Linear Algebra
Done Again*

John McCuan

October 16, 2020

Linear algebra is the study of linear functions

$$L : V \to W. \tag{1}$$

The property making a function **linear** is

$$L(av + bw) = aLv + bLw. \tag{2}$$

This property can be broken into two properties:

$$L(v + w) = L(v) + L(w) \tag{3}$$

and

$$L(av) = aL(v). \tag{4}$$

*This is a sarcastic reference to Sheldon Axler's well-known linear algebra text *Linear Algebra Done Right*. This is the way I think linear algebra "should" be presented. I wonder what professor Axler would think. Actually, it is not really clear to me how linear algebra should be presented. I have come to think of linear algebra as a relatively easy subject, but it is very often difficult for students. It was difficult for me to figure out what was going on when I learned it. The text and presentation might have been part of the problem. The key, I think, was motivation, and I did not really understand linear algebra until I understood it as a tool for the approximation of other more complicated, nonlinear, functions. This point of view is not the basis of any textbook on linear algebra, as far as I am aware. My favorite text has come to be *Linear Algebra* by Charles Curtis, though my presentation is radically different from that of Curtis. Another (truly famous) text is that of Gilbert Strang. One can also watch Strang's lectures on the internet—MIT open courseware. In some sense, Strang has set the standard for linear algebra instruction. Again, my presentation is radically different from Strang. The famous mathematician Peter Lax once took it upon himself to present linear algebra as a tool for doing computer graphics. That sounds like a really great idea, but the execution (surprisingly since Lax is a master at mathematical exposition) was disappointing.

The first of these two properties (3) is called **additivity**. The second (4) is called **multiplicative homogeneity** of order one. Together the two properties are equivalent to (2) which is called **linearity**.

Exercise 1 1. *Show that linearity implies/includes additivity.*

2. *Show that linearity implies/includes homogeneity.*

3. *Find an additive function which is not homogeneous.*

4. *Find a homogeneous function which is not additive.*

5. *Show that together additivity and homogeneity imply linearity.*

For any of what I've said above to make any sense at all, the domain V and co-domain W of the function L must be **vector spaces**. I'll discuss the technicalities of what it means to be a vector space below, but I'll guess you have a pretty good (intuitive) idea already. Look at the scaling property (4):

$$L(av) = aL(v).$$

The symbols a and v are playing rather different roles here. The symbol v represents a **vector** (whatever that is) and a represents a "number" or, more properly, a **scalar**. So, when you have a vector space you have scalars and vectors and enough algebra so that the properties above make sense. Again, I'll give the details presently, but the basic idea should be, more or less, clear: We are interested in functions L which are additive over vectors and homogeneous with respect to scaling.

1 Vector Spaces

In order to have a vector space V, which you can think of as a particular kind of set—the kind of set that can be the domain of a linear function—you also need a second set, which is a set of scalars. This second set needs to be a **field**. I'll tell you, technically, what a field of scalars is below, but again, just imagine for a moment that you know, and if you're not that imaginative, then think of the real numbers $\mathbb{R}$ as the field.

We say we have a vector space V **over** a field. This means we have a way to add any two elements of V, which are called **vectors** and a way to combine a scalar and a vector. That is, **scaling** is a function which takes a scalar a and a vector v and gives you back a vector av. I'll talk about this a little more below, but you should note that I've already started listing the properties of a vector space: A **vector space** V over a field F is a set V for which

1. There exists an operation of addition, that is a function $+ : V \times V \to V$ taking pairs (v, w) of vectors to their sum, written $+(v, w) = v + w$.

2. There exists a scaling which is an operation combining a scalar and a vector to give back a vector:
$$(a, v) \mapsto av.$$

   Exercise 2 *What is the domain and co-domain for scaling?*

   The other properties are these

3. Addition is **commutative**: $v + w = w + v$ whenever $v, w \in V$.

4. Addition is **associative**: $v + (w + z) = (v + w) + z$ whenever $v, w, z \in V$.

5. There exists a special vector called the **zero vector** $\mathbf{0} \in V$. This vector has the property that
$$v + \mathbf{0} = \mathbf{0} + v = v \qquad \text{for every } v \in V.$$
   The zero vector is called the **additive identity** and this property is called the property of the additive identity.

   Exercise 3 *Prove that the additive identity is unique, i.e., there can only be one zero vector. Hint: Assume there are two of them, and then show the two you have are equal to each other.*

6. Each vector $v \in V$ has an **additive inverse**, that is for each $v \in V$, there exists another vector $-v$ for which
$$v + (-v) = \mathbf{0}.$$
   We can also write the sum $v + (-v)$ simply as $v - v$.

   Exercise 4 *Show that additive inverses are unique.*

7. Scaling is associative: $a(bv) = (ab)v$ whenever $v \in V$ and a and b are scalars.

8. $0v = \mathbf{0}$ and $1v = v$ for any vector $v \in V$ where 0 is the additive identity in the field and 1 is the multiplicative identity in the field.

9. $a(v + w) = av + aw$ for $v, w \in V$ and a a scalar.

10. $(a + b)v = av + bv$ for $v \in V$ and a and b scalars.

Properties 1,3,4,5, and 6 are expressing the fact that V is required to be a **commutative group** under addition. (I'll come back to the notion of a group below.) Property 1 is sometimes called closure under addition. Property 2 is sometimes called closure under scaling. Property 9 is the distributive property of scaling across (sums of) vectors. Property 10 says that a vector distributes across a sum of scalars. Together properties 9 and 10 are called the distributive properties.

Exercise 5 *Show that the condition $0v = \mathbf{0}$ in property 8 can be omitted, i.e., prove that this must hold using the other properties.*

Exercise 6 *Look up the Wikipedia page giving the definition of a vector space and note that it lists only eight properties. Why do I have ten instead? Consider the following property:*

(a) V is a commutative group under addition.

How many properties would you need to list if you started with this as your first property?

Two important scalar fields (for us) are $\mathbb{R}$ and $\mathbb{C}$. The two simplest vector spaces to keep in mind are

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_1, x_2, \ldots, x_n \in \mathbb{R}\}$$

which is a field over $\mathbb{R}$ and

$$\mathbb{C}^n = \{(z_1, z_2, \ldots, z_n) : z_1, z_2, \ldots, z_n \in \mathbb{C}\}$$

which is a field over $\mathbb{C}$.

Incidentally, here is a definition (in case you missed it above):

Definition 1 *A* **vector** *is an element of a vector space.*

1.1 A Little Abstract Algebra

A vector space has one internal operation (addition) and a kind of external operation, namely scaling, in which a field "acts" on the vector space. That is, this second operation of scaling is a function $\cdot : F \times V \to V$ where F is the field. We of course write $\cdot(a, v) = av$.

A **field** has **two internal operations**. One is called addition and the other multiplication. Note: There is not really multiplication in a vector space, in general. There is scaling which is sometimes thought of as multiplying scalars with vectors, but that is not quite the same thing as a genuine multiplication. Thus, in a field F, there is addition $+ : F \times F \to F$ and multiplication $\cdot : F \times F \to F$.

The properties defining a field have a lot in common with those of a vector space. In fact, every field is also a vector space over itself. Here are the properties defining a field F:

1. F is a commutative group under addition.

2. $F \setminus \{0\}$ is a commutative group under multiplication.

3. Multiplication is distributive across addition, i.e.,

$$a(b + c) = ab + ac \quad \text{for every } a, b, c \in F.$$

To make everything utterly explicit, I should tell you the definition of a group.

Definition 2 *A* **group** *G is a set with an operation $* : G \times G \to G$ satisfying the following properties:*

1. *(associative) $g * (h * k) = (g * h) * k$ whenever $g, h, k \in G$.*

2. *(identity element) There exists an element $e \in G$ for which $e * g = g * e = g$ for every $g \in G$.*

3. *(inverses) For each $g \in G$, there is an element $g^{-1} \in G$ for which $g^{-1} * g = g * g^{-1} = e$.*

If a group happens to satisfy the following property:

$$g * h = h * g \qquad \text{for every } h, g \in G,$$

then the group is called a **commutative group**. *(This property is called...you guessed it...the commutative property.)*

Exercise 7 *Give an example of a group which is not commutative. Hint: Think about the set of square matrices under matrix multiplication.*

You might be surprised to be informed that there are other fields besides $\mathbb{R}$ and $\mathbb{C}$. Actually, there are a lot of them. One with which you are familiar is

$$\mathbb{Q} = \{p/q : p \in \mathbb{Z} \text{ and } q \in \mathbb{N}\},$$

the field of **rational numbers**. Here we are writing $\mathbb{N} = \{1, 2, 3, \ldots\}$ for the set of natural numbers and $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\}$ for the set of integers. The integers $\mathbb{Z}$ form a group under addition, and multiplication is also well-defined, but $\mathbb{Z}$ is not a group under multiplication. $\mathbb{N}$ is not even a group under addition.

Exercise 8 *Why isn't $\mathbb{Z}$ a group under multiplication? Is there an operation of multiplication on $\mathbb{Z}\backslash\{0\}$?*

Returning to the rational field, we can construct another field

$$\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}.$$

This is a field which is intermediate between $\mathbb{Q}$ and $\mathbb{R}$. It is the smallest field in which the polynomial $p(x) = x^2 - 2$ in $\mathbb{Q}[x]$ "splits" or factors.

Let me unpack the last sentence a little bit. The set $\mathbb{Q}[x]$ is the collection of all polynomials in x with rational coefficients. The elements look like

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = \sum_{j=0}^{n} a_j x^j.$$

This set $\mathbb{Q}[x]$ is most often called the **polynomial ring** over $\mathbb{Q}$. A **ring** is another kind of set that is important in abstract algebra. I won't get into it, but you can think of it as a poor man's field.

Exercise 9 *Why/how does $\mathbb{Q}[x]$ fail to be a field? Show $\mathbb{Q}[x]$ is a vector space over $\mathbb{Q}$.*

Considered as a polynomial in $\mathbb{Q}[x]$, the polynomial $x^2 - 2$ is irreducible, that is, it does not factor. But as a polynomial in $\mathbb{Q}[\sqrt{2}][x]$ we can write

$$x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2}).$$

We can't write this in $\mathbb{Q}[x]$ because $\sqrt{2}$ is not allowed as a coefficient. Of course we can also factor $p(x) = x^2 - 2$ in $\mathbb{R}[x]$, but $\mathbb{R}$ is not the *smallest* field that allows this polynomial to split. Fields like $\mathbb{Q}[\sqrt{2}]$ are called **Galois** extensions, named after the famous (and very tragic) French mathematician Évariste Galois.

Exercise 10 *What is the smallest field extension over which the polynomial $p(x) = x^2 + 1$ splits when considered as a polynomial in $\mathbb{R}[x]$?*

Exercise 11 *Show that $\mathbb{Q}[\sqrt{2}]$ is a vector space over $\mathbb{Q}$. What is the dimension of this vector space?*

1.2 Linear Combinations, Spans, and Linear Independence

I will go ahead and record these definitions/notions/ideas here.

The fundamental expression $av + bw$ appearing in the definition of linearity is an example of a **linear combination**. A linear combination is constructed by scaling (two) vectors and then adding the results. One can also construct a linear combination of any finite collection of vectors using the same approach: Scale each vector and then add up the results:

$$\sum_{j=1}^{k} a_j v_j.$$

The scalars $a_1, a_2, \ldots, a_k$ in a linear combination are called the coefficients.

The set of all linear combinations of a subset of a vector space is called the **span** of that subset. That is, given an set $S \subset V$ where V is a vector space

$$\text{span}(S) = \left\{ \sum_{j=1}^{k} a_j v_j : a_1, a_2, \ldots, a_k \in F, v_1, v_2, \ldots, v_k \in S \right\}$$

is called the span of the set S.

Exercise 12 *Show that the span of any subset of a vector space is also a vector space.*

If V is a vector space and $V = \text{span}(S)$ for some subset $S \subset V$, then S is said to be a **spanning set** for V. In the special case where $S = \{v\}$ is a subset of a vector space containing a single vector, the span of S is sometimes denoted by

$$\text{span}\{v\} = \langle v \rangle.$$

In this case, if v is the zero vector, then $\langle v \rangle = \{\mathbf{0}\}$ is the null (zero dimensional) vector subspace of V. Otherwise, if $v \neq 0$, then $\langle v \rangle = \{av : a \in F\}$ is said to be one-dimensional.

Definition 3 *A set of vectors $S \subset V$ in a vector space V is* **linearly independent** *if whenever one has a linear combination of vectors in S for which*

$$\sum_{j=1}^{k} a_j v_j = \mathbf{0},$$

then all the coefficients must be zero. A subset $S \subset V$ is **linearly dependent** *if there exist scalars $a_1, a_2, \ldots, a_k$ which are* **not all zero** *and vectors $v_1, v_2, \ldots, v_k \in S$ such that*

$$\sum_{j=1}^{k} a_j v_j = \mathbf{0}.$$

Technically, the notions of linear independence and linear dependence apply to sets of vectors. Keeping this in mind, can keep you out of trouble sometimes. For example, students often say "the vector $(0, 0, 1) \in \mathbb{R}^3$ is linearly independent of the vectors $(1, 0, 0)$ and $(0, 1, 0)$," which is a statement that really doesn't make any sense. This is especially the case if such a statement comes in the form "the vector $(0, 0, 1)$ is linearly independent," which is inevitably the form in which such a monstrosity presents itself. What such a student really means is that $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is linearly independent, i.e., is a linearly independent **set**. To be proper, if you want to talk about linear independence or linear dependence you need to talk about a **set** of vectors. A single vector can **never** be linearly independent or linearly dependent. Having said that, there is one harmless abuse of language in which we will indulge. When a finite collection of vectors $\{v_1, v_2, \ldots, v_n\}$ is linearly independent (or linearly dependent), we will sometimes say "the **vectors** $v_1, v_2, \ldots, v_n$ are linearly independent," (or linearly dependent). In such a usage, the set $\{v_1, v_2, \ldots, v_n\}$ can be assumed to be implied.

Exercise 13 *Show that any subset of a linearly independent set is linearly independent.*

2 Subspaces, Bases, and Dimension

Here are the formal definitions:

Definition 4 *A subset Σ of a vector space V is a* **subspace** *if Σ is itself a vector space with respect to the same operations making V a vector space.*

Exercise 14 *(When is a subset a subspace?) Show that if $\Sigma \subset V$ with*

(a) *V is a vector space,*

(b) *Σ is closed under addition, and*

(c) *Σ is closed under scaling,*

then Σ satisfies all the other conditions to be a vector space with respect to the same operations used in V.

Definition 1 *A subset $\mathcal{B}$ of a vector space V is a* **basis** *for V if every vector $v \in V$ can be written uniquely as*

$$v = \sum_{j=1}^{k} a_j v_j \tag{5}$$

for some distinct vectors $v_1, v_2, \ldots, v_k \in \mathcal{B}$ and scalars $a_1, a_2, \ldots, a_k$.

There are two really important things to recognize about the expression (5). The first one is that the sum involves only finitely many *distinct* vectors in $\mathcal{B}$. The second is that by uniqueness, we are saying both that the vectors from $\mathcal{B}$ that are used are unique, and the coefficients are uniquely determined. (The condition also requires existence of course.)

Exercise 15 *Show that the zero vector can* **never** *be in a basis. (You showed in Exercise 14 that the zero vector is* **always** *in a subspace.)*

Exercise 16 *Show that if*

(a) *$\mathcal{B}$ is a basis for a vector space V,*

(b) *$v_1, v_2, \dots v_k$ are distinct vectors in $\mathcal{B}$,*

(c) *$w_1, w_2, \dots w_\ell$ are distinct vectors in $\mathcal{B}$, and*

(d) *There exist scalars $a_1, a_2, \dots, a_k, b_1, b_2, \dots, b_\ell$ such that*

$$\sum_{j=1}^{k} a_j v_j = \sum_{j=1}^{\ell} b_j w_j, \tag{6}$$

then $k = \ell$,

$$\{v_1, v_2, \dots, v_k\} = \{w_1, w_2, \dots, w_k\},$$

and, the corresponding coefficients of corresponding basis elements in (6) are the same. Hint: Use induction.

Exercise 17 *Find a basis for the polynomial ring $\mathbb{Q}[x]$ considered as a vector space over $\mathbb{Q}$.*

When there exists a basis $\mathcal{B}$ for a vector space V and $\mathcal{B}$ has finitely many elements, then we say V is **finite dimensional**. When this happens, then every basis for V will have exactly the same number of elements, and that number is called the **dimension**:

Definition 2 *The* **dimension** *of a finite dimensional vector space is the number of elements in a basis.*

If the vector space V contains a nonzero vector but does not have a finite basis, then we say V is **infinite dimensional**. There is one other possibility: If $V = \{\mathbf{0}\}$, then we say V is **zero dimensional**.

