# Matthew's Great Question

John McCuan

April 4, 2022

Given the construction of a dual map

$$L' : W' \to V' \qquad \text{by} \qquad L'\psi = \psi \circ L$$

associated with a linear map $L : V \to W$:

> *Can you give me an example of a dual map, so I can understand what this construction means?*

This is a really good question, and I suggest we consider three examples associated with the following linear maps.

**Example 1** $L : \mathbb{R}^1 \to \mathbb{R}^2$ by

$$Lx = \begin{pmatrix} 2x \\ x \end{pmatrix}.$$

**Example 2** $L : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$L\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 \\ 3x_2 \\ x_1 + 5x_2 \end{pmatrix}.$$

**Example 3** $L : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$L\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 6x_1 + 10x_2 \\ 21x_1 + 35x_2 \\ 3x_1 + 5x_2 \end{pmatrix}.$$

Before I consider the dual maps associated with each of these examples of linear functions I would like to make some preliminary comments and observations intended to render the details of the examples easier to appreciate.

First of all, it will be noted that all three examples fall into the class of examples of linear maps

$$L : \mathbb{R}^n \to \mathbb{R}^m$$

with $m > n$. On the one hand, this will serve to make our considerations more concrete and (in some sense) easier to understand. There are two significant structures that our choice of the finite dimensional real Euclidean spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ as domain and codomain introduces. Both are quite familiar:

**(i)** The space $\mathbb{R}^n$ has a natural **standard basis**

$$\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$$

consisting of the (column) vectors $\mathbf{e}_j \in \mathbb{R}^n$ with zero in every entry and a 1 in the $j$-th entry. And the similar thing holds for $\mathbb{R}^m$. If we need to distinguish between $\mathbf{e}_j \in \mathbb{R}^n$ and $\mathbf{e}_j \in \mathbb{R}^m$, for example if $j = 2$, $n = 3$, and $m = 4$ we might wish to distinguish between

$$\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \in \mathbb{R}^3 \qquad \text{and} \qquad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \in \mathbb{R}^4,$$

we can use the cumbersome superscript notation

$$\mathbf{e}_j^{\mathbb{R}^n} \qquad \text{and} \qquad \mathbf{e}_j^{\mathbb{R}^m}$$

for the former and the latter respectively.

**(ii)** The space $\mathbb{R}^n$ has a natural **inner product** and associated inner product structure.

Both of these features, which are not necessarily present in other vector spaces, are fairly familiar. The presence of standard basis vectors has as it's primary consequence a simplification in the familiar matrix multiplication associated with linear functions on these spaces. The inner product has some details associated with it which may be unfamiliar, and these will be considered in some detail below.

Overall, specific examples in this class of examples allow one to see all the details of the dual map construction in a concrete setting which perhaps clarifies the abstract dual mapping. On the other hand, the specific examples involve (and in a certain sense introduce) a great deal more structure and a good deal more complication along with that structure.

The examples we have chosen all have $m > n$. The reason for this is primarily to illustrate the relation of the dual map construction with the effort to solve (or otherwise analyze) an equation

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

where $\mathbf{b}$ is a given vector in $\mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^n$ is considered unknown. This is of particular interest in the case $\mathbf{b} \notin \mathrm{Im}(L)$ so that the equation is not actually solvable, but the familiar **least squares approximation** procedure can be considered.

# 1 Big Picture Part I

We have a linear function $L : V \to W$ and we would like to understand $L$. The subspaces $\mathcal{N}(L)$ in $V$ and $\mathrm{Im}(L)$ in $W$ are a good place to start. We know

$$L\Big|_{\mathcal{N}(L)} : \mathcal{N}(L) \to \{\mathbf{0}\}.$$

That is, $L$ sends every vector in the subspace $\mathcal{N}(L)$ to the zero vector $\mathbf{0} = \mathbf{0}_W$ in $W$. It may not seem like the function $L$ is doing anything interesting on $\mathcal{N}(L)$, but the identification of the set/space where this uninteresting mapping action takes place… is important and interesting.

The next question might be:

*What is L doing elsewhere (outside of $\mathcal{N}$),*
*and how can we organize/understand that?*

We will assume generally in this discussion that $V$ and $W$ are finite dimensional, though many of the questions we might ask can also be asked when $V$ and/or $W$ are infinite dimensional. In the finite dimensional case, the fundamental (dimension) theorem of linear maps is a useful tool:

$$\dim \mathrm{Im}(L) = \dim V - \dim \mathcal{N}(L).$$

In a rough sense there are $\dim \mathrm{Im}(L)$ "dimensions" of $V$ (or within $V$) where the (nonzero) action of $L$ takes place. Perhaps ideally we would have a subspace $\tilde{V}$ of dimension $k = \dim \operatorname{Im}(L)$ for which

$$V = \mathcal{N}(L) \oplus \tilde{V}$$

and we could say "all the action of $L$ takes place on $\tilde{V}$." In a certain sense

$$L\big|_{\tilde{V}} : \tilde{V} \to \operatorname{Im}(L)$$

isolates the "interesting part of $L$. You may remember that Axler's Exercise 3B12 (Problem 8 of Assignment 5) involved finding a subspace with the same properties as $\tilde{V}$ in this discussion.

A couple comments are in order:

1. We should not expect the subspace $\tilde{V}$ to be unique. There may be many different subspaces $\tilde{V}$ which capture the nonzero action of $L$, or more properly on which exclusively nonzero action of $L$ takes place. In fact, this informal phrasing might not be completely clear either. Note that if we can identify such a subspace $\tilde{V}$ and assuming there is a nontrivial null space $\mathcal{N}(L)$, there are still lots of vectors $z + v \notin \tilde{V}$ with $z \in \mathcal{N}(L)$, $v \in \tilde{V}$, and

   $$L(z + v) = L(v) \neq \mathbf{0}_W.$$

   Thus, there is still a lot of nonzero action off of (or outside of) $\tilde{V}$. But $\tilde{V}$ is somehow representative of the nonzero action of $L$. We would expect

   $$L\big|_{\tilde{V}} : \tilde{V} \to \operatorname{Im}(L)$$

   to be an isomorphism, and indeed it would be since $\dim \tilde{V} = \dim \operatorname{Im}(L)$, and this restriction is clearly onto.

2. Rather than thinking the action of $L$ on $\mathcal{N}(L)$ is "not interesting," it is perhaps better to think of it as just "different." Perhaps then, for the sake of "diversity," let's just say we are thinking to (at least try to) "split up" $V$ into two parts $V = \mathcal{N}(V) \oplus \tilde{V}$ in such a way that $\tilde{V}$ is isomorphic to $\operatorname{Im}(L)$. Notice that $W \setminus \operatorname{Im}(L)$ is really sort of out of the picture. Obviously, there should be some kind of mandated inclusion or affirmative action for $W \setminus \operatorname{Im}(L)$, but until $L$ decides to pay some attention to this neglected and less fortunate part of $W$, it is not clear mathematically what to do.

$$
\begin{tikzcd}
\tilde{V} \arrow[rrr, "L"] \arrow[dr] & & & \mathrm{Im}(L) \arrow[dl] \\
& V \arrow[r, "L"] & W & \\
\mathcal{N}(L) \arrow[ur] \arrow[rrr, "L"] & & & \{\mathbf{0}\} \arrow[ul]
\end{tikzcd}
$$

Figure 1: Splitting up $V$ in order to understand $L$. All the diagonal arrows are injections; the top mapping is an isomorphism.

3. Finally, it may be noted that we are still considering a relatively "course" view of the linear function $L$. The dimension theorem may be viewed as giving even "courser" information; it only tells us dimensions. Here we are attempting to identify a particular subspace on which the (nonzero) action of $L$ takes place. We are not saying anything specific about the nature of that action aside from that it constitutes an isomorphism.

