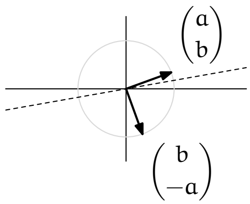
$$\text{Rep}_{\mathcal{E}_2,\mathcal{E}_2}(t) = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$$

The geometric description of these two cases is easy. Let θ be the counterclockwise angle between the x-axis and the image of $\vec{e}_1$. The first matrix above represents, with respect to the standard bases, a *rotation* of the plane by θ radians.



$$\begin{pmatrix} x \\ y \end{pmatrix} \overset{t}{\longmapsto} \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}$$

The second matrix above represents a *reflection* of the plane through the line bisecting the angle between $\vec{e}_1$ and $t(\vec{e}_1)$.



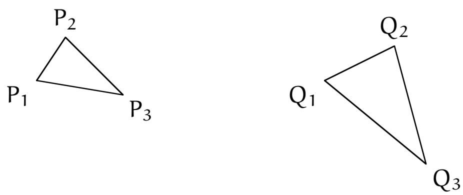
$$\begin{pmatrix} x \\ y \end{pmatrix} \stackrel{t}{\longmapsto} \begin{pmatrix} x\cos\theta + y\sin\theta \\ x\sin\theta - y\cos\theta \end{pmatrix}$$

(This picture shows $\vec{e}_1$ reflected up into the first quadrant and $\vec{e}_2$ reflected down into the fourth quadrant.)

Note: in the domain the angle between $\vec{e}_1$ and $\vec{e}_2$ runs counterclockwise, and in the first map above the angle from $t(\vec{e}_1)$ to $t(\vec{e}_2)$ is also counterclockwise, so it preserves the orientation of the angle. But the second map reverses the orientation. A distance-preserving map is *direct* if it preserves orientations and *opposite* if it reverses orientation.

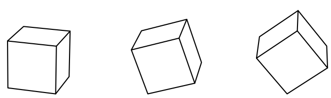
With that, we have characterized the Euclidean study of congruence. It considers, for plane figures, the properties that are invariant under combinations of (i) a rotation followed by a translation, or (ii) a reflection followed by a translation (a reflection followed by a non-trivial translation is a *glide reflection*).

Another idea encountered in elementary geometry, besides congruence of figures, is that figures are *similar* if they are congruent after a change of scale. The two triangles below are similar since the second is the same shape as the first but 3/2-ths the size.



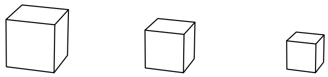
From the above work we have that figures are similar if there is an orthonormal matrix T such that the points $\vec{q}$ on one figure are the images of the points $\vec{p}$ on the other figure by $\vec{q} = (kT)\vec{v} + \vec{p}_0$ for some nonzero real number k and constant vector $\vec{p}_0$.

Although these ideas are from Euclid, mathematics is timeless and they are still in use today. One application of the maps studied above is in computer graphics. We can, for example, animate this top view of a cube by putting together film frames of it rotating; that's a rigid motion.



Frame 1 Frame 2 Frame 3

We could also make the cube appear to be moving away from us by producing film frames of it shrinking, which gives us figures that are similar.



Frame 1: Frame 2: Frame 3:

Computer graphics incorporates techniques from linear algebra in many other ways (see Exercise 4).

A beautiful book that explores some of this area is [Weyl]. More on groups, of transformations and otherwise, is in any book on Modern Algebra, for instance [Birkhoff & MacLane]. More on Klein and the Erlanger Program is in [Yaglom].

Exercises

1 Decide if each of these is an orthonormal matrix.

(a) $\begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ -1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$

(b) $\begin{pmatrix} 1/\sqrt{3} & -1/\sqrt{3} \\ -1/\sqrt{3} & -1/\sqrt{3} \end{pmatrix}$

(c) $\begin{pmatrix} 1/\sqrt{3} & -\sqrt{2}/\sqrt{3} \\ -\sqrt{2}/\sqrt{3} & -1/\sqrt{3} \end{pmatrix}$

2 Write down the formula for each of these distance-preserving maps.

(a) the map that rotates $\pi/6$ radians, and then translates by $\vec{e}_2$

(b) the map that reflects about the line $y = 2x$

(c) the map that reflects about $y = -2x$ and translates over 1 and up 1

3 **(a)** The proof that a map that is distance-preserving and sends the zero vector to itself incidentally shows that such a map is one-to-one and onto (the point in the domain determined by d_0, d_1, and d_2 corresponds to the point in the codomain determined by those three). Therefore any distance-preserving map has an inverse. Show that the inverse is also distance-preserving.

(b) Prove that congruence is an equivalence relation between plane figures.

4 In practice the matrix for the distance-preserving linear transformation and the translation are often combined into one. Check that these two computations yield the same first two components.

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \qquad \begin{pmatrix} a & c & e \\ b & d & f \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

(These are *homogeneous coordinates*; see the Topic on Projective Geometry).

5 **(a)** Verify that the properties described in the second paragraph of this Topic as invariant under distance-preserving maps are indeed so.

(b) Give two more properties that are of interest in Euclidean geometry from your experience in studying that subject that are also invariant under distance-preserving maps.

(c) Give a property that is not of interest in Euclidean geometry and is not invariant under distance-preserving maps.