MAT224: Linear Algebra II

Arnav Patil

University of Toronto

Contents

0 Preface and Acknowledgements

This notes package is based on *A Course in Linear Algebra* by Damiano and Little, which is the primary text for MAT224: Linear Algebra II at the University of Toronto. I self-learned this course over summer 2024 because I wanted to sharpen my linear algebra before heading into second-year studies in ECE, especially since I struggled with MAT188 in Fall 2023. This text also is more proof-based, and discusses a few other topics not covered in MAT188. Nevertheless, the interested reader may find my notes for MAT188 here [click me!].

To stay consistent with the text, I will be including the various math statements that are present in the book.

Definition:
An explanation of the mathematical meaning of the word.

Theorem:
A statement that has been proven to be true.

Proposition:
A less important but nonetheless interesting true statement.

Lemma:
A true statement used in proving other true statements.

Corollary:
A true statement that is a simple deduction from a theorem or proposition.

Furthermore, I will include most proofs as part of this notes package. While the start of the proof is explicitly marked, the end of the proof is denoted by ■, a common symbol in mathematical texts.

1 Vector Spaces

1.1 Vector Spaces

For the purposes of these notes, I will assume that the reader is familiar with the concept of a 'vector' as well as vector addition and scalar multiplication. First, we will start with a formal definition for a real vector space.

Definition:
A real vector space is a set of vector V together with:

1. An operation called *vector addition*, and
2. An operation called *scalar multiplication*.

There are also eight **axioms** (fundamental statements that we assume to be true) of math that these vectors and functions must satisfy:

1. $(\vec{x}+\vec{y})+\vec{z}=\vec{x}+(\vec{y}+\vec{z})$,
2. $\vec{x}+\vec{y}=\vec{y}+\vec{x}$,
3. There exists a vector $\vec{0}\in V$ with the property $\vec{x}+\vec{0}=\vec{x}$,
4. For every vector $\vec{x}$, there exists $\vec{x}+-\vec{x}=\vec{0}$,
5. $c(\vec{x}+\vec{y})=c\vec{x}+c\vec{y}$,
6. $(c+d)\vec{x}=c\vec{x}+d\vec{x}$,
7. $(cd)\vec{x}=c(d\vec{x})$, and
8. $1\vec{x}=\vec{x}$.

We see that in $\mathbb{R}^n$there is only one additive identity, which is the zero vector $\vec{0}$. There is also only one additive inverse for each vector in a vector space. This leads us to the following:

Proposition:
Let V be a vector space. Then, we see:

1. The zero vector $\vec{0}$ is unique,
2. For all $\vec{x}\in V, 0\vec{x}=\vec{0}$,
3. For each $\vec{x}\in\vec{V}$, the additive inverse $-\vec{x}$ is unique, and
4. For all $\vec{x}\in\vec{V}$ and all $c\in\mathbb{R}^n$, $(-c)\vec{x}=-(c\vec{x})$.

Now we will go and prove each of the above statements.

1. Suppose $\vec{0}$ and $\vec{0}'$ both satisfy axiom 3 above. Then $\vec{0}+\vec{0}'=\vec{0}$, since $\vec{0}'$ is an additive identity. Conversely, $\vec{0}+\vec{0}'=\vec{0}'+\vec{0}=\vec{0}'$, which gives us $\vec{0}=\vec{0}'$, or there is only one additive identity in V.
   (a) Assuming we have two examples of a given object, then proving those objects are the same, is one common way of proving that something is unique.
2. We have $0\vec{x}=(0+0)\vec{x}=0\vec{x}+0\vec{x}$ by axiom 6. If we add the inverse of $0\vec{x}$ to both sides, we obtain $\vec{0}=0\vec{x}$.

3. Using the same logic as part a), we use axioms 1, 3, and 4 to give $\vec{x}+-\vec{x}+(-\vec{x})' = (\vec{x}+-\vec{x})+(-\vec{x})' = \vec{0}+(-\vec{x})' = (-\vec{x})'$. On the other hand, axiom 2 gives us $\vec{x}+-\vec{x}+(-\vec{x})' = \vec{x}+(-\vec{x})'+-\vec{x} = (\vec{x}+(-\vec{x})')+-\vec{x} = \vec{0}+-\vec{x} = -\vec{x}$. Therefore, we have $-\vec{x} = (-\vec{x})'$ and the additive inverse of $\vec{x}$ is unique.

4. $c\vec{x}+(-c)\vec{x} = (c+-c)\vec{x} = 0\vec{x} = \vec{0}$ by axiom 6. Hence $(-c)\vec{x}$ is the additive inverse of $c\vec{x}$ by part c). Therefore, we can prove $(-c)\vec{x} = -(c\vec{x})$. ■

1.2 Subspaces

We define vector sum and scalar multiplication of $C(\mathbb{R})$ as usual for a function. If $f, g \in C(\mathbb{R})$ and $c \in \mathbb{R}$, then both $(f+g)(x)$ and $cf(x)$ are functions defined for all $\vec{x} \in \mathbb{R}$. We claim this set, $C(\mathbb{R})$ is a vector space. Because of this, we may summarize the two following facts:

Lemma:
Let $f, g \in C(\mathbb{R})$, and let $c \in \mathbb{R}$. Then,

1. $f, g \in C(\mathbb{R})$, and
2. $cf \in C(\mathbb{R})$.

Proof:

1. By the limit sum rule from calculus, for all $a \in \mathbb{R}$ we have:

$$\lim_{x\to a}(f+g)(x) = \lim_{x\to a}(f(x)+g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$$

Since f and g are continuous, this last expression equals $f(a)+g(a) = (f+g)(a)$. Hence $f+g$ is continuous.