If what we have outlined above is an acceptable approach toward understanding something about the linear function $L$, then we can think of the dual operator $L' : W' \to V'$ as a tool to identify the subspace $\tilde{V}$. Recall that $V$ and $V'$ are isomorphic, and $W$ and $W'$ are isomorphic as well. These identifications depend again on the fact that everything is finite dimensional. In a certain sense we would like to start with $\mathrm{Im}(L) \subset W$ and find an "inverse image" $\tilde{V} \subset V$. This, of course, is not really possible unless $L$ is an injection onto its image. Nevertheless, it turns out, we can think of $L' : W' \to V'$ as a sort of stand in for an inverse. One nice thing about this approach is that all of $W$ is brought back into the game. In particular, $L'$ allows us to define

$$
\begin{tikzcd}
V \arrow[r, "L"] \arrow[d, "\Phi"'] & W \arrow[d, "\Psi"] \\
V' & W' \arrow[l, "L'"']
\end{tikzcd}
$$

Figure 2: Using $L'$ to get back from $W$ into $V$ and identify the subspace $\tilde{V}$.

at least some kind of function $T : W \to V$. Let $\Phi : V \to V'$ and $\Psi : W \to W'$ be isomorphisms. Again these isomorphisms are not unique in general,[1] and we should expect this to correspond to the non-uniqueness of $\tilde{V}$. Nevertheless, it makes perfectly good sense to define

$$T : W \to V \qquad \text{by} \qquad T = \Phi^{-1} \circ L' \circ \Psi.$$

Let us come back to these "big picture" considerations after looking at some examples.

## 2 Examples

Let us start with some general observations about the class of examples $L : \mathbb{R}^n \to \mathbb{R}^m$. With the natural choice of basis (the standard unit basis vectors) there is also a natural choice of isomorphisms $\Phi : \mathbb{R}^n \to (\mathbb{R}^n)'$ and $\Psi : \mathbb{R}^m \to (\mathbb{R}^m)'$. These are equivalent to the correspondence(s)

$$\mathbf{e}_j = \mathbf{e}_j^{\mathbb{R}^n} \longleftrightarrow \phi_j \qquad \text{and} \qquad \mathbf{e}_j = \mathbf{e}_j^{\mathbb{R}^m} \longleftrightarrow \psi_j$$

where $\{\phi_1, \phi_2, \dots, \phi_n\}$ is the **dual basis** for $(\mathbb{R}^n)'$, and $\{\psi_1, \psi_2, \dots, \psi_m\}$ is the **dual basis** for $(\mathbb{R}^m)'$. Since an essentially identical situation prevails for the relation between $\mathbb{R}^n$ and the dual space $(\mathbb{R}^n)'$ as with $\mathbb{R}^m$ and its dual space $(\mathbb{R}^m)'$, except that the names of the dual basis elements are different, we will mention certain things only about $\mathbb{R}^n$ and assume the appropriate comments apply without specific mention to $\mathbb{R}^m$. For example, $\phi_j : \mathbb{R}^n \to \mathbb{R}$ is determined by

$$\phi_j(\mathbf{e}_k) = \delta_{jk} = \begin{cases} 0, & j \neq k \\ 1, & j = k, \end{cases}$$

so that for each $\mathbf{v} = (v_1, v_2, \dots, v_n)^T \in \mathbb{R}^n$ we have[2]

$$\Phi(\mathbf{v}) = \Phi\left(\sum_{j=1}^{n} v_j \mathbf{e}_j\right) = \sum_{j=1}^{n} v_j \phi_j.$$

Also with respect to the standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, there exists an $m \times n$ matrix $A$ with real entries for which

$$L\mathbf{v} = A\mathbf{v}.$$

[1] In particular, the isomorphisms we have constructed between a finite dimensional vectors space and its dual depend on a choice of basis in the vector space.

[2] I am using a superscript "$T$" here to denote the transpose so that $\mathbf{v}$ is a column vector. I am doing this simply to make the typesetting easier.

## 2.1 Least Squares Approximation

Let us recall also the basic outline of **least squares approximation**. We are given a vector $\mathbf{b} \in \mathbb{R}^m$. Since we have assumed $m > n$, it is very likely that $\mathbf{b} \notin \mathrm{Im}(L)$ and the equation

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

has no solution; let us assume this is the case. Then the standard approach to obtaining an "approximate solution" or "least squares (approximate) solution" is to take the transpose $A^T$ of the matrix $A$ which corresponds to a linear function $T : \mathbb{R}^m \to \mathbb{R}^n$ by

$$T\mathbf{w} = A^T\mathbf{w}$$

and consider the alternative equation

$$A^T A\mathbf{x} = T\mathbf{b} = A^t\mathbf{b}. \tag{1}$$

This alternative equation, as it turns out and as we will verify below, is always **consistent** or solveable, though possibly not uniquely. In general $A^T A$ is an $n \times n$ matrix and the following are equivalent:

**(i)** There is a unique solution $\mathbf{x}$ of the equation (1).

**(ii)** The matrix $A^T A$ is invertible.

**(iii)** The matrix $A$ has full rank, in this case $n$.

Of course, we haven't talked about the **rank** of a matrix yet, but Axler's discussion of duality is designed precisely for this purpose.

Assuming $A^T A$ is invertible, then the least squares approximation is given by

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}. \tag{2}$$

In this case, the **orthogonal projection** of $\mathbf{b}$ onto $\mathrm{Im}(L)$ is also given by

$$\mathrm{proj}_{\mathrm{Im}(L)}\mathbf{b} = A(A^T A)^{-1} A^T \mathbf{b} = \tilde{\mathbf{b}}. \tag{3}$$

Thus, we can solve the equation $L\mathbf{x} = \tilde{\mathbf{b}}$, and we call the solution the (least squares) approximate solution of $L\mathbf{x} = \mathbf{b}$. In cases where $A^T A$ is not invertible, there is no really nice formula for the approximate solution nor for the projection. One is reduced to using something like the $QR$ decomposition which is covered later. Of course, in practice you can just use Gaussian elimination or something to find all solutions of (1). In the first two examples mentioned above we have $A^T A$ invertible.

## 2.2 The Matrices in the Examples

**Example 1** To find the matrix corresponding to the linear function $L : \mathbb{R}^1 \to \mathbb{R}^2$ given by $Lx = (2x, x)^T$, we take the image of the standard unit basis vector $\mathbf{e}_1$ for $\mathbb{R}^1$ and put that image in the column. Thus, in matrix notation the (strange but) appropriate way to express the value of $L$ is given by

$$Lx = \begin{pmatrix} 2 \\ 1 \end{pmatrix} x.$$

The image is clearly $\mathrm{Im}(L) = \mathrm{span}\{(2,1)^T\} \subset \mathbb{R}^2$, and the least squares approximation technique proceeds as follows:

$$A^T = (2, 1)$$

so that given $\mathbf{w} = (w_1, w_2)^T \in \mathbb{R}^2$ we have

$$A^T\mathbf{w} = 2w_1 + w_2 \in \mathbb{R}^1.$$

Also,

$$A^T A = 5,$$

so the equation alternative to $Lx = Ax = \mathbf{b}$ is

$$A^T Ax = 5x = \tilde{b} = 2b_1 + b_2.$$

Solving this equation, we get

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b} = \frac{1}{5}(2b_1 + b_2).$$

And finally, we get the projection

$$\mathrm{proj}_{\mathrm{Im}(L)}\mathbf{b} = A(A^T A)^{-1} A^T \mathbf{b} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \frac{1}{5}(2b_1 + b_2) = \frac{2b_1 + b_2}{5} \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

It may be recalled that I had mentioned that the dual map, which we have not yet considered, is related to the transpose of $A$ and the projection operator. It may be noticed in this example that $A^T\mathbf{b} = \tilde{\mathbf{b}} = 2b_1 + b_2$ gives the dot product of $\mathbf{b}$ with the vector $(2,1)^T$ which is a piece of information which can be used to find the projection though this particular scalar is not the scalar of the column in $A$ giving the projection directly. Something more complicated is going on, which we will discuss below.

