6

Elementary Linear Transform Theory

Whether they are spaces of “arrows in space”, functions or even matrices, vector spaces quickly become boring if we don’t do things with their elements — move them around, differentiate or integrate them, whatever. And often, what we do with these generalized vectors end up being linear operations.

6.1 Basic Definitions and Examples

Let us suppose we have two vector spaces $\mathcal{V}$ and $\mathcal{W}$ (they could be the same vector space). Then we can certainly have a function $\mathcal{L}$ that transforms any vector $\mathbf{v}$ from $\mathcal{V}$ into a corresponding vector $\mathbf{w} = \mathcal{L}(\mathbf{v})$ in $\mathcal{W}$. This function is called a *linear transformation* (or *linear transform* or *linear operator*) from $\mathcal{V}$ into $\mathcal{W}$ if and only if

$$\mathcal{L}(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2) \;=\; \alpha_1\mathcal{L}(\mathbf{v}_1) \;+\; \alpha_2\mathcal{L}(\mathbf{v}_2)$$

whenever α_1 and α_2 are scalars, and $\mathbf{v}_1$ and $\mathbf{v}_2$ are vectors in $\mathcal{V}$. This expression, of course, can be expanded to

$$\mathcal{L}\left(\sum_k \alpha_k\mathbf{v}_k\right) \;=\; \sum_k \alpha_k\mathcal{L}(\mathbf{v}_k)$$

for any linear combination $\sum_k \alpha_k\mathbf{v}_k$ of vectors in $\mathcal{V}$. (Remember, we are still insisting on linear combinations having only finitely many terms.)

The *domain* of the operator is the vector space $\mathcal{V}$, and the *range* is the set of all vectors in $\mathcal{W}$ given by $\mathcal{L}(\mathbf{v})$ where $\mathbf{v} \in \mathcal{V}$. On occasion, we might also call $\mathcal{V}$ the “input space”, and $\mathcal{W}$ the “target space”. This terminology is not standard, but is descriptive.

Often, $\mathcal{V}$ and $\mathcal{W}$ will be the same. If this is the case, then we will simply refer to $\mathcal{L}$ as a linear transformation/transform/operator on $\mathcal{V}$.

It is also often true that $\mathcal{W}$ is not clearly stated. In such cases we can take $\mathcal{W}$ to be any vector space containing every $\mathcal{L}(\mathbf{v})$ for every $\mathbf{v} \in \mathcal{V}$. There is no requirement that every vector in $\mathcal{W}$ can be treated as $\mathcal{L}(\mathbf{v})$ for some $\mathbf{v} \in \mathcal{V}$.

Here are a few examples of linear transforms on a traditional three-dimensional vector space $\mathcal{V}$ with $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ being a standard basis. In each case, the operator is defined by explaining what it does to an arbitrary vector $\mathbf{v}$ in $\mathcal{V}$. Also given is a least one possible target space $\mathcal{W}$.

1. Any constant "magnification", say,

$$\mathcal{M}_2(\mathbf{v}) \;=\; 2\mathbf{v} \quad .$$

Here, $\mathcal{W} = \mathcal{V}$.

2. Projection onto the $\mathbf{k}$ vector,

$$\overrightarrow{\mathrm{pr}}_{\mathbf{k}}(\mathbf{v}) \quad .$$

Here, $\mathcal{W}$ is the one-dimensional space subspace of $\mathcal{V}$ consisting of all scalar multiples of $\mathbf{k}$. (Actually, you can also view all of $\mathcal{V}$ as $\mathcal{W}$.)

3. Projection onto the plane spanned by $\mathbf{i}$ and $\mathbf{j}$,

$$\mathcal{P}(v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) \;=\; \overrightarrow{\mathrm{pr}}_{\{\mathbf{i},\mathbf{j}\}}(\mathbf{v}) \;=\; v_1\mathbf{i} + v_2\mathbf{j} \quad .$$

Here, $\mathcal{W}$ is the plane spanned by $\mathbf{i}$ and $\mathbf{j}$. (Again, you can also view all of $\mathcal{V}$ as $\mathcal{W}$.)

4. The cross product with some fixed vector $\mathbf{a}$, say, $\mathbf{a} = 1\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$,

$$\mathcal{K}_{\mathbf{a}}(\mathbf{v}) \;=\; \mathbf{a} \times \mathbf{v} \quad .$$

Here, we can take $\mathcal{W}$ to be $\mathcal{V}$. Also, since the cross product with $\mathbf{a}$ is always perpendicular to $\mathbf{a}$, we can refine our choice of $\mathcal{W}$ to being the plane orthogonal to $\mathbf{a}$.

5. The dot product with some fixed vector $\mathbf{a}$, say, $\mathbf{a} = 1\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$,

$$\mathcal{D}_{\mathbf{a}}(\mathbf{v}) \;=\; \mathbf{a} \cdot \mathbf{v} \quad .$$

Here, it does not really make sense to view $\mathcal{V}$ as $\mathcal{W}$ since the result is not a vector in $\mathcal{V}$. Instead, since the result is a scalar (a real scalar, in fact), so $\mathcal{W} = \mathbb{R}$.