2. By the limit product rule, we have:

$$\lim_{x\to a}(cf)(x) = \lim_{x\to a} cf(x) = (\lim_{x\to a} c)\cdot(\lim_{x\to a} f(x)) = cf(a) = (cf)(a)$$

therefore cf is also continuous. ■

Proving the eight axioms proves unnecessary as we see that their properties are already established in the process of proving that the sets in the above lemma are vector spaces. This provides us with the following definition:

Definition:
Let V be a vector space and let $W \subseteq V$ be a subset. Then W is a subspace of V if W is a vector space itself under the operations of vector addition and scalar multiplication.

We can alternatively state the following requirements for a space W to be a vector subspace V:

1. The set W must be closed under vector addition,
2. The set W must be closed under scalar multiplication, and
3. The zero vector $\vec{0}$ must be contained within the set.

Examples [in the book] imply the following, which we state as a theorem:

Theorem:
Let V be a vector space. Then the intersection of any collection of subspaces of V is a subspace of V.

Proof: Consider a collection of subspaces of V. At the very minimum, the intersection of all these sets is nonempty since each subspace must contain the zero vector $\vec{0}$.
Another important application of this theorem is to show that the set of solutions to any systems of linear equations is a subspace of $\mathbb{R}^n$.

Corollary:
Let $a_{ij}(1 \leq i \leq m, 1 \leq j \leq n)$ be any real numbers and let $W = \{(x_1, ..., x_n) \in \mathbb{R}^n | a_{i1}x_1 + ... + a_{in}x_n = 0$ for all $i, 1\ eqi \leq m\}$. Then, W is a subspace of $\mathbb{R}^n$.

1.3 Linear Combinations

We will begin this section with some terminology.

Definition:
Let S be a subset of a vector space V.

1. A *linear combination* of vectors in S is any sum $a_1\vec{x}_1 + ... + a_n\vec{x}_n$,
2. If $S \neq \varnothing$, the set of all linear combinations of vectors in S is called the **span** of S, denoted $\text{Span}(S)$. If $S = \varnothing$ we define $\text{Span}(S) = \{\vec{0}\}$, and
3. If $W = \text{Span}(S)$, we sat S *spans or generates* W.

Simply put, the span of S is the set of all vectors that can be built with **all** linear combinations of the vectors.

We can generally say that the span of a set of vectors is a subspace of the vector space from where those vectors was chosen. This is summarized in the theorem below:

Theorem:
Let V be a vector space and let S be any subset of V. Then $\text{Span}(S)$ is a subspace of V.

Proof: $\text{Span}(S)$ is nonempty by definition (the zero vector). We can write $\vec{x} = a_1\vec{x}_1 + ... + a_n\vec{x}_n$ and $\vec{y} = b_1\vec{y}_1 + ... + b_m\vec{y}_m$. Then for any scalar c, we have:

$$
\begin{aligned}
c\vec{x} + \vec{y} &= c(a_1\vec{x}_1 + ... + a_n\vec{x}_n) + (b_1\vec{x}'_1 + ... + b_m\vec{x}'_m) \\
&= ca_1\vec{x}_1 + ... + ca_n\vec{x}_n + b_1\vec{x}'_1 + ... + b_m\vec{x}'_m
\end{aligned}
$$

Since these are all linear combinations of the vectors in the set S, we have $c\vec{x} + \vec{y} \in \text{Span}(S)$, and therefore, $\text{Span}(S)$ is a subspace of V. ■

Proving this theorem is important because every subspace of a given vector space can be constructed this way. It gives us the following definition.

Definition:
Let W_1 and W_2 be subspaces of a given vector space V. The sum of W_1 and W_2 is:

$$W_1 + W_2 = \{\vec{x} \in V | \vec{x} = \vec{x}_1 + \vec{x}_2, \text{ for some } \vec{x}_1 \in W_1 \text{ and } \vec{x}_2 \in W_2\}$$

Proposition:
In general, it can be proven that if $W_1 = \text{Span}(S_1)$ and $W_2 = \text{Span}(S_2)$, then $W_1 + W_2$ is the span of the union of S_1 and S_2.

Proof: We can show two sets are equal by proving that each is contained in the other.

To see $W_1 + W_2 \subseteq \text{Span}(S_1 \cup S_2)$, let $\vec{v} = \vec{v}_1 + \vec{v}_2$ where $\vec{v}_1 \in W_1$ and $\vec{v}_2 \in W_2$. Because the linear combination representation of both vectors is in either W_1 or W_2, we can show that $W_1 + W_2 \subseteq \text{Span}(S_1 \cup S_2)$.

Conversely, to see $\text{Span}(S_1 \cup S_2) \subseteq W_1 + W_2$, we write the linear combination of $\vec{v} \in \text{Span}(S_1 \cup S_2)$ as

$$\vec{v} = a_1\vec{x}_1 + ... + a_n\vec{x}_n + b_1\vec{y}_1 + ... + b_m\vec{y}_m$$

where each $\vec{x}_i \in S_1$ and $\vec{y}_i \in S_2$. Thus, we have $\vec{v} \in W_1 + W_2$. Since this is true for all $\vec{v} \in \text{Span}(S_1 \cup S_2)$, we can prove $\text{Span}(S_1 \cup S_2) \subseteq W_1 + W_2$.

With both parts of the proof complete, we can show that the sum of two subspaces is also a subspace. ■

Theorem:
Let W_1 and W_2 be subspaces of a vector space V. Then $W_1 + W_2$ is also a subspace of V.

