

Linear Algebra II: The Fundamental Structure of Linear Transforms

Nathan Ratliff

Oct 24, 2014

Abstract

This document explores the fundamental underlying structure of linear transforms, ultimately showing how all linear transformations can be represented by a correspondence between orthonormal vectors in the domain and co-domain and an associated stretching along those axes. The structural properties of the linear transform are straightforwardly encoded in this representation, and it suggests a natural pseudoinverse for any linear transform that, in coordinates, manifests as the well-known Moore-Penrose pseudoinverse. The discussion culminates in a calculation that reveals that this fundamental underlying structure is exposed in coordinates by the Singular Value Decomposition of the transform's matrix given a choice of orthonormal basis for both the domain and co-domain.

1 Some notes on terminology

We'll start out with some brief notes on terminology to ensure everyone's on the same page. Terminology is useful in that it helps us communicate about arcane topics and ideas, but for the most part it's just confusing. We have multiple terms from different contexts that mean the same thing, and terms that mean different things in different contexts. If you're confused about anything, make sure it's not just the terminology that's confusing. Here, we list off some of the terms that we throw around in linear algebra and standardize their usage.

When discussing a **linear map** $T : \mathcal{V} \rightarrow \mathcal{U}$ between vector spaces \mathcal{V} and \mathcal{U} , we often call it a **linear transform** as well. Both terms mean the same thing. The space of inputs into the map \mathcal{V} is known as the **domain** and the space of outputs that result from the mapping is known as the **co-domain**. Sometimes we additionally refer to the space of outputs as the **range** as well, but for the most part, we'll try to stick to "co-domain". Additionally, the **image** of a map is the set of all possible outputs. The image may be a strict subset of the co-domain (for instance, the transform that sends everything to the zero vector has just one element in its image, namely the zero vector). Finally, we often use the term **operator** to refer specifically to linear transforms that map from \mathcal{U} back to the same space \mathcal{U} . Some texts also refer to such a mapping whose domain

and co-domain are the same space as an *endomorphism*. We'll typically stick to the term *operator*.

Generally, good definitions of all these terms can be found on Wikipedia. Be quick to look things up or ask someone the definition, because it's important that these terms be recognized and understood precisely so there's no uncertainty in their interpretation. Math is hard enough without uncertainty obfuscating the ideas.

2 Introducing geometry to vector spaces

We've discussed the underlying structure of a linear space in detail (see Ratliff (2014)), but so far we have no notion of geometry. How do two vectors $|a\rangle$, and $|b\rangle$ from a vector space \mathcal{V} relate to one another. We know that we can scale them and add them together to get a new vector $\alpha|a\rangle + \beta|b\rangle \in \mathcal{V}$, but we have no geometric notion of, for instance, angle, between the vectors.

Or do we? (The answer is no, but lets step through the argument anyway.) Suppose we have a basis for the space $\mathcal{B}_{\mathcal{V}} = \{|u_i\rangle\}_{i=1}^n$. We can represent any two vectors $|a\rangle = \sum_{i=1}^n \alpha_i |u_i\rangle$ and $|b\rangle = \sum_{i=1}^n \beta_i |u_i\rangle$ by their vectors of coefficients $\mathbf{a} = (\alpha_1, \dots, \alpha_n)$ and $\mathbf{b} = (\beta_1, \dots, \beta_n)$. Both are elements of \mathbb{R}^n , and we know what angles look like in \mathbb{R}^n —they're represented in the dot product:

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n \alpha_i \beta_i = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta, \quad (1)$$

where θ is the angle between the two vectors. So we just need to normalize the two vectors of coefficients and calculate their dot product to get the cosine of the angle between the vectors! Done. Right?

Unfortunately, we can step through some examples to see that this definition of angle is inconsistent. It changes depending on the the type of basis we choose. Consider, for instance, the space \mathbb{R}^2 as the underlying vector space. Here, we already know how to measure angles between vectors using the dot product; we'll see that choosing the wrong basis makes the dot product no longer work. Choose, for example, the vectors $\mathcal{B} = \{(0, 1), (1, 1)\}$ as a basis for \mathbb{R}^2 . We know that the two canonical basis vectors $\mathbf{x} = (1, 0)$ and $\mathbf{y} = (0, 1)$ should be orthogonal to each other since $\mathbf{x} \cdot \mathbf{y} = (1, 0) \cdot (0, 1) = 1 \cdot 0 + 0 \cdot 1 = 0$. However, if we represent these vectors in this new basis \mathcal{B} , the coordinates transform to $\tilde{\mathbf{x}} = (-1, 1)$ and $\tilde{\mathbf{y}} = (1, 0)$. The traditional dot product between the two coordinate vectors in this basis is

$$\tilde{\mathbf{x}} \cdot \tilde{\mathbf{y}} = -1 \cdot 1 + 1 \cdot 0 = -1. \quad (2)$$

That's very different from 0. A change of basis broke the dot product. To be consistent with what we know is the definition of angle in \mathbb{R}^2 , we need to change how we compute the *transformed* dot product in the new coefficient space. One

can check that the right calculation in the new space is should be

$$\mathbf{x} \cdot \mathbf{y} = \tilde{\mathbf{x}}^T \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \tilde{\mathbf{y}}. \quad (3)$$

For our example, with $\tilde{\mathbf{x}} = (-1, 1)$ and $\tilde{\mathbf{y}} = (1, 0)$, we now get

$$\begin{bmatrix} -1 \\ 1 \end{bmatrix}^T \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0, \quad (4)$$

which is the number we were looking for.