6. The rotation $\mathcal{R}$ through some fixed angle ϕ about some fixed vector, say, $\mathbf{k}$.

?►*Exercise 6.1:* *Convince yourself that each of the above transformations are linear transformations.*

We can also have very nontrivial linear transformations on nontraditional vector spaces. For example, derivatives and the Laplace and Fourier transforms are linear transformations on certain vector spaces of functions. Later on, we will be studying differential linear transformations of functions which are associated with partial differential equations.

On the other hand, we can also have operators that are not linear. For example, the norm $\|\cdot\|$ is a nonlinear operator from any vector space with an inner product into $\mathbb{R}$.

?►*Exercise 6.2* ***a:*** *Verify that the norm is not a linear operator from any traditional vector space $\mathcal{V}$ into $\mathbb{R}$.*

b: *Come up with some other examples of nonlinear operators (i.e, functions from one vector space into another that are not linear).*

6.2 Matrices for Linear Transforms
Defining the Matrix in General

Suppose $\mathcal{L}$ is a linear operator from an N-dimensional vector space $\mathcal{V}$ to an M-dimensional vector space $\mathcal{W}$. Let

$$\mathcal{B}_{\mathcal{V}} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N\} \quad \text{be a basis for } \mathcal{V} \quad ,$$

and let

$$\mathcal{B}_{\mathcal{W}} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_M\} \quad \text{be a basis for } \mathcal{W} \quad .$$

Now, since each $\mathcal{L}(\mathbf{a}_k)$ is in $\mathcal{W}$, each $\mathcal{L}(\mathbf{a}_k)$ must be a linear combination of the $\mathbf{b}_j$'s. So, there are constants λ_{jk} such that, for $k = 1, 2, \ldots, N$,

$$\mathcal{L}(\mathbf{a}_k) = \sum_{j=1}^{M} \lambda_{jk}\mathbf{b}_j \qquad \text{for} \quad k = 1, 2, \ldots N \quad , \tag{6.1a}$$

which, we might note, can also be written as

$$\mathcal{L}(\mathbf{a}_k) = \sum_{j=1}^{M} \mathbf{b}_j\lambda_{jk} = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_N\end{bmatrix}\begin{bmatrix}\lambda_{1k}\\ \lambda_{2k}\\ \vdots \\ \lambda_{Mk}\end{bmatrix} \quad , \tag{6.1b}$$

or even as

$$|\mathcal{L}(\mathbf{a}_k)\rangle_{\mathcal{B}_{\mathcal{W}}} = \begin{bmatrix}\lambda_{1k}\\ \lambda_{2k}\\ \vdots \\ \lambda_{Mk}\end{bmatrix} \quad . \tag{6.1c}$$

Using these λ_{jk}'s, we then define the *matrix* $\mathbf{L}$ *for* $\mathcal{L}$ *with respect to these bases* to be the $M \times N$ matrix given by

$$[\mathbf{L}]_{jk} = \lambda_{jk} \qquad \text{for} \quad j = 1, 2, \ldots, M \text{ and } k = 1, 2, \ldots, N \quad .$$

That is, the $(j, k)^{\text{th}}$ entry in $\mathbf{L}$ is the j^{th} (scalar) component with respect to $\mathcal{B}_{\mathcal{W}}$ of the k^{th} vector in $\mathcal{B}_{\mathcal{V}}$,

$$\mathcal{L}(\mathbf{a}_k) = \sum_{j=1}^{M} [\mathbf{L}]_{jk}\mathbf{b}_j \qquad \text{for} \quad k = 1, 2, \ldots N \quad . \tag{6.2}$$

Equivalently, we can say that $\mathbf{L}$ is the $M \times N$ matrix such that

$$\begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_M\end{bmatrix}\mathbf{L} = \begin{bmatrix}\mathcal{L}(\mathbf{a}_1) & \mathcal{L}(\mathbf{a}_2) & \cdots & \mathcal{L}(\mathbf{a}_N)\end{bmatrix} \quad , \tag{6.3}$$

or, perhaps more simply, we can say that

$$\begin{bmatrix} k^{\text{th}} \\ \text{column} \\ \text{of } \mathbf{L}\end{bmatrix} = |\mathcal{L}(\mathbf{a}_k)\rangle_{\mathcal{B}_{\mathcal{W}}} \qquad \text{for} \quad k = 1, 2, \ldots, N \quad . \tag{6.4}$$

Any of these can be used to find $\mathbf{L}$ for a given $\mathcal{L}$.