Proof: If W_1 and W_2 are nonempty then it follows that $W_1 + W_2$ is also nonempty. Let $\vec{x}, \vec{y}$ be any two vectors in $W_1 + W_2$, which we can write as $\vec{x} = \vec{x}_1 + \vec{x}_2$ and $\vec{y} = \vec{y}_1 + \vec{y}_2$ where $\vec{x}_1, \vec{y}_1 \in W_1$ and $\vec{x}_2, \vec{y}_2 \in W_2$. Then:

$$c\vec{x} + \vec{y} = c(\vec{x}_1 + \vec{x}_2) + (\vec{y}_1 + \vec{y}_2)$$
$$= (c\vec{x}_1 + \vec{y}_1) + (c\vec{x}_2 + \vec{y}_2)$$

By the definition of a subspace, we can now show that $W_1 + W_2$ is a subspace of V. ■

Consider two subspaces of a vector space V: W_1 and W_2. Generally, $W_1 \cup W_2$ is not a subspace of V. For example, consider $W_1 = \text{Span}([1, 0])$ and $W_2 = \text{Span}([0, 1])$. $[1, 0] + [0, 1] = [1, 1]$, but we see that $[1, 1] \notin W_1 \cup W_2$. It then follows from the above proposition that $W_1 + W_2$ is a subspace of V containing $W_1 \cup W_2$.

1.4 Linear Dependence and Linear Independence

We will begin this section with an example to show why we frequently consider different spanning sets for the same subspace.

In $\mathbb{R}^3$, consider the subspace W spanned by:

$$S = \{(1, 2, 1), (0, -1, 3), (1, 0, 7)\}$$

We can describe this subspace in set builder notation as:

$$W = \{(x_1, x_2, x_3) \in \mathbb{R}^3 | (x_1, x_2, x_3) = a_1(1, 2, 1) + a_2(0, -1, 3) + a_3(1, 0, 7) \text{ for some } a_i \in \mathbb{R}\}$$

Now we ask ourselves: could there be a *subset* of S that also spans W? There is one such case:

$$(1, 2, 1) + 2(0, -1, 3) + (-1)(1, 0, 7) = (0, 0, 0)$$

In other words, we have a linear combination of the vectors in S that adds up to the zero vector, but whose coefficients are nonzero. This means we can solve for one of the vectors in terms of the others:

$$(1, 0, 7) = (1, 2, 1) + 2(0, -1, 3)$$

As a result, we see that every vector in S is also in the span of the set $S' = \{(1,2,1),(0,-1,3)\}$. Effectively, that makes one of the vectors in the original set S redundant.

Any time we have a linear combination of the form $a_1\vec{x}_1 + a_2\vec{x}_2 + ... + a_n\vec{x}_n = \vec{0}$, we are able to write $\vec{x}_n$ in terms of the rest of the variables. We define this as such:

Definition:
Let V be a vector space and S be a subset of V:

1. A **linear dependence** among vectors in S is an equation

$$a_1\vec{x}_1 + a_2\vec{x}_2 + ... + a_n\vec{x}_n = \vec{0}$$

   where at least one of the coefficients $a_i \neq 0$, and

2. The set S is said to be **linearly dependent** if there exists linear dependence in the vectors of S.

This leads us directly to another definition:

Definition:
A subset S of a vector space V is **linearly independent** if $a_1\vec{x}_1 + a_2\vec{x}_2 + ... + a_n\vec{x}_n = \vec{0}$ only when all $a_i = 0$.

Now we consider some examples to solidify the concepts discussed above.

1. Consider $S = \{(1,0,0),(0,1,0),(0,0,1)\}$ in $\mathbb{R}^3$. S must be linearly independent as

$$a_1(1,0,0) + a_2(0,1,0) + a_3(0,0,1) = (0,0,0)$$

   We see that $a_1 = a_2 = a_3 = 0$, so we the set is linearly independent.

2. Returning to part 1, take $S = \{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$ where $\vec{e}_i$ is a vector with 1 in its ith component and 0s in all the rest. Then, S is linearly independent in $\mathbb{R}^n$by the same argument as in part 1.

   (a) This set is a special set of vectors known as the **standard basis** of $\mathbb{R}^n$.

3. The empty subset, $\varnothing$ is linearly independent in any vector space. By the definition of linear independence, if there are any linear combinations, the coefficients must be zero. This requirement is "vacuously" satisfied as there are no linear combinations to check!

To conclude, the following proposition provides some properties of sets that follow from the definitions above.

Proposition:

1. Let S be a linearly dependent subset of a vector space V, and let S' be another subset of V that contains S. Then, S' is also linearly dependent, and

2. Let S be a linearly dependent subset of a vector space V and let S' be another subset of V that is contained in S. Then, S' is also linearly independent.

Proof: We will conduct this proof point-by-point.

1. Since S is linearly dependent, there exists a linear dependence among the vectors in S, and since S is contained in S', there is also a linear dependence among the vectors in S'.

2. Consider any equation $a_1\vec{x}_1 + a_2\vec{x}_2 + ... + a_n\vec{x}_n = \vec{0}$, where $\vec{x} \in S'$. Since S is contained in S', this may potentially be a linear dependence in S. However, since S is linearly independent, it follows that S' is also linearly independent. ■

1.5 Interlude on Solving Systems of Linear Equations

When working with vector spaces and subspaces, we find that most computational problems are some combination of the following:

1. Given a subspace W of a vector space V, defined by giving some condition on the vectors, find a set of vectors S such that $W = \text{Span}(S)$.