**Example 2** The same details and computations for the second example run as follows: The matrix of $L$ is given in

$$L\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

The image is the plane spanned by the (linearly independent) columns of the matrix

$$\mathrm{Im}(L) = \left\{ v_1 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 3 \\ 5 \end{pmatrix} : v_1, v_2 \in \mathbb{R} \right\}.$$

The least squares approximation technique proceeds as follows:

$$A^T = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 3 & 5 \end{pmatrix}$$

so that given $\mathbf{w} = (w_1, w_2, w_3)^T \in \mathbb{R}^3$ we have

$$A^T\mathbf{w} = \begin{pmatrix} 2w_1 + w_3 \\ 3w_2 + 5w_3 \end{pmatrix} \in \mathbb{R}^2.$$

This is the matrix of inner products of $\mathbf{w}$ with the columns of $A$, which may be viewed as crucial information for finding the projection of $\mathbf{w}$ onto $\mathrm{Im}(L)$.

$$A^T A = \begin{pmatrix} 5 & 5 \\ 5 & 34 \end{pmatrix},$$

so the equation alternative to $Lx = Ax = \mathbf{b}$ is

$$A^T A\mathbf{x} = \begin{pmatrix} 5 & 5 \\ 5 & 34 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \tilde{\mathbf{b}} = \begin{pmatrix} 2b_1 + b_3 \\ 3b_2 + 5b_3 \end{pmatrix}.$$

Solving this equation, we get

$$\mathbf{x} = (A^T A)^{-1} A^T\mathbf{b} = \frac{1}{145}\begin{pmatrix} 34 & -5 \\ -5 & 5 \end{pmatrix}\begin{pmatrix} 2b_1 + b_3 \\ 3b_2 + 5b_3 \end{pmatrix}.$$

The projection is given by

$$\begin{aligned}
\mathrm{proj}_{\mathrm{Im}(L)}\mathbf{b} &= A(A^TA)^{-1}A^T\mathbf{b} \\
&= \begin{pmatrix} 2 & 0 \\ 0 & 3 \\ 1 & 5 \end{pmatrix} \left[ \frac{1}{145} \begin{pmatrix} 34 & -5 \\ -5 & 5 \end{pmatrix} \begin{pmatrix} 2b_1 + b_3 \\ 3b_2 + 5b_3 \end{pmatrix} \right] \\
&= \frac{1}{145} \begin{pmatrix} 2(68b_1 - 15b_2 + 9b_2 \\ -6(2b_1 - 3b_2 - 4b_2) \\ 18b_1 + 60b_2 + 109b_3 \end{pmatrix}.
\end{aligned}$$

**Example 3** For

$$A = \begin{pmatrix} 6 & 10 \\ 21 & 35 \\ 3 & 5 \end{pmatrix},$$

the image is the line spanned by $\mathbf{u} = (2, 7, 1)^T$:

$$\mathrm{Im}(L) = \left\{ t \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Notice the first column of $A$ is $3\mathbf{u}$ and the second column of $A$ is $5\mathbf{u}$, so we can also write

$$\mathrm{Im}(L) = \left\{ 3v_1 \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} + 5v_2 \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} : v_1, v_2 \in \mathbb{R} \right\}.$$

The least squares approximation technique proceeds as follows:

$$A^T = \begin{pmatrix} 6 & 21 & 3 \\ 10 & 35 & 5 \end{pmatrix}$$

so that given $\mathbf{w} = (w_1, w_2, w_3)^T \in \mathbb{R}^3$ we have

$$A^T\mathbf{w} = \begin{pmatrix} 6w_1 + 21w_2 + 3w_3 \\ 10w_1 + 35w_2 + 5w_3 \end{pmatrix} \in \mathbb{R}^2.$$

Again, this is the matrix of inner products of $\mathbf{w}$ with the columns of $A$.

$$A^TA = \frac{1}{54} \begin{pmatrix} 9 & 15 \\ 15 & 25 \end{pmatrix},$$

so the equation alternative to $Lx = Ax = \mathbf{b}$ is

$$A^T A\mathbf{x} = \frac{1}{54}\begin{pmatrix} 9 & 15 \\ 15 & 25 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \tilde{b} = \begin{pmatrix} 3(2b_1 + 7b_2 + b_3) \\ 5(2b_1 + 7b_2 + b_3) \end{pmatrix}.$$

As expected the columns

$$\frac{1}{18}\begin{pmatrix} 3 \\ 5 \end{pmatrix} \qquad \text{and} \qquad \frac{5}{54}\begin{pmatrix} 3 \\ 5 \end{pmatrix}$$

of $A^T A$ form a linearly dependent set in $\mathbb{R}^2$; the matrix $A^T A$ has no inverse. Nevertheless, we see the alternative equation can be written as

$$\frac{1}{54}\begin{pmatrix} 9 & 15 \\ 15 & 25 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{x_1}{18}\begin{pmatrix} 3 \\ 5 \end{pmatrix} + \frac{5x_2}{54}\begin{pmatrix} 3 \\ 5 \end{pmatrix} = (2b_1 + 7b_2 + b_3)\begin{pmatrix} 3 \\ 5 \end{pmatrix}.$$

Therefore, any $\mathbf{x} = (x_1, x_2)^T \in \mathbb{R}^2$ for which

$$3x_1 + 5x_2 = 54(2b_1 + 7b_2 + b_3) \tag{4}$$

is a solution.

Note that the image of the function $T : \mathbb{R}^3 \to \mathbb{R}^2$ by $T(\mathbf{w}) = A^T\mathbf{w}$ is a particular line

$$\mathrm{Im}(T) = \mathrm{span}\left\{\begin{pmatrix} 3 \\ 5 \end{pmatrix}\right\} = \left\{t\begin{pmatrix} 3 \\ 5 \end{pmatrix} : t \in \mathbb{R}\right\}$$

in $\mathbb{R}^2$. As $\mathbf{b}$ ranges over $\mathbb{R}^3$ the values $T\mathbf{b}$ range over this line. On the other hand, given a particular $\mathbf{b} \in \mathbb{R}^3$, there is a unique solution $\mathbf{x} \in \mathbb{R}^2$ of (4) satisfying

$$\mathbf{x} = t\begin{pmatrix} 3 \\ 5 \end{pmatrix} \in \mathrm{Im}(T).$$

Substituting this form of $\mathbf{x}$ into (4) we find

$$34t = 54(2b_1 + 7b_2 + b_3) \qquad \text{or} \qquad t = \frac{27}{17}(2b_1 + 7b_2 + b_3),$$

and the solution

$$\mathbf{x} = \frac{27}{17}(2b_1 + 7b_2 + b_3)\begin{pmatrix} 3 \\ 5 \end{pmatrix}$$

is the unique solution $\mathbf{x}$ for which

$$\begin{aligned}
A\mathbf{x} &= \frac{27}{17}(2b_1 + 7b_2 + b_3)\begin{pmatrix} 6 & 10 \\ 21 & 35 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} 3 \\ 5 \end{pmatrix} \\
&= 54(2b_1 + 7b_2 + b_3)\begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} \\
&= \mathrm{proj}_{\mathrm{Im}(L)}\mathbf{b}.
\end{aligned}$$

## 2.3 Inner Products

Axler does not discuss inner products until Chapter 6, but we need to know about them now. Actually, the inner product on a Euclidean space $\mathbb{R}^n$ (or $\mathbb{R}^m$) is familiar as the standard **dot product**:

$$\mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{j=1}^{n} x_j y_j.$$

The existence of such a function, which abstractly means a **symmetric positive definite bilinear form**

$$\langle \,\cdot\, , \,\cdot\, \rangle : V \times V \to \mathbb{R}$$

on a vector space $V$, allows one to measure **angles** and **orthogonality** in particular. This has several important consequences among which are the following:

**(i)** Given a subspace $U$ in a vector space $V$ with an inner product the **orthogonal complement**

$$U^{\perp} = \{v \in V : \langle v, u \rangle = 0 \text{ for every } u \in U\}$$

is always a well-defined subspace.