Keep in mind that matrix $\mathbf{L}$ depends on the choice of bases $\mathcal{B}_{\mathcal{V}}$ and $\mathcal{B}_{\mathcal{W}}$. On occasion, it may be important to explicitly indicate which bases are being used. We will do that by denoting the matrix for $\mathcal{L}$ by

$$\mathbf{L}_{\mathcal{B}_{\mathcal{W}}\mathcal{B}_{\mathcal{V}}} \quad .$$

This will be simplified to

$$\mathbf{L}_{\mathcal{B}}$$

if the input and target spaces are the same and we are just using one basis $\mathcal{B}$ (so, in the above, $\mathcal{B}_{\mathcal{V}} = \mathcal{B}_{\mathcal{W}} = \mathcal{B}$).

!►*Example 6.1:* *Let's consider the cross product with some fixed vector* $\mathbf{a}$*, say,* $\mathbf{a} = 1\mathbf{i} + 2\mathbf{j}$,

$$\mathcal{K}_{\mathbf{a}}(\mathbf{v}) = \mathbf{a} \times \mathbf{v}$$

where $\mathcal{V} = \mathcal{W}$ *is a traditional vector space with standard basis* $\mathcal{B} = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$*. We already know* $\mathcal{K}_{\mathbf{a}}$ *is a linear operator on* $\mathcal{V}$*. Computing the "*$\mathcal{K}_{\mathbf{a}}$ *for each basis vector", we get*

$$\mathcal{K}_{\mathbf{a}}(\mathbf{i}) = \mathbf{a} \times \mathbf{i} = \big(1\mathbf{i} + 2\mathbf{j}\big) \times \mathbf{i} = \cdots = 0\mathbf{i} + 0\mathbf{j} - 2\mathbf{k} \quad ,$$

$$\mathcal{K}_{\mathbf{a}}(\mathbf{j}) = \mathbf{a} \times \mathbf{j} = \big(1\mathbf{i} + 2\mathbf{j}\big) \times \mathbf{j} = \cdots = 0\mathbf{i} + 0\mathbf{j} + 1\mathbf{k}$$

and

$$\mathcal{K}_{\mathbf{a}}(\mathbf{k}) = \mathbf{a} \times \mathbf{k} = \big(1\mathbf{i} + 2\mathbf{j}\big) \times \mathbf{k} = \cdots = 2\mathbf{i} - 1\mathbf{j} + 0\mathbf{k} \quad .$$

So,

$$|\mathcal{K}_{\mathbf{a}}(\mathbf{i})\rangle = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix} \quad , \quad |\mathcal{K}_{\mathbf{a}}(\mathbf{j})\rangle = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad , \quad |\mathcal{K}_{\mathbf{a}}(\mathbf{k})\rangle = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \quad .$$

But, as noted above (equation (6.4)), these are also the respective columns in the matrix $\mathbf{K}$ *for the operator* $\mathcal{K}_{\mathbf{a}}$*. Thus,*

$$\mathbf{K} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & -1 \\ -2 & 1 & 0 \end{bmatrix} \quad .$$

So What? (Using the Matrix)

So what? Well, let $\mathbf{v}$ be a vector in $\mathcal{V}$, and let $\mathbf{w} = \mathcal{L}(\mathbf{v})$. Since $\mathbf{v} \in \mathcal{V}$, it has (scalar) components v_1, v_2, ... and v_N with respect to the given basis for $\mathcal{V}$. So we can write

$$\mathbf{v} = \sum_{k=1}^{N} v_k \mathbf{a}_k \qquad \text{and} \qquad |\mathbf{v}\rangle = |\mathbf{v}\rangle_{\mathcal{B}_{\mathcal{V}}} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} \quad .$$

And since $\mathbf{w} = \mathcal{L}(\mathbf{v})$ is in $\mathcal{W}$, $\mathbf{w}$ has (scalar) components w_1, w_2, $\ldots$ and w_M with respect to the given basis for $\mathcal{W}$. So we can write

$$\mathbf{w} = \sum_{j=1}^{M} w_j \mathbf{b}_j \qquad \text{and} \qquad |\mathbf{w}\rangle = |\mathbf{w}\rangle_{\mathcal{B}_{\mathcal{W}}} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{bmatrix} \quad .$$

Using this with the linearity of the operator,

$$\begin{aligned} \sum_j w_j \mathbf{b}_j = \mathbf{w} = \mathcal{L}(\mathbf{v}) &= \mathcal{L}\left(\sum_k v_k \mathbf{a}_k\right) \\ &= \sum_k v_k \mathcal{L}(\mathbf{a}_k) \\ &= \sum_k v_k \sum_j \lambda_{jk} \mathbf{b}_j \\ &= \sum_j \sum_k \lambda_{jk} v_k \mathbf{b}_j = \sum_j \left[\sum_k [\mathbf{L}]_{jk} v_k \right] \mathbf{b}_j \quad . \end{aligned}$$