2. Given a set of vectors S and a vector $\vec{v} \in V$, determine if $\vec{x} \in \text{Span}(S)$.

3. Given a set of vectors S, determine if S is linearly dependent or linearly independent.

All three basic problems lead to the fundamental question of finding all solutions to a simultaneous system of linear equations. We begin with some definitions:

Definition:
A system of m equations in n unknowns $x_1, ..., x_n$ of the form

$$\begin{aligned} a_{11}x_1 + ... + a_{1n}x_n &= b_1 \\ a_{21}x_1 + ... + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + ... + a_{mn}x_n &= b_m \end{aligned}$$

is called a **system of linear equations**. The system is **homogeneous** if all $b_i = 0$, and **inhomogeneous** if otherwise. Every homogeneous system has the **trivial solution** $(0, ..., 0)$.

The reader is undoubtedly familiar with some methods to find solutions to far simpler systems. These methods include eliminating variables, multiplying equations to add/subtract from others, and so on. We implement a similar strategy here, keeping in mind not to lose solutions or gain new ones.

Definition:
Two systems of linear equations are said to be **equivalent** if every solution to one is a solution to the other, and vice versa.

This definition allows us to propose three basic operations that will yield an equivalent system when applied to a system of linear equations.

Proposition:
An equivalent system may be obtained from a given system of linear equations by:

1. Adding any multiple of any one equation to another equation,

2. Multiplying any one equation by a nonzero scalar and leaving the other equations unchanged, and

3. Interchanging the order of any two equations.

We may conduct elementary operations onto a system of linear equations to manipulate it into row echelon form, defined below:

Definition:
A system of linear equations is in **row echelon form** (REF) if satisfies the following:

1. The first nonzero coefficient counting from the left of the equation is 1. This term is known as the **leading term**.
2. For each i the coefficient of x_{ji} in each equation other than the ith.
3. For each i (for which equations i and $i+1$ have some nonzero coefficients) $j(i+1) \geq j(i)$.

The echelon form of a system is unique, but this is an irrelevant act so we (the textbook) will omit this proof.

The process of finding the echelon form of a system is known as **elimination** and is given as a step-by-step algorithm below:

1. Begin with the entire system of m equations.
2. Do the following for each i between 1 and m:
   (a) Among the equations numbered i to m in the modified system, pick one equation containing the variable with the smallest index. If there are multiple such equations, pick the one with the simplest leading coefficient.
   (b) If necessary, interchange this chosen equation and the ith equation.
   (c) If necessary, multiply the (new) ith equation such that its leading coefficient becomes 1.
   (d) If necessary, eliminate all other occurrences of the leading variable (of the ith equation) in other equations.
3. The process ends when we finish step 2d for $i = m$, or possibly sooner if no further computations are necessary.

To introduce the notion of different types of solutions, consider an example of the following system:

$$\begin{aligned} x_1 + x_2 &= 1 \\ 0 &= 1 \end{aligned}$$

If a system has an equation like $0 = 1$, then the system is labelled **inconsistent** and no solutions exist. If such an equation does not exist, then the system is **consistent** and it has either one or infinitely-many solutions.

Definition:

1. In an echelon form the variables appearing in leading terms are called the **basic variables** of the system.
2. All other variables (even those with zero coefficients) are called **free variables**.

Before moving into further discussion, let us take a look at an example. Consider the echelon form system:

$$\begin{aligned} x_1 \quad + x_3 \qquad\qquad &= 0 \\ x_2 - x_3 \quad + x_5 &= 0 \\ x_4 + x_5 &= 0 \end{aligned}$$

We have basic variables x_1, x_2, x_4 and free variables x_3, x_5. We can represent this as:

$$\begin{aligned} x_1 &= -x_3 \\ x_2 &= \quad x_3 - x_5 \\ x_4 &= \qquad\quad - x_5 \end{aligned}$$

If we set $x_3 = 1$ and $x_5 = 0$, then we get the solution $(-1, 1, 1, 0, 0)$, and if we set $x_3 = 0$ and $x_5 = 1$, then we get the solution $(0, -1, 0, -1, 1)$. Instead, if we set one variable to zero, and the other to t where $t \in \mathbb{R}$, then we see that the **parameterization** of the set of solutions is:

$$W = \{t_1(-1, 1, 1, 0, 0) + t_2(0, -1, 0, -1, 1) | t_1, t_2 \in \mathbb{R}\}$$

In general, we can find the parameterization of any set of solutions the same way. We can obtain all solutions by substituting t_i for the ith free variable in the system. Note that in this example, if $t_1 \neq 0$ or $t_2 \neq 0$ then we get a nontrivial solution for the system. The same is true for every homogeneous system with at least one free variable.

Corollary:
If $m \leq n$, every homogeneous system of m linear equations in n unknowns has a nontrivial solution.

Proof: By the REF theorem, every system of linear equations has a corresponding echelon form. Since the number of nontrivial equations only decreases, there will be more variables than equations, so the number of basic variables cannot be bigger than m. Therefore, there must be at least one free variable, which if we let it run over $\mathbb{R}$, then the resulting solution is nontrivial. ■

Now, we may discuss how the three basic problems outlines at the start of this section can be solved using the developed strategies.

1. Given a subspace W of $\mathbb{R}^n$defined as the set of solutions of a system of equations in n variables, to find a set S such that $W = \text{Span}(S)$, we can proceed as follows:
   (a) First, reduce the system to its echelon form,
   (b) Then, the set of solutions may be found exactly as in the above example.
   (c) A set S may be obtained by procedurally setting each free variable to 1 and the rest to 0.
2. Given a finite set S of vectors in $\mathbb{R}^n$, and a vector $\vec{x}$ in $\mathbb{R}^n$, to determine if $\vec{x} \in \text{Span}(S)$, set up the system of equations and reduce to REF.
   (a) If the system is inconsistent, then $\vec{x} \notin \text{Span}(S)$,
   (b) If the system is consistent, then the solution vectors of the system give the scalars in the linear combinations of the vectors of S.

