4 Linear Functions

4.1 Definition and examples

Definition 4.1. Let U and V be vector spaces, and let $L : U \to V$ be a function with domain U and codomain V. We say L is a *linear transformation* if:

1. Whenever $\mathbf{u}_1, \mathbf{u}_2 \in U$, then $L(\mathbf{u}_1 + \mathbf{u}_2) = L(\mathbf{u}_1) + L(\mathbf{u}_2)$.

2. Whenever $\mathbf{u} \in U$ and $r \in \mathbb{R}$, then $L(r\mathbf{u}) = rL(\mathbf{u})$.

Example 4.2. If A is a $m \times n$ matrix, then A gives us a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$, given by $A(\mathbf{x}) = A\mathbf{x}$. That is, our input is a (column) vector in $\mathbb{R}^n$, and our output is the vector in $\mathbb{R}^m$ we get by multiplying our column vector by our matrix.

Geometrically, a linear transformation can stretch, rotate, and reflect, but it cannot bend or shift.

Example 4.3. Consider the function from $\mathbb{R}^2$ to $\mathbb{R}^2$ given by a rotation of ninety degrees counterclockwise. We can see by drawing pictures that the sum of two rotated vectors is the rotation of the sum of the vectors, and that the rotation of a streched vector is the same as the strech of a rotated vector. So this is a linear transformation.

Example 4.4. A *translation* is a function defined by $f(\mathbf{x}) = \mathbf{x} + \mathbf{u}$ for some fixed vector $\mathbf{u}$. (Geometrically, it corresponds to sliding or translating your input in the direction and distance of the vector $\mathbf{u}$).

This is *not* a linear transformation. For instance, $f(r\mathbf{x}) = r\mathbf{x} + \mathbf{u} \neq r(\mathbf{x} + \mathbf{u}) = rf(\mathbf{x})$ unless $\mathbf{u} = \mathbf{0}$.

Example 4.5. The function $f(x) = x^2$ is not a linear transformation from $\mathbb{R}$ to $\mathbb{R}$, since $f(2x) = (2x)^2 = 4x^2 \neq 2x^2 = 2f(x)$.

Example 4.6. Define a function $L : \mathbb{R}^3 \to \mathbb{R}^2$ by $L(x, y, z) = (x + y, 2z - x)$. We check:

$$\begin{aligned}
L((x_1, y_1, z_1) + (x_2, y_2, z_2)) &= L(x_1 + x_2, y_1 + y_2, z_1 + z_2) \\
&= (x_1 + x_2 + y_1 + y_2, 2z_1 + 2z_2 - x_1 - x_2) \\
&= (x_1 + y_1, 2z_1 - x_1) + (x_2 + y_2, 2z_2 - x_2) \\
&= L(x_1, y_1, z_1) + L(x_2, y_2, z_2). \\
L(r(x, y, z)) &= L(rx, ry, rz) = (rx + ry, 2rz - rx) = \\
&= r(x + y, 2z - x) = rL(x, y, z).
\end{aligned}$$

Thus L is a linear transformation by definition.

Definition 4.7. Let $L : U \to V$ be a linear transformation. If $\mathbf{u} \in U$ is a vector, we say the element $L(\mathbf{u}) \in V$ is the *image* of $\mathbf{u}$.

If $S \subset U$ then we define the image of S to be the set $L(S) = \{L(\mathbf{u}) : \mathbf{u} \in S\}$ to be the set of images of elements of S. We say the image of the entire set U is the *image* of the function L.

The *kernel* of L is the set $\ker(L) = \{\mathbf{u} \in U : L(\mathbf{u}) = \mathbf{0}\}$ of elements of U whose image is the zero vector.

Another way of thinking about linear transformations is that they send lines to lines. In particular, the image of a subspace under a linear transformation is always a subspace—thus the image of a line will be either a point or a line.

Proposition 4.8. *Let $L : U \to V$ be a linear transformation, and let $S \subseteq U$ be a subspace of U. Then:*

1. $\ker(L)$ *is a subspace of U.*
2. *The image $L(S)$ of S is a subspace of V.*

Proof. 1. See homework 6.