So

$$\sum_j w_j \mathbf{b}_j = \mathbf{w} = \sum_j \left[\sum_k [\mathbf{L}]_{jk} v_k \right] \mathbf{b}_j \quad , \tag{6.5}$$

which means

$$w_j = \sum_k [\mathbf{L}]_{jk} v_k \qquad \text{for each} \quad j \quad .$$

But the righthand side of this is simply the formula for the j^{th} entry in the matrix product of $\mathbf{L}$ with $|\mathbf{v}\rangle$. So,

$$|\mathcal{L}(\mathbf{v})\rangle = |\mathbf{w}\rangle = \mathbf{L}\,|\mathbf{v}\rangle \quad ,$$

or, more explicitly,

$$|\mathcal{L}(\mathbf{v})\rangle_{\mathcal{B}_{\mathcal{W}}} = \mathbf{L}_{\mathcal{B}_{\mathcal{W}}\mathcal{B}_{\mathcal{V}}}\,|\mathbf{v}\rangle_{\mathcal{B}_{\mathcal{V}}} \quad .$$

What we have demonstrated is that:

1. Computationally, any linear transform can be completely described by a corresponding matrix (provided we have a basis for each of the spaces).

2. This matrix is easily constructed from the components with respect to the basis of the target space of the transforms of the basis vectors used for the "input space".

Just for the heck of it, take another look at equation (6.5), which I will slightly rewrite as

$$\mathcal{L}(\mathbf{v}) = \sum_{j=1}^{M} \mathbf{b}_j \left[\sum_{k=1}^{N} [\mathbf{L}]_{jk} v_k \right]$$

This is just the matrix product

$$\mathcal{L}(\mathbf{v}) = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_M \end{bmatrix} \mathbf{L}\,|\mathbf{v}\rangle \quad ,$$

or, more explicitly,

$$\mathcal{L}(\mathbf{v}) = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_M \end{bmatrix} \mathbf{L}_{\mathcal{B}_{\mathcal{W}}\mathcal{B}_{\mathcal{V}}}\,|\mathbf{v}\rangle_{\mathcal{B}_{\mathcal{V}}} \quad .$$

!►Example 6.2: *Again let's consider the cross product operator from example 6.1,*

$$\mathcal{K}_{\mathbf{a}}(\mathbf{v}) = \mathbf{a} \times \mathbf{v} \qquad \textit{with} \quad \mathbf{a} = 1\mathbf{i} + 2\mathbf{j} \quad .$$

In that example, we saw that the matrix of this operator with respect to the standard basis $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ *is*

$$\mathbf{K} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & -1 \\ -2 & 1 & 0 \end{bmatrix} \quad .$$

So, in general,

$$|\mathbf{a} \times \mathbf{v}\rangle = |\mathcal{K}_{\mathbf{a}}(\mathbf{v})\rangle = \mathbf{K}\,|\mathbf{v}\rangle \quad .$$

In particular, if $\mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + 5\mathbf{k}$ *, then*

$$|\mathbf{a} \times \mathbf{v}\rangle = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & -1 \\ -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 10 \\ -5 \\ -1 \end{bmatrix} \quad .$$

That is,

$$\mathbf{a} \times \mathbf{v} = 10\mathbf{i} - 5\mathbf{j} - 1\mathbf{k} \quad .$$

Special (but Common) Cases
Using an Orthonormal Basis of the Target Space

Let's now assume that we are being intelligent and using an orthonormal basis at least of the target space $\mathcal{W}$, and let's emphasize this by renaming the vectors in this basis as

$$\mathcal{B}_{\mathcal{W}} = \{\, \mathbf{e}_1,\ \mathbf{e}_2,\ \ldots,\ \mathbf{e}_M \,\} \quad .$$

Then equation (6.2) describing the entries of $\mathcal{L}(\mathbf{a}_k)$ becomes

$$\mathcal{L}(\mathbf{a}_k) = \sum_{m=1}^{M} [\mathbf{L}]_{mk}\mathbf{e}_m \qquad \text{for} \quad k = 1, 2, \ldots N \quad .$$

As we've done at least a couple of times before, we'll take the inner product of both sides with one of the $\mathbf{e}_m$'s and use the orthonormallity of this basis along with the linearity of $\mathcal{L}$:

$$\begin{aligned} \left\langle\, \mathbf{e}_j \,\middle|\, \mathcal{L}(\mathbf{a}_k) \,\right\rangle &= \left\langle\, \mathbf{e}_j \,\middle|\, \sum_{m=1}^{M} [\mathbf{L}]_{mk}\mathbf{e}_m \right\rangle \\ &= \sum_{m=1}^{M} [\mathbf{L}]_{mk} \left\langle\, \mathbf{e}_j \,\middle|\, \mathbf{e}_m \,\right\rangle \\ &= \sum_{m=1}^{M} [\mathbf{L}]_{mk}\delta_{jm} = [\mathbf{L}]_{jk} \quad . \end{aligned}$$