1.6 Bases and Dimension

Let us return to the scenario brought up at the beginning of Section 1.4. We see that if a set of vectors S spans a vector space V, there may be a subset of S, which we will label S' that may also span V. The lack of redundant vectors in S' implies that the vectors in S' are linearly independent. This forms the basis (pun intended) of our next definition:

Definition:
A subset S of a vector space V is called a **basis** of V if $V = \text{Span}(S)$ and S is linearly independent (there are no redundant vectors).

Let us take a look at a few examples to solidify our understanding:

1. Consider the standard basis $S = \{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$ in $\mathbb{R}^n$. S is a basis of $\mathbb{R}^n$by definition since it is linearly independent and every vector $(a_1, a_2, ..., a_n) \in \mathbb{R}^n$ is written as a linear combination:

$$(a_1, a_2, ..., a_n) = a_1\vec{e}_1 + ... + a_n\vec{e}_n$$

2. Let $V = P_n(\mathbb{R})$ and consider $S = \{1, x, x^2, ..., x^n\}$. Obviously, S spans V. Consider also:

$$a_0 + a_1x + a_2x^2 + ... + a_nx^n = 0$$

This holds true for all values only when all coefficients $a_i = 0$, meaning that S is linearly independent, and is indeed a basis for V.

3. Lastly, the empty set $\varnothing$ is also a basis which contains only the zero vector.

We see that given a basis S of a vector space V, there is only one way to profice a vector $\vec{x}$ as a linear combination of the vectors present in S. This is an important property that we will state and prove in the following theorem:

> **Theorem:**
> Let V, and let S be a nonempty subset of V. Then S is a basis of V iff every vector $\vec{x} \in V$ may be written uniquely as a linear combination of the vectors in S.

Proof: As for other "if and only if" proofs, we will prove the direct and reverse implication of the theorem.

Directly, if S is a basis, and $\vec{x} \in V$, then there must exist scalars and vectors in S such that $\vec{x} = a_1\vec{x}_1 + ... + a_n\vec{x}_n$. To prove that this representation is unique, assume there is a second representation of $\vec{x}$. Subtracting these two expressions, we see that:

$$\vec{0} = (\vec{x} = a_1\vec{x}_1 + ... + a_n\vec{x}_n) - (\vec{x} = a_1'\vec{x}_1 + ... + a_n'\vec{x}_n)$$

which implies that $a_i - a_i' = 0$, or $a_i = a_i'$. Thus, there can only be one set of scalars used to represent a given vector $\vec{x}$ as a linear combination of the vectors in S.

Conversely, assume that every vector in V may be written in one and only one way as a linear combination of the vectors in S. This implies immediately that $\text{Span}(S) = V$. To convince ourselves of this, consider:

$$a_1\vec{x}_1 + ... + a_n\vec{x}_n = \vec{0}$$

where $\vec{x}_i \in S_i$. Then, we also see that:

$$0\vec{x}_1 + ... + 0\vec{x}_n = \vec{0}$$

Therefore, by our hypothesis, it must be true that $a_i = 0$ for all i. Furthermore, the set S is linearly independent ■

To state once again, if we have any basis for a vector space, it is possible to express each vector in one and only one way as a linear combination of the vectors in the basis. The scalars appearing in the linear combination are known as the **coordinates** of the vector with respect to the basis.

This leads us to the next important question to ask ourselves: does every vector space have a basis? We begin with the following theorem:

> **Theorem:**
> Let V be a vector space that has a finite spanning set, and let S be a linearly independent subset of V. Then, there exists a basis of S' of V, with $S \subseteq S'$.

Generally, we instead say that every linearly independent set can be extended to a basis by adding additional vectors to the set. To prove the theorem above, we need the following conclusion:

Lemma:
Let S be a linearly independent subset of V and let $\vec{x} \in V$, but $\vec{x} \notin S$. Then, $S \cup \{\vec{x}\}$ is linearly independent iff $\vec{x} \notin \text{Span}(S)$.

Proof: Let us start by proving the lemma. Directly, suppose that $S \cup \{\vec{x}\}$ is linearly independent but $\vec{x} \in \text{Span}(S)$. Then we have an equation $\vec{x} = a_1\vec{x}_1 + ... + a_n\vec{x}_n$. We can rewrite this as $\vec{0} = (-1)\vec{x} + a_1\vec{x}_1 + ... + a_n\vec{x}_n$. Since the coefficient of $\vec{x} = 1 \neq 0$, this is a linear dependence, and thus the contradiction shows that $\vec{x} \notin \text{Span}(S)$.

Conversely, now let us suppose that $\vec{x} \notin \text{Span}(S)$. Recall that we are also assuming S is linearly independent, and we need to show that $S \cup \{\vec{x}\}$ is linearly independent. Consider any potential potential linear dependence in $S \cup \{\vec{x}\} = a\vec{x} + a_1\vec{x}_1 + ... + a_n\vec{x}_n$. If $a \neq 0$, then we could write:

$$\vec{x} = (-a_1/a)\vec{x}_1 + ... + (-a_n/a)\vec{x}_n$$

which contradicts our hypothesis. Hence, $a = 0$. Now, since S is linearly independent, it follows that all $a_i = 0$ so therefore, $S \cup \{\vec{x}\}$ is linearly independent. ■

Now, we are finally ready to prove the theorem above. Let $T = \{\vec{y}_1, ..., \vec{y}_n\}$ be a finite set that spans V, and $S = \{\vec{x}_1, ..., \vec{x}_k\}$ be a linearly independent set in V. The following process will produce a basis of V.