Indeed, our understanding of angle in \mathbb{R}^2 comes from the underlying space itself. We know what it means for two vectors to be at right angles to each other on a plane, so we choose two such vectors, usually horizontally and vertically aligned, to represent the coordinate axes in the space. As we'll see below, it's our use of an orthonormal basis that makes the simple dot product work as a representation of what we've already defined as the underlying notion of angle.

2.1 Inner product defined

So we can't just rely on simple dot products in coefficient spaces to define the notion of an angle between two vectors—that calculation changes depending on the choice of basis. It only makes sense when the underlying basis representation is orthogonal, which means that underneath everything we still need to define axiomatically what it means to measure an angle between two abstract vectors. Thus, we find ourselves requiring yet another definition to fully understand the geometry of a vector space.

An **inner product** is a mapping, typically denoted $\langle a | b \rangle$ that assigns a real number to every pair of vectors $|a\rangle, |b\rangle \in \mathcal{V}$, satisfying the following axioms:

1. **Symmetry.** $\langle a | b \rangle = \langle b | a \rangle$ for every pair of vectors.
2. **Linearity in arguments.** Linearity in the first argument $\langle \alpha a + \beta b | c \rangle = \alpha \langle a | c \rangle + \beta \langle b | c \rangle$, and by symmetry a corresponding linearity in the second argument.
3. **Positive definiteness.** For any non-zero vector $|a\rangle \in \mathcal{V}$, $\langle a | a \rangle > 0$. And for the zero vector, $\langle 0 | 0 \rangle = 0$.

This inner product between a vector and itself forms a **norm** on the space and denoted $\sqrt{\langle a | a \rangle} = \|a\|$. This norm automatically satisfies the usual axioms for a *size* measurement, so we typically use the inner product between a vector and itself to represent the squared size of the vector.

2.2 A quick note on the dual vector space

Formally, many texts (such as Hassani (2013)) define what's known as the dual vector space \mathcal{V}^* consisting of the collection of all linear functionals¹ on vectors.

¹Traditionally, much of abstract linear algebra is useful in the definition and analysis of infinite-dimensional spaces of functions. Vectors in these spaces are functions, so functions

These functions of vectors $\phi : \mathcal{V} \rightarrow \mathbb{R}$ are linear in the same sense as more general linear transforms $\phi(\alpha|a\rangle + \beta|b\rangle) = \alpha\phi|a\rangle + \beta\phi|b\rangle$.

A more rigorous text would then go through the machinery of proving that this dual space \mathcal{V}^* of linear functionals on \mathcal{V} is also a vector space, and indeed a finite-dimensional space of the same dimension n as the *primal* space \mathcal{V} . Moreover, given an inner product on the vector space, it turns out that these linear functions are actually the same functions as those defined by $\phi_a(\cdot) = \langle a | \cdot \rangle$ for some $|a\rangle$. In other words, a function on vectors can be linear only if its evaluation is really just an inner product between its argument as some predefined vector. Every linear functional ϕ_a defined on this vector space takes the form $\phi_a(|x\rangle) = \langle a | x \rangle$ for some $|a\rangle \in \mathcal{V}$.

This observation, that all linear functionals are just functions defined by taking an inner product with a fixed vector, leads to the convention of denoting each of the linear functionals ϕ_a by a notational modification of its associated vector $|a\rangle$. We write $\langle a | \in \mathcal{V}^*$ (called a “bra”) as a representation of the linear functional. That notation has the convenient property that the evaluation of the functional ϕ_a represented by bra $\langle a |$ is given by

$$\phi_a(|x\rangle) = \langle a | x \rangle. \tag{5}$$

See? A bra and ket come together to make a bra(c)ket. Right? And there we have it. The foundational notation of this form of abstract linear algebra is built on a joke... because physicists don’t get enough sleep.

For our purposes, rather than thinking of the dual space as a space of linear functionals, we’ll just treat them as a space of “transposed” vectors. For every ket $|a\rangle \in \mathcal{V}$ there is a corresponding bra $\langle a | \in \mathcal{V}^*$, and vice versa, with the understanding that bras $\langle a |$ and kets $|b\rangle$ operate on each other through inner product evaluation $\langle a | b \rangle$.

3 The nice properties of orthogonal bases

Now that we have a definition of angle (inner product), we can talk about special bases whose elements in some way least interfere with each other. One such basis is the **orthonormal basis**. An orthonormal basis $\mathcal{B} = \{|e_i\rangle\}_{i=1}^n$ is a basis whose elements are all normalized and mutually orthogonal. We typically write this relation as

$$\langle e_i | e_j \rangle = \delta_{ij}, \tag{6}$$

for each pair of vectors $|e_i\rangle, |e_j\rangle \in \mathcal{V}$. Here δ_{ij} is the Kronecker delta, which is 0 when $i \neq j$ and 1 when $i = j$. Specifically, this relation compactly encodes

of vectors would be functions of functions, and hence functionals. That terminology has carried through to the formal definition of finite-dimensional abstract vector spaces, too. So even though coefficient representations of finite-dimensional vectors are often associated with points in \mathbb{R}^n , we still call functions of vectors *functionals* in this context rather than simply *functions*.

both that $\|e_i\| = 1$ for each $|e_i\rangle$ (i.e. the vectors are normalized), and that $\langle e_i | e_j \rangle = 0$ when $|e_i\rangle \neq |e_j\rangle$ (i.e. they're mutually orthogonal).

