Introduction to Linear Algebra

Vishnu Iyer

October 30, 2016

Contents

1 Vectors and Vector Spaces

1.1 Vector Spaces

We begin our study of Linear Algebra by exploring the concept of a vector. You may recall from prior experience (especially in physics) that a vector is a mathematical object with magnitude and direction. Indeed, such an object is a vector, but it is hardly the definition of one. We abandon this definition for a more general one: a **vector** is an object which belongs to a structure known as a **vector space**. More formally, we give the following definition:

Definition. *A set of vectors V is a vector space if, for $\mathbf{u}$ $\mathbf{v}$ and $\mathbf{w} \in V$, the following axioms hold:*

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(ii) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$

(iii) One can define a unique zero vector $\mathbf{0}$ such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$

(iv) For each $\mathbf{u}$, there exists an inverse $-\mathbf{u}$ such that $\mathbf{u} + -\mathbf{u} = \mathbf{0}$ and $-\mathbf{u} = -1\mathbf{u}$

(v) For scalars c, d: $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

(vi) For a scalar c: $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

(vii) For scalars c, d: $c(d\mathbf{u}) = (cd)\mathbf{u}$

(viii) $1\mathbf{u} = \mathbf{u}$

(ix) V is closed under addition. That is, $\mathbf{u}, \mathbf{v} \in V \implies \mathbf{u} + \mathbf{v} \in V$.

(x) V is closed under scalar multiplication. That is, for scalar c, $\mathbf{u} \in V \implies cu \in V$.

Indeed, one can see that this extends the concept of vectors beyond the idea of a quantity with magnitude and direction. For example, consider the space of continuous functions over the real numbers. With a little insight, one can see that these, too form a vector space. A common example we will consider in this book is $\mathbb{R}^n$, or the collection of n-tuples of real numbers. This is often what comes to mind when referring to vector spaces, as its geometric description is rather intuitive (each element can be represented by an arrow in n-dimensional space pointing from the origin to the point with coordinates described by the n-tuple).

1.2 Subspaces

Often, we want to consider subsets of vectors within a larger vector space. A key notion of vector spaces within contained in other vector spaces can be found in the **subspace**.

Definition. *A subspace $S \subseteq V$ is a subset of V that satisfies the following:*

(i) S is nonempty. That is, $S \neq \emptyset$.

(ii) S is closed under addition. That is, $\mathbf{u}, \mathbf{v} \in S \implies \mathbf{u} + \mathbf{v} \in S$.

(iii) V is closed under scalar multiplication. That is, for scalar c, $\mathbf{u} \in V \implies cu \in S$.

Theorem 1.1. *The zero vector is necessarily contained in every subspace.*

Proof. Since every vector space is nonempty, it has at least one vector (call it $\mathbf{u}$). Consider the vector $0\mathbf{u}$. We know that $\mathbf{u} + -\mathbf{u} = (1 + -1)\mathbf{u} = 0\mathbf{u} = \mathbf{0}$ (by vector axioms iv and v). Therefore, for the space to be closed under scalar multiplication, $\mathbf{0}$ must be contained in it. □

Immediately, one can see that the set only containing the zero vector is a subspace. Furthermore, a vector space is always a subset of itself, since the vector space axioms contain the subspace axioms. These subspaces are often known as the **trivial subspaces**.

2 Eigenvalues and Eigenvectors

2.1 Introduction

It is often useful to think of vectors, which, when operated upon by a linear transformation yield some scalar multiple of themselves. A clear example is a vector in the kernel of a linear transformation, whose image is obviously 0 times itself. However, there are often cases where certain vectors become nonzero multiples of themselves after being operated upon by a given linear transformation. We call these **eigenvectors** of the linear transformation, and the factor by which its image is scaled is known as its **eigenvalue**. Note that this implies that the domain and codomain of the transformation are of the same dimension.

Definition. *An eigenvector of a linear transformation $T : V \to V$ is a nonzero vector $\mathbf{u}$ such that $T(\mathbf{u}) = \lambda\mathbf{u}$, where λ is some scalar, known as an eigenvalue of T associated with eigenvector $\mathbf{u}$.*

Indeed, there are several problems to which the theory of eigenvalues and eigenvectors can be applied effectively.