**ii** Given an inner product, one can always define a **norm** by

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

**(iii)** As a consequence of the fact that a norm is positive definite,[3] meaning $\langle v, v \rangle = 0$ implies $v = \mathbf{0}$, it is always the case that **orthogonal subspaces**, by which we

[3] We will discuss the abstract definitions of inner product and norm below and their relations.

mean subspaces $U$ and $S$ in $V$ for which $\langle u, s\rangle = 0$ for every $u \in U$ and $s \in S$, intersect in the zero subspace. This holds for orthogonal complements in particular:

$$U \cap U^{\perp} = \{\mathbf{0}\}.$$

As a consequence of this property the sum of orthogonal subspaces is always a direct sum.

**(iv)** Given a proper subspace $U$ of a vector space $V$ with an inner product, one can always write $V = U \oplus U^{\perp}$ and for each $v \in V$, there exist unique vectors $x \in U$ and $y \in U^{\perp}$ for which $v = x + y$.

I think we are in a position to establish the main property of the function $T : \mathbb{R}^m \to \mathbb{R}^n$ by

$$T\mathbf{w} = \Phi^{-1} \circ L' \circ \Psi \mathbf{w}$$

mentioned above.

**Theorem 1** *Given any vectors* $\mathbf{v} \in \mathbb{R}^n$ *and* $\mathbf{w} \in \mathbb{R}^m$ *and using the respective dot products* $\langle\ \cdot\ ,\ \cdot\ \rangle_{\mathbb{R}^n}$ *and* $\langle\ \cdot\ ,\ \cdot\ \rangle_{\mathbb{R}^n}$ *we have*

$$\langle T\mathbf{w}, \mathbf{v}\rangle_{\mathbb{R}^n} = \langle \mathbf{w}, L\mathbf{v}\rangle_{\mathbb{R}^m}.$$

*This corresponds to the familiar identity* $A^T\mathbf{w} \cdot \mathbf{v} = \mathbf{w} \cdot A\mathbf{v}$.

Proof:

$$\begin{aligned}
\langle T\mathbf{w}, \mathbf{v}\rangle_{\mathbb{R}^n} &= \left\langle \Phi^{-1} \circ L' \circ \Psi \left( \sum_{j=1}^{m} w_j \mathbf{e}_j^{\mathbb{R}^m} \right), \mathbf{v} \right\rangle_{\mathbb{R}^n} \\
&= \left\langle \Phi^{-1} \circ L' \left( \sum_{j=1}^{m} w_j \psi_j \right), \mathbf{v} \right\rangle_{\mathbb{R}^n} \\
&= \left\langle \Phi^{-1} \left( \sum_{j=1}^{m} w_j \psi_j \circ L \right), \mathbf{v} \right\rangle_{\mathbb{R}^n} \\
&= \left\langle \Phi^{-1} \left( \sum_{j=1}^{m} \sum_{\ell=1}^{n} w_j \psi_j \circ L(\mathbf{e}_\ell^{\mathbb{R}^n}) \phi_\ell \right), \sum_{r=1}^{n} v_r \mathbf{e}_r^{\mathbb{R}^n} \right\rangle_{\mathbb{R}^n} \\
&= \left\langle \sum_{j=1}^{m} w_j \sum_{\ell=1}^{n} \psi_j \circ L(\mathbf{e}_\ell^{\mathbb{R}^n}) \mathbf{e}_\ell^{\mathbb{R}^n}, \sum_{r=1}^{n} v_r \mathbf{e}_r^{\mathbb{R}^n} \right\rangle_{\mathbb{R}^n} \\
&= \sum_{j=1}^{m} w_j \sum_{\ell=1}^{n} \sum_{r=1}^{n} v_r \psi_j \circ L(\mathbf{e}_\ell^{\mathbb{R}^n}) \left\langle \mathbf{e}_\ell^{\mathbb{R}^n}, \mathbf{e}_r^{\mathbb{R}^n} \right\rangle_{\mathbb{R}^n} \\
&= \sum_{j=1}^{m} w_j \sum_{\ell=1}^{n} v_\ell \psi_j \circ L(\mathbf{e}_\ell^{\mathbb{R}^n}) \\
&= \sum_{j=1}^{m} w_j \sum_{\ell=1}^{n} v_\ell \psi_j \left( \sum_{r=1}^{m} \left\langle L(\mathbf{e}_\ell^{\mathbb{R}^n}), \mathbf{e}_r^{\mathbb{R}^m} \right\rangle_{\mathbb{R}^m} \mathbf{e}_r^{\mathbb{R}^m} \right) \\
&= \sum_{j=1}^{m} w_j \sum_{\ell=1}^{n} v_\ell \left\langle L(\mathbf{e}_\ell^{\mathbb{R}^n}), \mathbf{e}_j^{\mathbb{R}^m} \right\rangle_{\mathbb{R}^m} \\
&= \left\langle \sum_{j=1}^{m} w_j \mathbf{e}_j^{\mathbb{R}^m}, \sum_{r=1}^{m} \left( \sum_{\ell=1}^{n} v_\ell \left\langle L(\mathbf{e}_\ell^{\mathbb{R}^n}), \mathbf{e}_r^{\mathbb{R}^m} \right\rangle_{\mathbb{R}^m} \right) \mathbf{e}_r^{\mathbb{R}^m} \right\rangle_{\mathbb{R}^m} \\
&= \left\langle \sum_{j=1}^{m} w_j \mathbf{e}_j^{\mathbb{R}^m}, \sum_{\ell=1}^{n} v_\ell \sum_{r=1}^{m} \left\langle L(\mathbf{e}_\ell^{\mathbb{R}^n}), \mathbf{e}_r^{\mathbb{R}^m} \right\rangle_{\mathbb{R}^m} \mathbf{e}_r^{\mathbb{R}^m} \right\rangle_{\mathbb{R}^m} \\
&= \left\langle \mathbf{w}, \sum_{\ell=1}^{n} v_\ell L(\mathbf{e}_\ell^{\mathbb{R}^n}) \right\rangle_{\mathbb{R}^m} \\
&= \langle \mathbf{w}, L\mathbf{v}\rangle_{\mathbb{R}^m}. \qquad \square
\end{aligned}$$

With the crucial result of Theorem 1 we can identify the image of $T$:

**Theorem 2** $\mathrm{Im}(T) = \mathcal{N}(L)^\perp$ *and* $\mathcal{N}(T) = \mathrm{Im}(L)^\perp$.

Proof: Let us consider the second assertion first. If $\mathbf{w} \in \mathcal{N}(T)$, then $T\mathbf{w} = \mathbf{0}_{\mathbb{R}^n}$, so for every $\mathbf{u} \in \mathrm{Im}(L)$, there is some $\mathbf{x} \in \mathbb{R}^n$ with $L\mathbf{x} = \mathbf{u}$ and

$$0 = \langle T\mathbf{w}, \mathbf{x}\rangle_{\mathbb{R}^n} = \langle \mathbf{w}, L\mathbf{x}\rangle_{\mathbb{R}^m} = \langle \mathbf{w}, \mathbf{u}\rangle_{\mathbb{R}^m}.$$

Thus, $\mathcal{N}(T) \subset \mathrm{Im}(L)^\perp$.