Why do we care? One reason is that the coordinates in this basis are really easy to find. We can see this by considering the basis expansion of a vector $|v\rangle = \sum_{j=1}^n \alpha_j |e_j\rangle$. Since the basis is orthonormal, the inner product between that vector and a basis element is

$$\begin{aligned} \langle e_i | v \rangle &= \langle e_i | \sum_{j=1}^n \alpha_j |e_j\rangle = \sum_{j=1}^n \alpha_j \underbrace{\langle e_i | e_j \rangle}_{\delta_{ij}} \\ &= \alpha_i. \end{aligned} \tag{7}$$

This means we can calculate the coefficients of an orthonormal basis expansion simply by taking the inner product between the vector and the basis elements $\alpha_i = \langle e_i | v \rangle$.

Since the full component of the vector expansion along basis element $|e_i\rangle$ is given by $\alpha_i |e_i\rangle = (\langle e_i | v \rangle) |e_i\rangle$, we often rewrite the expression as

$$\alpha_i |e_i\rangle = \underbrace{(|e_i\rangle \langle e_i|)}_{\text{outer product}} |v\rangle, \tag{8}$$

where now we've introduced an operator, referred to as an **outer product** of a ket and a bra $P_i = |e_i\rangle \langle e_i|$, with the understanding that it operates on vectors as $(|e_i\rangle \langle e_i|) |v\rangle = \langle e_i | v \rangle |e_i\rangle$.

Using this notation, we can now give a representation of the **identity operator** in terms of the orthonormal basis elements $\{|e_i\rangle\}$:

$$\mathcal{I} = \sum_{i=1}^n P_i = \sum_{i=1}^n |e_i\rangle \langle e_i|. \tag{9}$$

Using the operational definition of these outer products given above, we can easily verify that multiplying a vector $|v\rangle$ by this identity operator just gives back the basis expansion of $|v\rangle$:

$$\begin{aligned} \mathcal{I}|v\rangle &= \sum_{i=1}^n (|e_i\rangle \langle e_i|) |v\rangle = \sum_{i=1}^n \langle e_i | v \rangle |e_i\rangle \\ &= |v\rangle. \end{aligned} \tag{10}$$

3.1 Coordinate representation and inner product preservation

We're now equipped to examine what dot products between coefficient representations look like in orthonormal basis expansions. Consider two vectors $|a\rangle$

and $|b\rangle$. The dot product between the vectors of coefficients is

$$\begin{bmatrix} \langle e_1 | a \rangle \\ \vdots \\ \langle e_n | a \rangle \end{bmatrix} \cdot \begin{bmatrix} \langle e_1 | b \rangle \\ \vdots \\ \langle e_n | b \rangle \end{bmatrix} = \sum_{i=1}^n \langle e_i | a \rangle \langle e_i | b \rangle = \sum_{i=1}^n \langle a | (|e_i\rangle \langle e_i|) | b \rangle \quad (11)$$

$$= \langle a | \underbrace{\left(\sum_{i=1}^n |e_i\rangle \langle e_i| \right)}_{\mathcal{I}} | b \rangle = \langle a | b \rangle. \quad (12)$$

In other words, **orthonormal coefficient representations are angle preserving** in the sense that the dot product between two coefficient vectors of expansions in an orthonormal basis correctly reflects the underlying inner product of the abstract vector space.

This is the same phenomenon we saw in the simple example of \mathbb{R}^2 above. There, dot products correctly encode angle because we implicitly assumed that the basis representation is orthogonal with respect to the underlying notion of inner product in the two-dimensional vector space (i.e. that our coordinate axes were at right angles to one another). Here we see that this relationship is true in general. Once we define the inner product on a vector space and choose a basis representation that's orthonormal with respect to that inner product, simple dot products between coefficient vectors correctly reflect the underlying notion of angle defined for the space.

3.2 Projections onto orthogonal vectors

Projecting onto the span of a set of orthonormal vectors $\mathcal{B}_{//} = \{|e_i\rangle\}_{i=1}^K$ where $K \leq n$ is easy. How can we see that? Lets take a look at it intuitively first. We can complete the collection of vectors by adding $\mathcal{B}_{\perp} = \{|e_i\rangle\}_{i=K+1}^n$ so that the full set of n vectors becomes an orthonormal basis $\mathcal{B}_{\text{full}} = \mathcal{B}_{//} \cup \mathcal{B}_{\perp}$. Every vector in the span of $\mathcal{B}_{//}$ is necessarily orthogonal to any vector in the span of \mathcal{B}_{\perp} . When we find the basis representation of a vector $|v\rangle = \sum_{i=1}^n \alpha_i |e_i\rangle$, we are automatically finding a decomposition of the vector in terms of a portion $|v_{//}\rangle = \sum_{i=1}^K \alpha_i |e_i\rangle$ in the span of $\mathcal{B}_{//}$ and a portion $|v_{\perp}\rangle = \sum_{i=K+1}^n \alpha_i |e_i\rangle$ in the span of \mathcal{B}_{\perp} . Since $|v_{\perp}\rangle$ is orthogonal to the entire linear space $\mathcal{B}_{//}$, it must be the smallest “residual” vector we can possibly make. For all other possible residual vectors, the residual is a combination of $|v_{\perp}\rangle$ and some other component $|r_{//}\rangle \in \mathcal{B}_{//}$, and it's total norm will necessarily be larger than the norm of just $|v_{\perp}\rangle$ alone (see Figure 1).

We can more rigorously prove it by analytically solving for the coefficients of the closest vector within the span of $\mathcal{B}_{//}$

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2} \left\| |v\rangle - \sum_{i=1}^K \alpha_i |e_i\rangle \right\|^2, \quad (13)$$

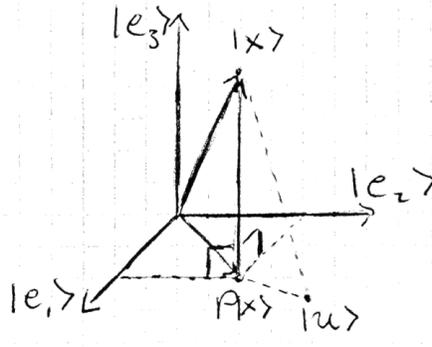


Figure 1: Depiction of a projection of a vector $|x\rangle$ onto a space spanned by orthonormal vectors $|e_1\rangle$ and $|e_2\rangle$. Since the vectors are orthonormal, the projection is simply the components of $|x\rangle$ along the two vectors. $|u\rangle$ depicts why, intuitively, other vectors in the span must have larger residual.

but we leave this as an exercise to the reader. The bottom line is that it's true: The projection of $|v\rangle$ onto a collection of orthonormal vectors $\mathcal{B}_\mathcal{V}$ is given by

$$P_{\mathcal{B}_\mathcal{V}} |v\rangle = \underbrace{\left(\sum_{i=1}^K |e_i\rangle \langle e_i| \right)}_{\text{Projection operator}} |v\rangle. \quad (14)$$

3.3 Matrix elements of linear transforms in orthogonal bases

There's one final reason why we like orthogonal bases that we want to mention. Generally, finding the matrix representation of a linear transform is kind of a pain. We need to push each of the basis vectors through the transform and then find the coefficients of their representation in the co-domain's basis (see Ratliff (2014)). But we just saw above in Section 3 that finding a coefficient in an orthonormal basis is easy. You simply take the inner product between the vector and the basis element. Putting this all together, that means finding the elements of the matrix representation of a linear transform should be relatively easy in orthogonal bases.

Consider a linear transform $T : \mathcal{V} \rightarrow \mathcal{U}$ mapping from an n -dimensional vector space \mathcal{V} to an m -dimensional vector space \mathcal{U} . Let $\mathcal{B}_\mathcal{V} = \{|a_i\rangle\}_{i=1}^n$ be an orthonormal basis for \mathcal{V} and let $\mathcal{B}_\mathcal{U}$ be an orthonormal basis for \mathcal{U} . Each basis element in the domain $|a_j\rangle \in \mathcal{B}_\mathcal{V}$ is mapped to a vector $|v_j\rangle = T|a_j\rangle$ by the linear transform. The coefficients of that vector in the co-domain's basis are simply $\langle b_i | v_j \rangle$ since the co-domain's basis $\mathcal{B}_\mathcal{U}$ is orthonormal. Putting this all together, we see that the matrix representation of the linear transform is given

by²

$$\mathbf{A} = \begin{bmatrix} \langle b_1|T|a_1\rangle & \langle b_1|T|a_2\rangle & \cdots & \langle b_1|T|a_n\rangle \\ \langle b_2|T|a_1\rangle & \langle b_2|T|a_2\rangle & \cdots & \langle b_2|T|a_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle b_m|T|a_1\rangle & \langle b_m|T|a_2\rangle & \cdots & \langle b_m|T|a_n\rangle \end{bmatrix}. \quad (15)$$

We can get the ij^{th} matrix element by just multiplying T on the left by $\langle b_i|$ and on the right by $|a_j\rangle$.

4 The geometry of linear transformations

In order to understand the behavior of operators constructed from outer products of orthonormal basis vectors $\mathcal{B}_{\mathcal{V}} = \{|e_i\rangle\}_{i=1}^n$, it's usually algebraically helpful to expand the vector that we're transforming in terms of those basis elements. For instance, suppose we have simply a projection operator that projects onto the space spanned by just the first K basis vectors $P_{1:K} = \sum_{i=1}^K |e_i\rangle\langle e_i|$. Using the expansion $|v\rangle = \sum_{i=1}^n \alpha_i |e_i\rangle$, we can pretty easily see that

$$\begin{aligned} P_{1:K}|v\rangle &= \left(\sum_{i=1}^K |e_i\rangle\langle e_i| \right) \sum_{j=1}^n \alpha_j |e_j\rangle \\ &= \sum_{i=1}^K \sum_{j=1}^n \underbrace{\langle e_i | e_j \rangle}_{\delta_{ij}} \alpha_j |e_i\rangle \\ &= \sum_{i=1}^K \alpha_i |e_i\rangle, \end{aligned} \quad (16)$$

as we suspected. Each outer product term $|e_i\rangle\langle e_i|$ simply selects the component of the input vector corresponding to $\langle e_i|$.

An interesting form of operator that we might consider is one that selects each of these components, but additionally scales them by some factors $\sigma_i \in \mathbb{R}$. We can consider this operation manually. Define a very simple operator T_{σ} that, for a given orthonormal basis $\mathcal{B}_{\mathcal{V}} = \{|e_i\rangle\}_{i=1}^n$, simply scales each of the components of a vector by the values σ_i . Specifically, if $|v\rangle = \sum_{i=1}^n \alpha_i |e_i\rangle$, this operator acts as follows

$$T_{\sigma}|v\rangle = \sum_{i=1}^n (\sigma_i \alpha_i) |e_i\rangle. \quad (17)$$

² Here, we use the notation $\langle b|T|a\rangle$ to denote the application of the bra $\langle b|$ to the transformed ket $T|a\rangle$. We haven't discussed the application of linear transforms on bras, but there's a way to apply the linear transform to the bra, denoted $\langle b|T$, to produce a transformed bra that's analogous to multiplying a matrix on the left with a transposed vector to produce another transposed vector. If we go through the motions of defining that fully, we'd see that we get the same result regardless of whether we apply the bra $\langle b|$ to the transformed ket $T|a\rangle$ or the transformed bra $\langle b|T$ to the original ket $|a\rangle$. Since it doesn't matter which side we transform first, we typically write it symmetrically, without parentheses, as we do here.

It literally just scales the i^{th} component by σ_i . How might we represent this operator? We saw before that an outer product term $|e_i\rangle\langle e_i|$ simply selects the i^{th} component of the vector and adds it to the output. What happens if we scale that outer product term by σ_i ? Specifically, consider an operator defined by $\sum_{i=1}^n \sigma_i |e_i\rangle\langle e_i|$. Such an operator acts on a vector $|v\rangle$ as

$$\begin{aligned} \left(\sum_{i=1}^n \sigma_i |e_i\rangle\langle e_i| \right) |v\rangle &= \sum_{i=1}^n \sigma_i \langle e_i | v \rangle |e_i\rangle \\ &= \sum_{i=1}^n (\sigma_i \alpha_i) |e_i\rangle, \end{aligned} \tag{18}$$

since $\alpha_i = \langle e_i | v \rangle$. The behavior of this operator is exactly the same operator as T_σ . Thus, we can conclude that T_σ can be algebraically represented as a sum of scaled outer product terms. Each outer product term $|e_i\rangle\langle e_i|$ (independently of all other terms, since they're formed from an orthonormal basis) selects the $|e_i\rangle$ component of the input vector, and stretches it by σ_i .

We can construct a whole collection of nice operators simply by laying down an orthonormal bases and defining how the space stretches (or reflects if $\sigma_i < 0$) along their axes. This seemingly simplistic construction runs even deeper than one might expect. One of the fundamental results of linear algebra demonstrates that, in fact, **all linear symmetric operators $T : \mathcal{V} \rightarrow \mathcal{V}$ can be represented in this form**. Here we give a somewhat vague definition of what it means to be a symmetric operator and refer instead to Hassani (2013) for a more complete definition.³ For our purposes, we simply draw an analogy to symmetric matrices. We consider an operator to be symmetric if all matrix representations, using the same basis for both the domain and co-domain, are symmetric.

Theorem 1. Spectral decomposition of an operator. *For every symmetric operator $T : \mathcal{V} \rightarrow \mathcal{V}$ there exists an orthonormal basis $\mathcal{B}_\mathcal{V} = \{|e_i\rangle\}_{i=1}^n$ such that the operator can be represented as*

$$T = \sum_{i=1}^n \lambda_i |e_i\rangle\langle e_i|, \tag{19}$$

for some scalars $\lambda_i \in \mathbb{R}$.

We denote the scalars here as λ_i rather than σ_i to emphasize their connection to Eigenvalues. Eigenvectors and Eigenvalues are typically developed axiomatically as the collection of vectors that don't change direction when transformed. That development is sound and mathematically nice, but it typically takes a while to get from there to the final reveal of the fundamental underlying structure of all linear operators in their spectral decomposition theorem. And by then,

³More formal definitions refer to a symmetric operator as being "self-Adjoint".

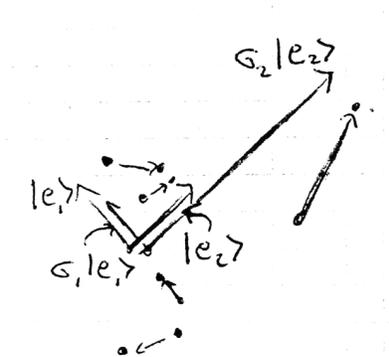


Figure 2: This figure depicts the stretching of a space along Eigenvectors of an operator along with the corresponding movement of select points in the space.

we, the exhausted readers, are usually already lost in a forest of intermediate results and details. In this document, rather than formally, building to this conclusion in a rigorous manner, we simply state it outright so that the reader can, as quickly as possible, understand what fundamental structure underlying all linear operators looks like. Underneath all the confusing details, every linear operator has a very simple form. There is always an orthogonal basis along which the operator simply stretches (and/or reflects) by the values λ_i . It so happens that we often decide to call these basis vectors $|e_i\rangle$ *Eigenvectors* and the scaling values λ_i *Eigenvalues*, but for our purposes what's more fundamental is the structure they entail.

The need for mathematical rigor prevents most mathematicians from talking about something until it's been proven from basic principles, and unfortunately that neurosis too often bleeds into pedagogical expositions obfuscating some of the most beautiful mathematical results in a mess of details. Regardless of whether you decide you can talk about it or not, it's still true. All linear operators are just stretchings along orthonormal bases. That's a very deep mathematical phenomenon; it should be the first thing that comes to mind when you think of linear operators, not the last.

4.1 A generalization to arbitrary linear transforms

We've actually been overly restrictive. This property holds more generally for any linear transform $T : \mathcal{V} \rightarrow \mathcal{U}$ between distinct linear spaces \mathcal{V} and \mathcal{U} of dimensions n and m , respectively. To see how, consider choosing orthonormal bases for both the domain $\mathcal{B}_{\mathcal{V}} = \{|a_i\rangle\}_{i=1}^n$ and the co-domain $\mathcal{B}_{\mathcal{U}} = \{|b_i\rangle\}_{i=1}^m$, and then pairing up the first K of those basis elements. $|a_1\rangle$ is paired with $|b_1\rangle$, $|a_2\rangle$ is paired with $|b_2\rangle$, and so on up to K . Within the span of these two K -dimensional subspaces, since we're dealing with orthonormal bases, the geometry within the spaces of coefficients actually means something (as we saw

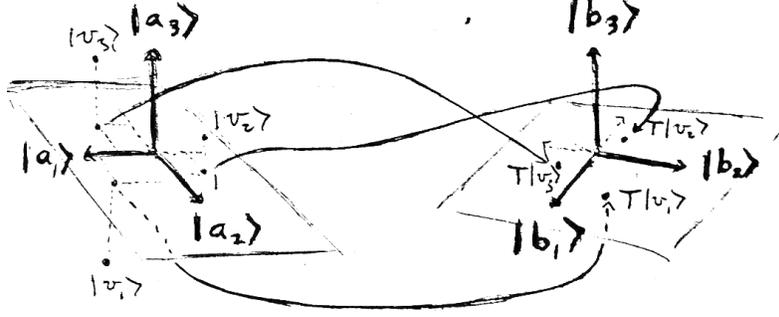


Figure 3: A depiction of the two-dimensional fundamental spaces of a rank 2 linear transform T mapping from its 3-dimensional domain \mathcal{V} on the left to its 3-dimensional co-domain \mathcal{U} on the right. Null space components input vectors are clipped off and the component lying within the fundamental subspace of the domain is transported to fundamental subspace of the co-domain.

above in Section 3.1). The coefficients are fully representative of the space. Nothing we do in one space differs in any way from what we do in the other space, as long as the coefficients are the same. It's as though, with this pairing of basis vectors, we've sewn the two spaces together and declared them to be one and the same.

Thus, any vector $|v\rangle = \sum_{i=1}^K \alpha_i |a_i\rangle$ in \mathcal{V} living in the span of the first K basis vectors can be transported trivially to \mathcal{U} by simply placing those coefficients $\alpha_1, \dots, \alpha_K$ on the first K basis vectors of \mathcal{U} to form a map

$$\sum_{i=1}^K \alpha_i |a_i\rangle \rightarrow \sum_{i=1}^K \alpha_i |b_i\rangle. \quad (20)$$

This *identification map*, which we'll denote as A , can be represented algebraically in a very similar form to what we saw above with the identity map:

$$A = \sum_{i=1}^K |b_i\rangle \langle a_i|. \quad (21)$$

When we apply this operator to a vector $|v\rangle = \sum_{i=1}^n \alpha_i |a_i\rangle$, we see that, operationally, for each term, it finds the coefficient $\alpha_i = \langle a_i | v \rangle$ along $|a_i\rangle$ and then simply places that coefficient on $|b_i\rangle$:

$$A|v\rangle = \sum_{i=1}^K \left(|b_i\rangle \langle a_i| \right) |v\rangle = \sum_{i=1}^K \langle a_i | v \rangle |b_i\rangle \quad (22)$$

$$= \sum_{i=1}^K \alpha_i |b_i\rangle. \quad (23)$$

Thus, we have exactly the same situation as we had before with operators and projections, except now we can transport vectors from one space to another by identifying corresponding basis elements. It reasons that we can additionally scale each one of these outer product terms by σ_i to also stretch the space along these corresponding axes during the transform. This augmentation gives us an operator of the form

$$A_\sigma = \sum_{i=1}^K \sigma_i |b_i\rangle\langle a_i|. \quad (24)$$

This operator glues two K -dimensional subspaces together by finding corresponding pairs of orthonormal vectors, and then treats the combined space in exactly the same way we did with the single space in the case of operators.

Perhaps even more fundamental than the above spectral decomposition theorem is the theorem that states that, miraculously, actually all linear transforms between two spaces can be represented in the form given by Equation 24. We state it here, again without proof, with the hope that this picture of orthonormal basis pairing and stretching should be the first thing that pops into your mind when you think of a linear transform.

Theorem 2. The Fundamental Theorem of Linear Algebra. *Let \mathcal{V} be an n -dimensional vector space and \mathcal{U} be an m -dimensional vector space. Then for any linear transform $T : \mathcal{V} \rightarrow \mathcal{U}$ between the spaces, there exists a set of $K \leq n, m$ mutually orthonormal vectors $\{|a_i\rangle\}_{i=1}^K$ in the domain \mathcal{V} and K mutually orthonormal vectors $\{|b_i\rangle\}_{i=1}^K$ in the co-domain \mathcal{U} such that we can represent the transform by*

$$T = \sum_{i=1}^K \sigma_i |b_i\rangle\langle a_i|, \quad (25)$$

for some scalars $\sigma_i > 0 \in \mathbb{R}$.

Here, we specify $\sigma_i \neq 0$ because we both emphasize that K has specific meaning in the map, as we see in Section 4.2. We also note that since the domain and co-domain's K orthonormal vectors need not be the same (indeed they're often in different spaces), we can always choose them so that all the values σ_i are positive.

Similar to the case of Eigenvalues and Eigenvectors, we typically give these these vectors and values names. We call the values σ_i **singular values** and the vectors $\{|b_i\rangle\}$ and $\{|a_i\rangle\}$ the left and right singular vectors, respectively.

4.2 Fundamental spaces of a linear transform

Importantly, this representation given in Equation 25 directly encodes many of the properties of a linear operator. First, the number of outer product terms

$K \leq m, n$ defines what we call the **rank** of the operator. It is the dimension of the image of the domain under the transformation. This property is easy to see because we know both that any output of the transformation must be in the span of $\{|b_i\rangle\}_{i=1}^K$, and we can construct any linear combination of those vectors $|u\rangle = \sum_{i=1}^K \beta_i |b_i\rangle$ by inputting the vector $|v\rangle = \sum_{i=1}^K \frac{\beta_i}{\sigma_i} |a_i\rangle$.

Moreover, this analysis already shows that between the subspaces defined by the corresponding sets of K orthonormal vectors, which we call the left and right **fundamental subspaces**, the mapping is really a bijection (i.e. invertible). If a vector $|v\rangle = \sum_{i=1}^K \alpha_i |a_i\rangle$ is already in the span of the K fundamental basis vectors, it's simply mapped to $T|v\rangle = \sum_{i=1}^K \sigma_i \alpha_i |b_i\rangle$, and we've lost no information. We can get back by simply dividing each coefficient by σ_i and replying the coefficients to the original basis vectors $|a_i\rangle$. The inverse map that performs this operation is

$$T^{-1} = \sum_{i=1}^K \frac{1}{\sigma_i} |a_i\rangle \langle b_i|. \quad (26)$$

This transform just performs the opposite change of basis and undoes the scaling by σ_i .

Good, so we now know that every linear map of rank $K \leq m, n$ has a pair of fundamental K -dimensional subspaces on which the map is a bijection with inverse given by Equation 26. What happens to vectors that are off those fundamental subspaces?

The first part of the operation is to find the components of the input vector in the K orthonormal basis elements $\{|a_i\rangle\}_{i=1}^K$. As we saw above, since these vectors are mutually orthonormal, this operation is equivalent to projecting the point onto the their span. In general, the operation of the linear transform is decompose the vector into two components $|v\rangle = |v_{\parallel}\rangle + |v_{\perp}\rangle$, one within the span of the fundamental subspace and the other perpendicular to that span, and then simply throw away the orthogonal component $|v_{\perp}\rangle$ and transform what remains using the bijective portion of the map.

In this case, we do lose information, we've throw away part of the vector. Many texts refer to this operation of throwing away the perpendicular components as a projection as we did in Section 3.2. But it's useful to think of it as simply an inability to encode that orthogonal information. Fundamentally, the linear transform is constructed of only K outer product terms; it simply can't represent anything within that orthogonal span, so it has no choice but to throw it away.

Similarly, these arguments suggest that Equation 26 also acts as a useful **pseudoinverse** of sorts. Within the fundamental subspace of the co-domain, it's a bijective mapping to the fundamental subspace of the domain. But off that fundamental subspace, it simply doesn't have the capacity to represent anything perpendicular to that fundamental subspace, so it simply throws away those portions of the vectors. It's an interesting exercise to see that, in coordinates, this natural pseudoinverse is actually the familiar Moore-Penrose pseudoinverse.

Now, what are these spaces orthogonal to the K -dimensional fundamental subspaces in the domain and co-domain. The domain \mathcal{V} is an n -dimensional vector space, so there are $n - K$ dimensions orthogonal to the fundamental subspace. It's easy to see that any vector within this $(n - K)$ -dimensional subspace orthogonal to the fundamental subspace is sent to zero by the linear transform. Intuitively, as we discussed above, those orthogonal portions are simply thrown away, so they must be sent to the zero vector. But more concretely, we can construct a collection of basis vectors $|a_i\rangle_{i=K+1}^n$ spanning that orthogonal space. Since they're all orthogonal to each of the vectors $\{|a_i\rangle_{i=1}^K\}$, all of the inner products $\langle a_i | a_j \rangle$ must be zero for $i = 1, \dots, K$ and $j = K + 1, \dots, n$. So it must be that

$$\begin{aligned} T \sum_{j=K+1}^n \alpha_j |a_j\rangle &= \left(\sum_{i=1}^K \sigma_i |b_i\rangle \langle a_i| \right) \left(\sum_{j=K+1}^n \alpha_j |a_j\rangle \right) \\ &= \sum_{i=1}^K \sum_{j=K+1}^n \sigma_i \alpha_j \underbrace{\langle a_i | a_j \rangle}_{=0} |b_i\rangle = |0\rangle. \end{aligned} \quad (27)$$

Thus, this $(n - K)$ -dimensional subspace orthogonal to the K -dimensional fundamental space must be what we call the **(right) Null space** of T . It's the space of all vectors sent to the zero vector by the linear transform T .