Exercise 18 *Let* $\mathbf{e}_j \in \mathbb{R}^n$ *denote the vector with* 1 *(the multiplicative identity in* $\mathbb{R}$*) in the* j*-th entry and zeros in all other entries. Show that* $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ *is a basis for* $\mathbb{R}^n$*. This is called the* **standard unit basis***. When* $n = 3$*, one often encounters* $\mathbf{e}_1 = \hat{\imath}$*,* $\mathbf{e}_2 = \hat{\jmath}$ *and* $\mathbf{e}_3 = \hat{k}$ *or some similar notation for the standard unit basis vectors.*

Exercise 19 *Show that a set* $S \subset V$ *where* V *is a vector space is a basis if and only if* S **is a linearly independent spanning set***. Thus, this is an alternative definition of the term basis.*

3 Real Transformations

A good first start at learning linear algebra is to think about linear functions $L : \mathbb{R}^n \to \mathbb{R}^m$ where the domain and co-domain are real Euclidean vector spaces.

The linear transformations $L : \mathbb{R}^1 \to \mathbb{R}^m$ are easy to understand. Such mappings are determined by a single image vector $L(1) \in \mathbb{R}^m$. In fact,

$$L(t) = tL(1) \qquad \text{for every } t \in \mathbb{R}.$$

Of course, a little more can be said. If $L(1) = 0$, then L is called the **null transformation**. And this particular transformation is always one to consider no matter what class of linear functions you are thinking about. That is, $L : V \to W$ by $L(v) = 0_W$ for all $v \in V$ is always linear. This is a reasonable place to mention a marginally related fact which might be called the **first general fact about linear functions**:

Given $L : V \to W$*, linear, it is always true that* $L(0_V) = 0_W$*.*

Here are some words to describe linear transformations $L : \mathbb{R}^1 \to \mathbb{R}^m$ under various conditions determining subclasses:

1. $m = 1$ so that the domain and co-domain of L are both $\mathbb{R}^1$.
   (a) If $-1 < L(1) < 1$, i.e., $|L(1)| < 1$, then L is called a **contraction**.
   (b) If $L(1) < 0$, then L is **orientation reversing**.
   (c) If $L(1) > 0$, then L is **orientation preserving**.
   (d) The particular map $L : \mathbb{R}^1 \to \mathbb{R}^1$ by $L(t) = -t$ is called a **change of orientation**.
2. If $m > 1$, and L is not the null transformation, then L is a parameterization of the line $\ell = \{tL(1) : t \in \mathbb{R}\} \subset \mathbb{R}^m$, and this is a parameterization with **speed** $|L(1)|$.

Drawing the pictures for these kinds of mappings is important, but I'll leave it to you.

The next step might be to consider $L : \mathbb{R}^2 \to \mathbb{R}^1$ (linear). At this point it is perhaps worthwhile for the purpose of organization to bring in the **second and third general facts about linear functions**:

Given $L : V \to W$, linear, it is always true that

$$\ker(L) = \{v \in V : L(v) = 0_W\} \qquad \textit{is a subspace of } V.$$

Given $L : V \to W$, linear, it is always true that

$$\mathrm{Im}(L) = \{L(v) \in W : v \in V\} \qquad \textit{is a subspace of } W.$$

The subspace $\ker(L)$ is called the **kernel** or **null space** of L. The subspace $\mathrm{Im}(L)$ is called the **image**.

These general facts can be quite helpful for us in considering $L : \mathbb{R}^2 \to \mathbb{R}^1$ because a Euclidean space only has so many kinds of subspaces. In particular, in view of these facts there are three obvious cases to consider

(a) $\ker(L) = \{\mathbf{0}\}$.

(b) $\ker(L)$ is a line $\mathrm{span}\{\mathbf{v}\} = \{t\mathbf{v} : t \in \mathbb{R}\}$ for some nonzero vector $\mathbf{v} \in \mathbb{R}^2$.

(c) $\ker(L) = \mathbb{R}^2$.

In the case of the last one, we have the null transformation of course.

The first possibility, case (a), is somewhat interesting. In this case, there are very few vectors $\mathbf{v} \in \mathbb{R}^2$ for which $L(\mathbf{v}) = 0$. In fact, there is only one such vector. When we talk about linear transformations on Euclidean spaces, there are some common, and pretty helpful, conventions. One of those is that we usually write our vectors as column vectors. For example, instead of $\mathbf{e}_1 = (1, 0) \in \mathbb{R}^2$ we will write

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{7}$$

This doesn't work so well in regular text, but it's rather nice to look at in a display like (7). There is also a more important reason which, if you don't know it already, you will know soon. In any case, we can usually get away with writing vectors in rows when we write them in regular text and in columns when we write/type them in displays without much comment. If we want to emphasize one orientation or another, then we can use a

transpose. That is, the vector $\mathbf{e}_1$ appearing in (7) could have been described in the line before (7) as $\mathbf{e}_1 = (1, 0)^T$. A second convention is that we often omit the parentheses of evaluation. Thus, the vector $L(\mathbf{v})$ appearing near the beginning of this paragraph might be expressed as $L\mathbf{v}$. In particular, we know that if $\ker(L) = \{0\}$, then

$$L\begin{pmatrix} 1 \\ 0 \end{pmatrix} = a \neq 0 \qquad \text{and} \qquad L\begin{pmatrix} 0 \\ 1 \end{pmatrix} = b \neq 0.$$

This means that

$$-\frac{b}{a}L\begin{pmatrix} 1 \\ 0 \end{pmatrix} + L\begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\frac{b}{a}a + b = 0. \tag{8}$$

But using linearity we can write the left side of this relation as

$$L\left(-\frac{b}{a}\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = L\begin{pmatrix} -b/a \\ 1 \end{pmatrix}.$$

We conclude that

$$\begin{pmatrix} -b/a \\ 1 \end{pmatrix} \in \ker(L)\backslash\{\mathbf{0}\}$$

which contradicts the fact that $\ker(L) = \{\mathbf{0}\}$. The conclusion of this little calculation is that **we can't have** $\ker(L) = \{\mathbf{0}\}$ for $L : \mathbb{R}^2 \to \mathbb{R}^1$ linear.

We can arrive at this same conclusion another way too. For linear functions $L : \mathbb{R}^n \to \mathbb{R}^m$ there is a theorem saying, roughly, that all the dimensions in the domain have to be accounted for. More precisely,

$$n = \dim\ker(L) + \dim\operatorname{Im}(L).$$

You have n dimensions in the domain and either those dimensions have to be *collapsed,* as represented by the dimensions in $\ker(L)$ or they have to be transformed into dimensions in the image. In our case $L : \mathbb{R}^2 \to \mathbb{R}^1$, we have $n = 2$, and the image is a subspace of $\mathbb{R}^1$ so that

$$\dim\operatorname{Im}(L) \leq 1.$$

Thus, if you try to make the assumption $\dim\ker(L) = 0$, then the dimensions just don't add up. This theorem is so important, I'll state it again: If $L : V \to W$ is linear and V is a finite dimensional space, then

$$\dim\operatorname{Dom}(L) = \dim\ker(L) + \dim\operatorname{Im}(L). \tag{9}$$

3.1 A Proof

It is not the objective of these notes (or of this course) to emphasize proofs, especially when the result is intuitively plausible. However, sometimes stated results, after contemplation, do not seem so obvious and can spark curiosity. In such cases some comments about the proof seem, at least, nice to think about. This is usually true with results that are very important. The dimension theorem in the previous section (which is also sometimes called the Fundamental Theorem of Linear Transformations) is important, and we could go over the proof. On the other hand, the result is intuitively compelling and, with a little effort, easy to remember. Also, it has just been stated so I think it is best to let students reading these notes ruminate over it for a while. Of course, if you're really interested, you can look up the proof—or, even better, try to prove it yourself. But I'm going to present a proof of a different result here.

I hope a result, which was not stated explicitly, from the section before last has been gnawing at your curiosity. The main result justifying the definition of **dimension** and most of what we believe about the important concept of **basis** is the following:

Theorem 1 *If $w_1, w_2, \ldots, w_{n+1}$ are $n+1$ vectors lying in the span of a set $\{v_1, v_2, \ldots, v_n\}$ of n vectors, then*

$$\{w_1, w_2, \ldots, w_{n+1}\} \qquad \textit{is a linearly } \textbf{dependent} \textit{ set.}$$

Proof: We will give a kind of inductive proof. It is also a proof by contradiction.

Let us assume we have collections $\{w_1, w_2, \ldots, w_{n+1}\}$ and $\{v_1, v_2, \ldots, v_n\}$ as in the hypotheses of the theorem, but that

$$\{w_1, w_2, \ldots, w_{n+1}\} \qquad \text{is a linearly } \textbf{independent} \text{ set.} \tag{10}$$

Our basic claim about this assumption is that if $n > 1$, then it is possible to find (some) n vectors which are linearly independent within the span of (some) $n-1$ vectors. This we state as a technical lemma:

Lemma 1 *If $n > 1$ and*

$$\{w_1, w_2, \ldots, w_{n+1}\} \subset \text{span}\{v_1, v_2, \ldots, v_n\}$$

is a linearly independent set, then there exist vectors $\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n$ and vectors $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_{n-1}$ for which

$$\{\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n\} \subset \text{span}\{\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_{n-1}\}$$

with $\{\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n\}$ linearly independent.

Proof of the lemma: We know each w_i for $i = 1, 2, \ldots, w_{n+1}$ is in the span of $v_1, v_2, \ldots, v_n$. That is, there are coefficients a_{ij} with $1 \leq i \leq n+1$ and $1 \leq j \leq n$ such that

$$w_i = \sum_{j=1}^{n} a_{ij} v_j \qquad \text{for } i = 1, 2, \ldots, n+1.$$

By reordering if necessary, we can assume the last coefficient $a_{n+1,n}$ in the linear combination giving w_{n+1} is nonzero. (If all the last coefficients were zero, then we could just throw away v_n and get $\{w_1, w_2, \ldots, w_n\} \subset \text{span}\{v_1, v_2, \ldots, v_{n-1}\}$, and we would be done because

any subset of a linearly independent set is linearly independent

as per Exercise 13. Therefore, we can write

$$w_{n+1} = \sum_{j=1}^{n-1} a_{n+1,j} v_j + a_{n+1,n} v_n,$$

and

$$v_n = \frac{1}{a_{n+1,n}} \left(w_{n+1} - \sum_{j=1}^{n-1} a_{n+1,j} v_j \right).$$

Using this we can write

$$w_1 = \sum_{j=1}^{n-1} a_{1j} v_j + \frac{a_{1n}}{a_{n+1,n}} w_{n+1} - \frac{a_{nj}}{a_{n+1,n}} \sum_{j=1}^{n-1} a_{n+1,j} v_j$$

and similarly for $w_2, w_3, \ldots, w_n$. As a consequence the vectors

$$w_1 - \frac{a_{1n}}{a_{n+1,n}} w_{n+1}, w_2 - \frac{a_{2n}}{a_{n+1,n}} w_{n+1}, \ldots, w_n - \frac{a_{nn}}{a_{n+1,n}} w_{n+1}$$

constitute n vectors lying in the span of $\{v_1, v_2, \ldots, v_{n-1}\}$. It remains to show these vectors are linearly independent. If

$$\sum_{j=1}^{n} \alpha_j \left(w_j - \frac{a_{jn}}{a_{n+1,n}} w_{n+1} \right) = \mathbf{0},$$

then

$$\sum_{j=1}^{n} \alpha_j w_j - \left(\sum_{j=1}^{n} \frac{\alpha_j a_{jn}}{a_{n+1,n}} \right) w_{n+1} = \mathbf{0}.$$

By the assumed linear independence of $\{w_1, w_2, \ldots, w_{n+1}\}$, this implies

$$\alpha_j = 0 \qquad \text{for } j = 1, 2, \ldots, n.$$

This completes the proof of the lemma.

Returning to the main proof and applying the lemma repeatedly, we obtain finally two vectors $\tilde{w}_1$ and $\tilde{w}_2$ with

$$\{\tilde{w}_1, \tilde{w}_2\} \quad \text{linearly independent} \tag{11}$$

and both $\tilde{w}_1$ and $\tilde{w}_2$ in the span of a single vector v. This means that for some scalars a and b, we have

$$\tilde{w}_1 = av \qquad \text{and} \qquad \tilde{w}_2 = bv.$$

It follows that $b\tilde{w}_1 - a\tilde{w}_2 = \mathbf{0}$. According to (11) we conclude $a = b = 0$. But this means $\tilde{w}_1 = \tilde{w}_2 = \mathbf{0}$, and the set $\{\tilde{w}_1, \tilde{w}_2\}$ is very far from linearly independent[1] contradicting what we have shown in (11). $\square$

3.2 Considering Transformations Without Coordinates

Of course, there still remains the middle case for $L : \mathbb{R}^2 \to \mathbb{R}^1$ in which $\dim \ker(L) = \dim \operatorname{Im}(L) = 1$. Our suspicions about this case should include the following:

1. The nonzero vector $\mathbf{v}$ for which $\ker(L) = \operatorname{span}\{\mathbf{v}\}$ should be important; maybe this vector, more or less, determines L, and...

2. Somehow, this case should be, more or less similar to $L_0 : \mathbb{R}^1 \to \mathbb{R}^1$ when $L_0(1) \neq \mathbf{0}$.

One thing that becomes evident, after you think about it, is that the vector $\mathbf{v}$, while important, cannot (and is not going to) play the role the vector 1 did for L_0. This is because $L\mathbf{v} = \mathbf{0}$. The important thing $\mathbf{v}$ is telling us is not about interesting scaling/1D linear mapping. The vector $\mathbf{v}$ is telling us (all) the vectors that get collapsed by L.

We need another vector to find the interesting scaling. We can choose **any** vector

$$\mathbf{w} \in \mathbb{R}^2 \backslash \langle \mathbf{v} \rangle.$$

If we think in terms of coordinates, there is an obvious choice for $\mathbf{w}$, namely, a vector orthogonal to $\mathbf{v}$. To be very specific, if we take $\mathbf{v} = (v_1, v_2)^T$, then we can take the counterclockwise rotation $\mathbf{v}^\perp = (-v_2, v_1)^T$ for the mapping vector. At this point, the crucial observation is the following:

[1] See, for example, Exercise 15.

Every vector $\mathbf{x} \in \mathbb{R}^2$ can be written uniquely as

$$\mathbf{x} = a\mathbf{w} + b\mathbf{v} \tag{12}$$

for some $a, b \in \mathbb{R}$, and

$$L\mathbf{x} = aL\mathbf{w}. \tag{13}$$

Hopefully, the similarity between the relation (13), namely, $L(a\mathbf{w} + b\mathbf{v}) = aL\mathbf{w}$ and $L_0(t) = tL(1)$ should be noticeable. According to (13) we can see that indeed, $\mathbb{R}^2$ is partitioned into lines parallel to $\mathbf{v}$ on which L is constant, and the behavior along the line through $\mathbf{w}$ is very similar to that of $L_0 : \mathbb{R}^1 \to \mathbb{R}^1$. The only difference is the question of **coordinates**, and we return to consider this question carefully later. For the moment, we consider the verification of (12) which leads us to somewhat different considerations.

3.3 Systems of Linear Equations

As we consider the assertion above concerning (12), we wish to show two things:

1. Given $\mathbf{x} \in \mathbb{R}^2$, there exist real numbers a and b for which $\mathbf{x} = a\mathbf{w} + b\mathbf{v}$, and
2. Given $\mathbf{x} \in \mathbb{R}^2$, there is precisely one pair $(a, b) \in \mathbb{R}^2$ for which $\mathbf{x} = a\mathbf{w} + b\mathbf{v}$.

That is, we wish to show existence and uniqueness. If we introduce coordinates using the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ for $\mathbb{R}^2$, then we can write (12) as

$$\begin{cases} x_1 = aw_1 + bv_1 \\ x_2 = aw_2 + bv_2. \end{cases} \tag{14}$$

This is an example of a system of two linear equations in two unknowns. Let's make sure we understand the identity of the quantities involved. The unknowns here are a and b. The coordinates of $\mathbf{v} = (v_1, v_2)^T \neq \mathbf{0}$ are considered known, though they will depend on the particular linear transformation $L : \mathbb{R}^2 \to \mathbb{R}$ we start with and the fact that the assumption $\ker(L) = \langle \mathbf{v} \rangle$ is satisfied. The vector $\mathbf{w} = (w_1, w_2)^T \notin \langle \mathbf{v} \rangle$ is also considered known, and if we wish to make a specific choice, we could choose $\mathbf{w} = \mathbf{v}^\perp = (-v_2, v_1)^T$. Finally, the vector $\mathbf{x} = (x_1, x_2)^T$ is considered known (but arbitrary) in this context.

Now, there are always **five different ways to think about a system of linear equations**, and I'm going to go through all of them for this particular system.

1. **The system written out**, namely (14). This is what Strang would call the "row picture," and one can start to see from (14) why we wish to think of the vectors $\mathbf{w}$, $\mathbf{v}$ and $\mathbf{x}$ as row vectors.

2. **The vector equation** is a minor modification of (14) obtained by writing the sides of the equation as vectors:

$$\begin{pmatrix} w_1 a + v_1 b \\ w_2 a + v_2 b \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}. \tag{15}$$

   From this we can see clearly why we want to think of $\mathbf{x}$, at least, as a column vector.

3. **The linear combination problem** is a minor modification of the vector equation (15) obtained by writing the left side as a linear combination:

$$a \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} + b \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{16}$$

   This is what Strang calls the **column picture**. It will also be noted that this particular system of linear equations arose in this particular form: $a\mathbf{w} + b\mathbf{v} = \mathbf{x}$.

4. **The matrix equation** is obtained by interpreting the vector equation (15) in a different way, namely as a matrix multiplied by the **unknown vector**, which we have not yet introduced. The unknown vector, is simply the column vector of unknowns, namely,

$$\begin{pmatrix} a \\ b \end{pmatrix}.$$

   In terms of this vector (15) can be written as

$$\begin{pmatrix} w_1 & v_1 \\ w_2 & v_2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{17}$$

   Notice that in the linear combination/column picture the unknowns, a and b, are viewed as coefficients. The coefficients are unknown. In this matrix equation the matrix

$$M = \begin{pmatrix} w_1 & v_1 \\ w_2 & v_2 \end{pmatrix}$$

   is called the **matrix of coefficients** and its entries are considered (known) coefficients. Note finally that this matrix equation is

$$M \begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{x}.$$

5. **The mapping picture** takes one more step based on the observation that there is a linear map associated with multiplication by the matrix M. This map is $L_2 : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$L_2(\mathbf{z}) = M\mathbf{z}$$

with $\mathbf{z}$ denoting a column vector in $\mathbb{R}^2$. The existence question is thus a question of wether or not the vector $\mathbf{x}$ is in the image of L_2. The uniqueness question becomes: Does $L_2(\mathbf{z}_1) = L_2(\mathbf{z}_2)$ imply $\mathbf{z}_1 = \mathbf{z}_2$? This mapping picture brings the study of systems of linear equations back to the study of linear functions.

It will be noted, that we came upon this system of equations trying to understand a linear function $L : \mathbb{R}^2 \to \mathbb{R}^1$, and the mapping picture tells us that we need to know something about a linear function $L_2 : \mathbb{R}^2 \to \mathbb{R}^2$, which is (at least nominally) more complicated. We will discuss linear mappings from $\mathbb{R}^2$ to $\mathbb{R}^2$ in detail later, but we need to know something about them now.

3.4 Considering Transformations With Coordinates

We can begin with the special case when we take $\mathbf{w} = \mathbf{v}^\perp$. In this case, the original system (14) becomes

$$\begin{cases} -v_2 a + v_1 b = x_1 \\ v_1 a + v_2 b = x_2. \end{cases}$$

It is easy to check, for example by elimination, that our claim holds with

$$a = \frac{v_1 x_2 - v_2 x_1}{v_1^2 + v_2^2} \qquad \text{and} \qquad b = \frac{v_1 x_1 + v_2 x_2}{v_1^2 + v_2^2}.$$

The denominators here are known to be nonzero because $\mathbf{v}$ is a nonzero vector. Similar manipulations can be done in the general case of (14) to obtain

$$a = \frac{v_2 x_1 - v_1 x_2}{w_1 v_2 - w_2 v_1} \qquad \text{and} \qquad b = \frac{w_1 x_2 - w_2 x_1}{w_1 v_2 - w_2 v_1}.$$

This will give us the desired unique solution of our claim (12) as long as the quantity $w_1 v_2 - w_2 v_1$ is nonzero. This quantity will be recognized as the **determinant** of the matrix M appearing in the matrix equation and mapping picture above. Thus, for our immediate purposes, it is enough to establish the following:

Lemma 2 *If* $\mathbf{v}$ *and* $\mathbf{w}$ *are linearly independent vectors in* $\mathbb{R}^2$*, then the determinant of the matrix* M *with* $\mathbf{v}$ *and* $\mathbf{w}$ *in the columns is nonzero.*

Both solutions for a and b given above are forms of **Cramer's rule**.

What we know about $L : \mathbb{R}^2 \to \mathbb{R}^1$

Returning to the consideration of $L : \mathbb{R}^2 \to \mathbb{R}^1$, linear with $\ker(L) = \langle \mathbf{v} \rangle$ for some $\mathbf{v} \neq \mathbf{0}$, we had one way to understand L above (without coordinates):

$$L(\mathbf{x}) = L(a\mathbf{w} + b\mathbf{v}) = aL\mathbf{w}$$

which can be interpreted as a kind of skew projection where $a\mathbf{w}$ is the projection of $\mathbf{x}$ onto $\mathbf{w}$ with respect to the line field determined by $\mathbf{v}$. (You should draw a picture of this.) The equation

$$\mathbf{x} = a\mathbf{w} + b\mathbf{v}$$

is uniquely solvable for a and b with a being the $\mathbf{w}$-coordinate of $\mathbf{x}$ with respect to the basis $\{\mathbf{w}, \mathbf{v}\}$.

In this first way of understanding L, which is pretty good, we have not mentioned the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ for $\mathbb{R}^2$. In fact, it is rather difficult for us to think about the Euclidean spaces without thinking of the standard basis, but we should try because there are many vector spaces (even of dimension two over $\mathbb{R}$) for which the identification of a corresponding basis is far from obvious. One example is $P_1[x] = \{a_1x + a_0 : a_1, a_0 \in \mathbb{R}\}$ the vector space of polynomials of order no more than one over $\mathbb{R}$. You might think $\{x, 1\}$ is a basis corresponding to $\{\mathbf{e}_1, \mathbf{e}_2\}$ for $\mathbb{R}^2$, but there is really no reason to think that. It is true that $\{x, 1\}$ is a basis for $P_1[x]$, but there is no picture suggesting any natural orthogonality.

Exercise 20 *Consider the* **inner product** *on* $P_2[x] = \{a_2x^2 + a_1x + a_0 : a_2, a_1, a_2 \in \mathbb{R}\}$ *given by*

$$\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)\,dx.$$

(a) *Calculate* $\langle x^2, 1 \rangle$. *If you get a nonzero answer, then* x^2 *and* 1 *are not orthogonal/perpendicular.*

(b) *The inner product gives a* **norm** *or distance to the origin by*

$$\|p\| = \sqrt{\langle p, p \rangle}.$$

Calculate $\|1\|$. *If you get something other than* 1 *(the scalar), then* 1 *(the polynomial) is not a unit vector.*

What we know about a linear function $L : \mathbb{R}^2 \to \mathbb{R}^1$ without coordinates may seem a bit vague. After all, the vector $\mathbf{v}$ is only determined up to a nonzero constant scaling, and the vector $\mathbf{w}$ can be almost any vector at all. For a general mapping of a two-dimensional vector space into a one-dimensional vector space, this is about all one can say. If the domain is Euclidean space $\mathbb{R}^2$ and the co-domain is Euclidean space $\mathbb{R}^1$, we usually say more.

A second way to understand $L : \mathbb{R}^2 \to \mathbb{R}^1$

With (standard) coordinates we can write $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$ and

$$L\mathbf{x} = x_1 L\mathbf{e}_1 + x_2 L\mathbf{e}_2.$$

This last expression may be recognized as having the form of a **Euclidean inner product** or dot product:

$$L(\mathbf{x}) = \mathbf{x} \cdot \mathbf{w} \qquad \text{with } \mathbf{w} \text{ a new vector } (L\mathbf{e}_1, L\mathbf{e}_2)^T.$$

Another alternative is to consider the matrix multiplication

$$(L\mathbf{e}_1, L\mathbf{e}_2)\mathbf{x} = (L\mathbf{e}_1, L\mathbf{e}_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{18}$$

where $\mathbf{w}^T = (L\mathbf{e}_1, L\mathbf{e}_2)$ is a row vector called the **mapping vector** or **the matrix of** L with respect to the standard basis. This is a good point to try to emphasize something:

> *Technically, whenever we write down the matrix of a linear transformation, we* **must** *have a specific basis in mind for both the domain and the co-domain. Without these specified, there is no matrix associated with a linear transformation. We are simply so used to using the standard Euclidean bases whenever we see $L : \mathbb{R}^n \to \mathbb{R}^m$, that it is difficult for us to wrap our mind around this, technically correct, observation.*

In (18) the result of the matrix multiplication is, technically, a 1×1 matrix. This matrix, however, is naturally identified with the single scalar entry it contains, and one often writes

$$L\mathbf{x} = (L\mathbf{e}_1, L\mathbf{e}_2)\mathbf{x}. \tag{19}$$

This is a minor abuse of notation because the quantity on the left $L\mathbf{x} \in \mathbb{R}$ is a scalar while the object on the right is a 1×1 matrix in $M_1(\mathbb{R})$ the set of all square one-by-one matrices with real entries. But the important point is the following: In order to write (19) we have used the bases

$$\mathcal{D} = \{\mathbf{e}_1, \mathbf{e}_2\} \quad \text{for the domain } \mathbb{R}^2 \qquad \text{and} \qquad C = \{1\} \quad \text{for the co-domain } \mathbb{R}.$$

Let's go a little further with this. Recall our two suspicions about this case should include the following:

1. The nonzero vector $\mathbf{v}$ for which $\ker(L) = \operatorname{span}\{\mathbf{v}\}$ should be important; maybe this vector, more or less, determines L, and...

2. Somehow, this case should be, more or less similar to $L_0 : \mathbb{R}^1 \to \mathbb{R}^1$ when $L_0(1) \neq \mathbf{0}$.

Evidently, in coordinates, our function is determined by the mapping vector $(L\mathbf{e}_1, L\mathbf{e}_2)$. What is the relation between this vector and $\mathbf{v}$? What is the relation between this vector and scaling? Recall that $\mathbf{v}$ was, in general, a nonzero vector spanning the kernel. Note that $(L\mathbf{e}_1, L\mathbf{e}_2)$ is a nonzero vector. (Why?). First of all, we claim $(-L\mathbf{e}_2, L\mathbf{e}_1)^T$ spans $\ker(L)$. This means that no matter which spanning vector $\mathbf{v}$ we choose, we will get a multiple of $(-L\mathbf{e}_2, L\mathbf{e}_1)^T$. That is, up to scaling, the mapping vector $(L\mathbf{e}_1, L\mathbf{e}_2)$ tells us $\mathbf{v}$, and up to scaling $\mathbf{v}$ gives the mapping vector. Let us rewrite (19) as

$$L\mathbf{x} = \sqrt{(L\mathbf{e}_1)^2 + (L\mathbf{e}_2)^2}\, \frac{(L\mathbf{e}_1, L\mathbf{e}_2)}{\sqrt{(L\mathbf{e}_1)^2 + (L\mathbf{e}_2)^2}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{20}$$

This suggests the choice

$$\left\langle \frac{(L\mathbf{e}_2, -L\mathbf{e}_1)^T}{\sqrt{(L\mathbf{e}_1)^2 + (L\mathbf{e}_2)^2}} \right\rangle = \ker(L). \tag{21}$$

In fact, taking

$$\mathbf{w} = \frac{(L\mathbf{e}_1, L\mathbf{e}_2)^T}{\sqrt{(L\mathbf{e}_1)^2 + (L\mathbf{e}_2)^2}} \qquad \text{and} \qquad \mathbf{v} = \frac{(L\mathbf{e}_2, -L\mathbf{e}_1)^T}{\sqrt{(L\mathbf{e}_1)^2 + (L\mathbf{e}_2)^2}},$$

(20) becomes

$$L\mathbf{x} = \alpha \operatorname{comp}_{\mathbf{w}}(\mathbf{x}) \qquad \text{where } \alpha = \sqrt{(L\mathbf{e}_1)^2 + (L\mathbf{e}_2)^2} > 0$$

and $\operatorname{comp}_{\mathbf{w}} \mathbf{x}$ is the component of the Euclidean (orthogonal) projection of $\mathbf{x}$ onto $\mathbf{w} = \mathbf{v}^\perp$.

This gives a rather complete picture of $L : \mathbb{R}^2 \to \mathbb{R}^1$: There are two cases, either L is the null mapping or $\dim \ker(L) = 1$. In the latter case, there is a unit vector $\mathbf{v}$ spanning $\ker(L)$. Taking a unit vector $\mathbf{w}$ orthogonal to $\mathbf{v}$, the function L is given by a scaling of the component of the orthogonal projection onto $\mathbf{w}$. With a correct choice of orientation between $\mathbf{v}$ and $\mathbf{w}$, the mapping is given by $L\mathbf{x} = \alpha\mathbf{x} \cdot \mathbf{w}$ where $\alpha = |(L\mathbf{e}_1, L\mathbf{e}_2)|$ is a positive scaling factor.

Exercise 21 *What happens if* $\mathbf{w}$ *and* $\mathbf{v}$ *do not have the correct orientation, i.e.,* $\mathbf{w} = -\mathbf{v}^\perp$*?*

3.5 Components and Projections

We have already seen that given a linearly independent set/basis $\{\mathbf{v}, \mathbf{w}\} \subset \mathbb{R}^2$, there is a unique solution $(a, b)^T$ of the equation

$$\mathbf{x} = a\mathbf{v} + b\mathbf{w} \qquad \text{for any vector } \mathbf{x} \in \mathbb{R}^2.$$

The numbers a and b are sometimes called the **components** of $\mathbf{x}$ along $\mathbf{v}$ and $\mathbf{w}$ respectively with respect to the basis $\{\mathbf{v}, \mathbf{w}\}$. When the vectors $\mathbf{v}$ and $\mathbf{w}$ are orthonormal, i.e., orthogonal to each other and both having unit length, then a special notation is used.

Exercise 22 *Show that when* $\{\mathbf{v}, \mathbf{w}\}$ *is an orthonormal basis, then any vector* $\mathbf{x} \in \mathbb{R}^2$ *satisfies*

$$\mathbf{x} = (\mathbf{x} \cdot \mathbf{v})\, \mathbf{v} + (\mathbf{x} \cdot \mathbf{w})\, \mathbf{w} \tag{22}$$

so that, for example, the component along $\mathbf{v}$ *is independent of* $\mathbf{w}$.

In view of the representation (22) it makes sense to denote the dot product $\mathbf{x} \cdot \mathbf{v}$ by

$$\mathrm{comp}_{\mathbf{v}}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{v}$$

whenever $\mathbf{v}$ is a given unit vector (even without any discussion of an orthonormal basis). Similarly, the vector $\mathrm{comp}_{\mathbf{v}}(\mathbf{x})\, \mathbf{v} = (\mathbf{x} \cdot \mathbf{v})\, \mathbf{v}$ is called the **projection** or **orthogonal projection** of $\mathbf{x}$ onto $\mathbf{v}$. It should be emphasized that in order to define a component with respect to an arbitrary basis, knowledge of the entire basis $\{\mathbf{v}, \mathbf{w}\}$ is required. Whenever we use the special notation $\mathrm{comp}_{\mathbf{v}}(\mathbf{x})$ or $\mathrm{proj}_{\mathbf{v}}(\mathbf{x})$, then we are considering the orthogonal component and orthogonal projection determined by the dot product.

The notation for components and projections just introduced is not universal, and I am not sure it is the most ideal. I will try, however, to be consistent with the notation and terminology as I've introduced it here.

Other coordinates and other matrices

When we discussed linear functions $L_1 : \mathbb{R} \to \mathbb{R}$ we used simply the basis $\{1\}$ for $\mathbb{R}$. We could have used, the basis $\{w\}$, however, with w an arbitrary nonzero vector. Taking this basis for the domain and the co-domain, we write, instead of $L_1 x = x L_1(1)$,

$$L_1 x = L_1 \left(\frac{x}{w}\, w \right) = \frac{x}{w}\, L_1 w = \frac{L_1 w}{w}\, \frac{x}{w}\, w.$$

Thus, the matrix of L_1 with respect to $\{w\}$ (used as the basis for the domain and the co-domain) is

$$M_1 = \left(\frac{L_1 w}{w} \right).$$

Exercise 23 *Assuming* $L_1 : \mathbb{R}^1 \to \mathbb{R}^1$ *is linear and nonzero, find the matrix of* L_1 *with respect to the choices*

$$\mathcal{D} = \{w\} \qquad \text{and} \qquad C = \{Lw\}$$

for bases of the domain and co-domain respectively.

Exercise 24 *What is the matrix of* **any** *linear transformation* $L_1 : \mathbb{R}^1 \to \mathbb{R}^1$ *with respect to the basis* $\{1\}$ *(for both the domain and co-domain*[2]*)? In what way was the restriction that* L_1 *was* **nonzero** *used in Exercise 23.*

Now, let's take the case $L : \mathbb{R}^2 \to \mathbb{R}^1$ with $\ker(L) = \langle \mathbf{v} \rangle$ and $\mathbf{v} \neq \mathbf{0}$. If $\mathbf{w} \in \mathbb{R}^2 \backslash \langle \mathbf{v} \rangle$, then the matrix/mapping vector of L with respect to $\mathcal{D} = \{\mathbf{w}, \mathbf{v}\}$ and $C = \{1\}$ is

$$M = (L\mathbf{w}, 0).$$

3.6 More on Systems of Equations: Gaussian Elimination

Using the row picture, that is just the system of linear equations as it is given, to determine the answers to questions of existence and uniqueness of solutions is called **elimination**. Gauss came up with a systematic way to determine if a solution exists and, if so, how many solutions there are along with an explicit formula for those solutions. Gauss' method is relatively straightforward, but it gives some extremely useful information.

Gauss' method applied to the system in (14) looks (roughly) like this: We start with

$$\begin{cases} w_1 a + v_1 b &= x_1 \\ w_2 a + v_2 b &= x_2. \end{cases}$$

Multiplying the first equation by w_2/w_1 and subtracting from the second equation, we get

$$\begin{cases} w_1 a + v_1 b &= x_1 \\ (v_2 - w_2 v_1/w_1) b &= x_2 - w_2 x_1/w_1. \end{cases} \tag{23}$$

In principle, there might have been a problem here if we had $w_1 = 0$. In that case, there should be a step in which one switches the two equations, or exchanges the rows. Presumably, some equation has a nonzero coefficient in front of the first unknown, or else that unknown is not really part of this system of equations. If $w_1 \neq 0$, we are now in a position to "read off" the solutions, or at least know what happens.

Notice that the remaining (possibly nonzero coefficient) in the second equation is

$$\frac{w_1 v_2 - w_2 v_1}{w_1} = \frac{\det M}{w_1}$$

[2]Note: When the co-domain and the domain are the same vector space, then it is often convenient (and to some degree a convention) to take the same basis for both. When the domain and co-domain are different vector spaces this is not possible.

where $\det M = w_1v_2 - w_2v_1$ is the determinant of the coefficient matrix from the matrix equation (17). If this value is zero, then there are two possibilities. The first is that $x_2 - w_2x_1/w_1$ is nonzero. In this case, the system has no solution. The second possibility is that $x_2 - w_2x_1/w_1$ is zero. In this case, under the assumption $w_1 \neq 0$, there are infinitely many solutions having the form

$$\left\{ \begin{pmatrix} a \\ b \end{pmatrix} : a = \frac{1}{w_1}(x_1 - v_1 b) \text{ with } b \text{ arbitrary} \right\}.$$

In fact, neither of these possibilities happens. To see this, first observe that $\{\mathbf{w}, \mathbf{v}\}$ spans a subspace of $\mathbb{R}^2$ which has dimension at least two, and hence dimension exactly two. Thus, there is at least one solution of the system (14). This means, according to the column picture that the linaer transformation $L_2 : \mathbb{R}^2 \to \mathbb{R}^2$ by $L\mathbf{z} = M\mathbf{z}$ is onto. That is, $\dim \operatorname{Im} L_2 = 2$. Therefore, by the dimension theorem, $\dim \ker L_2 = 0$ and $\ker L_2 = \{\mathbf{0}\}$. This means L_2 is one-to-one and onto. (If we had $\tilde{\mathbf{z}} \neq \mathbf{z} = (a, b)^T$ with $L_2\tilde{\mathbf{z}} = L_2\mathbf{z} = \mathbf{x}$, then we would have $L_2(\tilde{\mathbf{z}} - \mathbf{z}) = \mathbf{0}$ with $\tilde{\mathbf{z}} - \mathbf{z} \neq \mathbf{0}$, that is, we would have obtained a nonzero element of $\ker(L_2)$ and a contradiction.) Finally, then, you can check the following:

1. The inverse function $L_2^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ is **linear**, and
2. The matrix associated to L_2^{-1} with respect to the standard basis is a $2x2$ matrix M^{-1} for which $MM^{-1} = I$ is the 2×2 identity matrix.
3. The determinant of the product $\det(MM^{-1})$ is the product of the determinants:

$$\det(M)\det(M^{-1}) = 1.$$

Therefore, $\det M = w_1v_2 - w_2v_1 \neq 0$. Therefore, we read off the solution given above by Cramer's rule.

Exercise 25 *Solving (23) for b first, and then substituting your expression for b in the first equation, show the Cramer's rule solution is obtained.*

Exercise 26 *Carry out the details of Gaussian elimination for (14) under the assumption* $w_1 = 0$. *Notice that in this case,* $v_2 \neq 0$ *and* $w_1 \neq 0$. *(Why?)*

Systems of linear equations are usually presented in the following general form:

$$\left\{ \begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ & \vdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m. \end{array} \right. \tag{24}$$

...here should follow a discussion of general systems of linear equations, Gaussian elimination, and some topics concerning matrices.

The five different ways to look at the system of linear equations (24) are as follows:

1. The row picture (24).

2. The vector equation:

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

   The vector

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

   is called the **inhomogeneity**, especially when $\mathbf{b} \neq \mathbf{0} \in \mathbb{R}^m$.

3. The matrix equation is

$$A\mathbf{x} = \mathbf{b}$$

   where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

   is the **coefficient matrix** which has size $m \times n$,

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

   is the unknown or **solution vector**, and the multiplication $A\mathbf{x}$ is carried out in the usual way with the dot product of the i-th row of A with $\mathbf{x}$ giving the i-th entry of the product vector on the left in the vector equation.

4. The linear combination problem (or column picture) is

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

5. The mapping picture is $L\mathbf{x} = \mathbf{b}$ where $L : \mathbb{R}^n \to \mathbb{R}^m$ by $L(\mathbf{x}) = A\mathbf{x}$.

It should be clear that systems of linear equations go hand in hand with matrices. The elementary row operations of Gaussian elimination obviously keep the variables in $\mathbf{x}$ in order and can be applied only to the coefficients and the inhomogeneity. Putting these relevant coefficients together in the same matrix (with the inhomogeneity as the last column) gives the **augmented matrix**.

Each elementary row operation corresponds to left multiplication by an invertible $m \times m$ matrix. obtained by applying the row operation in question to the $m \times m$ identity matrix. Thus, when $m = 2$,

1. Interchanging/switching the two rows:

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

   (This matrix is its own inverse.)

2. Scale a row by a nonzero constant:

$$\begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix}$$

   with inverses

$$\begin{pmatrix} 1/\alpha & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} 1 & 0 \\ 0 & 1/\alpha \end{pmatrix}.$$

3. Replace the second row by the second row plus a scaling of the first:

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

   This has inverse

$$\begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix}.$$

Exercise 27 *Write down the elementary row operation matrices for general m.*

The **column space** of a matrix with m rows is the subspace of $\mathbb{R}^m$ spanned by the columns. If such a matrix is used to define a linear function, then the image of the linear function is the column space. The dimension of the column space of a matrix is called the (column) rank of the matrix.

The **row space** of a matrix with k columns is the subspace of $\mathbb{R}^k$ spanned by the rows of the matrix, or equivalently the transposes of the row vectors. The dimension of the row space of a matrix is called the (row) rank of the matrix.

The following facts about the rank of a matrix and Gaussian elimination are basic:

1. Neither the row rank nor the column rank changes under an elementary row operation. That is, if M is a matrix, and R is an elementary row operation matrix, then the (row and column) rank(s) of M and the (row and column rank(s) of RM are the same.

2. After Gaussian elimination, obtained by a sequence of elementary row operations applied to M,
$$E = R_\ell R_{\ell-1} \cdots R_2 R_1 M$$
is matrix reduced to **row echelon** form (where each successive row has more initial zeros than the previous row).

3. It is clear that a row echelon form matrix E the same row rank as column rank.

4. The column rank and the row rank of a matrix are equal. This common number is called the **rank** of the matrix.

5. If a matrix $M = (\mu_{ij})$ has been reduced to row echelon form E using elementary row operations, then each **nonzero initial element** of a row in E determines a distinct column index. Each such index is called a **pivot** index with the corresponding column called a **pivot column**.

6. The collection of pivot columns are clearly a basis for the column space of E.

7. Let Γ be the collection of the pivot indices in the reduced matrix E. Then the set of columns
$$\left\{ \quad \begin{pmatrix} \mu_{1j} \\ \mu_{2j} \\ \vdots \\ \mu_{mj} \end{pmatrix} \quad : \quad j \in \Gamma \quad \right\}$$
of M with pivot indices from Γ is a basis for the column space of M.

This process tells you just about everything about a system of linear equations. In particular, there exists a solution if the rank of the augmented matrix is the same as the rank of A. Associated with a system of linear equations is a homogeneous system of linear equations $A\mathbf{x} = \mathbf{0}$. This **associated homogeneous system** always has a sulution $\mathbf{x} = \mathbf{0}$, and the set of solutions Σ_0 is a subspace. (It is the kernel of L.)

Exercise 28 *Show that if $L : V \to W$ is a linear function and $L(x) = y$ has one solution σ_0, then the set of solutions*

$$\Sigma = \{x \in V : L(x) = y\}$$

has the form

$$\Sigma = \{\sigma_0 + \sigma : \sigma \in \ker(L)\}.$$

The case of a system of linear equations when $n = m$ and A is a square matrix is special. In this case, L is one-to-one and onto (a bijection) if and only if the determinant of A is nonzero. This follows from the **product formula for determinants**:

$$\det(R_\ell R_{\ell-1} \cdots R_2 R_1 A) = \det(R_\ell)\det(R_{\ell-1}) \cdots \det(R_2)\det(R_1)\det(A).$$

(And if L is a bijection, then A row reduces to the identity matrix.)

3.7 The matrix of a linear function

Whenever we are talking about a matrix, **coordinates** are under consideration, that is, a specific **choice of basis** has been made. Very often this important choice is not even mentioned. In such cases, one assumes the coordinates are with respect to the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k\}$ for (vectors in) $\mathbb{R}^k$. But there are other possibilities. Given a linear function $L : \mathbb{R}^n \to \mathbb{R}^m$, like the one associated with a coefficient matrix in a system of linear equations—but ignoring the matrix, we can take any basis $\mathcal{D} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ for $\mathbb{R}^n$ and any basis $C = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ for $\mathbb{R}^m$ and get a matrix for L as follows: Recall that the function L is determined by specifying

$$L(\mathbf{v}_j) \qquad \text{for } j = 1, 2, \ldots, n.$$

This is because

$$L\left(\sum_{j=1}^{n} x_j \mathbf{v}_j\right) = \sum_{j=1}^{n} x_j L\mathbf{v}_j.$$

Each of the vectors $L(\mathbf{v}_j)$ can be expressed in terms of the basis C as

$$L\mathbf{v}_j = \sum_{i=1}^{m} c_{ij}\mathbf{w}_i. \tag{25}$$

Thus, we can say the matrix of L with respect to the bases $\mathcal{D}$ and C is

$$\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix}$$

Now, the **coordinate expression** for $\mathbf{v}_1$ in terms of $\mathcal{D}$ is

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{e}_1 \in \mathbb{R}^n.$$

That is, when you express $\mathbf{v}_1$ as a linear combination of the vectors in $\mathcal{D} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, then the first coefficient is $x_1 = 1$, and all the others are zero. In order to keep things straight here, we might write

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{\mathcal{D}}$$

for $\mathbf{v}_1$ in coordinates. If the matrix

$$\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix}_{\mathcal{D},C}$$

is multiplied by $\mathbf{e}_1 \in \mathbb{R}^n$ in the usual way, we get the first column

$$\begin{pmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{pmatrix} = \begin{pmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{pmatrix}_C$$

meaning $L\mathbf{v}_1$ is the vector

$$\sum_{i=1}^{m} c_{i1}\mathbf{w}_1$$

in accord with (25).

4 Summary: Pause before the (little) storm

You should now understand everything for a linear function $L : \mathbb{R}^1 \to \mathbb{R}^1$ and just about everything for a linear function $L : \mathbb{R}^2 \to \mathbb{R}^1$. You should be able to "see" exactly how these functions work. Certainly you should be able to "compute" anything having to do with such a function if one is actually given to you explicitly.

Exercise 29 *If $L : \mathbb{R}^1 \to \mathbb{R}^1$ and the scaling factor $L(1) = -1/3$, then what is*

$$\lim_{j\to\infty} L^j(7)$$

where $L^j(t)$ means "compose L on t over and over again j times."

Exercise 30 *If $L : \mathbb{R}^2 \to \mathbb{R}^1$ with $L(1,0) = 3$ and $L(0,1) = -7$, then find* $\ker(L)$.

All this was not very difficult, though some things may have seemed unnecessarily complicated and strangely unfamiliar. There was a reason for that.

In the mean time, you should have reviewed most of what you know/knew about linear algebra including the main facts about vector spaces, systems of linear equations, matrices, and determinants. In particular,

1. A set of vectors S is **linearly independent** if each vector in the span of S has a unique representation as a linear combination of vectors in S.

2. A set of vectors is **linearly dependent** if the zero vector can be written as a linear combination of (some of) those vectors with the coefficients in the linear combination nonzero.

3. A linear independent set which spans a vector space is called a **basis**.

4. If a vector space has a finite set as a basis, then every basis for that vector space has the same number of elements, and that number is called the **dimension** of the vector space. If a vector space V does not have a finite basis, then either $V = \{\mathbf{0}\}$ or V is called infinite dimensional.

5. Given a linear function $L : V \to W$ where V is a finite dimensional vector space

$$\dim V = \dim \ker(L) + \dim \operatorname{Im}(L).$$

   This is called **the dimension theorem**.

6. Given a basis $\{v_1, v_2, \ldots, v_n\}$ for a vector space V and **any** vectors $w_1, w_2, \ldots, w_n$ in a vector space W, a unique linear function $L : V \to W$ is determined by the conditions $L(v_j) = w_j$ for $j = 1, 2, \ldots, n$. In fact, if $x = \sum a_j v_j$, then

$$L(x) = \sum_{j=1}^{n} a_j L(v_j).$$

7. A system of linear equations is constructed using a **coefficient matrix**, an **unknown/solution vector**, and an **inhomogeneity** vector.

8. The **column rank** of the coefficient matrix is the dimension of the space spanned by the columns of the matrix and the **row rank** is the dimension of the space spanned by the rows. These two numbers are the same for every coefficient matrix and their common value is called the **rank**. (The same can be said of every matrix.)

9. The rank of a matrix can be determined by **Gaussian elimination** which involves **elementary row operations** and produces an upper triangular matrix in a special form (called **row echelon form**) which is **row equivalent** to the original matrix (and has the same rank). The rank of a matrix in row echelon form is the number of nonzero rows in the matrix.

10. The columns of the first nonzero entries in the nonzero rows of a matrix in row echelon form are called the **pivot columns**. These columns span the column space. If T is an upper triangular matrix in row echelon form with pivot columns T_j for $j \in \Gamma$ (where $\Gamma \subset \{1, 2, \ldots, n\}$ is some indexing set) and A is any $n \times n$ matrix row equivalent to T, then $\{A_j : j \in \Gamma\}$ is a basis for the column space of A (where A_j denotes the j-th column of A).

11. Every **homogeneous** system has the zero vector as a solution. A non-homogeneous system has a solution if and only if the coefficient matrix and the **augmented matrix**, constructed by appending the inhomogeneity to the coefficient matrix, have the same rank.

12. The set of solution vectors Σ_0 of a homogeneous system of m linear equations in n unknowns is the kernel of a linear function, which is a vector space. This space is called the **solution space**, and it has dimension $n - \text{rank}(A)$ where A is the coefficient matrix. (You should know why this is true.)

13. If a linear system has a solution σ_1, then the set of solutions Σ is given by

$$\Sigma = \{\sigma_1 + \sigma : A\sigma = \mathbf{0}\} = \{\sigma_1 + \sigma : \sigma \in \Sigma_0\}$$

where Σ_0 is the solution space of the **associated homogeneous system**.

14. Systems of linear equations in which the unknown/solution vector and the inhomogeneity are in the same vector space are special. The coefficient matrix in this case is square. A system of n linear equations in n unknowns has a unique solution if and only if $\det A \neq 0$ where A is the (square) coefficient matrix.

15. Determinants are only defined for square matrices. If A and B are square matrices of the same size, then

$$\det(AB) = \det(A)\det(B).$$

This is called **the product formula for determiants**.

In addition, a few things have been emphasized that perhaps you didn't know before. Among these are:

1. A linear function and a matrix are not the same thing. You can have a linear function without a matrix. However, any time you have a matrix, you have a (very specific) linear function.

2. Everything about a system of linear equations or matrices can be understood (more generally) in terms of linear functions.

3. On the other hand, the consideration of coordinates, systems of linear equations, and matrices can be a useful in understanding certain linear functions.

We are now going to consider linear functions $L : \mathbb{R}^2 \to \mathbb{R}^2$. It will not be so easy to "see" what all of these transormations do, but we should be able to compute just about anything.

5 Dot Products and Projections

Our next step is to discuss linear functions $L : \mathbb{R}^2 \to \mathbb{R}^2$. As can be expected, there will be new kinds of linear functions to understand. In the study of linear functions $L : \mathbb{R}^2 \to \mathbb{R}^1$ we were able to relate the dot product/geometry of elements in the domain to the values taken by the function. This kind of analysis will also be important for $L : \mathbb{R}^2 \to \mathbb{R}^2$, and we point out some additional details concerning that geometry here.

You are probably familiar with the notion of **orthogonal projection**. Let's review that before we consider more general projections. Given a nonzero vector $\mathbf{v} \in \mathbb{R}^2$, the projection of $\mathbf{x}$ onto $\langle \mathbf{v} \rangle$ is given by

$$\text{proj}_{\langle \mathbf{v} \rangle}(\mathbf{x}) = \text{projection}_{\mathbf{v}/|\mathbf{v}|}(\mathbf{x}) = \frac{\mathbf{x} \cdot \mathbf{v}}{|\mathbf{v}|} \frac{\mathbf{v}}{|\mathbf{v}|}.$$

Whenever we use this notation, we mean an orthogonal projection. Notice that it only takes one vector $\mathbf{v}$ to determine an orthogonal projection

$$\text{proj}_{\langle \mathbf{v} \rangle} : \mathbb{R}^2 \to \mathbb{R}^2.$$

Also, you can see from the definition that this function is a linear function.

Exercise 31 *Draw a picture illustrating the orthogonal projection function.*

Let us assume, for a moment, that $\mathbf{v}$ is a unit vector. If we have any nonzero vector $\mathbf{v}_0$, we can obtain a unit vector in the same direction by taking $\mathbf{v} = \mathbf{v}_0/|\mathbf{v}_0|$. We also recall the important relation between the norm of a vector in $\mathbb{R}^2$ and the dot product:

$$|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v}.$$

If $|\mathbf{v}| = 1$, then the projection simplifies to

$$\text{proj}_{\mathbf{v}}(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{v})\, \mathbf{v}.$$

The scalar coefficient given by the dot product $\mathbf{x} \cdot \mathbf{v}$ is called the **component** of $\mathbf{x}$ along $\mathbf{v}$. Of course, we again mean the **orthogonal component**. This value is denoted by

$$\text{comp}_{\mathbf{v}}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{v} \qquad (|\mathbf{v}| = 1).$$

Similarly, even when $\mathbf{v} \neq 0$ is not a unit vector, we can talk about the component of $\mathbf{x}$ along the subspace $\langle \mathbf{v} \rangle$.

Exercise 32 *Find a formula for the component of* $\mathbf{x}$ *along the subspace* $\langle \mathbf{v} \rangle$ *when* $\mathbf{v} \neq 0$ *does not have unit length so that*

$$\mathrm{proj}_{\langle \mathbf{v} \rangle}(\mathbf{x}) = \mathrm{comp}_{\langle \mathbf{v} \rangle}(\mathbf{x})\, \mathbf{v}.$$

This same number $\mathrm{comp}_{\langle \mathbf{v} \rangle}(\mathbf{x})$ would be the **coordinate** of $\mathbf{x}$ with respect to the basis $\{\mathbf{v}, \mathbf{w}\}$ if $\mathbf{w}$ is orthogonal to $\mathbf{v}$. Such a basis is called an **orthogonal basis**.

Exercise 33 *Write down a linear system for the coordinates*

$$\left(\mathrm{comp}_{\langle \mathbf{v} \rangle}(\mathbf{x}), \mathrm{comp}_{\langle \mathbf{w} \rangle}(\mathbf{x})\right)^T$$

of $\mathbf{x}$ *with respect to an orthogonal basis* $\{\mathbf{v}, \mathbf{w}\}$.

Solution: Starting with $\mathbf{x} = a\mathbf{v} + b\mathbf{w}$, we can take the dot product with $\mathbf{v}$ and $\mathbf{w}$ to obtain

$$\left\{ \begin{array}{rcl} |\mathbf{v}|^2 a & = & \mathbf{x} \cdot \mathbf{v} \\ |\mathbf{w}|^2 b & = & \mathbf{x} \cdot \mathbf{w}. \end{array} \right.$$

Recall that $\mathbf{v}$ and $\mathbf{w}$ being orthogonal means precisely that $\mathbf{v} \cdot \mathbf{w} = 0$. A basis $\{\mathbf{v}, \mathbf{w}\}$ of $\mathbb{R}^2$ is said to be an **orthonormal** basis if $\mathbf{v}$ and $\mathbf{w}$ are orthogonal unit vectors. We now generalize the concepts of projection and component to obtain a very convenient construction for the analysis of linear functions.

Parallel Projection

Recall that given any basis $\{\mathbf{v}, \mathbf{w}\}$ for $\mathbb{R}^2$ and any vector $\mathbf{x} \in \mathbb{R}^2$, there are unique numbers a and b for which

$$\mathbf{x} = a\mathbf{v} + b\mathbf{w}.$$

This is a (linear) vector equation for $(a, b)^T$ and we used the equivalent matrix equation with respect to the standard basis to discuss the solvability above. That matrix equation, in this case, is

$$\begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x_2 \\ x_2 \end{pmatrix}$$

where x_1 and x_2 are the coordinates of $\mathbf{x}$ with respect to the standard basis. The matrix M given by

$$M = \begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix}$$

is an example of a **change of basis matrix**. We will have more to say about such matrices later, but notice that one feeds into the associated linear function the coordinates $(a, b)^T$ of the vector $\mathbf{x}$ with respect to the basis $\{\mathbf{v}, \mathbf{w}\}$ and is given the coordinates (x_1, x_2) of $\mathbf{x}$ with respect to the standard basis. Also, the inverse matrix M^{-1} represents the reverse change of basis. Thus, change of basis is associated with invertible linear functions, in this case linear transformations of $\mathbb{R}^2$ (to itself).

We could look at obtaining the coordinates $(a, b)^T$ in a different way. If we take the dot product of the equation $a\mathbf{v} + b\mathbf{w} = \mathbf{x}$ with the vectors $\mathbf{v}$ and $\mathbf{w}$ we get

$$\begin{cases} (\mathbf{v}\cdot\mathbf{v})\, a + (\mathbf{v}\cdot\mathbf{w}), b &= \mathbf{x}\cdot\mathbf{v} \\ (\mathbf{v}\cdot\mathbf{w})\, a + (\mathbf{w}\cdot\mathbf{w}), b &= \mathbf{x}\cdot\mathbf{w} \end{cases}$$

or

$$\begin{pmatrix} \mathbf{v}\cdot\mathbf{v} & \mathbf{v}\cdot\mathbf{w} \\ \mathbf{v}\cdot\mathbf{w} & \mathbf{w}\cdot\mathbf{w} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \mathbf{x}\cdot\mathbf{v} \\ \mathbf{x}\cdot\mathbf{w} \end{pmatrix}. \tag{26}$$

This equation is also always uniquely solvable when $\{\mathbf{v}, \mathbf{w}\}$ is a basis for $\mathbb{R}^2$.

Exercise 34 *Write down the solution of (26) using Cramer's rule and draw a picture indicating the geometric significance of a and b.*

The numbers a and b are naturally called the **coefficients** or **components** of $\mathbf{x}$ with respect to the basis $\{\mathbf{v}, \mathbf{w}\}$. These are not orthogonal components, and they are sometimes called **parallel components**. This idea is naturally extended by saying that the **parallel projection** of $\mathbf{x}$ onto $\mathbf{v}$ is the vector $a\mathbf{v}$. Notice that the parallel projection of a vector $\mathbf{x}$ onto $\mathbf{v}$ cannot be determined from the vectors $\mathbf{x}$ and $\mathbf{v}$ in $\mathbb{R}^2$ alone. The *parallel* direction $\mathbf{w}$ is also required. Consequently, this kind of projection is somewhat more complicated than orthogonal projection and does not have a standard notation. Nevertheless, the geometric idea of parallel projection is very important, and we can attempt to introduce a (descriptive and colorful) notation for it. Given a basis $\{\mathbf{v}, \mathbf{w}\}$ for $\mathbb{R}^2$, let us denote by

$$\mathrm{proj}_{\mathbf{v}}^{\|\mathbf{w}}(\mathbf{x})$$

the unique vector $a\mathbf{v}$ obtained by expressing $\mathbf{x}$ as $\mathbf{x} = a\mathbf{v} + b\mathbf{w}$. This may be read: "The projection of $\mathbf{x}$ onto $\mathbf{v}$ with respect to the parallel direction $\mathbf{w}$." Similarly, we can express the coefficient as

$$a = \mathrm{component}_{\mathbf{v}}^{\|\mathbf{w}}(\mathbf{x})$$

so that

$$\mathrm{proj}_{\mathbf{v}}^{\|\mathbf{w}}(\mathbf{x}) = \mathrm{component}_{\mathbf{v}}^{\|\mathbf{w}}(\mathbf{x})\, \mathbf{v}.$$

That is "$a\mathbf{v}$ is the $\mathbf{v}$-component of $\mathbf{x}$ with respect to the basis $\{\mathbf{v}, \mathbf{w}\}$."

Exercise 35 *Show that*

$$\mathrm{proj}_{\mathbf{v}}^{\|\mathbf{w}} : \mathbb{R}^2 \to \mathbb{R}^2$$

and

$$\mathrm{comp}_{\mathbf{v}}^{\|\mathbf{w}} : \mathbb{R}^2 \to \mathbb{R}^1$$

are linear functions.

Finally, let us note that we can always obtain, from any basis of $\mathbb{R}^2$, a basis for $\mathbb{R}^2$ consisting of unit vectors. This simplifies the discussion of parallel projection. In particular, the system (26) becomes

$$\begin{pmatrix} 1 & \mathbf{v}\cdot\mathbf{w} \\ \mathbf{v}\cdot\mathbf{w} & 1 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \mathbf{x}\cdot\mathbf{v} \\ \mathbf{x}\cdot\mathbf{w} \end{pmatrix}.$$

Therefore,

$$a = \frac{(\mathbf{x}\cdot\mathbf{v}) - (\mathbf{v}\cdot\mathbf{w})(\mathbf{x}\cdot\mathbf{w})}{1-(\mathbf{v}\cdot\mathbf{w})^2} = \mathbf{x}\cdot\frac{\mathbf{v}-(\mathbf{v}\cdot\mathbf{w})\mathbf{w}}{1-(\mathbf{v}\cdot\mathbf{w})^2} \qquad \text{and} \qquad b = \mathbf{x}\cdot\frac{\mathbf{w}-(\mathbf{v}\cdot\mathbf{w})\mathbf{v}}{1-(\mathbf{v}\cdot\mathbf{w})^2}.$$

The vector $\mathbf{v} - (\mathbf{v}\cdot\mathbf{w})\mathbf{w}$ is orthogonal to $\mathbf{w}$. In particular,

$$\mathbf{w}^{\perp} = \pm\frac{\mathbf{v}-(\mathbf{v}\cdot\mathbf{w})\mathbf{w}}{|\mathbf{v}-(\mathbf{v}\cdot\mathbf{w})\mathbf{w}|},$$

and $\mathbf{v} - (\mathbf{v}\cdot\mathbf{w})\mathbf{w}$ is the orthogonal projection of $\mathbf{v}$ onto $\mathbf{w}^{\perp}$. We can compute the norm in the denominator as follows:

$$\begin{aligned} |\mathbf{v}-(\mathbf{v}\cdot\mathbf{w})\mathbf{w}|^2 &= 1 - 2(\mathbf{v}\cdot\mathbf{w})^2 + (\mathbf{v}\cdot\mathbf{w})^2 \\ &= 1-(\mathbf{v}\cdot\mathbf{w})^2. \end{aligned}$$

This suggests a reinterpretation of the parallel projection. Namely,

$$\mathrm{comp}_{\mathbf{v}}^{\|\mathbf{w}}(\mathbf{x}) = \frac{1}{\sqrt{1-(\mathbf{v}\cdot\mathbf{w})^2}}\,\mathbf{x}\cdot\frac{\mathrm{projection}_{\mathbf{w}^{\perp}}(\mathbf{v})}{\sqrt{1-(\mathbf{v}\cdot\mathbf{w})^2}}.$$

This means parallel projection onto $\mathbf{v}$ (with respect to the parallel field determined by $\mathbf{w}$) is given by a scaling of orthogonal projection onto a unit vector

$$\mathbf{z} = \frac{\mathrm{projection}_{\mathbf{w}^{\perp}}(\mathbf{v})}{\sqrt{1-(\mathbf{v}\cdot\mathbf{w})^2}}$$

in the direction of $\mathbf{w}^{\perp}$.

Exercise 36 *Notice that the coefficient matrix in (26) is symmetric, meaning that the matrix is its own transpose. Show that if A is a symmetric 2×2 matrix, then*

$$A\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot A\mathbf{y} \qquad \textit{for every } \mathbf{x},\mathbf{y}\in\mathbb{R}^2.$$

6 Linear Automorphisms of $\mathbb{R}^2$: Eigenvectors

So far, we should understand all linear transformations $L : \mathbb{R}^1 \to \mathbb{R}^m$ and $L : \mathbb{R}^2 \to \mathbb{R}^1$ pretty completely. The first class of linear functions was very easy to understand because only scaling was involved with the fundamental relation

$$L(t) = tL(1)$$

which told us everything. The transformations $L : \mathbb{R}^2 \to \mathbb{R}^1$ were a little more complicated, but we used the clever idea of classifying them according to the dimension of the kernel (and the image) subspaces. In this case, it should be noted that **we were saved significant complication because the case** $\ker(L) = \{\mathbf{0}\}$**, in which the transformation was one-to-one, was not possible**. First of all, this was simply because there was not room in the co-domain $\mathbb{R}$ to do anything interesting with all of $\mathbb{R}^2$. Put another way, such a transformation is forced to **collapse a one-dimensional subspace**, and consequently nothing very exciting happens in that direction. We will not be able to take advantage of that simplification for $L : \mathbb{R}^2 \to \mathbb{R}^2$.

Of course, we can still begin by classifying linear transformations $L : \mathbb{R}^2 \to \mathbb{R}^2$ according to the dimension of the subspaces $\ker(L)$ and $\mathrm{Im}(L)$:

(a) $\dim \ker(L) = 2$, and L is the null transformation.

(b) $\dim \ker(L) = 1$. (A one-dimensional subspace $\langle \mathbf{v} \rangle$ is collapsed.)

(c) $\dim \ker(L) = 0$. (This can happen now with $\dim \mathrm{Im}(L) = 2$ by the dimension theorem.)

We can consider this a preliminary classification. An interacting secondary classification is obtained by considering the possibility of finding **eigenvalues and eigenvectors**. The definition is crucial:

Definition 5 *Given $L : \mathbb{R}^2 \to \mathbb{R}^2$, a* **nonzero** *vector $\mathbf{v} \in \mathbb{R}^2$ is called an* **eigenvector** *for L if there is some real number λ for which*

$$L(\mathbf{v}) = \lambda \mathbf{v}. \tag{27}$$

Given an eigenvector, the number λ is called an **eigenvalue**. *The set of all eigenvectors (along with the zero vector),*

$$V_\lambda = \{\mathbf{v} \in \mathbb{R}^2 : L(\mathbf{v}) = \lambda \mathbf{v}\}$$

*is called the λ-***eigenspace**. *(It is a subspace of the domain of L.)*

Given an eigenvector $\mathbf{v}$ with corresponding eigenvalue λ, we say $(\lambda, \mathbf{v})$ is an **eigenvalue-eigenvector pair**.

This definition can obviously be extended for $L : \mathbb{R}^n \to \mathbb{R}^n$ but would make no sense for $L : \mathbb{R}^n \to \mathbb{R}^m$ when n and m are different. The condition (27) is familiar and important. The word "eigen" in German means "own" and (27) expresses the fact that the linear function when applied to $\mathbf{v}$ "stays in its own space" that is the space/line $\langle \mathbf{v} \rangle$ spanned by $\mathbf{v}$. Put another way, $\langle \mathbf{v} \rangle$ is **invariant** under L, so $\mathbf{v}$ might also be called an invariant vector, though it's really $\langle \mathbf{v} \rangle$ that is invariant.

It should not be missed, however, that the requirement $\mathbf{v} \neq \mathbf{0}$ is also important. The vector $\mathbf{0}$ always satisfies (27) for every λ, but $\mathbf{0}$ is not an eigenvector.

The number λ can be zero. Thus, the kernel is an **eigenspace** associated with the eigenvalue $\lambda = 0$ (whenever there is a nonzero vector in the kernel). Case (b) above says, in particular, that there exists an eigenvector. The first lesson to learn about $L : \mathbb{R}^2 \to \mathbb{R}^2$ is the following:

> *If there exists an eigenvector, then we can "see" and understand everything about the linear map $L : \mathbb{R}^2 \to \mathbb{R}^2$.*

It may take some time to see this statement fully, but let's start with case (b) and see that it is an accurate statement in that case.

Before we do that, perhaps some examples are in order. A **simple rotation** of the plane is a linear function. To be specific counterclockwise rotation by an angle θ is given by

$$L\mathbf{x} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{28}$$

Note that the images of $\mathbf{e}_1$ and $\mathbf{e}_2$ are the columns in the matrix. This linear transformation has **no eigenvectors** unless $\theta = k\pi$ for some $k \in \mathbb{Z}$.

A **homogeneous scaling** given by

$$L\mathbf{x} = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda I \mathbf{x} \tag{29}$$

is also linear. This is perhaps the simplest linear function that is not the null map or the identity (though the class includes the null map with $\lambda = 0$ and the identity when $\lambda = 1$). When $\lambda = -1$, this is a rotation by angle π. For a homogeneous scaling, every nonzero vector is an eigenvector.

Exercise 37 *Show that if $L : \mathbb{R}^2 \to \mathbb{R}^2$ is linear and every nonzero vector in $\mathbb{R}^2$ is an eigenvector of L (associated with some eigenvalue), then $L = \lambda\,\mathrm{id}$ for some λ.*

In Exercise37 the symbol id = $\mathrm{id}_{\mathbb{R}^2}$ represents the **identity mapping**. That is, $\mathrm{id} : \mathbb{R}^2 \to \mathbb{R}^2$ by $\mathrm{id}(\mathbf{x}) = \mathbf{x}$.

If $L : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$L\mathbf{x} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{30}$$

where λ_1 and λ_2 may be different real numbers, then L is said to be a **diagonal** linear map. If $\lambda_1 \neq \lambda_2$, then not every nonzero vector is an eigenvector, but $\{\mathbf{e}_1, \mathbf{e}_2\}$ is **a basis of eigenvectors**. This is also (when $\lambda_1 \neq \lambda_2$) an example of an **anisotropic scaling**.

Exercise 38 *Consider* $\partial B_r(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^2 : |\mathbf{x}| = r\}$ *for* $r > 0$.

(a) *If L is a simple rotation of the plane, then*

$$L(\partial B_r(\mathbf{0})) = \{L\mathbf{x} : \mathbf{x} \in \partial B_r(\mathbf{0})\} = \partial B_r(\mathbf{0}).$$

(b) *If L is a homogeneous scaling, what is* $L(\partial B_r(\mathbf{0}))$*?*

(c) *If* $L : \mathbb{R}^2 \to \mathbb{R}^2$ *is a diagonal linear map, then show* $L(\partial B_r(\mathbf{0}))$ *is an ellipse, and give the equation/relation defining that ellipse.*

Given $\theta \notin \{k\pi : k \in \mathbb{Z}\}$, the vectors $(\cos\theta, \sin\theta)^T$ and $(-\sin\theta, \cos\theta)^T$ (together) are an orthonormal basis for $\mathbb{R}^2$. The linear map $L : \mathbb{R}^2 \to \mathbb{R}^2$ determined by

$$L\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \lambda_1 \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \qquad \text{and} \qquad L\begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} = \lambda_2 \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \tag{31}$$

is very similar to a diagonal map. This is an example of a kind of linear map we will consider later which is called **diagonalizable**. Recall that the mapping specifications (31) can also be indicated by

$$\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \mapsto \lambda_1 \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix} \mapsto \lambda_2 \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

Exercise 39 *Suppose* $\{\mathbf{v}, \mathbf{w}\}$ *is a basis for* $\mathbb{R}^2$, *and* $L : \mathbb{R}^2 \to \mathbb{R}^2$ *is determined by the conditions*

$$\mathbf{v} \mapsto \lambda_1 \mathbf{v} \qquad \text{and} \qquad \mathbf{w} \mapsto \lambda_2 \mathbf{w}$$

for some nonzero real numbers λ_1 *and* λ_2. *Under what conditions is* $L(\partial B_1(\mathbf{0}))$ *an ellipse?*

If $\{\mathbf{v}, \mathbf{v}^\perp\}$ is an orthonormal basis for $\mathbb{R}^2$ and $\{\mathbf{w}, \mathbf{w}^\perp\}$ is another orthonormal basis for $\mathbb{R}^2$, then there are two linear transformations $L_\pm$ determined by the conditions

$$L_+ : \mathbf{v} \mapsto \mathbf{w} \quad \text{and} \quad \mathbf{v}^\perp \mapsto \mathbf{w}^\perp$$

and

$$L_- : \mathbf{v} \mapsto \mathbf{w}^\perp \quad \text{and} \quad \mathbf{v}^\perp \mapsto \mathbf{w}.$$

Exercise 40 *Show that L_+ is a rotation and L_- is not a rotation.*

In summary, we have suggested the following examples:

1. orthogonal projection
2. parallel projection
3. isotropic scaling
4. anisotropic scaling (at least in the case of orthogonal eigenvectors)
5. rotation

It may be noted also that rotation can be combined freely with (isotropic) scaling (using composition in either order). This gives an example which will be used later.

Let us begin with our classification of linear functions $L : \mathbb{R}^2 \to \mathbb{R}^2$ according to the dimension of $\ker(L)$. We know that if $\dim \ker(L) = 2$, then L is the null map, and there is nothing to say. Again, let us refer to the situation in which $\dim \ker(L) = 1$ as "case (b)." In this case, $\ker(L) = \langle \mathbf{v} \rangle$ for some nonzero vector $\mathbf{v} \in \mathbb{R}^2$ and (now) $\mathrm{Im}(L) = \langle L\mathbf{w} \rangle \subset \mathbb{R}^2$ has dimension one (by the dimension theorem) for any $\mathbf{w} \in \mathbb{R}^2 \backslash \langle \mathbf{v} \rangle$.

Exercise 41 *Give the details of why we know that any vector $\mathbf{w} \in \mathbb{R}^2 \backslash \langle \mathbf{v} \rangle$ is a nonzero vector with $L\mathbf{w} \neq \mathbf{0}$.*

I think we are in a position to make at least a vague guess at this point.

> Setting $\alpha = |L\mathbf{w}|/|\mathbf{w}|$, we guess L is a scaling by α of the component projection onto some vector $\mathbf{z} \in \mathbb{R}^2 = \mathrm{Dom}(L)$, transferred to $\langle L\mathbf{w} \rangle$. That is, we expect
>
> $$L\mathbf{x} = \alpha \, \mathrm{comp}_{\mathbf{z}}(\mathbf{x}) \, \frac{L\mathbf{w}}{|L\mathbf{w}|} \qquad \text{for } \alpha = |L\mathbf{w}|/|\mathbf{w}| \text{ and some } \mathbf{z} \in \mathbb{R}^2 \text{ with } |\mathbf{z}| = 1.$$

Exercise 42 *Draw a picture associated with case (b) indicating the vectors $\mathbf{v}$, $\mathbf{w}$, and $L\mathbf{w}$. Draw scaling marks on $\mathbf{w}$ and $L\mathbf{w}$ to indicate the guess above.*

If we are able to take $|\mathbf{w}| = 1$ (which is likely), then the proposed scaling factor is just $|L\mathbf{w}|$. For example, we may begin by taking our null eigenvector $\mathbf{v}$ with $|\mathbf{v}| = 1$ and then take $\mathbf{w}$ with $\mathbf{w}^\perp = \mathbf{v}$ as before. In this case we will have $|\mathbf{w}| = 1$, and we can write any vector $\mathbf{x}$ as

$$\mathbf{x} = a\mathbf{w} + b\mathbf{v}$$

so that

$$Lx = aL\mathbf{w} = a\,|L\mathbf{w}|\,\frac{L\mathbf{w}}{|L\mathbf{w}|}. \tag{32}$$

Thus, we see the proposed scaling factor does make a relatively natural immediate appearance with this choice, and it remains to see the projection.

As we consider the picture associated with Exercise 42 and contemplate the fact that the domain and co-domain are the same space here and, hence, can be compared to one another, two subcases present themselves:

CASE 1 $L\mathbf{w} \in \langle \mathbf{v} \rangle = \ker(L)$.

CASE 2 $L\mathbf{w} \notin \langle \mathbf{v} \rangle = \ker(L)$.

In CASE 1, a very interesting thing happens. Not only is it the case that $L\mathbf{w} \in \langle \mathbf{v} \rangle$ but

$$L\mathbf{x} \in \langle \mathbf{v} \rangle \qquad \text{for every } \mathbf{x} \in \mathbb{R}^2.$$

This is easy to see since we can write $\mathbf{x} = a\mathbf{w} + b\mathbf{v}$, so

$$L\mathbf{x} = aL\mathbf{w} \in \langle \mathbf{v} \rangle.$$

Thus, we can take $\mathbf{z} = \mathbf{w}$ to be a vector with $\mathbf{w}^\perp = \mathbf{v}$ and $|\mathbf{v}| = 1$ as suggested above. With this choice $\mathbf{x} = (\mathbf{x} \cdot \mathbf{w})\mathbf{w} + (\mathbf{x} \cdot \mathbf{v})\mathbf{v}$ and

$$\begin{aligned} L\mathbf{x} &= (\mathbf{x} \cdot \mathbf{w})\, L\mathbf{w} \\ &= |L\mathbf{w}|\,(\mathbf{x} \cdot \mathbf{w})\,\frac{L\mathbf{w}}{|L\mathbf{w}|} \\ &= \pm\alpha\, \mathrm{comp}_{\mathbf{w}}(\mathbf{x})\,\mathbf{v}. \end{aligned}$$

This is almost our guess. It **is** our guess up to a sign.

If $L\mathbf{w}/|L\mathbf{w}| = \mathbf{v}$, then

$$L\mathbf{x} = \alpha\, \mathrm{comp}_{\mathbf{w}}(\mathbf{x})\,\mathbf{v},$$

and our guess holds exactly. The other possibility is that $L\mathbf{w}/|L\mathbf{w}| = -\mathbf{v}$. Then we can choose $\mathbf{z} = -\mathbf{w}$ and write

$$L\mathbf{x} = \alpha\, \mathrm{comp}_{\mathbf{z}}(\mathbf{x})\,\mathbf{v}.$$

Again, our guess holds. This is an interesting linear function which we should understand completely now.

Exercise 43 *Draw the mapping picture (with scaling dots) for L is case (b), CASE 1.*

Exercise 44 *Show that in case (b), CASE 1,* $L^2 \equiv 0$.

In case (b), CASE 2, we claim there is **another eigenvector** corresponding to a different eigenvalue $\lambda \neq 0$. This is easy to see in this case, though you may know more sophisticated reasons we will talk about later. Here we know by assumption that $L^2\mathbf{w} \neq \mathbf{0}$, because $L\mathbf{w} \notin \ker(L)$. On the other hand, we can write $L\mathbf{w} = c\mathbf{w} + d\mathbf{v}$ for some scalars c and d. This means, $L\mathbf{w} \neq \mathbf{0}$ and $L(L\mathbf{w}) = cL\mathbf{w}$. That is, $L\mathbf{w}$ is an eigenvector. Again, the constant c is nonzero because $L\mathbf{w} \notin \ker(L)$.

This means, there is a basis $\{\mathbf{v}, \mathbf{w}\}$ of $\mathbb{R}^2$ consisting of eigenvectors. To find $\mathbf{w}$ just take any $\mathbf{w}_0 \in \mathbb{R}^2 \backslash \langle \mathbf{v} \rangle$ and set $\mathbf{w} = L\mathbf{w}_0$. At this point, we can see geometrically $L : \mathbb{R}^2 \to \mathbb{R}^2$ with $\ker(L) = \langle \mathbf{v} \rangle \neq \{\mathbf{0}\}$ and $L(\mathbb{R}^2) = \{L\mathbf{x} : \mathbf{x} \in \mathbb{R}^2\} = \langle \mathbf{w} \rangle$.

Exercise 45 *Draw the mapping picture (with scaling dots) for* $L : \mathbb{R}^2 \to \mathbb{R}^2$ *in case (b), CASE 2.*

You might think at this point that we can simply take $\mathbf{z} = \mathbf{w}$, the eigenvector with nonzero eigenvalue, and verify our guess (at least up to a sign). It turns out that is not always the case. Let us write again, as usual, $\mathbf{x} = a\mathbf{w} + b\mathbf{v}$ with $\mathbf{v}$ and $\mathbf{w}$ unit (eigen)vectors corresponding to the eigenvalues $\lambda \neq 0$ and $\lambda_0 = 0$ respectively. Then

$$L\mathbf{x} = aL\mathbf{w} = \lambda\, a\, \mathbf{w}. \tag{33}$$

The question is: Is the coefficient a the projection of $\mathbf{x}$ onto some unit vector $\mathbf{z}$? That is, is there a vector $\mathbf{z}$ for which

$$a = \mathbf{x} \cdot \mathbf{z} = \text{comp}_{\mathbf{z}}(\mathbf{x})\,?$$

We found these coefficients a and b above by resorting to the standard basis coordinates and writing $\mathbf{w} = (w_1, w_2)^T$ and $\mathbf{v} = (v_1, v_1)^T$. We could do that again, but let's do something a little different (that does not depend on the standard basis). Taking the dot product of the equation $\mathbf{x} = a\mathbf{w} + b\mathbf{v}$ with $\mathbf{w}$ and $\mathbf{v}$, we obtain a system of equations:

$$\begin{cases} \quad a \quad + (\mathbf{v} \cdot \mathbf{w})b = \mathbf{x} \cdot \mathbf{w} \\ (\mathbf{v} \cdot \mathbf{w})a + \quad b \quad = \mathbf{x} \cdot \mathbf{v}. \end{cases}$$

As usual, there is a linear function and mapping picture associated with this system which we see by writing

$$\begin{pmatrix} 1 & \mathbf{v}\cdot\mathbf{w} \\ \mathbf{v}\cdot\mathbf{w} & 1 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \mathbf{x}\cdot\mathbf{w} \\ \mathbf{x}\cdot\mathbf{v} \end{pmatrix}.$$

The coefficient matrix in this equation has determinant $1-(\mathbf{v}\cdot\mathbf{w})^2 > 0$ because $\mathbf{v}$ and $\mathbf{w}$ are not parallel. Therefore, there is a unique solution from Cramer's rule

$$a = \frac{\mathbf{x}\cdot\mathbf{w} - (\mathbf{v}\cdot\mathbf{w})(\mathbf{x}\cdot\mathbf{v})}{1-(\mathbf{v}\cdot\mathbf{w})^2} = \mathbf{x}\cdot\frac{\mathbf{w}-(\mathbf{v}\cdot\mathbf{w})\mathbf{v}}{1-(\mathbf{v}\cdot\mathbf{w})^2}. \tag{34}$$

The vector $\mathbf{w}-(\mathbf{v}\cdot\mathbf{w})\mathbf{v}$ appearing here is an interesting vector. Notice that this vector is orthogonal to $\mathbf{v}$ because

$$\mathbf{v}\cdot[\mathbf{w}-(\mathbf{v}\cdot\mathbf{w})\mathbf{v}] = 0.$$

Also, $(\mathbf{v}\cdot\mathbf{w})\mathbf{v} = \mathrm{proj}_{\mathbf{v}}(\mathbf{w})$ is the projection of $\mathbf{w}$ on $\mathbf{v}$. Therefore, this vector is the *projection of* $\mathbf{w}$ *orthogonal to* $\mathbf{v}$. If we compute the norm (squared) of this vector we get

$$|\mathbf{w}-(\mathbf{v}\cdot\mathbf{w})\mathbf{v}|^2 = [\mathbf{w}-(\mathbf{v}\cdot\mathbf{w})\mathbf{v}]\cdot[\mathbf{w}-(\mathbf{v}\cdot\mathbf{w})\mathbf{v}] = 1-(\mathbf{v}\cdot\mathbf{w})^2.$$

This suggests that we consider the unit vector $\mathbf{z}$ given by

$$\mathbf{z} = \frac{\mathbf{w}-(\mathbf{v}\cdot\mathbf{w})\mathbf{v}}{\sqrt{1-(\mathbf{v}\cdot\mathbf{w})^2}}.$$

Now (33) becomes

$$Lx = \frac{\lambda}{\sqrt{1-(\mathbf{v}\cdot\mathbf{w})^2}}\,(\mathbf{x}\cdot\mathbf{z})\,\mathbf{w} = \frac{\lambda}{\sqrt{1-(\mathbf{v}\cdot\mathbf{w})^2}}\,\mathrm{comp}_{\mathbf{z}}\mathbf{x}\;\mathbf{w}.$$

Thus, we see that our guess was correct to a point, but not quite completely correct. The scaling factor is not $|L\mathbf{w}|$ but rather

$$\alpha = \frac{\lambda}{\sqrt{1-(\mathbf{v}\cdot\mathbf{w})^2}} = \pm\frac{|L\mathbf{w}|}{\sqrt{1-(\mathbf{v}\cdot\mathbf{w})^2}}.$$

Exercise 46 *Consider two cases and show that a corrected version of our guess (with a corrected scaling factor) does hold.*

The reason our guess failed was, essentially, that it assumed the projection was onto $\mathbf{z} = \mathbf{w}$. As our calculation showed, the vector $\mathbf{z}$ needed to be orthogonal to $\mathbf{v}$, and in the case where the other eigenvector $\mathbf{w}$ is orthogonal to the first eigenvector $\mathbf{v}$, then the guess also works because $\mathbf{v} \cdot \mathbf{w} = 0$, and the correction factor becomes one. Our intuition was reasonably good, however, because if we had taken a non-orthogonal projection/component determined by $\mathbf{v}$, then we would have arrived at a correct guess as well, but the guess would have been more complicated.

Exercise 47 *Express a linear function* $L : \mathbb{R}^2 \to \mathbb{R}^2$ *with* $\ker(L) = \langle \mathbf{v} \rangle \neq \{\mathbf{0}\}$ *and a second eigenvector* $\mathbf{w}$ *in terms of a component along* $\mathbf{w}$.

Solution: The component of $\mathbf{x}$ along $\mathbf{w}$ with respect to the eigenbasis $\{\mathbf{v}, \mathbf{w}\}$ is given precisely by the value of a in (34). Thus, for $\mathbf{x} = a\mathbf{w} + b\mathbf{v}$ we have

$$L\mathbf{x} = \lambda \, a \, \mathbf{w}.$$

This is saying precisely that the value of L is given by a scaling of the (non-orthogonal) component of $\mathbf{x}$ along $\mathbf{w}$ with respect to the basis $\{\mathbf{v}, \mathbf{w}\}$ scaled by λ.

Let us return briefly to the part of the analysis of case (b), CASE 2 in which we obtained the second eigenvector $\mathbf{w}$. We expect to solve the mapping problem $L\mathbf{w} = \lambda\mathbf{w}$ for some nonzero vector $\mathbf{w} \in \mathbb{R}^2 \backslash \langle \mathbf{v} \rangle$. There are a couple important general facts one should keep in mind here. Recall, first of all, that by the dimension theorem a linear function $L : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one and onto if L is either one-to-one or L is onto. Also, these are equivalent to the condition that no nonzero vector $\mathbf{v}$ satisfies $L\mathbf{v} = \mathbf{0}$. It is relatively easy to see by the product theorem for determinants that this condition (of L being a bijection) is also equivalent to the condition

$$\det A \neq 0$$

where A is a matrix for which $L\mathbf{x} = A\mathbf{x}$ with respect to any particular basis. More generally, if A and B are matrices for which $L\mathbf{x} = A\mathbf{x}$ when $\mathbf{x}$ and $L\mathbf{x}$ are written in coordinates with respect to one basis and $L\mathbf{x} = B\mathbf{x}$ when $\mathbf{x}$ and $L\mathbf{x}$ are written in coordinates with respect to another basis, then there is an $n \times n$ matrix Q (a change of basis matrix) with an inverse matrix Q^{-1} for which the conjugation

$$B = QAQ^{-1}$$

holds. Since $QQ^{-1} = I$ is the identity matrix, $\det(QQ^{-1}) = \det Q \det Q^{-1} = 1$, and

$$\det B = \det(QAQ^{-1}) = \det Q \det A \det Q^{-1} = \det A.$$

This means the number $\det A$ is independent of which basis with respect to which one expresses L in terms of matrix multiplication. This number only depends on L, and we can call it $\det(L)$.

Please note this well: There is such a thing as **the determinant of a linaer transformation** $L : \mathbb{R}^n \to \mathbb{R}^n$. This is, in principle, different from the determinant of a matrix. The determinant of a linear transformation is the number you get whenever you choose a basis for $\mathbb{R}^n$ and express L in terms of matrix multiplication by $L(\mathbf{x}) = A\mathbf{x}$, and then take the usual matrix determinant of A.

Furthermore, the linear transformation $L : \mathbb{R}^n \to \mathbb{R}^n$ is a bijection if and only if $\det(L) \neq 0$.

Next, this has an important relation to eigenvalues and eigenvectors. When we are looking for an eigenvector $\mathbf{w}$ of $L : \mathbb{R}^n \to \mathbb{R}^n$, we are looking for a nonzero vector $\mathbf{w}$ satisfying $L\mathbf{w} = \lambda\mathbf{w}$ (for some λ. This can be written

$$(L - \lambda\,\mathrm{id})(\mathbf{w}) = \mathbf{0}.$$

That is, the linear function $L - \lambda\,\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ is **not** a bijection, so

$$\det(L - \lambda\,\mathrm{id}) = 0. \tag{35}$$

The relation appearing in (35) is notable in that it does not involve the (sought after and unknown) eigenvector $\mathbf{w}$. In fact, the condition (35) is an n-th order polynomial equation for the eigenvalue λ called the **characteristic equation**. The expression $\det(L - \lambda\,\mathrm{id}) \in P_n[\lambda]$, viewed as an n-th order polynomial equation in λ, is called the **characteristic polynomial**.

The following are also worth noting:

1. Not all linear functions $L : \mathbb{R}^n \to \mathbb{R}^n$ have eigenvalue-eigenvector pairs.

   Exercise 48 *Compute the characteristic equation for a rotation of* $\mathbb{R}^2$.

2. If $L : \mathbb{R}^2 \to \mathbb{R}^2$ and
   $$\det(L - \lambda\,\mathrm{id}) = (\lambda - \lambda_1)(\lambda - \lambda_2)$$
   for some real numbers λ_1 and λ_2, then
   $$(L - \lambda_1\,\mathrm{id}) \circ (L - \lambda_2\,\mathrm{id})(\mathbf{x}) \equiv \mathbf{0} \qquad \text{for every } \mathbf{x} \in \mathbb{R}^2, \tag{36}$$
   where the symbol "$\circ$" indicates function composition. This says, roughly, that "If you evaluate the characteristic polynomial on the transformation L, then you get the zero transformation." It is a special case of what is called the **Cayley-Hamilton Theorem**.

   Exercise 49 *Prove the Cayley-Hamilton theorem in the special case expressed by (36).*

Exercise 50 *Explain (to yourself) what it means to evaluate a polynomial*

$$p(x) = \sum_{j=0}^{k} a_j x^j$$

on a linear function $L : \mathbb{R}^n \to \mathbb{R}^n$*. In particular, you should note that powers are interpreted in such an evaluation in terms of iteration, and you should understand what it means to evaluate a constant polynomial* $p(x) = c$ *on* L*.*

3. The polynomial of least order $m = m(x)$ for which $m(L) : \mathbb{R}^n \to \mathbb{R}^n$ is the zero mapping is called the **minimal polynomial** for L. The minimal polynomial divides the characteristic polynomial and is unique up to a scaling.

4. This last note is not immediately directly related, but this is a good place to get it out of the way: If

   (a) $L : \mathbb{R}^n \to \mathbb{R}^m$ is given by $L\mathbf{x} = A\mathbf{x}$ with respect to some given bases $\mathcal{B}_n$ and $\mathcal{B}_m$ of $\mathbb{R}^n$ and $\mathbb{R}_m$ respectively, and

   (b) $J : \mathbb{R}^m \to \mathbb{R}^k$ is given by $J\mathbf{y} = B\mathbf{y}$ with respect to the basis $\mathcal{B}_m$ for $\mathbb{R}^m$ and the basis $\mathcal{B}_k$ of $\mathbb{R}^k$,

   then there is a linear function

   $$J \circ L : \mathbb{R}^n \to \mathbb{R}^k : \mathbb{R}^n \to \mathbb{R}^k \qquad \text{by} \qquad J \circ L(\mathbf{x}) = J(L(\mathbf{x}))$$

   called the **composition** of J on L. This linear function is given by $J \circ L\mathbf{x} = BA\mathbf{x}$ where BA is the matrix product of the $k \times m$ matrix B and the $m \times n$ matrix A.

Let us assume we have finished with case (b) in which $\ker(L)$ is one-dimensional.

Exercise 51 *Discuss the eigenvectors associated with orthogonal projection and parallel projection. You may wish to look at the summary discussion of these functions in Section 8 below.*

We turn now to case (c) in which $\dim\ker(L) = 0$ and $L : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear bijection. We will use standard coordinates. In this case, the vector $L\mathbf{e}_1$ is a nonzero vector. It follows that

$$L\mathbf{e}_1 = |L\mathbf{e}_1| \begin{pmatrix} \cos\theta_0 \\ \sin\theta_0 \end{pmatrix}$$

for some unique $\theta_0 \in [0, 2\pi)$. It is easy to see then that $P \circ L : \mathbb{R}^2 \to \mathbb{R}^2$ where

$$P : \mathbb{R}^2 \to \mathbb{R}^2 \qquad \text{by} \qquad P\mathbf{x} = \begin{pmatrix} \cos\theta_0 & \sin\theta_0 \\ -\sin\theta_0 & \cos\theta_0 \end{pmatrix} \mathbf{x}$$

is clockwise rotation by the angle θ_0 satisfies

$$P \circ L\mathbf{e}_1 = \alpha\mathbf{e}_1$$

with $\alpha = |L\mathbf{e}_1| > 0$. Thus, the composition $P \circ L$ has an eivenvector, namely, $\mathbf{e}_1$.

Theorem 2 *Given any linear function $L : \mathbb{R}^2 \to \mathbb{R}^2$ with $\det(L) \neq 0$, there is a unique angle θ_0 such that $L = P^{-1} \circ L_0$ where $P^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ is counterclockwise rotation by θ_0 and $L_0 : \mathbb{R}^2 \to \mathbb{R}^2$ satisfies $L_0\mathbf{e}_1 = \alpha\mathbf{e}_1$ for $\alpha = |L\mathbf{e}_1| > 0$.*

With respect to the standard basis the matrix of $L_0 = P \circ L$ has the form

$$A = \begin{pmatrix} \alpha & a_{12} \\ 0 & a_{22} \end{pmatrix}.$$

The characteristic polynomial of L_0 is $p(\lambda) = (\lambda - \alpha)(\lambda - a_{22})$. This means a_{22} is also an eigenvalue. Thus, we have eigenvalue-eigenvector pairs $(\alpha, \mathbf{e}_1)$ and $(\beta, \mathbf{w})$ with $\beta = a_{22}$. This does not mean, however, that $\{\mathbf{e}_1, \mathbf{w}\}$ is necessarily a linearly independent set and, thus, a basis for $\mathbb{R}^2$.

Exercise 52 *Show $a_{22} \neq 0$ (in case (c)). Find an eigenvector $\mathbf{w}$ associated with the eigenvalue $\beta = a_{22}$ and explain when $\{\mathbf{e}_1, \mathbf{w}\}$ is a basis for $\mathbb{R}^2$ (and when this is not the case).*

Recall that the first eigenvalue of L_0 is given by $\alpha = |L\mathbf{e}_1|$. We may be interested in an expression for the eigenvalue $\beta = a_{22}$ in terms of the original linear transformation L. If the matrix of L with respect to the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ is

$$M = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix}$$

then our observation about the relation between compositions and matrix multiplication tells us

$$\begin{pmatrix} \alpha & a_{12} \\ 0 & a_{22} \end{pmatrix} = \begin{pmatrix} \cos\theta_0 & \sin\theta_0 \\ -\sin\theta_0 & \cos\theta_0 \end{pmatrix} \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix}$$

so that

$$\beta = a_{22} = -\mu_{12}\sin\theta_0 + \mu_{22}\cos\theta_0.$$

If $\beta = 0$, then $\det(P)\det M = \det A = 0$ (which is a contradiction because we are considering case (c) in which $\dim\ker(L) = 2$ so $\det M \neq 0$).

In order to delineate cases associated with the simplified matrix L_0, it is perhaps natural to ask first the question:

> Is it possible that the eigenvector $\mathbf{w}$ associated with the nonzero eigenvalue $\beta = a_{22}$ is $\mathbf{e}_1$?

The answer to this question is "yes," but only under rather special circumstances. Expressing $\mathbf{w}$ in standard coordinates as $(w_1, w_2)^T$, we have the condition

$$(\alpha - a_{22})w_1 + a_{12}w_2 = 0. \tag{37}$$

Consequently, we can take

$$\mathbf{w}_0 = \begin{pmatrix} a_{12} \\ a_{22} - \alpha \end{pmatrix} \tag{38}$$

as an initial eigenvector corresponding to $\beta = a_{22}$. Clearly, in order for $\mathbf{w}_0$ to be a multiple of $\mathbf{e}_1$, one must have $\beta = a_{22} = \alpha$.

On the other hand, we see that our assertion concerning the vector $\mathbf{w}_0$ in (38) was a bit too hasty. We may now delineate three cases:

(i) $a_{22} = \alpha = |L\mathbf{e}_1|$ and $a_{12} = 0$.

(ii) $a_{22} = \alpha$ and $a_{12} \neq 0$.

(iii) $a_{22} \neq \alpha$.

In the first case $\mathbf{w}_0$ given in (38) is the zero vector and, hence, is not an eigenvector. On the other hand, every nonzero vector $\mathbf{w} = (w_1, w_2)^T$ satisfies (37) in this case, and L_0 is an isotropic scaling by $\alpha = |L\mathbf{e}_1|$. Therefore, $L = P^{-1} \circ (\alpha\, \mathrm{id}_{\mathbb{R}^2}) = \alpha P^{-1}$ is a counterclockwise rotation by the angle θ_0 with an isotropic scaling (it does not matter in which order these are applied), and is thus a simple example of a linear function considered above.

In case (ii) we encounter a new kind of linear transformation worth considering in detail.

6.1 Jordan Shear

If the matrix for $L_0 = P \circ L$ (with respect to the standard basis) has the form

$$A = \begin{pmatrix} \alpha & a_{12} \\ 0 & \alpha \end{pmatrix}$$

with $\alpha = |L\mathbf{e}_1| > 0$ and $a_{12} \neq 0$, then two different normalizations by scaling are standard. Writing $L_0 = a_{12}J$ where

$$J\mathbf{x} = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}\mathbf{x} \tag{39}$$

(in standard coordinates) and $\lambda = \alpha/a_{12}$, the linear function $J : \mathbb{R}^2 \to \mathbb{R}^2$ is called a **Jordan transformation** (in standard form).

If, on the other hand, we write $L_0 = \alpha\Sigma$ where

$$\Sigma\mathbf{x} = \begin{pmatrix} 1 & \sigma \\ 0 & 1 \end{pmatrix}\mathbf{x}$$

(in standard coordinates) and $\sigma = a_{12}/\alpha = \tan\phi$, the linear function $\Sigma : \mathbb{R}^2 \to \mathbb{R}^2$ is called a (standard) **linear shear** of $\mathbb{R}^2$. It may be mentioned at this point that Mathematica has in its library of standard functions a function

$$\texttt{ShearingTransform} = \texttt{ShearingTransform}[\phi, \mathbf{v}, \mathbf{n}]$$

which is said to "shear by ϕ radians along the direction $\mathbf{v}$ normal to the direction $\mathbf{n}$." Taking $\mathbf{v} = \mathbf{e}_1$ and $\mathbf{n} = \mathbf{e}_2$, `ShearingTransform` is precisely an implementation of our standard shear transformation Σ.

Exercise 53 *What does* `ShearingTransform` *do for other values of the vectors* $\mathbf{v}$ *and* $\mathbf{n}$*?*

The eigenspace

$$\{\mathbf{v} : J\mathbf{v} = \lambda\mathbf{v}\} = \langle \mathbf{e}_1 \rangle$$

associated with J is one-dimensional. Perhaps the easiest way to see the effect of the Jordan transformation on the plane is to consider the images of the vertical lines $\{(x, y) : y \in \mathbb{R}\}$ and x is fixed. Each maps to the line through $\alpha\, x\, \mathbf{e}_1$ parallel to the vector $\mathbf{e}_1 + \alpha\mathbf{e}_2 = (1, \alpha)^T$. Thus, the image of the L^∞ ball

$$B_1^\infty(\mathbf{0}) = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} : \max\{|x|, |y|\} < 1 \right\}$$

is a parallelogram skewed/sheared left for $\lambda < 0$, decreasing in height as λ increases to zero, and converging to the segment $\{(u, 0)^T : |u| \leq 1\}$ as λ tends to zero from the left. Squares concentric with the boundary of the L^∞ ball have images proportionally concentric to the image of $\partial B_1^\infty(\mathbf{0})$ giving a relatively satisfactory depiction of the Jordan transformation.

Careful consideration of the image of the unit circle (with respect to the Euclidean metric), however, suggests that this basic understanding leaves some important properties of the Jordan transformation obscure.

Image of the unit circle under Jordan transformation of $\mathbb{R}^2$

$$J(\partial B_1(\mathbf{0})) = \left\{ J\begin{pmatrix} x \\ y \end{pmatrix} : x^2 + y^2 = 1 \right\} = \left\{ \begin{pmatrix} u \\ v \end{pmatrix} : \left| J^{-1}\begin{pmatrix} u \\ v \end{pmatrix} \right| = 1 \right\}.$$

Thus, a first step in understanding this set is to observe

$$J^{-1}\begin{pmatrix} u \\ v \end{pmatrix} = \frac{1}{\lambda^2}\begin{pmatrix} \lambda & -1 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix},$$

so we obtain the quadratic relation

$$\lambda^2 u^2 - 2\lambda uv + (1 + \lambda^2)v^2 = \lambda^4 \tag{40}$$

for $(u, v) \in J(\partial B_1(\mathbf{0}))$. It is standard in the treatment of such quadratic relations to introduce a change of variables by rotation. Let us introduce a new orthonormal basis

$$\left\{ \begin{pmatrix} \cos\psi \\ \sin\psi \end{pmatrix}, \begin{pmatrix} -\sin\psi \\ \cos\psi \end{pmatrix} \right\}.$$

With respect to this new basis we have

$$\begin{pmatrix} u \\ v \end{pmatrix} = \xi \begin{pmatrix} \cos\psi \\ \sin\psi \end{pmatrix} + \eta \begin{pmatrix} -\sin\psi \\ \cos\psi \end{pmatrix}.$$

Substituting these values for u and v in the relation (40) we find

$$\begin{aligned}(\lambda^2 + \sin^2\psi - \lambda\sin 2\psi)\xi^2 \\ + (\sin 2\psi - 2\lambda\cos 2\psi)\xi\eta \\ + (\lambda^2 + \cos^2\psi - \lambda\sin 2\psi)\eta^2 = \lambda^4.\end{aligned}$$

The choice

$$\psi = \psi_1 = \frac{1}{2}\tan^{-1}(2\lambda) \tag{41}$$

reduces the relation (40) in terms of the new coordinates ξ and η to

$$(\lambda^2 + \sin^2\psi - \lambda\sin 2\psi)\xi^2 + (\lambda^2 + \cos^2\psi - \lambda\sin 2\psi)\eta^2 = \lambda^4. \tag{42}$$

At this point we recall and use the identities

$$\cos^2\psi = \frac{1}{2}(\cos^2\psi + 1 - \sin^2\psi) = \frac{1}{2}(1 + \cos 2\psi)$$

and

$$\sin^2\psi = \frac{1}{2}(\sin^2\psi + 1 - \cos^2\psi) = \frac{1}{2}(1 - \cos 2\psi)$$

to write (42) as

$$\left[\lambda^2 + \frac{1}{2}(1 - \cos 2\psi) - \lambda \sin 2\psi\right]\xi^2 + \left[\lambda^2 + \frac{1}{2}(1 + \cos 2\psi) - \lambda \sin 2\psi\right]\eta^2 = \lambda^4.$$

Finally, because $\tan 2\psi_1 = 2\lambda$, we know

$$\cos\psi_1 = \frac{1}{\sqrt{1 + 4\lambda^2}} \qquad \text{and} \qquad \sin\psi_1 = \frac{2\lambda}{\sqrt{1 + 4\lambda^2}}.$$

With these substitutions our relation becomes

$$\frac{\xi^2}{\frac{1}{2}(2\lambda^2 + 1 + \sqrt{4\lambda^2 + 1})} + \frac{\eta^2}{\frac{1}{2}(2\lambda^2 + 1 - \sqrt{4\lambda^2 + 1})} = 1.$$

which is the standard form of an ellipse (in ξ, η coordinates) with major semi-axis horizontal and having length

$$a = \sqrt{\frac{1}{2}(2\lambda^2 + 1 + \sqrt{4\lambda^2 + 1})}$$

and minor semi-axis vertical and having length

$$b = \sqrt{\frac{1}{2}(2\lambda^2 + 1 - \sqrt{4\lambda^2 + 1})}.$$

Since our change of coordinates was simply a rotation, we see that the image $J(\partial B_1(\mathbf{0}))$ is an ellipse with major semi-axis of length a along the line of inclination $\psi = \psi_1$ given by (41) and minor semi-axis of length b along the line with inclination $\psi_1 + \pi/2$.

Exercise 54 *Show that the expression*

$$2\lambda^2 + 1 - \sqrt{4\lambda^2 + 1}$$

determining the length of the minor semi-axis of the elliptical image of $\partial B_1(\mathbf{0})$ is actually a positive number.

Exercise 55 *Show that the relation (40) can be expressed as*

$$\left\langle B^T B \begin{pmatrix} u \\ v \end{pmatrix}, \begin{pmatrix} u \\ v \end{pmatrix} \right\rangle = 1$$

where B is the matrix of J^{-1} with respect to the standard basis and the inner product here is the Euclidean dot product. Compute the matrix product $B^T B$ and note that you obtain a symmetric matrix. An expression $\langle S v, v\rangle$ (where S is a fixed symmetric transformation) considered as a real valued function on a vector space V containing the vector v is called a **quadratic form** *on V.*

Exercise 56 *Notice that if $\lambda = 0$, then the Jordan transformation falls back into case (b) of linear functions with one-dimensional kernel. What is $J : \mathbb{R}^2 \to \mathbb{R}^2$ in this case (in terms of projection)?*

Exercise 57 *Make an animation of the image $J(B_1(\mathbf{0}))$ under the Jordan transformation $J : \mathbb{R}^2 \to \mathbb{R}^2$ depending on the eigenvalue λ. Indicate the axes of inclination $\psi_1 = \psi_1(\lambda)$ and $2\psi_1$ as well as the images $J\mathbf{e}_1$ and $J\mathbf{e}_2$.*

Exercise 58 *Adapt the discussion above to apply to the image $\Sigma(B_1(\mathbf{0}))$ of the unit circle under the standard linear shear. What is the eigenvalue for Σ?*

To summarize case (ii) of case (c) in which $L : \mathbb{R}^2 \to \mathbb{R}^2$ is an automorphism for which an initial rotation P (clockwise by angle ϕ determined by $(\cos\phi, \sin\phi)^T = L\mathbf{e}_1/|L\mathbf{e}_1|$) results in a scaling

$$P \circ L = L_0 = (\mu_{12}\cos\phi - \mu_{22}\sin\phi)J$$

of the Jordan transformation, our original transformation

$$L = (\mu_{12}\cos\phi - \mu_{22}\sin\phi)P^{-1} \circ J$$

may be understood as a scaled rotation of Jordan shearing of $\mathbb{R}^2$ with (nonzero) eigenvalue

$$\lambda = \frac{|L\mathbf{e}_1|}{\mu_{12}\cos\phi - \mu_{22}\sin\phi}.$$

The shear angle of this Jordan transformation is

$$\phi = \tan^{-1}\left(\frac{1}{\lambda}\right) = \tan^{-1}\left(\frac{\mu_{12}\cos\phi - \mu_{22}\sin\phi}{|L\mathbf{e}_1|}\right)$$

and the additional rotation determined by the elliptical image of the unit circle is

$$\psi_1 = \frac{1}{2}\tan^{-1}(2\lambda) = \frac{1}{2}\tan^{-1}\frac{2|L\mathbf{e}_1|}{\mu_{12}\cos\phi - \mu_{22}\sin\phi}.$$

It remains to consider automorphisms falling into case (iii) in which we have two distinct eigenvalues.

6.2 anisotropic scaling

7 Additional Topics

7.1 Geometry: Inner Product Spaces

7.2 Vector Spaces of Functions

8 Summary

8.1 orthogonal projection

Let $\mathbf{v}$ be any unit vector in $\mathbb{R}^n$. Then projection $\text{proj}_{\langle \mathbf{v} \rangle^\perp} : \mathbb{R}^n \to \mathbb{R}^n$ onto $\langle \mathbf{v} \rangle^\perp$ is given by

$$\text{projection}_{\langle \mathbf{v} \rangle^\perp}(\mathbf{x}) = \mathbf{x} - (\mathbf{x} \cdot \mathbf{v})\, \mathbf{v}.$$

Note $\langle \mathbf{v} \rangle^\perp$ is the **orthogonal subspace** determined by $\langle \mathbf{v} \rangle$ and is given by

$$\langle \mathbf{v} \rangle^\perp = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{v} = 0\}.$$

If we take $\mathbf{v}$ as any nonzero vector, then $\mathbf{v}/|\mathbf{v}|$ is a unit vector, and

$$\text{projection}_{\langle \mathbf{v} \rangle^\perp}(\mathbf{x}) = \mathbf{x} - \left(\frac{\mathbf{x} \cdot \mathbf{v}}{|\mathbf{v}|^2} \right) \mathbf{v}.$$

If we take the special case $\mathbb{R}^2$, then the vector $\mathbf{v}^\perp$ is well-defined and

$$\mathbf{x} = \left(\frac{\mathbf{x} \cdot \mathbf{v}}{|\mathbf{v}|^2} \right) \mathbf{v} + \left(\frac{\mathbf{x} \cdot \mathbf{v}^\perp}{|\mathbf{v}|^2} \right) \mathbf{v}^\perp$$

for any vector $\mathbf{x} \in \mathbb{R}^2$. Therefore,

$$\text{projection}_{\langle \mathbf{v} \rangle^\perp}(\mathbf{x}) = \left(\frac{\mathbf{x} \cdot \mathbf{v}^\perp}{|\mathbf{v}|^2} \right) \mathbf{v}^\perp.$$

8.2 parallel projection

If we take any vector $\mathbf{v} \subset \mathbb{R}^n$ and an $(n-1)$-dimensional subspace W with $\mathbf{v} \notin W$, then every vector $\mathbf{x} \in \mathbb{R}^n$ has a unique representation as

$$\mathbf{x} = a\mathbf{v} + \mathbf{w} \tag{43}$$

for some scalar a and some $\mathbf{w} \in W$. The proof of this fact is as follows:

First notice that $\operatorname{span}(\{\mathbf{v}\} \cup W)$ is an n-dimensional subspace of $\mathbb{R}^n$. Therefore this subspace is $\mathbb{R}^n$ and every vector $\mathbf{x} \in \mathbb{R}^n$ can be expressed in the form indicated in (43).

Notice that the intersection of $\operatorname{span}(\mathbf{v}) = \langle \mathbf{v} \rangle$ and W is $\{\mathbf{0}\}$. In fact, if $c\mathbf{v} \in W$, then either $c = 0$ or $(1/c)(c\mathbf{v}) = \mathbf{v} \in W$ (which is a contradiction). On the other hand, if

$$\mathbf{x} = a\mathbf{v} + \mathbf{w} = \tilde{a}\mathbf{v} + \tilde{\mathbf{w}}$$

for some scalar $\tilde{a}$ and some $\tilde{\mathbf{w}} \in W$, then

$$(a - \tilde{a})\mathbf{v} = \tilde{\mathbf{w}} - \mathbf{w}$$

so this vector is clearly in the intersection $\langle \mathbf{v} \rangle \cap W$. $\square$

The vector $\mathbf{w}$ in (43) is the **parallel projection** of $\mathbf{x}$ onto the subspace W parallel to $\mathbf{v}$.

In the case where $n = 2$, so that $\mathbf{v} \in \mathbb{R}^2$, the $n - 1 = 1$ dimensional subspace is spanned by a single vector $\mathbf{w}$. Consequently, the parallel projection of $\mathbf{x}$ onto $\langle \mathbf{w} \rangle$ parallel to $\mathbf{v}$ is given by a linear function $L : \mathbb{R}^2 \to \mathbb{R}^2$ with $\dim \ker(L) = 1$ and formula

$$L\mathbf{x} = b\mathbf{w} \qquad \text{where} \qquad \mathbf{x} = a\mathbf{v} + b\mathbf{w}.$$

If we assume also that $|\mathbf{v}| = |\mathbf{w}| = 1$, then we obtain the system of equations

$$\begin{cases} \phantom{\mathbf{v}\cdot\mathbf{w}\,} a & + & \mathbf{v}\cdot\mathbf{w}\, b & = & \mathbf{x}\cdot\mathbf{v} \\ \mathbf{v}\cdot\mathbf{w}\, a & + & b & = & \mathbf{x}\cdot\mathbf{w} \end{cases}$$

for a and b.

Exercise 59 *Use Cramer's rule to find formulas for a and b to conclude*

$$L\mathbf{x} = \mathbf{x} \cdot \frac{\mathbf{w} - (\mathbf{v}\cdot\mathbf{w})\,\mathbf{v}}{1 - (\mathbf{v}\cdot\mathbf{w})^2}\,\mathbf{w}.$$

If we compute $|\mathbf{w} - (\mathbf{v}\cdot\mathbf{w})\,\mathbf{v}|^2$, we find the value to be

$$|\mathbf{w} - (\mathbf{v}\cdot\mathbf{w})\,\mathbf{v}|^2 = 1 - (\mathbf{v}\cdot\mathbf{w})^2.$$

Therefore, the unit vector

$$\mathbf{u} = \frac{\mathbf{w} - (\mathbf{v}\cdot\mathbf{w})\,\mathbf{v}}{\sqrt{1 - (\mathbf{v}\cdot\mathbf{w})^2}} = \pm\mathbf{v}^{\perp}$$

and

$$
\begin{aligned}
L\mathbf{x} &= \operatorname*{comp}_{\mathbf{u}} \mathbf{x} \, \frac{1}{\sqrt{1 - (\mathbf{v} \cdot \mathbf{w})^2}} \, \mathbf{w} \\
&= (\operatorname*{proj}_{\mathbf{u}} \mathbf{x} \cdot \mathbf{u}) \, \alpha \, \mathbf{w}
\end{aligned}
$$

where the scaling factor α is given by

$$\alpha = \frac{1}{\sqrt{1 - (\mathbf{v} \cdot \mathbf{w})^2}}.$$

Thus, as discussed above, parallel projection in $\mathbb{R}^2$ is a scaling of orthogonal projection.

Exercise 60 *When is* $\mathbf{u} = \mathbf{v}^\perp$*?*

Exercise 61 *Review/evaluate the assertion that when* $L : \mathbb{R}^2 \to \mathbb{R}^2$ *is linear and satisfies* $\dim \ker(L) = 1$*, then L is (essentially) a scaling of a projection.*

Exercise 62 *Describe the eigenvalues associated with an orthogonal projection and a parallel projection.*

8.3 scaling (isotropic)

If $L : \mathbb{R}^n \to \mathbb{R}^n$ by $L\mathbf{x} = \alpha\mathbf{x}$, then either $\dim \ker(L) = 0$ (when $\alpha \neq 0$) or $\dim \ker(L) = n$. If $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is any basis, then the matrix of L with respect to this basis is

$$\alpha I$$

where I is the $n \times n$ identity matrix. Notice that

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{\mathcal{B}} \mapsto \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{\mathcal{B}} = \alpha \mathbf{v}_1$$

is the first column of the matrix (and the other columns are obtained similarly).

For an isotropic scaling every vector is an eigenvector.

8.4 scaling (anisotropic)