To see the reverse inclusion, we assume $\mathbf{w} \in \mathrm{Im}(L)^\perp$. This means for every $\mathbf{x} \in \mathbb{R}^n$ we have

$$0 = \langle L\mathbf{x}, \mathbf{w}\rangle_{\mathbb{R}^m} = \langle \mathbf{x}, T\mathbf{w}\rangle_{\mathbb{R}^n}.$$

In particular, taking $\mathbf{x} = T\mathbf{w}$, we get

$$\|T\mathbf{w}\|^2_{\mathbb{R}^n} = \langle T\mathbf{w}, T\mathbf{w}\rangle_{\mathbb{R}^n} = 0,$$

so $T\mathbf{w} = 0$, and $\mathbf{w} \in \mathcal{N}(T)$. This establishes the second assertion.

If $\mathbf{v} \in \mathrm{Im}(T)$, then $\mathbf{v} = T\mathbf{w}$ for some $\mathbf{w} \in \mathbb{R}^m$. Therefore, for any $\mathbf{z} \in \mathcal{N}(L)$ we have by Theorem 1

$$\langle \mathbf{v}, \mathbf{z}\rangle_{\mathbb{R}^n} = \langle T\mathbf{w}, \mathbf{z}\rangle_{\mathbb{R}^n} = \langle \mathbf{w}, L\mathbf{z}\rangle_{\mathbb{R}^m} = \langle \mathbf{w}, \mathbf{0}_{\mathbb{R}^m}\rangle_{\mathbb{R}^m} = 0.$$

Therefore, $\mathbf{v} \in \mathcal{N}(L)^\perp$.

It is, generally speaking, rather more difficult to show the reverse inclusion. We'll give two proofs. The first is relatively easy, but is strictly appicable to the situation of finite dimensional vector spaces: We know, so far, that $\mathrm{Im}(T) \subset \mathcal{N}(L)^\perp$, that is, the image of $T$ is a subspace of $\mathcal{N}(L)^\perp$. But

$$\begin{aligned}
\dim \mathrm{Im}(T) &= m - \dim \mathcal{N}(T) \\
&= m - \dim \mathrm{Im}(L)^\perp \\
&= m - [m - \dim \mathrm{Im}(L)] \\
&= \dim \mathrm{Im}(L) \\
&= n - \dim \mathcal{N}(L) \\
&= \dim \mathcal{N}(L)^\perp.
\end{aligned}$$

Thus, since $\mathrm{Im}(T)$ is a subspace of $\mathcal{N}(L)^\perp$ with the same dimension as $\mathcal{N}(L)^\perp$, the subspace $\mathrm{Im}(T)$ must be all of the larger space $\mathcal{N}(L)^\perp$. Notice we've used here also the second assertion of the theorem $\mathcal{N}(T) = \mathrm{Im}(L)^\perp$ as well as the fact that the dimension of a subspace $U \subset V$ and the dimension of its orthogonal complement $U^\perp$ add up to the dimension $V$.

Here is a second proof which is rather more difficult but based on techniques which can be applied in some cases when the vector spaces involved are infinite dimensional (inner product spaces). Let $\mathbf{x} \in \mathcal{N}(L)^\perp$. Rather than showing $\mathbf{x} \in \mathrm{Im}(T)$ directly, we will show

$$\mathbf{x} \in \mathrm{Im}(T)^{\perp\perp}$$

the **double orthogonal complement**. We will then show $\mathrm{Im}(T)^{\perp\perp} \subset \mathrm{Im}(T)$.

**Lemma 1** $\mathrm{Im}(T)^\perp \subset \mathcal{N}(L) \subset \mathcal{N}(L)^{\perp\perp}$.

Proof: If $\mathbf{v} \in \mathrm{Im}(T)^\perp$, then for each $\mathbf{w} \in \mathbb{R}^m$ we have

$$0 = \langle \mathbf{v}, T\mathbf{w} \rangle_{\mathbb{R}^n} = \langle L\mathbf{v}, \mathbf{w} \rangle_{\mathbb{R}^m}.$$

In particular, taking $\mathbf{w} = L\mathbf{v}$, we have

$$\|L\mathbf{v}\|^2_{\mathbb{R}^m} = \langle L\mathbf{v}, L\mathbf{v} \rangle_{\mathbb{R}^m} = 0,$$

so $L\mathbf{v} = 0$ and $\mathbf{v} \in \mathcal{N}(L)$.

**Note:** We have shown something stronger than $\mathrm{Im}(T)^\perp \subset \mathcal{N}(L)\mathcal{N}(L)^{\perp\perp}$. We have shown $\mathrm{Im}(T)^\perp \subset \mathcal{N}(L)$, and it is a general fact that any vector subspace $U$ in an **any** inner product space $V$ satisfies $U \subset U^{\perp\perp}$. In fact, if $u \in U$, then

$$\langle u, v \rangle = 0 \qquad \text{for every } v \in U^\perp,$$

but this means precisely that $u \in (U^\perp)^\perp = U^{\perp\perp}$. $\square$

**Note:** It is a general strategy when trying to show an inclusion $U \subset W$ of subspaces $U$ and $V$ in an inner product space $V$ to show $W^\perp \subset U^\perp$ instead. This is the way I phrased Lemma 1, but as noted, we have proved something stronger.

**Corollary 1** $\mathcal{N}(L)^\perp \subset \mathrm{Im}(T)^{\perp\perp}$.

Proof: If $\mathbf{x} \in \mathcal{N}(L)^\perp$, then

$$\langle \mathbf{x}, \mathbf{v} \rangle_{\mathbb{R}^n} = 0 \qquad \text{for every } \mathbf{v} \in \mathcal{N}(L).$$

By Lemma 1 we know $\mathrm{Im}(T)^\perp \subset \mathcal{N}(L)$, so

$$\langle \mathbf{x}, \mathbf{v} \rangle_{\mathbb{R}^n} = 0 \qquad \text{for every } \mathbf{v} \in \mathrm{Im}(T)^\perp.$$

This is precisely what it means to have $\mathbf{x} \in (\mathrm{Im}(T)^\perp)^\perp = \mathrm{Im}(T)^{\perp\perp}$. $\square$

**Lemma 2** $\mathrm{Im}(T)^{\perp\perp} \subset \mathrm{Im}(T)$.

Proof: This is the inclusion that does not always hold. We will again give two proofs. Both proofs use the fact that $\mathbb{R}^n$ and $\mathbb{R}^m$ are finite dimensional, but neither use dimension (or the dimension theorem) directly. The first uses bases: Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a basis for $\mathrm{Im}(T)$. Extend this basis to a basis

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$$

for $\mathbb{R}^n$ with each of the vectors $\mathbf{v}_{k+1}, \mathbf{v}_{k+2}, \ldots, \mathbf{v}_n \in \mathrm{Im}(T)^{\perp}$. It is clear that

$$\mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \subset \mathrm{Im}(T)^{\perp\perp}.$$

In particular, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a linearly independent set in $\mathrm{Im}(T)^{\perp\perp}$. If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is not a spanning set, we can find some

$$\mathbf{w}_{k+1} \in \mathrm{Im}(T)^{\perp\perp} \backslash \, \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}.$$

We can also extend $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{w}_{k+1}\}$ to a basis

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k, \mathbf{w}_{k+1}, \ldots, \mathbf{w}_n\}$$

for $\mathbb{R}^n$. Then we have

$$\mathbf{w}_{k+1} = \sum_{j=1}^{n} a_j \mathbf{v}_j,$$

and

$$\mathbf{w}_{k+1} - \sum_{j=1}^{k+1} a_j \mathbf{v}_j = \sum_{j=k+2}^{n} \mathbf{v}_j = \sum_{j=k+1}^{n} b_j \mathbf{w}_j$$

for some coefficients $a_j$ and $b_j$. It follows that

$$\sum_{j=1}^{k} a_j \mathbf{v}_j + (a_{k+1} + b_{k+1} - 1)\mathbf{w}_{k+1} + \sum_{j=k+2}^{n} b_j \mathbf{w}_j = \mathbf{0}.$$

It follows that $a_1 = a_2 = \cdots = a_k = 0$, $a_{k+1} + b_{k+1} = 1$, and $b_{k+2} = \cdots = b_n = 0$. Returning to the expression for $\mathbf{w}_{k+1}$ in terms of the basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, we see

$$\mathbf{w}_{k+1} = \sum_{j=k+1}^{n} a_j \mathbf{v}_j \in \mathrm{Im}(T)^{\perp}.$$

But then $\mathbf{w}_{k+1} \in \mathrm{Im}(T)^{\perp} \cap \mathrm{Im}(T)^{\perp\perp}$, and consequently,

$$\|\mathbf{w}_{k+1}\|^2 = \langle \mathbf{w}_{k+1}, \mathbf{w}_{k+1} \rangle = 0.$$

This is a contradiction since we assumed $\mathbf{w}_{k+1} \notin \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$.