2.2 Characteristic Polynomials

Given the definition of eigenvectors, a question is imminent: given a linear transformation, how can one determine its eigenvalues and eigenvectors? Assume that $\mathbf{u}$ is an eigenvector of some linear transformation T and λ is its associated eigenvalue. Then, by definition, we have:

$$T(\mathbf{u}) = \lambda\mathbf{u}$$

$$T(\mathbf{u}) - \lambda Id(\mathbf{u}) = \mathbf{0}$$

Since T and Id are linear, we can reduce this system to:

$$(T - \lambda Id)(\mathbf{u}) = \mathbf{0}$$

Recall that, by definition, $\mathbf{u}$ must be nonzero. Therefore, the kernel of the linear transformation $T - \lambda Id$ has a nonzero element and is not invertible. We are now left with the question: for what values of λ is $T - \lambda Id$ not invertible? The answer is simple if we consider the standard matrices of T and Id, call them A and I, respectively. We must find the values of λ for which:

$$det(A - \lambda I) = 0$$

Thus, by the definition of the determinant of a matrix, we are left with a polynomial $p(\lambda) = det(A - \lambda I)$, which we call the **characteristic polynomial**.

Definition. *Given a linear transformation $T : V \to V$ and its standard matrix A, the characteristic polynomial is the polynomial given by $det(A - \lambda I)$, where I is the identity matrix.*

One can immediately see that the roots of the polynomial are exactly the eigenvalues of T. Now, the multiplicity of the roots of the polynomial may vary. This is often called the **algebraic multiplicity** of the eigenvalue corresponding to that root of the polynomial.

2.3 Eigenvectors

Now that we have a method of finding the eigenvalues of a linear transformation through the characteristic polynomial, we can proceed to find the corresponding eigenvectors. Given that for an eigenvector $\mathbf{u}$ of T with corresponding eigenvalue λ, $(T - \lambda Id)\mathbf{u} = 0$, we have that $\mathbf{u}$ is contained in the kernel of $T - \lambda I$. In fact, every member of this kernel is an eigenvector of T with eigenvalue λ. Therefore, the process of eigenvalue calculation can be reduced to the process of finding the kernel.

Since any member of the kernel of $T - \lambda Id$ is an eigenvalue and the kernel is a vector space, we can define an **eigenspace** of T.

Definition. *For a linear transformation $T : V \to V$, a subspace $U \subseteq V$ is an eigenspace (corresponding to the eigenvalue λ) if $\forall \mathbf{u} \in U$, $T(\mathbf{u}) = \lambda \mathbf{u}$*

We don't need to check that this is a subspace because, as mentioned previously, we can equate it to the kernel of a linear transformation, which we have already proven is a subspace.

Theorem 2.1. *Eigenvectors corresponding to distinct eigenvalues are linearly independent.*

Proof. In the case where one of the eigenvalues is zero, the proof is trivial. If the eigenvectors were linearly dependent, then the one with a nonzero eigenvalue would be part of the kernel of T, which means its eigenvalue is 0, which is a contradiction. Say the linear transformation T has two unique nonzero eigenvalues, λ_1 and λ_2, which correspond to eigenvectors **u** and **v**, respectively. Assume, for the sake of contradiction, that **u** and **v** are linearly dependent. That is, $\mathbf{u} = c\mathbf{v}$ for some scalar c. Then,

$$T(\mathbf{u}) = T(c\mathbf{v})$$

$$\lambda_1 \mathbf{u} = c\lambda_2 \mathbf{v}$$

$$\mathbf{u} = \frac{c\lambda_2}{\lambda_1}\mathbf{v}$$

Since $\lambda_1 \neq \lambda_2$, $\dfrac{c\lambda_2}{\lambda_1} \neq c$. However, this is a contradiction, because we assumed that $\mathbf{u} = c\mathbf{v}$. □

Since an eigenspace of T is is a subspace of V, some may be tempted to find a basis to represent it. Indeed, this idea is useful in many theoretical and practical applications. In fact, in the next section, we will investigate deeply the idea of an eigenvector basis. However, now we are just concerned about the *dimension* of said eigenspace. This quantity is often known as the **geometric multiplicity** of the eigenvalue.