Similarly, in the opposite direction, we see that the $(m - K)$ -dimensional subspace orthogonal to the left fundamental subspace in \mathcal{U} is also the space sent to zero by the pseudoinverse⁴ defined above in Equation 26. So it, too, is what we know of as the **(left) Null space**.

So simply from the underlying structure of the linear transform, as exposed by the Fundamental Theorem of Linear Algebra, we immediately can see that a linear transform of rank K forms a bijection between left and right fundamental subspaces, each of dimension K , and has associated left and right Null spaces of dimensions $(m - K)$ and $(n - K)$, respectively. Moreover, we can easily write down a pseudoinverse for that extends the inverse defined on the left fundamental subspace to the rest of the co-domain exposing an intrinsic symmetry in the action of the linear transform.

5 Manifestations of structure in coordinates

All of the above discussion has been abstract, focusing on the theoretical structure underlying any linear transform. But thus far, a lot of this structure has been intangible. Nothing in that we've covered so far has revealed how we might access that structure. This section remedies that.

⁴Formally, we define the left Null space in terms of the vectors that are sent to zero under the *adjoint* of the transform T^\dagger , but we haven't discussed the notion of an adjoint transform or the "dagger" operation, so we discuss it here in terms of the natural pseudoinverse, which is both intuitive and equivalent for these purposes. See Hassani (2013) for details.

As always, when we want linear transforms to become tangible, we need to choose bases and calculate the matrix of the linear transform. Suppose we're given an orthonormal basis for the domain $\mathcal{B}_{\mathcal{V}} = \{|v_i\rangle\}_{i=1}^n$ and an orthonormal basis for the co-domain $\mathcal{B}_{\mathcal{U}} = \{|u_i\rangle\}_{i=1}^m$. We saw in Section 3.3 that under orthonormal bases the ij^{th} element of the matrix \mathbf{A} is given by $A_{ij} = \langle u_i | T | v_j \rangle$. Using the expansion in Equation 25 we see that the ij^{th} matrix element takes the form

$$\begin{aligned} A_{ij} &= \langle u_i | \left(\sum_{k=1}^K \sigma_k |b_k\rangle \langle a_k| \right) | v_j \rangle = \sum_{k=1}^K \sigma_k \langle u_i | b_k \rangle \langle a_k | v_j \rangle \\ \Rightarrow \mathbf{A} &= \sum_{k=1}^K \sigma_k \mathbf{u}_k \mathbf{v}_k^T, \end{aligned}$$

where

$$\mathbf{u}_k = \begin{bmatrix} \langle u_1 | b_k \rangle \\ \langle u_2 | b_k \rangle \\ \vdots \\ \langle u_m | b_k \rangle \end{bmatrix} \quad \text{and} \quad \mathbf{v}_k = \begin{bmatrix} \langle v_1 | a_k \rangle \\ \langle v_2 | a_k \rangle \\ \vdots \\ \langle v_n | a_k \rangle \end{bmatrix}.$$

Note that \mathbf{u}_k is the coordinate vector of $|b_k\rangle$ represented in the co-domain basis $\mathcal{B}_{\mathcal{U}}$. As discussed in Section 3.1, since this basis is orthonormal, simple dot products between coefficient vectors in this basis correctly calculate the inner product between the underlying vectors. In this case, that tells us the coefficient vectors \mathbf{u}_k are all mutually orthonormal with respect to the dot product since they are coordinate representations of mutually orthonormal vectors. Similarly, the vectors \mathbf{v}_k are coordinate representations of $|a_k\rangle$ in the chosen domain basis $\mathcal{B}_{\mathcal{V}}$, which means they're all mutually orthonormal as well with respect to the simple dot product.

Collecting these vectors up into full matrix form gives

$$\mathbf{A} = \mathbf{U} \mathbf{S} \mathbf{V}^T$$

where

$$\mathbf{U} = \left[\begin{array}{c|ccc|c} & & & & \\ \mathbf{u}_1 & \dots & \mathbf{u}_K & & \\ & & & & \end{array} \right] \quad \mathbf{S} = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_K & \end{pmatrix} \quad \mathbf{V} = \left[\begin{array}{c|ccc|c} & & & & \\ \mathbf{v}_1 & \dots & \mathbf{v}_K & & \\ & & & & \end{array} \right]$$

This representation is the familiar Singular Value Decomposition (SVD) of the matrix. Since the SVD is unique, these observations tell us that embedded within the SVD of a matrix are the coordinate representations \mathbf{u}_i and \mathbf{v}_j of the left and right underlying singular vectors $|b_i\rangle$ and $|a_j\rangle$ along with the same singular values σ_i found in Equation 25. Computational methods for calculating the SVD expose the fundamental structure of the underlying abstract linear transform in the chosen coordinates.

References

- Hassani, Sadri. *Mathematical Physics: A Modern Introduction to its Foundations*. Springer, 2nd edition, 2013.
- Ratliff, Nathan. Linear algebra I: Matrix representations of linear transforms, 2014. Lecture notes: Mathematics for Intelligent Systems series.