The second proof uses the direct sum decomposition $\mathbb{R}^n = \mathrm{Im}(T) \oplus \mathrm{Im}(T)^{\perp}$ much more directly. Let $\mathbf{x} \in \mathrm{Im}(T)^{\perp\perp}$, then we can write

$$\mathbf{x} = \mathbf{y} + \mathbf{z} \qquad \text{with } \mathbf{y} \in \mathrm{Im}(T) \text{ and } \mathbf{z} \in \mathrm{Im}(T)^{\perp}.$$

Since $\mathrm{Im}(T) \subset \mathrm{Im}(T)^{\perp\perp}$ and $\mathrm{Im}(T)^{\perp\perp}$ is a subspace, we see that

$$\mathbf{z} = \mathbf{x} - \mathbf{y} \in \mathrm{Im}(T)^{\perp\perp}.$$

But this means $\langle \mathbf{z}, \mathbf{v} \rangle = 0$ for every $\mathbf{v} \in \mathrm{Im}(T)^{\perp}$. In particular, since $\mathbf{z} \in \mathrm{Im}(T)^{\perp}$ we must have

$$\|\mathbf{z}\|^2 = \langle \mathbf{z}, \mathbf{z} \rangle_{\mathbb{R}^n} = 0.$$

Therefore, $\mathbf{z} = \mathbf{0}$ and $\mathbf{x} = \mathbf{y} \in \mathrm{Im}(T)$. $\square$

In summeary we have the following result:

**Corollary 2** $\mathcal{N}(L)^{\perp} \subset \mathrm{Im}(T)$.

This result completes the proof of Theorem 2. To review we include the details of two proofs of Corollary 2: If $\mathbf{x} \in \mathcal{N}(L)^{\perp}$, then

$$\mathbf{x} = \mathbf{y} + \mathbf{z} \qquad \text{for } \mathbf{y} \in \mathrm{Im}(T) \text{ and } \mathbf{z} \in \mathrm{Im}(T)^{\perp}.$$

Then $\mathbf{z} = \mathbf{x} - \mathbf{y} \in \mathrm{Im}(T)^{\perp\perp}$ using the note in the proof of Lemma 1 according to which $\mathbf{y} \in \mathrm{Im}(T) \subset \mathrm{Im}(T)^{\perp\perp}$. This means

$$\langle \mathbf{z}, \mathbf{v} \rangle_{\mathbb{R}^n} = 0 \qquad \text{for every } \mathbf{v} \in \mathrm{Im}(T)^{\perp}.$$

This applies in particular to $\mathbf{v} = \mathbf{z}$, so

$$\|\mathbf{z}\|^2 = \langle \mathbf{z}, \mathbf{z} \rangle = 0.$$

We conclude $\mathbf{x} = \mathbf{y} \in \mathrm{Im}(t)$.

For the second proof we note that

$$\begin{aligned} \mathcal{N}(L)^{\perp} &\subset \mathrm{Im}(T)^{\perp\perp} && \text{by Corollary 1} \\ &\subset \mathrm{Im}(T) && \text{by Lemma 2.} \qquad \square \end{aligned}$$

## 2.4 Examples

**Example 1** The linear function $L : \mathbb{R}^1 \to \mathbb{R}^2$ by $L(x) = (2x, x)$ is an isomorphism onto $\operatorname{span}\{(2,1)^T\}$. The mapping $T$ is given by

$$T \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = (2,1) \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = 2w_1 + w_2,$$

and this function is onto $\mathbb{R}^1$. There is no alternative choice for $\tilde{V}$ in this case since, in particular, $\mathcal{N}(L) = \{0\}$. Restricting $T$ to $\operatorname{Im}(L) = \operatorname{span}\{(2,1)^T\}$, we get an isomorphism and we see the composition

$$TL : \mathbb{R} \to \mathbb{R} \qquad \text{by} \qquad TLx = 5x$$

is an isomorphism. In particular, given any $\mathbf{b} = \beta(2,1)^T \in \operatorname{Im}(L)$, there is a unique $x \in \mathbb{R}$ for which $Lx = \mathbf{b}$, namely, $x = \beta$. This solution may also be obtained as

$$x = (TL)^{-1}T\mathbf{b} = \frac{1}{5}\ (2,1)\ \beta(2,1)^T = \beta.$$

More generally, given any $\mathbf{b} \in \mathbb{R}^2$ we can consider

$$x = (TL)^{-1}T\mathbf{b} = \frac{1}{5}\ (2,1)\ \mathbf{b}.$$

The vector

$$\tilde{\mathbf{b}} = Lx = \frac{2b_1 + b_2}{5} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \operatorname{proj}_{\operatorname{Im}(L)}\mathbf{b}$$

as noted above. The dimension of the domain of $L$ is too small for anything particularly interesting to happen here, and also $\mathcal{N}(L) = \{0\}$, so nothing very interesting is happening, but these are the details. The formula for the projection is somewhat interesting.

**Example 2** The same details and computations for the second example run as follows: The image of $T : \mathbb{R}^3 \to \mathbb{R}^2$ is again the entire domain $\mathbb{R}^2$. Furthermore, the restriction of $T$ to

$$\operatorname{Im}(L) = \left\{ v_1 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 3 \\ 5 \end{pmatrix} : v_1, v_2 \in \mathbb{R} \right\}$$

is an isomorphism as is the composition $TL : \mathbb{R}^2 \to \mathbb{R}^2$. As computed above with different names

$$TL\mathbf{x} = \begin{pmatrix} 5 & 5 \\ 5 & 34 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

and given $\mathbf{b} \in \mathrm{Im}(L)$ there is a unique $\mathbf{x} \in \mathbb{R}^2$ given by

$$\mathbf{x} = (TL)^{-1}T\mathbf{b} = \frac{1}{145} \begin{pmatrix} 34 & -5 \\ -5 & 5 \end{pmatrix} \begin{pmatrix} 2b_1 + b_3 \\ 3b_2 + 5b_3 \end{pmatrix}$$

with $L\mathbf{x} = \mathbf{b}$. More generally, if $\mathbf{b} \in \mathbb{R}^3$ we can consider

$$\mathbf{x} = (TL)^{-1}T\mathbf{b} = \frac{1}{145} \begin{pmatrix} 34 & -5 \\ -5 & 5 \end{pmatrix} \begin{pmatrix} 2b_1 + b_3 \\ 3b_2 + 5b_3 \end{pmatrix}$$

and we will have

$$L\mathbf{x} = \frac{1}{145} \begin{pmatrix} 2(68b_1 - 15b_2 + 9b_2 \\ -6(2b_1 - 3b_2 - 4b_2) \\ 18b_1 + 60b_2 + 109b_3 \end{pmatrix} = \mathrm{proj}_{\mathrm{Im}(L)}\mathbf{b}.$$