Thus,

> *If the basis being used for the target space,*
>
> $$\mathcal{B}_{\mathcal{W}} = \{ \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_M \} \quad ,$$
>
> *is orthonormal, then the* $(jk)^{th}$ *entry of the corresponding matrix* $\mathbf{L}$ *for a linear operator* $\mathcal{L}$ *is*
>
> $$[\mathbf{L}]_{jk} = \left\langle \mathbf{e}_j \,\middle|\, \mathcal{L}(\mathbf{a}_k) \right\rangle \quad . \tag{6.6}$$

Target Space Equals Input Space

In a great many cases, we will be interested in a linear operator $\mathcal{L}$ just on a vector space $\mathcal{V}$ (i.e., $\mathcal{L}$ just transforms each vector in $\mathcal{V}$ to another vector in $\mathcal{V}$). Then we might as well use a single orthonormal basis

$$\mathcal{B}_{\mathcal{V}} = \mathcal{B}_{\mathcal{W}} = \mathcal{B} = \{ \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_M \} \quad .$$

In this case, when we are being explicit about the basis used for the matrix $\mathbf{L}$, we use

$$\mathbf{L}_{\mathcal{B}} \qquad \text{instead of} \qquad \mathbf{L}_{\mathcal{B}_{\mathcal{W}},\mathcal{B}_{\mathcal{V}}} \quad .$$

The above formula for the $(j,k)^{\text{th}}$ entry in this matrix reduces a little further to

$$[\mathbf{L}]_{jk} = \left[\mathbf{L}_{\mathcal{B}}\right]_{jk} = \left\langle \mathbf{e}_j \,\middle|\, \mathcal{L}(\mathbf{e}_k) \right\rangle \quad . \tag{6.7}$$

Linear Functionals

A *linear functional* is simply a linear operator for which the target space is the space of all scalars, $\mathbb{R}$ or $\mathbb{C}$. The traditional vector operators described in problems ***A7*** and ***B7***, and the integral operator described in problems ***C2*** and ***D*** of *Homework Handout V* are linear functionals, and an interesting theorem (the Riesz theorem) relating linear functionals to inner products is developed in problem ***K*** in the same homework set.

Infinite Dimensional Spaces

See problem *D* of *Homework Handout V*.

6.3 Change of Basis for Transform Matrices

Suppose we have some linear transform $\mathcal{L}$ on an N-dimensional vector space $\mathcal{V}$ (with $\mathcal{V}$ being both the input and the target space).[1] Let $\mathcal{A}$ and $\mathcal{B}$ be two orthonormal bases for this space. Then $\mathcal{L}$ has matrices $\mathbf{L}_{\mathcal{A}}$ and $\mathbf{L}_{\mathcal{B}}$ with respect to bases $\mathcal{A}$ and $\mathcal{B}$. Our goal now is to figure out how to find matrix $\mathbf{L}_{\mathcal{B}}$ from matrix $\mathbf{L}_{\mathcal{A}}$ using the "change of basis" matrices $\mathbf{M}_{\mathcal{AB}}$ and $\mathbf{M}_{\mathcal{BA}}$ from the Big Theorem (theorem 5.3).

[1] For fun, you might try redoing this section assuming $\mathcal{L}$ is a linear transform from one vector space $\mathcal{V}$ into another (possibly different) vector space $\mathcal{W}$. It's not that hard.

To do this, let $\mathbf{v}$ be any vector in $\mathcal{V}$, and observe that

$$\begin{aligned}
\mathbf{L}_{\mathcal{B}} \left|\mathbf{v}\right\rangle_{\mathcal{B}} &= \left|\mathcal{L}(\mathbf{v})\right\rangle_{\mathcal{B}} \\
&= \mathbf{M}_{\mathcal{B}\mathcal{A}} \left|\mathcal{L}(\mathbf{v})\right\rangle_{\mathcal{A}} \\
&= \mathbf{M}_{\mathcal{B}\mathcal{A}} \left[\mathbf{L}_{\mathcal{A}} \left|\mathbf{v}\right\rangle_{\mathcal{A}}\right] = \mathbf{M}_{\mathcal{B}\mathcal{A}} \left[\mathbf{L}_{\mathcal{A}} \left[\mathbf{M}_{\mathcal{A}\mathcal{B}} \left|\mathbf{v}\right\rangle_{\mathcal{B}}\right]\right] \quad .
\end{aligned}$$

After cutting out the middle and dropping the needless brackets, we have

$$\mathbf{L}_{\mathcal{B}} \left|\mathbf{v}\right\rangle_{\mathcal{B}} = \mathbf{M}_{\mathcal{B}\mathcal{A}}\mathbf{L}_{\mathcal{A}}\mathbf{M}_{\mathcal{A}\mathcal{B}} \left|\mathbf{v}\right\rangle_{\mathcal{B}} \qquad \text{for every} \quad \mathbf{v} \in \mathcal{V} \quad ,$$

telling us that

$$\mathbf{L}_{\mathcal{B}} = \mathbf{M}_{\mathcal{B}\mathcal{A}}\mathbf{L}_{\mathcal{A}}\mathbf{M}_{\mathcal{A}\mathcal{B}} \quad .$$

This, along with the corresponding formula for computing $\mathbf{L}_{\mathcal{A}}$ from $\mathbf{L}_{\mathcal{B}}$, is important enough to be called a theorem (even though it is really just a corollary of Big Theorem 5.3).

Theorem 6.1 (Change of Orthonormal Bases for Linear Transforms)
Assume $\mathcal{V}$ is a vector space with orthonormal bases $\mathcal{A}$ and $\mathcal{B}$, and let $\mathcal{L}$ be a linear operator on $\mathcal{V}$. Then $\mathbf{L}_{\mathcal{A}}$ and $\mathbf{L}_{\mathcal{B}}$, the matrices of $\mathcal{L}$ with respect to basis $\mathcal{A}$ and basis $\mathcal{B}$, respectively, are related by

$$\mathbf{L}_{\mathcal{B}} = \mathbf{M}_{\mathcal{B}\mathcal{A}}\mathbf{L}_{\mathcal{A}}\mathbf{M}_{\mathcal{A}\mathcal{B}} \qquad \text{and} \qquad \mathbf{L}_{\mathcal{A}} = \mathbf{M}_{\mathcal{A}\mathcal{B}}\mathbf{L}_{\mathcal{B}}\mathbf{M}_{\mathcal{B}\mathcal{A}} \tag{6.8}$$

where $\mathbf{M}_{\mathcal{A}\mathcal{B}}$ and $\mathbf{M}_{\mathcal{B}\mathcal{A}}$ are the change of bases matrices described in theorem 5.3.

The formulas

$$\mathbf{M}_{\mathcal{B}\mathcal{A}}\mathbf{L}_{\mathcal{A}}\mathbf{M}_{\mathcal{A}\mathcal{B}} \qquad \text{and} \qquad \mathbf{M}_{\mathcal{A}\mathcal{B}}\mathbf{L}_{\mathcal{B}}\mathbf{M}_{\mathcal{B}\mathcal{A}}$$

are said to define *similarity transforms* of matrices $\mathbf{L}_{\mathcal{A}}$ and $\mathbf{L}_{\mathcal{B}}$ respectively. In short, a *similarity transform* of a matrix is simply the computations for converting a matrix for some linear operator with respect to one basis into the corresponding matrix with respect to another basis for the same operator.

In common practice, the notation $\mathbf{M}_{\mathcal{A}\mathcal{B}}$ and $\mathbf{M}_{\mathcal{B}\mathcal{A}}$ is not particularly common. Instead, you might have, say, the matrix we are calling $\mathbf{M}_{\mathcal{B}\mathcal{A}}$ denoted by $\mathbf{U}$. Keeping in mind $\mathbf{M}_{\mathcal{A}\mathcal{B}}$ and $\mathbf{M}_{\mathcal{B}\mathcal{A}}$ are adjoints of each other, the above similarity transforms would then be written as

$$\mathbf{U}\mathbf{L}_{\mathcal{A}}\mathbf{U}^{\dagger} \qquad \text{and} \qquad \mathbf{U}^{\dagger}\mathbf{L}_{\mathcal{B}}\mathbf{U} \quad ,$$

and the above theorem would be written as

Theorem 6.1′ (Change of Orthonormal Bases for Linear Transforms – Alt. Version)
Assume $\mathcal{V}$ is a vector space with orthonormal bases

$$\mathcal{A} = \{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N\} \qquad \text{and} \qquad \mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N\} \quad ,$$

and let $\mathcal{L}$ be a linear operator on $\mathcal{V}$. Then $\mathbf{L}_{\mathcal{A}}$ and $\mathbf{L}_{\mathcal{B}}$, the matrices of $\mathcal{L}$ with respect to basis $\mathcal{A}$ and basis $\mathcal{B}$, respectively, are related by

$$\mathbf{L}_{\mathcal{B}} = \mathbf{U}\mathbf{L}_{\mathcal{A}}\mathbf{U}^{\dagger} \qquad \text{and} \qquad \mathbf{L}_{\mathcal{A}} = \mathbf{U}^{\dagger}\mathbf{L}_{\mathcal{B}}\mathbf{U}$$

where $\mathbf{U}$ is the unitary matrix such that

$$\left|\mathbf{v}\right\rangle_{\mathcal{B}} = \mathbf{U} \left|\mathbf{v}\right\rangle_{\mathcal{A}} \qquad \text{for each} \quad \mathbf{v} \in \mathcal{V} \quad .$$

Don't forget that all the above was based on our bases $\mathcal{A}$ and $\mathcal{B}$ being orthonormal. If they are not orthonormal, then arguments very similar to that given for theorem 6.1 lead to the following theorem. Since we plan to limit ourselves to orthonormal bases, we probably won't need this theorem. It is included here just for completeness, and for those who may be interested.

Theorem 6.2 (Change of General Bases for Linear Transforms)
Assume $\mathcal{V}$ is a vector space with bases

$$\mathcal{A} = \{ \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N \} \qquad \text{and} \qquad \mathcal{B} = \{ \mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_N \} \quad ,$$

and let $\mathcal{L}$ be a linear operator on $\mathcal{V}$. Then $\mathbf{L}_\mathcal{A}$ and $\mathbf{L}_\mathcal{B}$, the matrices of $\mathcal{L}$ with respect to basis $\mathcal{A}$ and basis $\mathcal{B}$, respectively, are related by

$$\mathbf{L}_\mathcal{B} = \mathbf{M}\mathbf{L}_\mathcal{A}\mathbf{M}^{-1} \qquad \text{and} \qquad \mathbf{L}_\mathcal{A} = \mathbf{M}^{-1}\mathbf{L}_\mathcal{B}\mathbf{M}$$

where $\mathbf{M}$ is the invertible matrix such that

$$|\mathbf{v}\rangle_\mathcal{B} = \mathbf{M}\,|\mathbf{b}\rangle_\mathcal{A} \qquad \text{for each} \quad \mathbf{v} \in \mathcal{V} \quad .$$

?►*Exercise 6.3 (very optional):* *Derive/prove the above theorem.*

6.4 Adjoints of Operators

Let $\mathcal{L}$ be a linear operator on a vector space $\mathcal{V}$ having an inner product $\langle \cdot | \cdot \rangle$. We define the *adjoint of* $\mathcal{L}$, denoted by $\mathcal{L}^\dagger$, to be the linear operator on $\mathcal{V}$ satisfying

$$\left\langle \mathcal{L}^\dagger(\mathbf{v}) \;\middle|\; \mathbf{a} \right\rangle = \langle \mathbf{v} \mid \mathcal{L}(\mathbf{a}) \rangle \qquad \text{for every} \quad \mathbf{a}, \mathbf{v} \in \mathcal{V}$$

"provided it exists". Actually, it's not hard to prove it exists when $\mathcal{V}$ is finite-dimensional (we'll do that in a minute). If $\mathcal{V}$ is not finite-dimensional, then $\mathcal{L}^\dagger$ may or may not exist.

It also is not hard to show that, if it exists, then $\mathcal{L}^\dagger$ is also a linear operator on $\mathcal{V}$:

?►*Exercise 6.4:* *Let $\mathcal{L}$ be a linear operator on a vector space $\mathcal{V}$ having inner product $\langle \cdot | \cdot \rangle$.*

a: *Let α and β denote two arbitrary scalars, and let $\mathbf{v}$, $\mathbf{w}$ and $\mathbf{c}$ denote three arbitrary vectors in $\mathcal{V}$. Using the properties of inner products and definition of $\mathcal{L}^\dagger$, show that*

$$\left\langle \mathcal{L}^\dagger(\alpha\mathbf{v} + \beta\mathbf{w}) \;\middle|\; \mathbf{c} \right\rangle = \left\langle \alpha\mathcal{L}^\dagger(\mathbf{v}) + \beta\mathcal{L}^\dagger(\mathbf{w}) \;\middle|\; \mathbf{c} \right\rangle \quad .$$

b: *Why does the last exercise verify that $\mathcal{L}^\dagger$ is a linear operator?*

Let us now assume $\mathcal{V}$ is finite dimensional and $\mathbf{L}$ is the matrix for $\mathcal{L}$ with respect to some orthonormal basis

$$\{ \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N \} \quad .$$

Finding the matrix for $\mathcal{L}^\dagger$ turns out to be easy — just use formula (6.7), the conjugate symmetry of the inner product and the definition of the adjoint operator:

$$
\begin{aligned}
(j,k)^{\text{th}} \text{ entry in the matrix for } \mathcal{L}^\dagger &= \left\langle \mathbf{e}_j \;\middle|\; \mathcal{L}^\dagger(\mathbf{e}_k) \right\rangle \\
&= \left\langle \mathcal{L}^\dagger(\mathbf{e}_k) \;\middle|\; \mathbf{e}_j \right\rangle^* \\
&= \left\langle \mathbf{e}_k \;\middle|\; \mathcal{L}(\mathbf{e}_j) \right\rangle^* \;=\; \left([\mathbf{L}]_{kj}\right)^* \;=\; [\mathbf{L}^\dagger]_{jk} \quad .
\end{aligned}
$$

So the matrix for $\mathcal{L}^\dagger$, the adjoint of operator $\mathcal{L}$, is $\mathbf{L}^\dagger$, the adjoint of matrix $\mathbf{L}$. What a surprise!

(Side note: The existence of $\mathcal{L}^\dagger$ actually follows from the above computations for its matrix and the fact that $\mathbf{L}^\dagger$ certainly exists.)

Unsurprisingly, we refer to a linear operator $\mathcal{L}$ as being *self adjoint* or *Hermitian* if $\mathcal{L}^\dagger$ exists and equals $\mathcal{L}$. By the above, an operator (on a finite-dimensional vector space) is self adjoint if and only if its matrix with respect to any orthonormal basis is self adjoint. We will be discussing these operators more extensively in a week or so.

?►*Exercise 6.5:* *Take a look at the linear operators defined in Homework Handout V problems **A1** (a magnification), **A2** (a projection), **A3** (a different projection), **A5** (a 2-dimensional rotation), **B3** (another projection), **B4** (a 3-dimensional rotation) and **B8** (a cross product), and do the following:*

a: *Find the matrix for the adjoint.*

b: *Identify those operators that are Hermitian.*

c: *Verbally describe what the adjoint operator does to inputted vectors.*

Finally, let me throw in one more little exercise, the result of which may be useful later:

?►*Exercise 6.6:* *Let $\mathcal{L}$ be a linear operator on $\mathcal{V}$ that has an adjoint. Show that*

$$
\left\langle \mathbf{a} \;\middle|\; \mathcal{L}^\dagger(\mathbf{v}) \right\rangle = \langle\, \mathcal{L}(\mathbf{a}) \mid \mathbf{v} \,\rangle \qquad \text{for every} \quad \mathbf{a}, \mathbf{v} \in \mathcal{V} \quad .
$$

Hermitian operators are important because they often describe "linear processes" in which you expect certain symmetries. Later, we will develop the Sturm-Liouville theory and use it to find solutions to partial differential equations. This theory is really a theory of Hermitian differential operators.

6.5 Inner-Product Preserving Operators and Unitary/Orthogonal Matrices

Though our main interest with unitary matrices is how they arise in the "change of basis formulas", their relation with the matrices for "inner product preserving" linear operators should be noted, at least in exercise:

?►*Exercise 6.7:* *Let $\mathcal{L}$ be a linear operator on a finite-dimensional vector space $\mathcal{V}$ with inner product $\langle \cdot \mid \cdot \rangle$, and let $\mathbf{L}$ be its matrix with respect to some orthonormal basis for $\mathcal{V}$.*

a: *Assume* $\mathbf{L}$ *is a unitary matrix. Show that, if*

$$\mathbf{x} = \mathcal{L}(\mathbf{a}) \qquad \textit{and} \qquad \mathbf{y} = \mathcal{L}(\mathbf{b}) \quad ,$$

then

$$\langle\, \mathbf{x} \mid \mathbf{y} \,\rangle \;=\; \langle\, \mathbf{a} \mid \mathbf{b} \,\rangle \quad .$$

In other words, "$\mathcal{L}$ preserves the inner product."

b: *Assume* $\mathbf{L}$ *is a unitary matrix. Show that, if*

$$\mathbf{x} = \mathcal{L}(\mathbf{a}) \quad ,$$

then

$$\|\mathbf{x}\| \;=\; \|\mathbf{a}\| \quad .$$

In other words, "$\mathcal{L}$ preserves the norm." (Don't forget the problem just done.)

c: *Now assume that "$\mathcal{L}$ preserves the inner product." That is,*

$$\langle\, \mathbf{x} \mid \mathbf{y} \,\rangle \;=\; \langle\, \mathbf{a} \mid \mathbf{b} \,\rangle \qquad \textit{whenever} \quad \mathbf{x} = \mathcal{L}(\mathbf{a}) \quad \textit{and} \quad \mathbf{y} = \mathcal{L}(\mathbf{b}) \quad .$$

Show that $\mathbf{L}$ *is unitary.*[2]

6.6 On Confusing Matrices and Linear Operators — A Brief Rant

You are likely to see many cases where a matrix for an operator and the operator are treated as the same thing. This is unfortunate and leads to too much confusion, silly computations, and a lack of understanding as to what the hell is going on. It complicates matters and leads to the idea that much of mathematical physics is just the blind manipulation of symbols according to arcane laws. Avoid doing this yourself where possible.

The matrix of an operator depends on the basis. The operator does not. The matrix describes the operator in terms of some particular basis. If there is only one basis, and everyone knows what it is, then there is no great harm in confusing the operator and its matrix with respect to that basis. If no basis is given, however, then the description of the operator as a matrix is vacuous (typically, though, someone is assuming a "standard basis" without telling you). Worse yet are the cases where more than one basis is floating around — especially if no one tells you that multiple bases are involved. For example, you may recall "diagonalizing matrices". Did you realize that half the issue was the *finding* of a basis for which the matrix of the underlying operator is particularly simple? Or did you think it was just an exercise to pass the time?

Yes, sometimes everyone knows what everyone really means, and we can use verbal and written shortcuts. The problem is, of course, that, more often than not, not everyone knows what everyone else really means.

[2] You can even show that $\mathbf{L}$ is unitary if you simply know that it preserves the norm. That's a bit harder to verify.