Linear Algebra

Dilip Thiagarajan

August 6, 2016

Contents

Preface

This reading is not meant as a comprehensive take at learning linear algebra and the associated mathematics necessary, but as an extension that may be helpful to simplify what should be taught that is necessary in all topics to linear algebra. As such, I consider this writing to be a helpful extension to traditional textbooks, and will therefore not write any exercises to complement each section unless requested to.

My main goal in writing this is to clarify to myself the motivation of all the things I've learned in linear algebra, as well as to introduce to myself and the reader specific interesting applications of linear algebra. At the time of the beginning of this writing, I have completed two linear algebra courses, and am embarking on learning how to apply machine learning and deep learning. As such, the applications in this book may present a heavy bias towards those topics, but there are many more fruitful applications of linear that are useful in today's world.

In almost all topics brought up in this reading, unless explicitly specified otherwise, I will be using finite sets in analysis. The study of infinite sets in linear is an interesting one, but one that I am not completely familiar with, and one who's application in conventional use is not particularly clear yet.

Some of the explicit assumptions I assume the reader has made include the following (constantly being added to):

- Basic knowledge of proof by induction
- Terminology such as: the solution of a matrix equation being consistent, the rank of a matrix, and reduced row echelon forms.

Chapter 1

Sets, Relations, and Modular Arithmetic

1.1 Sets

For our purposes, we will define a **set** as a collection of distinct elements. In practice, we write sets as follows:

$$\{1, 2, 3, 4\}$$

where the above set has 4 elements, or it's cardinality is 4. The following examples of sets, and are useful sets that will be seen in almost all of the subsequent chapters.

$$\mathbb{N} = \{0, 1, 2, 3, 4, ...\}$$

$$\mathbb{Z} = \{..., -1, 0, 1, 2, 3, 4, 5, ...\}$$

We can write sets of other types of elements, not necessarily like above. For example, we can write the set of ordered pairs of natural numbers as follows:

$$\mathbb{N} \times \mathbb{N} = \{(0, 0), (1, 0), (0, 1), (2, 0), ...\}$$

If a collection of elements is contained within a set, we say that collection of elements is a **subset**. For example, we note that the natural numbers are a subset of the integers. We can write this symbolically as

$$\mathbb{N} \subseteq \mathbb{Z} \text{ or } \mathbb{N} \subset \mathbb{Z}$$

where $\subseteq$ indicates that the collection of elements may comprise the whole set, and $\subset$ indicates that the collection of elements strictly does not comprise the whole set.

We see subsets appear in the **powerset** of a set. Suppose $S = \{1, 2, 3\}$. The powerset of S is as follows:

$$2^S = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

where $\emptyset$ indicates the set $\{\}$, or the **empty set**. Note that the cardinality of the set 2^S, or $|2^S|$, is equivalent to $2^{|S|}$. The proof is left as an exercise to the reader (hint: induction).

1.2 Relations

Relations, at the most basic level, can be thought of as a binary operation that outputs true or false on the input. In practice, we think of relations as describing whether there is an interaction between two elements, like the make of two cars, or the school two students go to. For sets, this same idea holds.

Building upon this idea, let's consider the set of all students in some state, and call this set S. We can build a set P, where the elements of P are ordered pairs (a, b), where a and b are students that go to the same school. Intuitively, looking at P wholistically, we see that $P \subseteq S \times S$ must be true.

This is how we'll formally define a relation R; namely, $R = P \subseteq S \times S$, where $\forall (a, b) \in P, aRb$

1.2.1 Equivalence Relations

For our purposes, we'll only be interested in **equivalence relations**, which are special kinds of relations. More specifically, equivalence relations are relations with the following properties:

1. Reflexive: If $a \in S$, then aRa.
2. Symmetric: If $a, b \in S$ such that aRb, then bRa.
3. Transitive: If $a, b, c \in S$ such that aRb and bRc, then aRc.

Going back to our example of students in some state, we see that the relation "going to the same school" is an equivalence relation, because:

1. Every student goes to the same school as himself.
2. If student 1 goes to the same school as student 2, then student 2 clearly goes to the same school as student 1.
3. If student 1 goes to the same school as student 2, and student 2 goes to the same school as student 3, then student 1 and student 3 go to the same school as well.