**Example 3** In this case the image of $L$ is one dimensional, and it follows that $L$ has a non-trivial null space, so the situation should be more interesting. We recall for reference that in this example

$$L(\mathbf{v}) = (3v_1 + 5v_2) \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix}$$

where $\mathbf{v} = (v_1, v_2)^t \in \mathbb{R}^2$. The image of $T$ is determined as follows: For $\mathbf{w} = (w_1, w_2, w_3)^T \in \mathbb{R}^3$ we have

$$T\mathbf{w} = \begin{pmatrix} 6w_1 + 21w_2 + 3w_3 \\ 10w_1 + 35w_2 + 5w_3 \end{pmatrix} = (2w_1 + 7w_2 + w_3) \begin{pmatrix} 3 \\ 5 \end{pmatrix}.$$

Therefore, $\mathrm{Im}(T) = \mathrm{span}\{(3,5)^T\}$ and

$$\mathcal{N}(L) = \mathrm{Im}(T)^\perp = \mathrm{span}\left\{ \begin{pmatrix} -5 \\ 3 \end{pmatrix} \right\}.$$

It is now possible to choose any one-dimensional subspace $\tilde{V} = \mathrm{span}\{\mathbf{y}\}$ in $\mathbb{R}^2$ with $\mathbf{y} \notin \mathcal{N}(L)$ to obtain a direct sum decomposition $\mathcal{N}(L) \oplus \tilde{V}$ and isomorphisms

$$L\Big|_{\tilde{V}} : \tilde{V} \to \mathrm{Im}(L) \qquad \text{and} \qquad T\Big|_{\mathrm{Im}(L)} : \mathrm{Im}(L) \to \mathrm{Im}(T). \tag{5}$$

The composition $TL : \tilde{V} \to \mathrm{Im}(T)$ is also an isomorphism, and we can find formulas for these mappings as follows:

$$\begin{aligned} L(\alpha\mathbf{y}) &= (3y_1 + 5y_2)\alpha \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} \qquad \text{with} \\ \left(L\Big|_{\tilde{V}}\right)^{-1} \left(t \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix}\right) &= \frac{t}{3y_1 + 5y_2}\mathbf{y}, \end{aligned} \tag{6}$$

$$\begin{aligned} T\Big|_{\mathrm{Im}(L)} \left(t \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix}\right) &= 54t \begin{pmatrix} 3 \\ 5 \end{pmatrix} \qquad \text{with} \\ \left(T\Big|_{\mathrm{Im}(L)}\right)^{-1} \left(\beta \begin{pmatrix} 3 \\ 5 \end{pmatrix}\right) &= \frac{\beta}{54} \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix}, \end{aligned} \tag{7}$$

and

$$\begin{aligned} TL(\alpha\mathbf{y}) &= 54(3y_1 + 5y_2)\alpha \begin{pmatrix} 3 \\ 5 \end{pmatrix} \qquad \text{with} \\ \left(TL\Big|_{\mathrm{Im}(T)}\right)^{-1} \left(\beta \begin{pmatrix} 3 \\ 5 \end{pmatrix}\right) &= \frac{\beta}{54(3y_1 + 5y_2)}\mathbf{y}. \end{aligned} \tag{8}$$

Given any $\mathbf{b} = t(2, 7, 1)^T \in \mathrm{Im}(L)$, we can find a unique $\mathbf{x} \in \tilde{V} = \mathrm{span}\{\mathbf{y}\}$ for which $L\mathbf{x} = \mathbf{b}$, namely,

$$\mathbf{x} = \left(L\Big|_{\tilde{V}}\right)^{-1} \left(t \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix}\right) = \frac{t}{3y_1 + 5y_2}\mathbf{y}.$$

More generally, given any $\mathbf{b} \in \mathbb{R}^3$ we can consider

$$\begin{aligned}
\mathbf{x} &= \left( TL\Big|_{\mathrm{Im}(T)} \right)^{-1} T\mathbf{b} \\
&= \left( TL\Big|_{\mathrm{Im}(T)} \right)^{-1} \left( (2b_1 + 7b_2 + b_3) \begin{pmatrix} 3 \\ 5 \end{pmatrix} \right) \\
&= \frac{2b_1 + 7b_2 + b_3}{54(3y_1 + 5y_2)} \mathbf{y}.
\end{aligned}$$

We then have

$$L\mathbf{x} = \frac{2b_1 + 7b_2 + b_3}{54} \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} = \mathrm{proj}_{\mathrm{Im}(L)} \mathbf{b}.$$

If we take the special choice $\mathbf{y} = T(2,7,1)^T = 54(3,5)^T$, then the discussion following (5) simplifies as follows:

The one-dimensional subspace $\tilde{V} = \mathrm{span}\{54(3,5)^T\} = \mathcal{N}(L)^\perp = \mathrm{Im}(T)$, so the associated isomorphisms become

$$L\Big|_{\mathrm{Im}(T)} : \mathrm{Im}(T) \to \mathrm{Im}(L) \qquad \text{and} \qquad T\Big|_{\mathrm{Im}(L)} : \mathrm{Im}(L) \to \mathrm{Im}(T). \tag{9}$$

The composition $TL : \mathrm{Im}(T) \to \mathrm{Im}(T)$ is also an isomorphism and an **operator** or automorphism, and the mappings (6-8) are given by

$$L(54\alpha(3,5)^T) = (54)(34)\alpha \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} \qquad \text{with}$$

$$\left( L\Big|_{\mathrm{Im}(T)} \right)^{-1} \left( t \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} \right) = \frac{t}{34} \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \tag{10}$$

$$T\Big|_{\mathrm{Im}(L)} \left( t \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix} \right) = 54t \begin{pmatrix} 3 \\ 5 \end{pmatrix} \qquad \text{with}$$

$$\left( T\Big|_{\mathrm{Im}(L)} \right)^{-1} \left( \beta \begin{pmatrix} 3 \\ 5 \end{pmatrix} \right) = \frac{\beta}{54} \begin{pmatrix} 2 \\ 7 \\ 1 \end{pmatrix}, \tag{11}$$

which is identical to (7), and

$$TL\left(54\alpha\begin{pmatrix}3\\5\end{pmatrix}\right) = (54)^2(34)\alpha\begin{pmatrix}3\\5\end{pmatrix} \qquad \text{with}$$

$$\left(TL\big|_{\mathrm{Im}(T)}\right)^{-1}\left(\beta\begin{pmatrix}3\\5\end{pmatrix}\right) = \frac{\beta}{(54)(34)}\begin{pmatrix}3\\5\end{pmatrix}. \tag{12}$$

Given any $\mathbf{b} = t(2,7,1)^T \in \mathrm{Im}(L)$, we can find a unique $\mathbf{x} \in \mathrm{Im}(T) = \mathrm{span}\{54(3,5)^T\}$ for which $L\mathbf{x} = \mathbf{b}$, namely,

$$\mathbf{x} = \left(L\big|_{\mathrm{Im}(T)}\right)^{-1}\left(t\begin{pmatrix}2\\7\\1\end{pmatrix}\right) = \frac{t}{34}\begin{pmatrix}3\\5\end{pmatrix}.$$

More generally, given any $\mathbf{b} \in \mathbb{R}^3$ we can consider

$$\begin{aligned}\mathbf{x} &= \left(TL\big|_{\mathrm{Im}(T)}\right)^{-1} T\mathbf{b}\\ &= \left(TL\big|_{\mathrm{Im}(T)}\right)^{-1}\left((2b_1+7b_2+b_3)\begin{pmatrix}3\\5\end{pmatrix}\right)\\ &= \frac{2b_1+7b_2+b_3}{(54)(34)}\begin{pmatrix}3\\5\end{pmatrix}.\end{aligned}$$

We then have

$$L\mathbf{x} = \frac{2b_1+7b_2+b_3}{54}\begin{pmatrix}2\\7\\1\end{pmatrix} = \mathrm{proj}_{\mathrm{Im}(L)}\mathbf{b}.$$