Theorem 2.2. *The geometric multiplicity of an eigenvalue λ never exceeds its algebraic multiplicity.*

Proof. To show this, we will first prove the following lemma:

Lemma 2.2.1. *The eigenvalues of a matrix A are not changed when the matrix is operated upon by the elementary operations of row swapping and adding a scalar multiple of one row to another.*

Proof. We know that if λ is an eigenvalue of A, then $det(A - \lambda I) = 0$. □

□

2.4 Diagonalization

We have discussed change of basis of a linear transformation before. However, there are certain bases for which calculation of the linear transformation becomes very simple. One example is **diagonalization**.

Definition. *A linear transformation $T : V \to V$ is diagonalizable if there exists a basis for V consisting entirely of eigenvectors of T. We call this basis an eigenbasis.*

Why is this type of linear transformation diagonalizable? The answer is obvious when one considers the following alternative definition of a diagonalizable linear transformation.

Definition. *Let $T : V \to V$ be a linear transformation with standard matrix A. If $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D, then T called diagonalizable.*

Theorem 2.3. *The two above definitions of diagonalization are equivalent.*

Proof. Let $\lambda_1, \lambda_2, ...\lambda_n$ be the eigenvalues of the linear transformation T, and let $\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n$ be the corresponding eigenvectors. Let A be the standard matrix of T. Assume that $\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n$ are linearly independent. Then, by the first definition, T is diagonalizable. Now, consider the matrix $P = [\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n]$, which is invertible because its columns are linearly independent, and the diagonal matrix D with diagonal entries $\lambda_1, \lambda_2, ...\lambda_n$. Then, we submit that $A = PDP^{-1}$. To show this, we have:

$$\begin{aligned} det(A - \lambda_i I) &= det(PDP^{-1} - P\lambda_i P^{-1}) \\ &= det(P(D - \lambda_i I)P^{-1}) \\ &= det(P)det(P^{-1})det(D - \lambda_i I) \\ &= det(D - \lambda_i I) \end{aligned}$$

Now, since the last matrix has a 0 in the diagonal and the determinant of a diagonal matrix is the product of the diagonal entries, we have that $det(A - \lambda_i I) = 0$ for all eigenvalues of A. □

It is a simple calculation to prove that the matrix PDP^{-1} actually does evaluate to A.

3 Inner Products

3.1 Introduction

Some readers may be familiar with the dot product of two vectors, especially in $\mathbb{R}^n$. This calculation introduces several useful ideas, such as the concept of an angle between two vectors and whether or not the vectors are perpendicular to each other. In fact, the notion of a dot product, more formally called the **inner product**, can b

3.2 Inner Product Spaces

We begin our discussion with a formal definition of an inner product.

Definition. *An inner product in a vector space V (over some scalar field $\mathbb{F}$) is some function $< \cdot, \cdot >: V \times V \to \mathbb{F}$ which satisfies the following properties:*

(i) The inner product is conjugate symmetric. That is, $< \mathbf{u}, \mathbf{v} >=< \mathbf{v}, \mathbf{u} >^*$

(ii) $< \mathbf{u}, \mathbf{u} >\geq 0$*, with equality if and only if* $\mathbf{u} = \mathbf{0}$

(iii) The inner product is linear in the first argument. That is, $< c_1\mathbf{u} + c_2\mathbf{v}, \mathbf{w} >= c_1 < \mathbf{u}, \mathbf{w} > +c_2 < \mathbf{v}, \mathbf{w} >$.

One can easily verify that the dot product of two vectors in $\mathbb{R}^n$, defined as the sum of the products of the respective entries, satisfies all these axioms, and we leave this as an exercise to the reader.

Inner products are very useful within the context of vector spaces. We now define the term **inner product space**.

Definition. *An inner product space is a vector space equipped with an additional function which satisfies the properties of the inner product.*

As we have discussed previously, the vector space $\mathbb{R}^n$ equipped with the dot product is an inner product space. For an example of an inner product space outside of $\mathbb{R}^n$, consider $C^\infty(a,b)$, the vector space of functions infinitely differentiable over the open interval (a,b). We can define what is known as the $L2$ inner product, given by

$$< f, g >= \int_a^b f^*(t)g(t)dt$$