1.3 Equivalence Classes

Recall that sets may consist of sets of elements as well. We now introduce the concept of sets modulo an equivalence relation. To clarify what this means, we'll start with an example: consider the set $\mathbb{Z}/2\mathbb{Z}$, which is read "the set of integers mod 2". Intuitively, we can somewhat tell that this will just separate the integers into even and odd numbers, but how is this formally defined?

The equivalence relation here is actually the "mod 2" mentioned - we can partition the integers into sets where each set consists of elements such that each of those elements mod 2 is the same thing. More formally, let's consider $\mathbb{Z}/m\mathbb{Z}$, and each set S in this:

$$\forall e \in S, e \bmod m \equiv n | 0 \leq n < m - 1$$

where the n for each S is distinct, i.e. for two different sets in $\mathbb{Z}/m\mathbb{Z}$, the result of taking an element in one set mod m will be different from the result of taking an element in the other set mod m.

In general, given a set S and an equivalence relation R, we say that S/R, or the set S modulo R, is the set of l **equivalence classes** $e_i, 0 \leq i < l$ such that:

1. For all $a_1, a_2 \in e_i$, $a_1 R a_2$ holds.
2. For all $a_i \in e_i, a_j \in e_j, i \neq j$, $a_i R a_j$ does not hold.

Equivalence classes can be written as

$$e_i = [a_i], \text{ where } a_i \in e_i$$

where we see a_i is a representative of the equivalence class e_i.

Operations on equivalence classes can be defined, but we must be careful in saying that they are **well-defined**. Addition and multiplication of equivalence classes in numerical sets such as $\mathbb{Z}/m\mathbb{Z}$ work out fine, but exponentiation in such sets is a bit intricate, and requires a slight modification to become well-defined.

To show that an operation on equivalence classes is well-defined for our numerical sets, we must show that the result of the operation on arbitrary equivalence classes is the same, regardless of which representative is used to represent the equivalence class. For example, using this idea, we can show that multiplication and addition are well-defined (this is left as an exercise to the reader; keep in mind that numbers can be written out as something like $a = qm + r$, where $0 \leq r < m$, and that any two elements belonging to the same equivalence class must be congruent mod m).

Chapter 2

Groups, Rings and Fields

The concept of mathematical sets is important in all fields of mathematics, and is especially true in linear algebra given the ubiquity of vector spaces and subspaces. As such, the idea of fields, a close analog to vector spaces, is quite important to mention in a proper reading of the topic, and thus, we begin by introducing groups, upon which we build rings, which we consequently use to build a field.

2.1 Groups

A **group** of elements is a set of elements with an associated binary operation on those elements. Formally, suppose the set of elements G is equipped with the operation $f : G \times G \to G$. Then, we say that G is a group.

2.1.1 Properties and Axioms

The operation f must satisfy the following four properties:

1. Closure: For all $x, y \in G$, $f(x, y) \in G$.
2. Associativity: For all $x, y, z \in G$, $f(x, f(y, z)) = f(f(x, y), z)$.
3. Identity: There exists $i \in G$ such that $f(i, x) = f(x, i) = x$, for all $x \in G$.
4. Inverse: For all $x \in G$, there exists $x_i \in G$ such that $f(x, x_i) = f(x_i, x) = i$.

We say that a group G is **abelian** if $ab = ba$ for all $a, b \in G$. For convenience, henceforth, we will write $f(x, y)$ as $x \cdot y$ or xy. Given these properties, we arrive at some noteworthy conclusions regarding the elements of the group G. One of those is the following property:

Cancellation Property: Given $a, b, c \in G$, if $ab = ac$, then $b = c$; if $ac = bc$, then $a = b$. The proof comes immediately from the application of the inverse and closure axioms.

There is a lot more useful applications of groups in other branches of mathematics, especially in abstract algebra, but for our purposes, knowing these definitions and this one property will be sufficient to build up into more intricate mathematical sets.

2.2 Rings

A **ring** of elements is a set of elements with two associated binary operations, which often generalize to the addition and multiplication, and through these generalizations, give way to rings of elements that aren't necessarily numerical in nature, including polynomials, vectors, functions, and matrices.