1. Start by setting $S' = S$.
2. For each $\vec{y}_i \in T$ do the following:
   (a) If $S' \cup \{\vec{y}_i\}$ is linearly independent, replace S' with $S' \cup \{\vec{y}_i\}$.
   (b) Go on to the next $\vec{y}_i$.

To better understand why this works, note that we are only including the set $\vec{y}_i$ such that $S' \cup \{\vec{y}_i\}$ is linearly independent at all times. Then, note that every $\vec{y}_i \in T$ is in the span of the final set S', since that set contains all the vectors $\vec{y}$ that are adjoined to the original S. Since T spans V, and every vector in T is also in $\text{Span}(S')$, it follows that S' spans V as well. Therefore, S' is a basis for V. ■

We see that every vector space with a finite spanning set has a basis which is a finite set of vectors. It then seems that the number of vectors in a basis is roughly a measure of how "big" the space is.

Theorem:
Let V be a vector space and let S be a spanning set for V which has m elements. Then, no linearly independent set in V can have more than m elements.

We can now show that the number of vectors in a basis does not vary with the space V.

Corollary:
Let V be a vector space and let S and S' be two bases of V, with m and m' elements, respectively. Then, $m = m'$.

Proof: Since S spans V and S' is linearly independent, we have that $m \geq m'$. Since the inverse is also true, we see that $m = m'$.

With the above corollary in hand, we may now set some definitions:

Definition:

1. If V is a vector space with some finite basis, we say V is **finite-dimensional**.
2. Let V be a finite-dimensional vector space. The **dimension** of V, denoted as $\dim(V)$, is the number of vectors in a (hence all) basis of V.
3. If $V = \{\vec{0}\}$, we define $\dim(V) = 0$.

Let us again look at some examples:

1. For each $n, \dim(\mathbb{R}^n) = n$ since the standard basis $\{\vec{e}_1, ..., \vec{e}_n\}$ contains n vectors.
2. The vector spaces $P(\mathbb{R}), C^1(\mathbb{R}),$ and $C(\mathbb{R})$ are not finite-dimensional. We label these spaces **infinite-dimensional**.
3. $\dim(P_n(\mathbb{R})) = n + 1$, since a basis for $P_n(\mathbb{R})$ is the set $\{1, x, x^2, ..., x^n\}$, which contains $n + 1$ functions in all.

Corollary:
Let W be a subspace of a finite-dimensional vector space V. Then, $\dim(W) \leq \dim(V)$, and $\dim(W) = \dim(V)$ iff $W = V$.

The method for solving systems of linear equations presented in Section 1.5 also provides a method to compute the dimension of W. When we reduce the system to echelon form, then find the spanning set by setting all but one free variables to 0, we are left with a basis for W. This leads us directly to the following corollary.

Corollary:
Let W be subspace of $\mathbb{R}^n$defined by a system of homogeneous linear equations. Then $\dim(W)$ equals the number of free variables in the corresponding echelon form of the system.

Theorem:
Let W_1 and W_2 be finite-dimensional subspaces of a vector space V. Then,

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2)$$

2 Linear Transformations

We will now shift the focus of our discussion from "the objects of linear algebra," which we have established are vector spaces (not vectors) to functions between these spaces, called linear transformations. These functions must preserve properties of vector spaces to count as linear transformations.

2.1 Linear Transformations

Let us begin with some notation for functions: A function T from vector space V to W is denotes as $T : V \to W$. This transformation takes in each vector $\vec{v} \in V$ and transforms it into a vector $\vec{w} \in W$. This leads us to the following definition:

Definition:
A function $T : V \to W$ is called a **linear mapping** or a **linear transformation** if:

1. $T(\vec{u}+\vec{v}) = T(\vec{v}) + T(\vec{u})$ for all $\vec{v}$ and $\vec{u} \in V$, and
2. $T(a\vec{v}) = aT(\vec{v})$ for all $\vec{v} \in V$.

If we ask ourselves where the zero vector is mapped to in such a transformation, we have:

$$\begin{aligned} T(\vec{0}_V) &= T(\vec{0}_V + \vec{0}_V) \\ &= T(\vec{0}_V) + T(\vec{0}_V) \end{aligned}$$

By adding the additive inverse to both sides of the equation, we obtain:

$$\vec{0}_W = T(\vec{0}_V)$$

which essentially states that **the zero vector in the domain vector space always gets mapped to the zero vector in the target vector space**.

Some topics of calculus may be considered as linear mappings of functions as well. Let us discuss a few examples.

First, we will consider differentiation. Let V be the vector space $C^\infty(\mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ for which derivatives of all orders are defined. Let $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ be the mapping of each function to its respective derivative. By using the sum and scalar multiple rules for derivatives, we see that D satisfies the properties of linear transformations.

Now, we will consider definite integration. Let V denote the vector space $C[a, b]$ of continuous functions on the closed interval $[a, b] \subset \mathbb{R}$ and let $W = \mathbb{R}$. Then we can show $\text{Int} : V \to W$ by $\text{Int}(f) = \int_a^b f(x)dx \in \mathbb{R}$.