# 3 Big Picture Part II

The assertions

$$\mathcal{N}(T) = \mathrm{Im}(L)^{\perp} \qquad \text{and} \qquad \mathrm{Im}(T) = \mathcal{N}(L)^{\perp}$$

of Theorem 2 should be compared to Axler's results

$$\mathcal{N}(L') = \mathrm{Im}(L)^{a} \qquad \text{and} \qquad \mathrm{Im}(L') = \mathcal{N}(L)^{a}$$

on the dual map. Notice the annihilator plays the role of the orthogonal complement, and we have shown the correspondence is complete: The corresponding spaces are essentially "the same" in the sense that they are images of one another under the canonical isomorphisms. The resulting "splitting up" of $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$ takes place directly using orthogonal complements and the corresponding splitting of the dual spaces takes places using the dual map and annihilators. See Figure 3.

$$
\begin{tikzcd}
\tilde{V} = \mathcal{N}(L)^{a} \arrow[rrrr, "L"] \arrow[dr] \arrow[ddddd, "\Phi"'] & & & & \mathrm{Im}(L) \arrow[dl] \arrow[ddddd, "\Psi"] \\
& V \arrow[rr, "L"] \arrow[ddd, "\Phi"'] & & W \arrow[ddd, "\Psi"] & \\
& & \mathcal{N}(L) \arrow[ul] & & \\
& & \mathcal{N}(L') = a(\mathrm{Im}(L)) \arrow[dr] & & \\
& V' & & W' \arrow[ll, "L'"'] & \\
\tilde{V}' = \mathrm{Im}(L') = a(\mathcal{N}(L)) \arrow[ur] & & & & \tilde{W}' \arrow[llll, "L'"'] \arrow[ul]
\end{tikzcd}
$$

$'$

Figure 3: The Big Picture. The restriction at the top is of primary interest. It is an isomorphism from the subspace $\tilde{V}$ of $V$ which is the **algebraic complement** of $\mathcal{N}(L)$ onto $\mathrm{Im}(L)$ in $W$. Again, all diagonal arrows correspond to injections. In infinite dimensional spaces various parts of this diagram may cease to function. The construction of $L'$ is always possible, but the isomorphisms $\Phi$ and $\Psi$ may not be available to tie the dual map back to the original linear map $L$ in a nice way. If $V$ and $W$ are any inner product spaces, there are still injections $\Phi$ and $\Psi$ into the respective dual spaces, but $\Phi$ and $\Psi$ may not be surjective, i.e., there may be no Riesz representation. In general for inner product spaces, if the image of $L'$ lies in the image of $\Phi$ in $V'$, then it should be possible to use $L'$ to construct a specific subspace $\tilde{V}$ in $V$ so that the restriction of $L$ to $\tilde{V}$ is an isomorphism onto $\mathrm{Im}(L)$. In this case, the algebraic complement obtained should be the **orthogonal complement** $\mathcal{N}(L)^{\perp}$.

We have mentioned that there are other ways to split $V = \mathbb{R}^n$ up in our examples so that $\mathbb{R}^n = \mathcal{N}(L) \oplus \tilde{V}$. The identification or choice of $\tilde{V}$ as

$$\tilde{V} = \mathrm{Im}(T) = \Phi^{-1}\,\mathrm{Im}(L')$$

has (at least) two advantages. First the composition

$$T \circ L\Big|_{\mathrm{Im}(T)} : \mathrm{Im}(T) \to \mathrm{Im}(T)$$

is an **operator**, i.e., a linear map with the same domain and co-domain, as opposed to

$$T \circ L\Big|_{\tilde{V}} : \tilde{V} \to \mathrm{Im}(T).$$

In either case, this restricted composition is an isomporphism. The second advantage is simply that the mapping $T : \mathbb{R}^m \to \mathbb{R}^n$ gives an easy way to identify a subspace $\tilde{V}$ that will accomplish the desired "splitting up" of $V = \mathbb{R}^n$.

In infinite dimensional cases, it turns out that $V'$ is never isomorphic to $V$. See Assignment 11 for some suggestsions that $V'$ is typically a "larger" vector space when $V$ is infinite dimensional. It may still be the case that there are some canonical injections $\Phi$ and $\Psi$ from $V$ into $V'$ and from $W$ into $W'$ respectively. In these cases, some version of the discussion about our examples may apply.

Another topic that should be considered is the possibility of how the above discussion adapts (or fails to adapt) to the situation when $V$ and $W$ are complex vector fields.

In the general case, there is no isomorphism between $V$ and $V'$, and consequently, the diagram/connection illustrated in Figure 3 and our discussion above "breaks apart." There is in general no mapping $T : W \to V$ corresponding to $L' : W' \to V'$. Nevertheless, there is always a mapping $L' : W' \to V'$, and one can think of $\mathrm{Im}(L')$ and $\mathcal{N}(L')$ with the relations

$$\mathcal{N}(L') = \mathcal{a}(\mathrm{Im}(L)) \cong \mathrm{Im}(L)^{\perp} = \mathrm{Im}(L)^{\mathcal{a}}$$

and

$$\mathrm{Im}(L') = \mathcal{a}(\mathcal{N}(L)) \cong \mathcal{N}(L)^{\perp} = \mathcal{N}(L)^{\mathcal{a}}$$

as "weak imitations" for the spaces $\mathrm{Im}(T)^{\perp}$ and $\mathcal{N}(L)^{\perp}$. As $U^{\perp}$ is called the orthogonal complement of a subspace $U$ in a vector space $V$, we might call the annihilator $\mathcal{a}(W)$ the "algebraic complement" of a subspace $W \subset V'$ of the dual space $V'$.

As a final remark, the dual spaces $V'$ of linear functionals defined by Axler are sometimes called **algebraic dual spaces**. This terminology is in contrast to what are sometimes called **analytic dual spaces** (or also just simply dual spaces in other contexts). Those contexts include, for example, when one has an inner product (and an induced/associated norm). In that case, there is a notion of **continuity**, and instead of the algebraic dual space $V'$ we have considered, one restricts to the smaller (dual) space of **continuous linear functionals**

$$C^0(V \to F) \cap \mathcal{L}(V \to F).$$

This set constitutes the analytic dual space of $V$. In fact, the approach Axler is taking to linear algebra (especially in Section 3F concerning duality) may be viewed as using techniques from the subject of **linear algebra on infinite dimensional spaces**, that is the study of linear functions $L : X \to Y$ where $X$ and $Y$ are infinite dimensional spaces. As Axler mentioned, usually linear algebra, as a subject, is synonomous with the study of linear functions on finite dimensional vector spaces, as suggested by the title of Halmos' book. This more difficult subject of linear algebra on infinite dimensional spaces is usually called **functional analysis**. As functional analysis is generally a more difficult subject than plain-Jane linear algebra, one usually makes additional assumptions, like the presence of inner products and norms. And one also assumes in functional analysis that linear functionals are continuous. As an aside, it does turn out that cases where the smaller analytic dual is isomorphic to the vector space $V$ are of interest. In particular, a normed space $V$ (in functional analysis) for which the analytic double dual is isomorphic to the original space $V$ is called a **reflexive space**. This may also be viewed as a kind of "weak imitation" of our discussion of the double orthogonal complement above.

In any case, it is quite interesting that Axler is using a kind of "functional analysis" approach to linear algebra in Section 3F on **duality**, which is often not even mentioned as a topic in linear algebra. The real significance of this is that he gets a proof of what might be considered the **first really difficult theorem in linear algebra** that the column rank and the row rank of a matrix are the same, and he gets this proof without saying much of anything at all about matrices. Usually, one has to go through a lengthy and complicated discussion of Gaussian elimination and pivots and such things to prove row rank and column rank are equal. Duality gives you this result essentially for "free," that is without much complication but just for putting in the effort of learning a few natural definitions and following your nose through some functional analytic "abstract nonsense."