2.2.1 Properties and Axioms

In essence, a ring is an abelian group with a second binary operation. The two operations must interact in a specific way, as detailed by the following axioms. Letting the first operation be addition, and the second operation be multiplication for a ring R:

1. Associativity of addition: for any $a, b, c \in R$, $(a + b) + c = a + (b + c)$
2. Commutativity of addition: for any $a, b \in R$, $a + b = b + a$
3. Additive Identity: there exists an element $0 \in R$ such that for all $a \in R$, $a + 0 = 0 + a = a$.
4. Additive Inverses: for all $a \in R$, there exists $a_i \in R$ such that $a + a_i = 0$.
5. Distributivity of multiplication over addition: for any $a, b, c \in R$, $(a + b)c = ac + bc$.
6. Associativity of multplication: for any $a, b, c \in R$, $a(bc) = (ab)c$.
7. Multiplicative identity: there exists an element $1 \in R$ such that for all $a \in R$, $a(1) = 1(a) = a$.

We note that while it's possible to define subtraction within a ring, it is not possible to define division in a ring unless all multiplicative inverses are within the ring itself, i.e. for all $a \neq 0 \in R$, there exists $a_i \in R$ such that $a(a_i) = 1$.

A ring is commutative when it's multiplication operation is commutative. As an example, the set of natural numbers and the set of integers are both rings. Additionally, the integers modulo some number n, denoted $\mathbb{Z}/n\mathbb{Z}$, will also be a ring.

2.3 Fields

A **field** is a set of elements with two associated binary operations, again which often generalize to addition and multiplication.

2.3.1 Properties and Axioms

A field is just a commutative ring where all the multiplicative inverses exist within the ring itself. The natural numbers and the integers do not make fields, but the set of rational numbers and the set of reals do make fields. Below are the relevant axioms for a field R:

1. Associativity of addition: for any $a, b, c \in R$, $(a + b) + c = a + (b + c)$
2. Commutativity of addition: for any $a, b \in R$, $a + b = b + a$
3. Additive Identity: there exists an element $0 \in R$ such that for all $a \in R$, $a + 0 = 0 + a = a$.

4. Additive Inverses: for all $a \in R$, there exists $a_i \in R$ such that $a + a_i = 0$.
5. Distributivity of multiplication over addition: for any $a, b, c \in R$, $(a + b)c = ac + bc$.
6. Associativity of multplication: for any $a, b, c \in R$, $a(bc) = (ab)c$.
7. Multiplicative identity: there exists an element $1 \in R$ such that for all $a \in R$, $a(1) = 1(a) = a$.
8. Commutativity of multiplication: for any $a, b \in R$, $ab = ba$.
9. Multiplicative Inverses: For each nonzero $a \in R$, there exists $a_i \in R$ such that $a(a_i) = 1$.

From here, we can arrive at a few interesting and useful properties.

Lemma: Let F be a field. Then, for any element $a \in F$, $a \cdot 0 = 0 \cdot a = 0$.
Proof: To see $a \cdot 0 = 0 \cdot a = 0$, we use the fact that $0 + 0 = 0$ by the additive identity axiom. Multiplying this by a, we get:

$$a(0 + 0) = a(0)$$

which, by the distributivity axiom, gives us:

$$0 \cdot a + 0 \cdot a = 0 \cdot a \rightarrow 0 \cdot a = 0$$

which also implies that $a \cdot 0 = 0$ by the commutativity of the multiplication operation.

Lemma: A field F has the cancellation property, i.e. if $a, b, c \in F$ such that $ab = ac$, then either $a = 0$ or $b = c$.
Proof: Rearranging the equation and using the distributive property, we have that $ab - ac = 0 \rightarrow a(b - c) = 0$. From this, we can arrive at one of two conclusions: either $a = 0$, or $a \neq 0$, in which case we multiply both sides by a^{-1} to get $b - c = 0 \rightarrow b = c$.

Lemma: Let p be a prime number. Then $\mathbb{Z}/p\mathbb{Z}$, the integers modulo p, is a field.
Proof: We know already that $\mathbb{Z}/n\mathbb{Z}$ is a commutative ring, so we only really need to check that every nonzero element of the set has a multiplicative inverse in the set.

Consider arbitrary $[a] \neq [0] \in \mathbb{Z}/p\mathbb{Z}$. We know immediately that p will not divide a, and because p is prime, the GCD of p and a must be 1, which means they are coprime. As a result, we can show that there exists $[x]$ such that $[a][x] = [1]$ (the proof of this is left to the reader), and thus, $[x]$ is a multiplicative inverse of a, and thus this property will hold for all $[a] \neq [0]$.

Definition: Let F be a field. If p is the smallest positive integer such that $p = 0_F$, we say F has **characteristic** p; if there is no positive integer $p = 0_F$, then we say F has characteristic 0.

As suggested by the notation, the characteristic of a field will always be prime or 0. To show this, we need to show that if F has nonzero characteristic p, then p really is prime. Suppose $n = a \cdot b$, where a, b are positive integers. Then, clearly $n_F = a_F \cdot b_F$ (i.e. the associated equivalence classes). However, because $n_F = 0$ by definition, we have $a_F \cdot b_F = 0$, which means either $a_F = 0$ or $b_F = 0$. Without loss of generality, suppose $a_F = 0$. Then, $a_F \geq n_F$ because n_F is the smallest positive integer with $n_F = 0$, and $a_F \leq n_F$ because a_F is a factor of n_F, so $a_F = n$, so $b_F = 1$, and so the factorization of n must be a product of n and 1, indicating that n must be prime.

Chapter 3

Spans and Linear Independence

With the notion of these mathematical sets, we can consider the nature of a finite set more wholistically, especially when the set consists of n-dimensional vectors. More specifically, we'll look at the space (which can be thought of as an infinite set) a set of vectors can span, whether a set of vectors are *linearly independent*, and how to describe a set that spans a space as simply as possible. These will prove useful to build upon for future concepts, such as vector spaces, and more.

3.1 Spans

Consider a set of m n-dimensional vectors $v_i \in \mathbb{R}^k$, i.e.

$$S = \{v_1, v_2, ..., v_{m-1}, v_m\}$$

We say that the linear span, or **span** of this set S is the following infinite set:

$$\text{Span of S} = L[S] = \{\sum_{i=1}^{n} a_i v_i, a_i \in \mathbb{R}, 1 \leq i, j \leq m\}$$

Note that the coefficients a, b are from the same set the vectors are built on, namely, $\mathbb{R}$. This applies for any mathematical sets (i.e. including complex numbers, integers, and so on).

Theorem: Once again, consider a set S as defined above, and let A be a matrix with these vectors as columns. Then, v is in the span of S iff $Ax = v$ is consistent. **Proof**: Consider the forward statement, i.e. assume v is in the span of S. Then, there exist corresponding a_i such that

$$\sum_{i=1}^{n} a_i v_i = v$$

and thus, $Ax = v$ must be consistent, with x merely being the coefficients a_i.

Now consider the backwards statement. If $Ax = v$ is consistent, then we chose the coefficients a_i corresponding to x, and come to the same conclusion.

We can come to a couple of interesting theorems just from the terminology we have so far.

Theorem: The following statements are equivalent regarding an $m \times n$ matrix A:

1. The span of the columns A is $\mathbb{R}^m$.
2. The equation $Ax = b$ is consistent for every $b \in \mathbb{R}^m$.
3. The rank of A is m.
4. The reduced row echelon form of A has no zero rows.

Because spans are infinite in cardinality, we can see that the span of any collection of vectors in the span of a set S is still contained in that same span. More formally:

Theorem: For any finite set S, the following statements are true:

- S is contained in $L[S]$.
- If S' is any finite set contained in $L[S]$, then $L[S']$ is also contained in $L[S]$.
- For any vector z in $\mathbb{R}^k$, $L[S] = L[S \cup \{z\}]$ iff $z \in L[S]$.

3.2 Linear Independence

Once again, we consider a set S, consisting of m n-dimensional vectors in $\mathbb{R}^k$. We say that S is **linearly independent** iff the only linear combination of v_i that results in the 0 vector is the trivial linear combination. More specifically:

$$\sum_{i=1}^{m} a_i v_i = 0 \text{ iff } a_i = 0$$

Consequently, we say that S is **linearly dependent** if there exists a non-trivial linear combination of v_i that results in the 0 vector, i.e. $a_i \neq 0$ for some i (can be more than one of the coefficients).

From what we've developed thus far, we can characterize the conditions that are equivalent to linear independence.

Theorem: The following statements regarding an $m \times n$ matrix A are equivalent:

1. The columns of A are linearly independent.
2. The system $Ax = b$ has at most one solution for $b \in \mathbb{R}^m$.
3. The nullity of A is 0, i.e. the dimension of the null space of A is 0.
4. The rank of A is n.
5. The columns of the reduced row echelon form of A are distinct standard vectors in $\mathbb{R}^m$.
6. The only solution to $Ax = 0$ is $x = 0$.

For future reference, the system $Ax = 0$ is referred to as the homogenous system. In addition, it's easy to see that any set of vectors containing the zero vector is linearly dependent, and any set containing just one non-zero vector is a linearly independent set.

The concept that ties together spans and linear independence will be introduced in the next chapter.

Chapter 4

Vector Spaces, Subspaces, Bases, and Quotient Spaces

Having defined these mathematical sets, we now introduce the idea of sets, that later on come with more geometric notions like length, and distance between elements. Vector spaces give us a better idea of the geometric notion behind vectors themselves, rather than thinking of them as a string of n numbers. Additionally, they often show up in conventional applications, and are thus useful to look at.

4.1 Vector Spaces

4.2 Subspaces

4.3 Bases

We now arrive at what ties linear independence and spans together: **bases**. In essence, bases are finite sets of vectors that are the minimal description of a set (finite or infinite) set of vectors. More formally, a basis for a vector space S is a linearly independent set B such that $L[B] = S$.

As an example, we consider the Cartesian coordinate plane, which is just $\mathbb{R}^2$. The x and y axes form a basis for $\mathbb{R}^2$, namely, the vectors $\begin{bmatrix}1 & 0\end{bmatrix}^T$ and $\begin{bmatrix}0 & 1\end{bmatrix}^T$.

There are essentially three important ideas regarding bases that are incredibly useful for all of the more complex theory, and are thus formalized below (all without proof, which is left as an exercise to the reader):
Theorem - Reduction Principle: Let V be a subspace of $\mathbb{R}^n$ spanned by a finite set S. Then, there is a basis for V contained in S.
Theorem - Extension Principle: Let V be a subspace of $\mathbb{R}^n$. Every linearly independent subset of V is contained in a basis for V.
Theorem - Basis Theorem: Let V be a subspace of $\mathbb{R}^n$. Then, V has a basis, and every basis of V has the same number of vectors.

Note that from our previous analysis of linear independence, any single non-zero vector in V is a linearly independent set, and thus, the extension principle allows us to find a basis on essentially any starting vector. Then, suppose we had two bases for V, B_1 and B_2, one with j vectors, and the other with k vectors. Knowing that the number of linearly independent vectors in V is at most the number of vectors in a spanning set, we have $j \leq k$ and $k \leq j$, i.e. $j = k$.

Now, using our definition of a basis, we can say that the dimension of a subspace V (other than $\{0\}$) is the number of vectors in any basis of V. The dimension of $\{0\}$ is defined to be 0 by convention. With this, we can enumerate our formal conditions:

Theorem: Let V be a subspace of $\mathbb{R}^n$ with dimension k. Then, any two of the following conditions on a subset S of V imply that S is a basis for V:

1. S is linearly independent.
2. S spans V.
3. S has exactly k vectors.

4.4 Quotient Spaces

Chapter 5

Linear Transformations and the Isomorphism Theorems

5.1 Nilpotent Transformations

5.2 Projection Transformations

Chapter 6

Matrices and Linear Systems

Chapter 7

Applications

At this point, we bring up some interesting applications that require only the knowledge of solving linear systems using basic row reduction operations.

7.1 Discrete Dynamics

7.2 Markov Chains

7.3 Stochastic Matrices

Chapter 8

Determinants, Invertibility, and Eigen-theory

In this chapter, we'll introduce the determinant function, which is a special function (in its alternating and mulitinear characteristic) that allows us to introduce another perspective of linear transformations. More specifically, we'll look at how transformations can be inverted (i.e. when they are bijective), and see how this may be useful in developing the idea of similar transformations.

8.1 Determinants

8.2 Invertibility

8.3 Eigenvalues and Eigenvectors

8.4 Diagonalization and Similarity

8.5 Spectral Value Decomposition

Chapter 9

Inner Products

Chapter 10

Adjoints, Spectral Theorem, Principal Axis Theorem

Chapter 11

Jordan and Rational Canonical Forms

asdf

11.1 Invariant Subspaces

11.2 Jordan Canonical Forms

11.3 Rational Canonical Forms

11.4 Applications

Chapter 12

Application to Differential Equations

Chapter 13

The Similarity Problem