Before we proceed, let us review some facts. If two vectors in a plane $\vec{a} = (a_1, a_2)$ and $\vec{b} = (b_1, b_2)$ form two sides of a triangle, the third side is formed by their vector sum $\vec{b} - \vec{a}$. By the Pythagorean Theorem, we can find the lengths of all sides as:

$$\begin{aligned} ||\vec{a}|| &= \sqrt{a_1^2 + a_2^2} \\ ||\vec{b}|| &= \sqrt{b_1^2 + b_2^2} \\ ||\vec{b} - \vec{a}|| &= \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2} \end{aligned}$$

By applying the law of cosines, we see that:

$$||\vec{b} - \vec{a}||^2 = ||\vec{a}||^2 + ||\vec{b}||^2 - 2||\vec{a}|| \cdot ||\vec{a}|| \cdot \cos(\theta)$$

With a little manipulating and substituting the lengths defined above, we can end with the expression:

$$a_1 b_1 + a_2 b_2 = ||\vec{a}|| \cdot ||\vec{b}|| \cdot \cos(\theta)$$

The quantity $a_1 b_1 + a_2 b_2$ has a special name: the **inner (or dot) product** of vectors $\vec{a}$ and $\vec{b}$. The dot product is also denoted by: $\langle \vec{a}, \vec{b} \rangle$. Thus, we have our following proposition:

Proposition:
If $\vec{a}$ and $\vec{b}$ are nonzero vectors in $\mathbb{R}^2$, then the angle θ is defined as:

$$\cos(\theta) = \frac{\langle \vec{a}, \vec{b} \rangle}{||\vec{a}|| \cdot ||\vec{b}||}$$

This leads us directly to the following.

Corollary:
If $\vec{a} \neq \vec{0}$ and $\vec{b} \neq \vec{0}$ are vectors in $\mathbb{R}^2$, the angle is a right angle iff $\langle \vec{a}, \vec{b} \rangle = 0$

Proof: Since $\pm\pi/2$ are the only angle (up to addition of integer multiples of π for which the cosine is 0, the result follows immediately. ■

Let us now consider rotation through an angle θ. We may define a function $R_\theta : V \to V$:

$R_\theta(\vec{v})$ = the vector obtained by rotating vector $\vec{v}$ through an angle θ while maintaining its length.

If $\vec{w} = R_\theta(\vec{v})$ then the expression for $\vec{w}$ in terms of its length and angle is:

$$\vec{w} = ||\vec{v}||(\cos(\varphi + \theta), \sin(\varphi + \theta))$$

where φ is the angle $\vec{v}$ makes with its coordinate axis. With some trigonometric manipulation, we end up with:

$$\begin{aligned}\vec{w} &= ||\vec{v}||(\cos(\varphi) \cdot \cos(\theta) - \sin(\varphi) \cdot \sin(\theta), \cos(\varphi) \cdot \sin(\theta) + \sin(\varphi) \cdot \cos(\theta)) \\ &= (v_1 \cos(\theta) - v_2 \sin(\theta), v_1 \sin(\theta) + v_2 \cos(\theta))\end{aligned}$$

This expression for $R_\theta(\vec{v})$ allows us to see if R_θ is a linear transformation.

$$\begin{aligned}R_\theta(a\vec{u} + b\vec{v}) &= R_\theta((au_1 + bv_1, au_2 + bv_2)) \\ &= ((au_1 + bv_1)\cos(\theta) - (au_2 + bv_2)\sin(\theta), (au_1 + bv_1)\sin(\theta) + (au_2 + bv_2)\cos(\theta)) \\ &= (a(u_1 \cos(\theta) - u_2 \sin(\theta)) + b(v_1 \cos(\theta) - v_2 \sin(\theta)), a(u_1 \sin(\theta) + u_2 \cos(\theta)) + b(v_1 \sin(\theta) + v_2 \cos(\theta)) \\ &= aR_\theta(\vec{u}) + bR_\theta(\vec{v})\end{aligned}$$

Thus R_θ is a linear transformation. ■

Proposition:
If $T : V \to W$ is a linear transformation and V is finite, then T is uniquely determined by its values on the members of a basis of V.

Proof: If S and T are linear transformation that take the same values on each vector in a bases for V, then $S = T$. Let $\{\vec{v}_1, ..., \vec{v}_k\}$ be a basis for V, and let $T(\vec{v}_i) = S(\vec{v}_i)$. If $\vec{v} = a_1\vec{v}_1 + ... + a_k\vec{v}_k$, then:

$$\begin{aligned}T(\vec{v}) &= T(a_1\vec{v}_1 + ... + a_k\vec{v}_k) \\ &= a_1T(\vec{v}_1) + ... + a_kT(\vec{v}_k) \text{ (since } T \text{ is linear)} \\ &= a_1S(\vec{v}_1) + ... + a_kS(\vec{v}_k) \\ &= a_1S(\vec{v}_1) + ... + a_kT(\vec{v}_k) \text{ (since } S \text{ is linear)} \\ &= S(\vec{v})\end{aligned}$$

And therefore, S and T must be equal mappings from V to W. ■

We can further this proposition by saying if $\{\vec{v}_1, ..., \vec{v}_k\}$ is a basis for V and $\{\vec{w}_1, ..., \vec{w}_k\}$ are k vectors in W, then we can define a linear transformation $T : V \to W$ by setting $T(\vec{v}_i) = \vec{w}_i$.

2.2 Linear Transformations Between Finite-Dimensional Vector Spaces

Let V be a subspace with $\dim(V) = k$ and basis $\{\vec{v}_1, ..., \vec{v}_k\}$ and W be subspace with $\dim(W) = l$ and basis $\{\vec{w}_1, ..., \vec{w}_l\}$. Each vector $T(\vec{v}_j)$ can be expressed uniquely in terms of $\{\vec{w}_1, ..., \vec{v}_l\}$, with l scalars: $a_{1j}, a_{2j}, ..., a_{lj}$, such that $T(\vec{v}_j) = a_{1j}\vec{w}_1 + a_{2j}\vec{w}_2 + ... + a_{lj}\vec{w}_l$. Thus, T is determined by the $l \cdot k$ scalars a_{ij} where $i = 1, .., l$ and $j = 1, ..., k$.