2. We use the subspace theorem:

   (a) We wish to show that $\mathbf{0} \in L(S)$. We claim in particular that $L(\mathbf{0}) = \mathbf{0}$: that is, the image of the zero vector in U must be the zero vector in V. Recall that $0 \cdot \mathbf{v} = \mathbf{0}$ for any $\mathbf{v} \in V$, so we have

   $$L(\mathbf{0}) = L(0 \cdot \mathbf{0}) = 0L(\mathbf{0}) = \mathbf{0}.$$

   Thus since S is a subspace we have $\mathbf{0} \in S$ and thus $\mathbf{0} \in L(S)$.

   (b) Suppose $\mathbf{v} \in L(S)$ and $r \in \mathbb{R}$. Then there is some $\mathbf{u} \in S$ with $L(\mathbf{u}) = \mathbf{v}$, and since S is a subspace we know that $r\mathbf{u} \in S$. Thus

   $$r\mathbf{v} = rL(\mathbf{u}) = L(r\mathbf{u}) \in L(S).$$

   (c) Suppose $\mathbf{v}_1, \mathbf{v}_2 \in L(S)$. Then there exist $\mathbf{u}_1, \mathbf{u}_2 \in S$ such that $L(\mathbf{u}_1) = \mathbf{v}_1$ and $L(\mathbf{u}_2) = \mathbf{v}_2$. Since S is a subspace we know that $\mathbf{u}_1 + \mathbf{u}_2 \in S$. Then

   $$\mathbf{v}_1 + \mathbf{v}_2 = L(\mathbf{u}_1) + L(\mathbf{u}_2) = L(\mathbf{u}_1 + \mathbf{u}_2) \in L(S).$$

□

Corollary 4.9. *If $L : U \to V$ is a linear transformation, then the image of L is a subspace of V.*

Example 4.10. In our geometric example of a ninety degree counterclockwise rotation, the kernel is just the origin—nothing gets mapped to the origin except the origin. The image is the entire plane.

Example 4.11. If A is a matrix, then the linear transformation of A has a kernel precisely equal to the nullspace of A, since the nullspace is the set of $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$.

A has an image precisely equal to the column space of A, since we know by proposition 3.40 the column space of A is precisely the set of $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ has a solution.

In particular, we see that the rank-nullity theorem implies that the dimension of the kernel plus the dimension of the image is the dimension of the domain. (You can think of this as saying every dimension of the domain either gets killed, or gives you a dimension of image).

Example 4.12. Let $\mathcal{D}([a, b], \mathbb{R})$ be the space of differentiable functions from the closed interval $[a, b]$ to the real line. Define the derivative operator $D : \mathcal{D}([a, b], \mathbb{R}) \to \mathcal{D}([a, b], \mathbb{R})$ by $D(f) = f'$. First we claim that D is a linear operator: we have that $D(f+g) = (f+g)' = f' + g' = D(f) + D(g)$, and $D(rf) = (rf)' + rf' = rD(f)$.

The kernel of D is the space of constant functions, which is a one-dimensional subspace. The image of D is actually a little hard to see, but it's actually the set of all continuous functions on $[a, b]$.

In other contexts we might write $\frac{d}{dx}$ instead of D for this linear transformation.

Example 4.13. Let $\mathcal{C}([a, b], \mathbb{R})$ be the set of all continuous functions on the closed interval $[a, b]$. The (indefinite) integral isn't quite a linear transformation, since there's an ambiguity in choice of constant. (This is what we mean when we say something is "not well defined": if I tell you to give me the integral of x^2, you can't give me a specific function back so my question is not precise enough).

But the function $I(f) = \int_a^x f(t)\,dt$ is a linear transformation, since $\int_a^x (f+g)(t)\,dt = \int_a^x f(t)\,dt + \int_a^x g(t)\,dt$ and $\int_a^x rf(t)\,dt = r\int_a^x f(t)\,dt$. In this case the choice of a as the basepoint resolves the earlier ambiguity.

The kernel of I is the trivial vector space containing only the zero function. The image is again a bit hard to see, but works out to be the space of differentiable functions with the property that $F(a) = 0$.

This last examples shows an important principle: our derivative and integral linear transformations (almost) undo each other. This is a very important property and we will look at it on its own in 5.1.

4.2 The Matrix of a Linear Transformation

Some linear transformations are easy to represent, because they come from matrices. In this subsection we will see that in fact *all* linear transformations (of finite-dimensional vector spaces) come from matrices, and see how we can obtain these matrices.

In essence, we can represent a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ with a matrix because we have a system of coordinates for $\mathbb{R}^n$ and $\mathbb{R}^m$; the matrix tells us what happens to each coordinate.

Example 4.14. Let $A = \begin{bmatrix} 3 & 5 & 1 \\ 2 & -1 & 3 \end{bmatrix}$ be a matrix, and thus a linear transformation $\mathbb{R}^3 \to \mathbb{R}^2$. Let's see what happens to each element of the standard basis for $\mathbb{R}^3$.

$$A\mathbf{e}_1 = \begin{bmatrix} 3 & 5 & 1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$A\mathbf{e}_1 = \begin{bmatrix} 3 & 5 & 1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$$

$$A\mathbf{e}_1 = \begin{bmatrix} 3 & 5 & 1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

We notice that the image of the standard basis elements are just the columns of the matrix! This isn't a coincidence; the columns of our matrix are telling us exactly where our basis vectors go.

Remark 4.15. This sort of argument is another way to see that the column space of a matrix is the set $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ has a solution.

Proposition 4.16. *Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there is an $m \times n$ matrix A such that $L(\mathbf{x}) = A\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}^n$.*

In particular, the jth column vector of A is given by $\mathbf{c}_j = L(\mathbf{e}_j)$.

Proof. According to the theorem statement, we know that $A = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_n \end{bmatrix}$. So we just need to check that this matrix gives us the linear transformation L.

First we show that our matrix does the right things on the standard basis vectors. We see that

$$A\mathbf{e}_j = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_j & \dots & \mathbf{c}_n \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{c}_j = L(\mathbf{e}_j).$$

Now let $\mathbf{u} \in \mathbb{R}^n$ be any vector. Then we know we can write $\mathbf{u} = \sum_{i=1}^n u_i \mathbf{e}_i$ since every element is some linear combination of basis vectors. Thus we have

$$\begin{aligned} A\mathbf{u} = A\left(\sum_{i=1}^n u_i \mathbf{e}_i\right) = \sum_{i=1}^n A u_i \mathbf{e}_i = \sum_{i=1}^n u_i A\mathbf{e}_i = \sum_{i=1}^n u_i L(\mathbf{e}_i) & \quad \text{by the previous computation} \\ = \sum_{i=1}^n L(u_i \mathbf{e}_i) & \quad \text{scalar multiplication} \\ = L\left(\sum_{i=1}^n u_i \mathbf{e}_i\right) & \quad \text{additivity} \\ = L(\mathbf{u}). & \end{aligned}$$

□

Example 4.17. Let's look at the linear transformation from earlier, of a 90 degree rotation counterclockwise. This is a transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$, so we can find a 2×2 matrix representing it. Let's call the map $R_{\pi/2}$.

By geometry, we see that $R_{\pi/2}(\mathbf{e}_1) = (0, 1) = \mathbf{e}_2$, and that $R_{\pi/2}(\mathbf{e}_2) = (-1, 0) = -\mathbf{e}_1$. Thus the matrix is $\begin{bmatrix} 0 & -1 \\ 1 & 0. \end{bmatrix}$

Let's generalize to any rotation; let R_θ be the rotation counterclockwise by θ. To see what happens we have to draw the unit circle; we compute that $R_\theta(\mathbf{e}_1) = (\cos\theta, \sin\theta)$, and $R_\theta(\mathbf{e}_2) = (\cos(\theta + \pi/2), \sin(\theta + \pi/2) = (-\sin(\theta), \cos(\theta)$. Thus the matrix of R_θ is $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

Example 4.18. Define a linear transformation $L : \mathbb{R}^2 \to \mathbb{R}^3$ by $L(x, y) = (x + y, x - y, 2x)$. First we should check that this is in fact a linear transformation, but I won't do that here.

We need to check the image of $\mathbf{e}_1$ and $\mathbf{e}_2$. We see that

$$L(\mathbf{e}_1) = L(1,0) = (1,1,2)$$
$$L(\mathbf{e}_2) = L(0,1) = (1,-1,0).$$

Thus the matrix of L is

$$A_L = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 0 \end{bmatrix}.$$

We can check this by computing

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x+y \\ x-y \\ 2x \end{bmatrix}$$

which is exactly what we should get.

We'd like to be able to do this to any vector space, or at least any finite dimensional one. We need some set of coordinates to let us matricize other linear transformations. Fortunately, we developed those in section 2: a set of coordinates is a basis.

Definition 4.19. If U is a vector space and $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for U, and $\mathbf{u} \in U$, we can write $\mathbf{u} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$. We define the *coordinate vector* of $\mathbf{u}$ with respect to E by

$$[\mathbf{u}]_E = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

The a_i are called the *coordinates* of $\mathbf{u}$ with respect to the basis E.

We here observe that every $\mathbf{u} \in U$ corresponds to exactly one coordinate vector with respect to E, and vice versa. We will discuss this in more detail in 5.1.

Example 4.20. Let $U = \mathcal{P}_3(x)$. Then $E = \{1, x, x^2, x^3\}$ is a basis for U. Also, $F = \{1, 1+x, 1+x^2, 1+x^3\}$ is a basis for U.

Let $f(x) = 1 + 3x + x^2 - x^3 \in U$. Then

$$[f]_E = \begin{bmatrix} 1 \\ 3 \\ 1 \\ -1 \end{bmatrix} \qquad [f]_F = \begin{bmatrix} -2 \\ 3 \\ 1 \\ -1 \end{bmatrix}.$$

These are two different vectors of real numbers, but they represent the *same* element of U, just in different bases.

Example 4.21. Let $U = \mathbb{R}^3$ and let $E = \{(1,0,0),(1,1,0),(1,1,1)\}$. Then if $\mathbf{u} = (1,3,2)$, then

$$[\mathbf{u}]_E = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}.$$

Remark 4.22. If B is the standard basis for $\mathbb{R}^n$, then any time we write a column vector there's an implicit $\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}_B$ that we just don't bother to write down.

Lemma 4.23. *If U is a vector space and $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for U, then the function $[\cdot]_E : U \to \mathbb{R}^n$ which sends $\mathbf{u}$ to $[\mathbf{u}]_E$ is a linear function.*

Proof. Let $\mathbf{u}, \mathbf{v} \in U$ and $r \in \mathbb{R}$. We can write

$$\mathbf{u} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$$
$$\mathbf{v} = b_1\mathbf{e}_1 + \cdots + b_n\mathbf{e}_n.$$

Then

$$[r\mathbf{u}] = [ra_1\mathbf{e}_1 + \cdots + ra_n\mathbf{e}_n] = (ra_1, \dots, ra_n) = r(a_1, \dots, a_n) = r[\mathbf{u}].$$
$$[\mathbf{u}+\mathbf{v}] = [(a_1+b_1)\mathbf{e}_1 + \cdots + (a_n+b_n)\mathbf{e}_n] = (a_1+b_1, \dots, a_n+b_n)$$
$$= (a_1, \dots, a_n) + (b_1, \dots, b_n) = [\mathbf{u}] + [\mathbf{v}].$$

Thus by definition, $[\cdot]_E$ is a linear transformation.

□

Theorem 4.24. *Let U and V be finite-dimensional vector spaces, with $E = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ a basis for U and $F = \{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ a basis for V. Let $L : U \to V$ be a linear transformation.*

Then there is a matrix A that represents L with respect to E and F, such that $L\mathbf{u} = \mathbf{v}$ if and only if $A[\mathbf{u}]_E = [\mathbf{v}]_F$. The columns of A are given by $\mathbf{c}_j = [L(\mathbf{e}_j)]_F$.

Remark 4.25. This looks really complicated, but it really just says that any linear transformation is determined entirely by what it does to the elements of some basis; if you have a basis and you know where your transformation sends each element of that basis, you know what it does to everything in your space.

In particular, if we have coordintes for our vector spaces, we can use a matrix to map one set of coordinates to the other, as if we were working in $\mathbb{R}^n$.

$$\begin{array}{ccc} U & \xrightarrow{L} & V \\ \downarrow{\scriptstyle [\cdot]_E} & & \downarrow{\scriptstyle [\cdot]_F} \\ \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^m \end{array} \qquad \begin{array}{ccc} \mathbf{u} & \xmapsto{L} & L(\mathbf{u}) \\ \downarrow{\scriptstyle [\cdot]_E} & & \downarrow{\scriptstyle [\cdot]_F} \\ [\mathbf{u}]_E & \xmapsto{A} & A[\mathbf{u}]_E = [L(\mathbf{u})]_F \end{array}$$

Proof. We just want to show that $A[\mathbf{u}]_E = [L(\mathbf{u})]_F$ for any $\mathbf{u} \in U$, where

$$A = [\mathbf{c}_1 \dots \mathbf{c}_n] = [[L(\mathbf{e}_1)]_F \dots [L(\mathbf{e}_n)]_F] .$$

Our proof is essentially the same as the proof of Proposition 4.16. Let $\mathbf{u} \in U$. Since E is a basis for U we can write $u = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$. Then we have

$$\begin{aligned} [L(\mathbf{u})]_F &= [a_1 L(\mathbf{e}_1) + \cdots + a_n L(\mathbf{e}_n)]_F = a_1 [L(\mathbf{e}_1)]_F + \cdots + a_n [L(\mathbf{e}_n)]_F \\ &= a_1\mathbf{c}_1 + \cdots + a_n\mathbf{c}_n; \\ A[\mathbf{u}]_E &= A [a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n]_E = A(a_1, \dots, a_n) = [\mathbf{c}_1 \dots \mathbf{c}_n] (a_1, \dots, a_n) \\ &= \mathbf{c}_1 a_1 + \cdots + \mathbf{c}_n a_n. \end{aligned}$$

Thus we have $[L(\mathbf{u})]_F = A [\mathbf{u}]_E$, so the matrix A does in fact represent the linear operator L. ☐

Example 4.26. Let $F = \{(1,1), (-1,1)\}$ be a basis for $\mathbb{R}^2$, and let $L : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $L(x,y,z) = (x - y - z, x + y + z)$. Find a matrix for L with respect to the standard basis in the domain and F in the codomain.

$$\begin{aligned} L(1,0,0) &= (1,1) = \mathbf{f}_1 \\ L(0,1,0) &= (-1,1) = \mathbf{f}_2 \\ L(0,0,1) &= (-1,1) = \mathbf{f}_2 \\ A &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}. \end{aligned}$$

Example 4.27. Let S be the subspace of $\mathcal{C}([a,b], \mathbb{R})$ spanned by $\{e^x, xe^x, x^2e^x\}$, and let D be the differentiation operator on S. Find the matrix of D with respect to $\{e^x, xe^x, x^2e^x\}$.

We compute:

$$\begin{aligned} D(e^x) &= e^x = \mathbf{s}_1 \\ D(xe^x) &= e^x + xe^x = \mathbf{s}_1 + \mathbf{s}_2 \\ D(x^2e^x) &= 2xe^x + x^2e^x = 2\mathbf{s}_2 + \mathbf{s}_3 \\ A &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Example 4.28. Let $E = \{(1,1,0),(1,0,1),(0,1,1)\}$ and $F = \{(1,0,0),(1,1,0),(1,1,1)\}$ be bases for $\mathbb{R}^3$, and define $L(x,y,z) = (x+y+z, 2z, -x+y+z)$. We can check this is a linear transformation.

To find the matrix of L with respect to E and the standard basis, we compute

$$\begin{aligned} L(1,1,0) &= (2,0,0) \\ L(1,0,1) &= (2,2,0) \\ L(0,1,1) &= (2,2,2). \end{aligned}$$

Thus the matrix with respect to E and the standard basis is

$$A = \begin{bmatrix} 2 & 2 & 2 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{bmatrix}.$$

If we want to find the matrix with respect to E and F, we observe that

$$\begin{aligned} L(1,1,0) &= (2,0,0) = 2(1,0,0) = 2\mathbf{f}_1 \\ L(1,0,1) &= (2,2,0) = 2(1,1,0) = 2\mathbf{f}_2 \\ L(0,1,1) &= (2,2,2) = 2(1,1,1) = 2\mathbf{f}_3. \end{aligned}$$

Thus the matrix is

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

We notice that this matrix is really simple; this is a "good" choice of bases for this linear transformation.

In contrast, let's look at the transformation $T(x, y, z) = (x, y, z)$. Then we have

$$\begin{aligned}
T(1,1,0) &= (1,1,0) = (1,1,0) = \mathbf{f}_2 \\
T(1,0,1) &= (1,0,1) = (1,0,0) - (1,1,0) + (1,1,1) = \mathbf{f}_1 - \mathbf{f}_2 + \mathbf{f}_3 \\
T(0,1,1) &= (0,1,1) = -(1,0,0) + (1,1,1) = -\mathbf{f}_1 + \mathbf{f}_3.
\end{aligned}$$

Thus the matrix of T with respect to E and F is

$$\begin{bmatrix} 0 & 1 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Thus this transformation, which is really simple with respect to the standard basis, is much more complicated with respect to these bases.

We'll talk a lot more about this choice of basis idea in section 5.