Proposition:
Let $T : V \to W$ be a linear transformations between finite spaces. If $\{\vec{v}_1, ..., \vec{v}_k\}$ is a basis for V and $\{\vec{w}_1, ..., \vec{w}_l\}$ a basis for W, then $T : V \to W$ is determined by the $l \cdot k$ scalars used to express $T(\vec{v}_j)$ in terms of $\vec{w}_1, ..., \vec{w}_l$.

We will now introduce matrices of scalars as a tool to simplify expressions of linear transformations:

Definition:
Let $a_{ij}, 1 \leq i \leq l$ and $1 \leq j \leq k$ be $l \cdot k$ scalars. The matrix with these entries is the rectangular array of l rows and k columns:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ & \vdots & & & \\ a_{l1} & a_{l2} & a_{l3} & \cdots & a_{lk} \end{bmatrix} \tag{1}$$

This is referred to as an "$l \times k$ matrix," and are denoted with capital letters.

If we have a linear transformation between finite spaces V and W, the transformation is determined by the choice of bases for V and W and the set of $l \times k$ scalars, where $k = \dim(V)$ and $l = \dim(W)$. Writing these scalars in the form of a matrix gives us a much easier way to study the transformation.

Definition:
Let $T : V \to W$ be a linear transformation between finite vector spaces V and W, with any bases α and β respectively. The matrix with scalars a_{ij} is known as the **matrix of linear transformation T with respect to bases α for V and β for W**. This matrix is denoted by $[T]_\alpha^\beta$.

We notice that $T(\vec{v}_j) = a_{1j}\vec{w}_1 + ... + a_{lj}\vec{w}_l$, the coefficients form the jth column of $[T]_\alpha^\beta$. Let us now study some examples of the previous concepts in action.

Let $T : V \to V$ be the identity transformation of a finite vector space to itself. With respect to **any** choice of basis, the matrix of I is the $k \times k$ matrix with 1 in each diagonal position and 0 in all other positions.

Return to the rotation transformation for some time; for an arbitrary vector $\vec{v} = (v_1, v_2)$, we have:

$$R_\theta(\vec{v}) = (v_1 \cos(\theta) - v_2 \sin(\theta), v_1 \sin(\theta) + v_2 \cos(\theta))$$

Thus, we may express the matrix of R_θ as:

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

In order to do computations using the matrix $[T]_\alpha^\beta$ to express $T(\vec{v})$, we must introduce the concept of multiplying a vector by a matrix:

Definition:
Let A be an $l \times k$ matrix, and let $\vec{x}$ be a column vector with k entries. Then, the product of vector $\vec{x}$ by matrix A is defined as the column vector with l entries:

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + ... + a_{1k}x_k \\ \vdots \\ a_{l1}x_1 + a_{l2}x_2 + ... + a_{lk}x_k \end{bmatrix}$$

and is denoted by $A\vec{x}$. We may also write it as:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1k} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2k} \\ & . & & & \\ & . & & & \\ & . & & & \\ a_{l1} & a_{l2} & a_{l3} & \cdots & a_{lk} \end{bmatrix} \begin{bmatrix} x_1 \\ . \\ . \\ . \\ x_k \end{bmatrix}$$

The ith entry of the product, $a_{i1}x_1 + ... + a_{ik}x_k$ can be thought of as the product of the ith row of A with $\vec{x}$. We can also make an important note: if the number of columns in A does not equal the number of entries in $\vec{x}$, then matrix multiplication is not defined.

Let us do an example:

$$\begin{aligned} \begin{bmatrix} 1 & 0 & 3 & 2 \\ 2 & 5 & 0 & 1 \\ 2 & -1 & 1 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 2 \\ -1 \end{bmatrix} &= 1 \cdot \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} + 0 \cdot \begin{bmatrix} 0 \\ 5 \\ -1 \end{bmatrix} + 2 \cdot \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} - 1 \cdot \begin{bmatrix} 2 \\ 1 \\ 9 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} 6 \\ 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 9 \end{bmatrix} \\ &= \begin{bmatrix} 5 \\ 1 \\ -5 \end{bmatrix} \end{aligned}$$

Matrix multiplication allows us to evaluate linear transformations in finite spaces in a procedural and concise manner. To evaluate $T(\vec{v})$

1. Compute the matrix of T with respect to the given bases for V and W,
2. Express the given vector $\vec{v}$ in terms of the basis for V, and
3. Multiply the coordinate vector of $\vec{v}$ by the matrix of T to obtain the new coordinate vector.

This can be aptly summed up in the following proposition.

Proposition:
Let $T : V \to W$ be a linear transformation between vector spaces V and W. Then, for each $\vec{v} \in V$:

$$[T(\vec{v})]_\beta = [T]_\alpha^\beta [\vec{v}]_\alpha$$

Let V and W be finite vector spaces and Let α be a basis for V and β a basis for W. Then, the assignment of a matrix to a linear transformation from V to W is bijective (both injective and surjective). This assignment of a matrix depends on the choices of α and β. If we consider another two bases α' and β', we can say that in general, $[T]_\alpha^\beta \neq [T]_{\alpha'}^{\beta'}$. We will explore this further in Section 2.7.

2.3 Kernel and Image

Let $T : V \to W$ be a linear transformation; in this section, we will define and discuss two subspaces, the kernel and image such that $\text{Ker}(T) \subset V$ and $\text{Im}(T) \subset W$. Understanding these subspaces unlocks a crucial, deeper understanding into the properties of T and solutions to systems of equations. We can find bases of these subspaces by using the elimination technique discussed in Section 1.5. Let us dive into some definitions: