Work cited: *Linear Algebra*, Friedberg et al. Theorem numbering and section numbers will follow the text.

§1. Vector spaces.

Axioms of a Vector Space.

A vector space is a set closed under addition and scalar multiplication that satisfies the following properties:

1. Commutativity of addition: $x + y = y + x \ \forall\ x, y \in V$
2. Associativity of addition: $(x + y) + z = x + (y + z) \ \forall\ x, y, z \in V$
3. Additive identity: $\exists\ 0_V$ with $0_V + x = x\ \forall\ x \in V$
4. Additive inverse: $\forall\ x \in V,\ \exists\ y \in V$ with $x + y = 0_V$
5. Associativity of scalar multiplication: $(\lambda\mu)x = \lambda(\mu x)\ \forall\ \lambda, \mu \in \mathbb{F},\ x \in V$
6. Distributivity (i): $\lambda(x + y) = \lambda x + \lambda y\ \forall\ \lambda \in \mathbb{F},\ x, y \in V$
7. Distributivity (ii): $(\lambda + \mu)x = \lambda x + \mu x\ \forall\ \lambda, \mu \in \mathbb{F},\ x \in V$
8. Multiplicative identity: $\exists\ 1 \in V$ with $1 \cdot x = x\ \forall\ x \in V$

Subspaces.

A *subspace* is a subset of a vector space V over $\mathbb{F}$ that satisfies the axioms of a vector space.

Theorem 1.3. As the set is a subset of V, some properties follow directly. The subset is a subspace if it satisfies the following properties:

1. Contains the additive identity (zero element)
2. Closed under addition
3. Closed under scalar multiplication

Subfields are subsets of a field that satisfy these same properties above. Note that the operations on the subset must be defined the same for both child and parent fields; integers modulo a prime are not subfields of the reals, for instance.

Theorem 1.4. Let V be a vector space. Let $\{W_i : i \in I\}$ be a family of subspaces of V, where $I \neq \varnothing$. Then

$$\bigcap_{i \in I} W_i = \{x : x \in W_i\ \forall\ i \in I\}$$

is again a subspace of V.

Proof.
1. For a given $i \in I$, since W_i is a subspace, $0 \in W_i\ \forall\ i \in I$. Thus, $0 \in \bigcap_{i \in I} W_i$.
2. Suppose $w_1, w_2 \in \bigcap_{i \in I} W_i$ are given. Consider $w_1 + w_2$. For each $i \in I$, $w_1, w_2 \in W_i$. Then $w_1 + w_2 \in W_i$ for each $W_i, \Rightarrow w_1 + w_2 \in \bigcap_{i \in I} W_i$.
3. Suppose $w \in \bigcap_{i \in I} W_i$ and $\lambda \in \mathbb{F}$, and consider λw. For each $i \in I, w \in W_i$. Therefore, as W_i is a subspace, $\lambda w \in W_i$. So $\lambda w \in \bigcap_{i \in I} W_i$.

□

Linear Combinations and Generating Sets.

We say that a subset $S \subseteq V$ *generates* or *spans* V if the set of all linear combinations of the elements of S equals V. We write this $\text{span}(S) = V$. Additionally, we define $\text{span}(\varnothing) = \{0\}$.

Let V be a vector space, $S = \{s_1, \ldots, s_n\} \subseteq V,\ S \neq \varnothing$. Then

$$\text{span}(S) = \left\{ \sum_{i=1}^{n} \lambda_i s_i : s_i \in S, \lambda_i \in \mathbb{F} \right\}$$

We call $\text{span}(S)$ the subspace *generated* by S.

Remark. For a vector space V, if S is a subspace of V, then $\text{span}(S) = S$.

Theorem 1.5. If S is a subset of a vector space V, then the set $W = \text{span}(S)$ is a subspace of V. Moreover, it is the smallest subspace of V containing S; that is, any subspace of V that contains S must contain $\text{span}(S)$.

Proof. We must verify W is a subspace of V. Certainly, $0 \in W$, since $\text{span}(\varnothing) = \{0\}$ and any larger set must generate 0 (by setting all scalars to 0 in the linear combination).

Now we check closure under the operations.

If y and z are elements of W, then y and z are linear combinations of elements of S. So there exist elements $x_1, x_2, \ldots, x_n$ and $w_1, \ldots, w_m$ in S such that $y = \sum_{i=1}^{n} a_i x_i$ and $z = \sum_{j=1}^{m} b_i w_i$ for some choice of scalars a_i and b_i. Then

$$\begin{aligned} y + z &= a_1 x_1 + \cdots + a_n x_n + b_1 w_1 + \cdots + b_m w_m \in S \\ cy &= c a_1 x_1 + \cdots + c a_n x_n \in S \end{aligned}$$

and as these are linear combinations of elements of S, they belong to W. Thus, W is a subspace of V.

Now consider a subspace of V, $W' \supseteq S$. Take any element of W, $y = \sum_{i=1}^{n} a_i x_i$, where $a_i \in \mathbb{F}$ and $x_n \in S$. Since $S \subseteq W'$, $x_1, \ldots, x_n \in W'$ which implies $y \in W'$ since W' is closed under addition and scalar multiplication as a subspace of V. Then we see that $W \subseteq W'$, so W must be the smallest subspace of V containing S.

□

Linear Dependence and Independence.

Let V be a vector space. Let $v_1, v_2, \ldots, v_n$ be a finite list of distinct vectors of V. We say that the list is *linearly dependent* if either of the following equivalent statements are true:

1. There is a $v_{i_0} \in \text{span}\{v_i : i \neq i_0\}$, e.g. some v_{i_0} is a linear combination of the rest.
2. For some $a_1, a_2, \ldots, a_n \in \mathbb{F} \setminus \{0\}, \sum_{i=1}^{n} a_i v_i = 0$.

A subset S of vector space V that is not linearly dependent is said to be *linearly independent*, e.g. the only way a linear combination of these vectors v_i to equal zero is if every scalar $a_i = 0$.

Note the following facts about linearly independent sets:

1. The empty set is linearly independent, since linearly dependent sets must be non-empty.
2. A set containing a single non-zero vector is always linearly independent.
3. For any vectors $x_1, x_2, \ldots, x_n$, we have $\sum_{i=0}^{n} a_i x_i = 0$ if and only if every $a_i = 0$. We call this the *trivial representation of 0* as a linear combination.

Theorem 1.6. Let V be a vector space with $S_1 \subseteq S_2 \subseteq V$. If S_1 is linearly dependent, then so is S_2.

Corollary: If $S_1 \subseteq S_2 \subseteq V$ and S_2 is linearly independent, then so is S_1.

Bases and Dimension.

A *basis* β is a linearly independent subset of V with $\text{span}(\beta) = V$.

Example.
Recalling that $\text{span}(\varnothing) = \{0\}$, and $\varnothing$ is linearly independent, $\varnothing$ is a basis for the vector space $\{0\}$.

Theorem 1.7. Every vector $v \in V$ can be expressed as a unique linear combination of all vectors in its basis, β. This follows directly from the definition of a basis.

Theorem 1.8. Let S be a linearly independent subset of a vector space V, and let x be an element of V that is not in S. Then $S \cup \{x\}$ is linearly dependent if and only if $x \in \text{span}(S)$.

Proof. If $S \cup \{x\}$ is linearly dependent, then there are vectors $x_1, \ldots, x_n \in \{x\} \cup S$ and non-zero scalars $a_1, \ldots, a_n$ such that $a_1 x_1 + \cdots + a_n x_n = 0$. Because S is linearly independent, one of the x_i, say x_1, equals x. Thus $a_1 x_1 + \cdots + a_n x_n = 0$, and so

$$x = a_1^{-1}(-a_2 x_2 - \cdots - a_n x_n).$$

Since x is a linear combination of $x_2, \ldots, x_n$, $x \in \text{span}(S)$.

Conversely, suppose that $x \in \text{span}(S)$. Then there exist vectors in S and scalars such that $x = a_1 x_1 + \cdots + a_n x_n$. Then

$$0 = a_1 x_1 + \cdots + a_n x_n + (-1)\, x,$$

so the set $\{x_1, \ldots, x_n, x\}$ is linearly dependent. Thus $S \cup \{x\}$ is linearly dependent.

□

Theorem 1.9. If a vector space V is generated by a finite set S_0, then a subset $S \subseteq S_0$ is a basis for V. Thus, V has a finite basis.

$\diamondsuit$ **Theorem 1.10: The Replacement Theorem.** Let V be a vector space generated by a set G containing exactly n elements. Let $S = \{y_1, \ldots, y_m\}$ be a linearly independent subset of V containing exactly m elements, where $m \leq n$. Then there exists a subset $S_1 \subseteq G$ containing exactly $n - m$ elements such that $S \cup S_1$ generates V.

Proof. We will prove this by induction on m. Our base case is $m = 0$, indicating $S = \varnothing$, so taking $S_1 = G$ clearly satisfies $\text{span}(S \cup S_1) = \text{span}(G) = V$.

Now assume the theorem is true for some $m < n$. We now prove the theorem for $m + 1$. Let $S = \{y_1, \ldots, y_m, y_{m+1}\}$ be a linearly independent subset of V containing exactly $m + 1$ elements. Since S is linearly independent, by the corollary to Thm. 1.6, we see that $\{y_1, \ldots, y_m\}$ is a linearly independent set. Then we may apply the inductive hypothesis to conclude that there exists a subset $\{x_1, \ldots, x_{n-m}\}$ of G such that $\{y_1, \ldots, y_m\} \cup \{x_1, \ldots, x_{n-m}\}$ generates V. Then there exist scalars $a_1, \ldots, a_m, b_1, \ldots, b_{n-m}$ such that

$$y_{m+1} = a_1 y_1 + \cdots + a_m y_m + b_1 x_1 + \cdots + b_{n-m} x_{n-m}$$

Observe that some b_i, say b_1, must be non-zero, because otherwise, this would imply that y_{m+1} is a linear combination of $y_1, \ldots, y_m$, contradicting our assumption that $\{y_1, \ldots, y_m, y_{m+1}\}$ is linearly independent. Then we can solve for x_1:

$$x_1 = (-b_1^{-1})(a_1 y_1 + \cdots + a_m y_m - y_{m+1} + b_2 x_2 + \cdots + b_{n-m} x_{n-m})$$

Hence, $x_1 \in \operatorname{span}(\{y_1, \ldots, y_m, y_{m+1}, x_2, \ldots, x_{n-m}\})$. Then it is clear that

$$\{y_1, \ldots, y_m, y_{m+1}, x_1, x_2, \ldots, x_{n-m}\} \subseteq \operatorname{span}(\{y_1, \ldots, y_m, y_{m+1}, x_2, \ldots, x_{n-m}\}).$$

Since $\{y_1, \ldots, y_m, x_1, \ldots, x_{n-m}\}$ generates V, by Thm. 1.5, the span of any subset of V containing these vectors is a subspace of V containing the span of this set. As this set generates V, it follows that

$$\operatorname{span}(\{y_1, \ldots, y_m, y_{m+1}, x_2, \ldots, x_{n-m}\}) = V$$

The set $\{x_2, \ldots, x_{n-m}\}$ has $(n-m) - 1 = n - (m+1)$ elements. So taking $S_1 = \{x_2, \ldots, x_{n-m}\}$ proves the theorem for the $m+1$ case.

$\square$

Corollary 1. Let V be a vector space with a basis β containing exactly n elements. Then any linearly independent subset of V containing exactly n elements is a basis for V.

Proof. Let $S = \{y_1, \ldots, y_n\}$ be a linearly independent subset of V containing exactly n elements. Applying the replacement theorem, we see there exists a subset $S_1 \subseteq \beta$ containing $n - n = 0$ elements such that $S \cup S_1$ generates V. As $S_1 = \varnothing$, $\operatorname{span}(S \cup S_1) = \operatorname{span}(S) = V$. Since S is also linearly independent, it must be a basis.

$\square$

Corollary 2. Let V be a vector space with a basis β containing exactly n elements. Then any subset of V containing more than n elements is linearly dependent. As such, any linearly independent subset of V must contain at most n elements.

Proof. Let S be a subset of V containing more than n elements. Let us assume that S is linearly independent. Let S_1 be any subset of S containing exactly n elements; by the replacement theorem, S_1 will be a basis for V by Cor. 1. Since $S_1 \subset S$, we can select an element $x \in S$ that is not an element of S_1. As S_1 is a basis for V, $x \in \operatorname{span}(S_1) = V$. Thus Thm. 1.8 implies that $S_1 \cup \{x\}$ is linearly dependent. But this contradicts our assertion that S is linearly indepedent, so it must be linearly dependent.

$\square$

This corollary can be applied to obtain a very useful formula.

Lagrange Interpolation.

Let $c_0, c_1, \ldots, c_n$ be distint scalars in an infinite field, $\mathbb{F}$. Define the polynomials $f_0(x), f_1(x), \ldots, f_n(x)$ by

$$f_i(x) = \frac{(x - c_0)\cdots(x - c_{i-1})(x - c_{i+1})\cdots(x - c_n)}{(c_i - c_0)\cdots(c_i - c_{i-1})(c_i - c_{i+1})\cdots(c_i - c_n)} = \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{x - c_k}{c_i - c_k}$$

It can be shown that for a linear combination of these $f_i(x)$, the polynomial

$$g = \sum_{i=0}^{n} b_i f_i$$

is the unique polynomial in $\mathcal{P}_n(\mathbb{F})$ such that $g(c_i) = b_i$. (See pages 51-53 of *Linear Algebra.*) So we can easily find a polynomial g of degree n given points $(c_0, b_0), (c_1, b_1), \dots, (c_n, b_n)$ we would like it to pass through.

Corollary 3. Let V be a vector space with a basis β containing exactly n elements. Then every basis for V contains exactly n elements.

A vector space V is called *finite-dimensional* if it has a basis consisting of a finite number of elements. The unique number of elements in each basis for V is called the *dimension* of V and is denoted $\dim(V)$. If a vector space is not finite-dimentional, then it is called *infinite-dimensional.*

Example. The vector space $\{0\}$ has dimension 0.

Example. The vector space $\mathbb{F}^n$ has dimension n.

Corollary 4. Let V be a vector space with $\dim(V) = n$, and let S be a subset of V that generates V and contains at most n elements. Then S is a basis for V and thus contains exactly n elements.

Proof. There exists a subset $S_1 \subseteq S$ such that S_1 is a basis for V by Thm. 1.9. By Cor. 3, S_1 contains exactly n elements. But $S_1 \subseteq S$ and S contains at most n elements. So $S = S_1$, and S is a basis for V.

$\square$

Corollary 5. Let β be a basis for a finite-dimensional vector space V, and let S be a linearly independent subset of V. Then there exists a subset $S_1 \subseteq \beta$ such that $S \cup S_1$ is a basis for V. Thus, every linearly independent subset of V can be extended to a basis for V.

Maximal Linearly Independent Subsets.

Let $\mathscr{C}$ be a collection of sets. A member $M \in \mathscr{C}$ is called the *maximal element* if no member of $\mathscr{C}$ properly contains M.

Let X be a set, and $\mathscr{C}$ be a collection of subsets of X. A subcollection of $\mathscr{C}$, say $\mathscr{T} \subset \mathscr{C}$ is called a *tower* (or *chain*) if for any two $T_1, T_2 \in \mathscr{T}$, either $T_1 \subseteq T_2$ or $T_2 \subseteq T_1$.

The Maximal Principle.

Let $\mathscr{C}$ be a collection of sets. If for every tower $\mathscr{T}$, there exists a $C \in \mathscr{C}$ such that $C \supset T \ \ \forall\, T \in \mathscr{T}$, then C is called an *upper bound* for $\mathscr{T}$, and furthermore, $\mathscr{C}$ will contain a maximal element $M \in \mathscr{C}$.

It would be useful to use this to rewrite the definition of a basis in terms of the maximal property. We first need another definition:

Let S be a subset of a vector space V. A *maximal linearly independent subset* of S is a subset $B \subseteq S$ satisfying both of the following conditions:

1. B is linearly independent.
2. Any subset of S that strictly contains B is linearly dependent.

Theorem 1.12. Let V be a vector space and S a subset that generates V. Then β is a maximal linearly independent subset of S if and only if β is a basis for V.

Proof. First, suppose β is a MLIS of S. Because β is linearly independent, we just have to prove that $\text{span}(\beta) = V$. Let us suppose that $S \not\subseteq \text{span}(\beta)$; then $\exists\ x \in S$ such that $x \notin \text{span}(\beta)$. Since Thm. 1.8 would then imply that $\beta \cup \{x\}$ is linearly independent, we have contradicted the maximality of β. Thus $S \subseteq \text{span}(\beta)$. Because $\text{span}(S) = V$, it follows that $\text{span}(\beta) = V$.

Conversely, suppose β is a basis for our vector space V. Note β is linearly independent by definition. Now consider $x \in V$ with $x \notin \beta$. By Thm. 1.8, $\beta \cup \{x\}$ is linearly dependent because $\text{span}(\beta) = V$. So β must be the MLIS of V.

$\square$

§2. Linear Transformations and Matrices.

Preamble.

Before we can consider linear transformations, we need some definitions.

We call a function $f : V \to W$ *injective* (or *one-to-one*) if for every $a, b \in V$, $f(a) = f(b)$ implies $a = b$. Effectively, f maps every $x \in V$ to a unique $y \in W$.

We call a function $g : V \to W$ *surjective* (or *onto*) if for every $y \in W$, there exists an $x \in V$ such that $g(x) = y$. Effectively, g maps some element to every $y \in W$ such that the image of g is W.

A function $h : V \to W$ is *bijective* if it is both injective and surjective. In this case, h maps every element of V to every element in W, and each pair of elements is unique. A function is *invertible* if and only if it is bijective; that is, the function's inverse is well-defined over the entire codomain of the original function.

Linear Transformations.

Let V and W be vector spaces over $\mathbb{F}$. We call a function $T : V \to W$ a *linear transformation* from V to W if, for all $x, y \in V$ and $c \in \mathbb{F}$, we have

1. $T(x + y) = T(x) + T(y)$ (additive)
2. $T(cx) = cT(x)$ (homogeneity)

It follows that every linear transformation must have $T(0) = 0$.

We define the *identity transformation* $I_V : V \to V$ by $I_V(x) = x$ for all $x \in V$ and the *zero transformation* $T_0 : V \to W$ by $T(x) = 0$ for all $x \in V$. These transformations are clearly linear.

The *null space* (or *kernel*) of $T : V \to W$ is defined as $N(T) = \{x \in V : T(x) = 0\}$; that is, the set of all elements in V that are mapped to 0.

The *range* (or *image*) of T is defined as $R(T) : \{T(x) : x \in V\}$; that is, the subset of W that T maps all elements of V to.

Example. The zero transformation's null space, $N(T_0) = V$, and its range, $R(T_0) = \{0\} \subseteq W$.

Theorem 2.1. Let V and W be vector spaces, and $T : V \to W$ be a linear transformation. Then $N(T)$ and $R(T)$ are subspaces of V and W, respectively.

Theorem 2.2. Let V and W be vector spaces, and $T : V \to W$ be linear. If $\beta = \{v_1, v_2, \ldots, v_n\}$ is a basis for V, then

$$R(T) = \operatorname{span}(T(\beta)) = \operatorname{span}(\{T(v_1), T(v_2), \ldots, T(v_n)\}).$$

This theorem allows us to quickly find a basis for the range, and thus its dimension.

We define the *rank* of T to be $\dim(R(T))$ and the *nullity* of T to be $\dim(N(T))$, denoted $\operatorname{rank}(T)$ and $\operatorname{nullity}(T)$ respectively.

◇ **Theorem 2.3: The Dimension Theorem.** Let V and W be vector spaces, and let $T : V \to W$ be linear. If V is finite-dimensional, then

$$\operatorname{nullity}(T) + \operatorname{rank}(T) = \dim(V).$$

Proof. Let us suppose that $\dim(V) = n$, $\dim(N(T)) = k$, and $\{v_1, \ldots, v_k\}$ is a basis for $N(T)$. By corollary 5 of the replacement theorem, we can extend this to a basis $\beta = \{v_1, \ldots, v_n\}$ for V. We claim that $S_1 = \{T(v_{k+1}), T(v_{k+2}), \ldots, T(v_n)\}$ is a basis for $R(T)$.

Consider

$$\begin{aligned} R(T) &= \operatorname{span}(\{T(v_1), T(v_2), \ldots, T(v_n)\}) \\ &= \operatorname{span}(\{0, \ldots, 0, T(v_{k+1}), \ldots, T(v_n)\}) \\ &= \operatorname{span}(\{T(v_{k+1}), \ldots, T(v_n)\}) \end{aligned}$$

Clearly, S_1 generates $R(T)$. But is the set linearly independent?

Well, let us suppose that

$$\sum_{i=k+1}^{n} b_i T(v_i) = 0 \text{ for some } b_{k+1}, b_{k+2}, \ldots, b_n \in \mathbb{F}$$

Since T is linear, this is equal to

$$T\left(\sum_{i=k+1}^{n} b_i T(v_i)\right) = 0$$

But then $\displaystyle\sum_{i=k+1}^{n} b_i v_i \in N(T)$, so there will exist some scalars $c_1, \ldots, c_k \in \mathbb{F}$ such that

$$\sum_{i=k+1}^{n} b_i v_i = \sum_{i=1}^{k} c_i v_i.$$

—in other words, since the result of this linear combination is in the null space, as it is mapped to 0, we must be able to express it as a linear combination of elements of the null space's basis.

Then

$$\begin{aligned} &\sum_{i=k+1}^{n} b_i v_i + \sum_{i=1}^{k} (-c_i) v_i = 0 \\ \Rightarrow &\sum_{i=1}^{n} a_i v_i = 0 \end{aligned}$$

for $\{a_1, \ldots, a_n\} = \{-c_1, \ldots, -c_k, b_{k+1}, \ldots, b_n\} \in \mathbb{F}$; and since $\beta = \{v_1, \ldots, v_n\}$ is a basis for V, it is linearly independent. Thus, we must have $a_i = 0 \ \forall\, i \ \Rightarrow\ b_i = 0 \ \forall\, i$. So, S_1 is linearly independent. As such, each $T(v_{k+1}), \ldots, T(v_n)$ is distinct; thus, $\operatorname{rank}(R(T)) = n - k$. $\square$

Theorem 2.4. Let V and W be vector spaces, and let $T : V \to W$ be linear. Then T is injective if and only if $N(T) = \{0\}$.

Theorem 2.5. Let V and W be vector spaces of equal, finite dimension, and let $T : V \to W$ be linear. Then the following properties are equivalent:

1. T is one-to-one.
2. T is onto.
3. $\operatorname{rank}(T) = \dim(V)$

In other words, if we have any of these properties, then T is bijective.

Theorem 2.6. Let V and W be vector spaces over $\mathbb{F}$, and suppose that $\{v_1, \ldots, v_n\}$ is a basis for V. For $w_1, \ldots, w_n \in W$, there exists exactly one linear transformation $T : V \to W$ such that $T(v_i) = w_i$ for $i = 1, 2, \ldots, n$.

Corollary. Let V and W be vector spaces, and suppose V has a finite basis $\{v_1, \ldots, v_n\}$. If $U, T : V \to W$ are linear and $U(v_i) = T(v_i)$ for $i = 1, 2, \ldots, n$, then $U = T$.

Sums of Subspaces.

If S_1 and S_2 are non-empty subsets of a vector space V, then the *sum* of S_1 and S_2, denoted $S_1 + S_2$, is the set $\{x + y : x \in S_1, y \in S_2\}$.

A vector space V is called the *direct sum* of W_1 and W_2 if W_1 and W_2 are subspaces of V such that $W_1 \cap W_2 = \{0\}$ and $W_1 + W_2 = V$. We denote the direct sum as $W_1 \oplus W_2 = V$.

A subspace $W \subseteq V$ is said to be *T-invariant* if $T(x) \in W$ for every $x \in W$; that is, $T(W) \subseteq W$, or W is closed under T.

Matrix Representation of a Linear Transformation.

Let V be a finite dimensional vector space. An *ordered basis* for V is a finite sequence of linearly independent vectors that generate V.

For a vector space $\mathbb{F}^n$, we define the *standard ordered basis* as $\{e_1, e_2, \ldots, e_n\} = \{(1, \ldots, 0), \ \ldots\ , \ (0, \ldots, 1)\}$, and $\{1, x, x^2, \ldots, x^n\}$ for $P_n(\mathbb{F})$.

Let $\beta = \{u_1, u_2, \ldots, u_n\}$ be an ordered basis for a finite-dimensional vector space V. For $x \in V$, let $a_1, a_2, \ldots, a_n$ be the unique scalars such that

$$x = \sum_{i=1}^{n} a_i u_i$$

We define the *coordinate vector of x relative to β* by

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

We can also represent a linear transformation in matrix form. The *matrix representation of $T : V \to W$ in the ordered bases β and γ* is denoted $A = [T]_\beta^\gamma$. If $V = W$ and $\beta = \gamma$, then we simply write $A = [T]_\beta$. We can write a linear transformation in matrix form as:

$$A = \Big[[T(u_1)]_\gamma \ \cdots \ [T(u_n)]_\gamma \Big]$$

Here, we have tranformed each element in the ordered basis β of V, corresponding to each column. Then each column vector is the tranformed vector expressed as a coordinate vector relative to γ, the ordered basis for W.

For $T, U : V \to W$, we define $T + U : V \to W$ as $(T + U)(x) = T(x) + U(x)$ and $aT : V \to W$ by $(aT)(x) = aT(x)$. These are just the typical defintions for addition and scalar multiplication of functions. However, we should note that the resulting functions from these operations are also linear.

Theorem 2.8. Let V and W be finite-dimensional vector spaces with ordered bases β and γ respectively, and let T, $U : V \to W$ be linear transformations. Then

1. $[T+U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$ and
2. $[aT]_\beta^\gamma = a[T]_\beta^\gamma$

Theorem 2.9. Let V, W, Z be vector spaces over $\mathbb{F}$ and let $T : V \to W$ and $U : W \to Z$ be linear. Then we define UT as $U \circ T$ and $UT : V \to Z$ is linear.

Matrix Multiplication.

It is the computation of the matrix representation of this composition of linear functions that motivates the definition of matrix multiplication. We won't go into this in detail here, and rather we will just define matrix multiplication. (For more detailed information, see page 87, *Linear Algebra*; in particular, **Theorem 2.10.**)

Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. We define the product of A and B, denoted AB, to be the $m \times p$ matrix whose entries are defined as

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} \text{ for } 1 \leq i \leq m \text{ (row index)}, \ 1 \leq j \leq p \text{ (column index)}$$

In other words, each entry (i, j) of the new matrix is determined by multiplying the ith row of A with the jth row of B element-wise.

Example.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 2\cdot 1 & 1\cdot 0 + 2\cdot 0 \\ 3\cdot 1 + 4\cdot 1 & 3\cdot 0 + 4\cdot 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 7 & 0 \end{bmatrix}$$

Let A be an $m \times n$ matrix over a field $\mathbb{F}$ (denoted $A \in M_{m\times n}(\mathbb{F})$). The *left-multiplication by A transformation* is defined as $L_A : \mathbb{F}^n \to \mathbb{F}^m$ with $L_A(x) = Ax$ for each column vector of $x \in \mathbb{F}^n$.

Example.
As seen in our last example, we could define $L_A(x)$ as left-multiplication by $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

Theorem 2.16. Matrix multiplication is associative; that is, for matrices A, B, C with $A(BC)$ defined, $A(BC) = (AB)C$.

Invertibility and Isomorphism.

Let V and W be vector spaces, and let $T : V \to W$ be linear. A function $U : W \to V$ is said to be an *inverse* of T if $TU = I_W$ and $UT = I_V$. If T has an inverse, we say that T is *invertible*, and if T is invertible, its inverse is unique and denoted by T^{-1}.

For invertible functions U and T, the following hold:

1. $(TU)^{-1} = U^{-1}T^{-1}$
2. $(T^{-1})^{-1} = T$; it should be obvious that T^{-1} is invertible.

With this knowledge, we can restate **Theorem 2.5** as:

Let V and W be vector spaces of equal, finite dimension, and let $T : V \to W$ be linear. Then T is invertible if and only if $\text{rank}(T) = \dim(V)$.

Theorem 2.17. Let V and W be vector spaces, and let $T : V \to W$ be linear and invertible. Then $T^{-1} : W \to V$ is linear.

Considering the matrix representation of a linear transformation, can we extend the concept of invertibility to matrices? The answer is yes.

We call an $n \times n$ matrix A *invertible* if there exists an $n \times n$ matrix B such that $AB = BA = I$. If A is invertible, then its inverse B is unique and is denoted A^{-1}.

Lemma. Let $T : V \to W$ be linear and invertible. Then V is finite-dimensional if and only if W is finite-dimensional, and in this case, $\dim(V) = \dim(W)$.

We will call on this lemma in our next few theorems.

Theorem 2.18. Let V and W be finite-dimensional vector spaces with ordered bases β and γ respectively, and let $T : V \to W$ be linear. Then T is invertible if and only if $[T]_\beta^\gamma$ is invertible, and furthermore, $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.

Corollary. Let A be an $n \times n$ matrix. Then A is invertible if and only if L_A is invertible. Furthermore, $(L_A)^{-1} = L_{A^{-1}}$.

Now, let V and W be vector spaces. We say that V is *isomorphic* to W if there exists a linear transformation $T : V \to W$ that is invertible. Such a linear transformation is called an *isomorphism* from V onto W.

Theorem 2.19. Let V and W be finite-dimensional vector spaces over the same field $\mathbb{F}$. V is isomorphic to W if and only if $\dim(V) = \dim(W)$.

§3. Matrix Operations and Systems of Equations.

Elementary operations and matrices.

We define three *elementary row (or column) operations* on the rows (columns) of an $m \times n$ matrix A:

1. Swapping any two rows (columns) of A: $R_i \longleftrightarrow R_j$
2. Multiplying any row (column) by a non-zero scalar: $\lambda R_i \to R_i$
3. Adding a scalar multiple of one row (column) to another row (column): $\lambda R_i + R_j \to R_j$

An $n \times n$ *elementary matrix* is the matrix obtained by applying any of these operations to I_n.

Theorem 3.1. Let $A \in M_{m\times n}(\mathbb{F})$, and suppose that B is obtained from A by applying an elementary row operation. Then there exists an $m \times m$ elementary matrix E such that $B = EA$. Furthermore, this matrix E is obtained by performing the same row operation that was performed on A on I_m.

Alternatively, suppose B is obtained from A by applying an elementary column operation. Then there exists an $n \times n$ matrix E such that $B = AE$, and E is obtained by performing this row operation on I_n.

With this, we now find ourselves able to represent row operations with left-multiplication, and column operations with right-multiplication on elementary matrices.

Reduced Row Echelon Form.

A matrix is said to be in *reduced row echelon form* (RREF) if:

1. Any row with non-zero entries preceed all rows in which all the entries are zero.
2. The first non-zero entry in each row is the only non-zero entry in that column.
3. The first non-zero entry in each row must be 1 and it must occur to the right of the first non-zero entry of rows preceeding it.

Example.

$$\begin{bmatrix} 3 & 2 & 6 & -2 & 1 \\ 1 & 1 & 1 & 0 & 3 \\ 1 & 2 & 1 & -1 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix}_{\text{RREF}}$$

Application. This matrix could be used to represent the solution for a set of linear equations:

$$\begin{array}{rcrcrcrcl} 3x_1 & + & 2x_2 & + & 6x_3 & - & 2x_4 & = & 1 \\ x_1 & + & x_2 & + & x_3 & & & = & 3 \\ x_1 & + & 2x_2 & + & x_3 & - & x_4 & = & 2 \end{array} \quad \Rightarrow \quad \begin{array}{rcl} x_1 + x_3 & = & 1 \\ x_2 & = & 2 \\ x_4 & = & 3 \end{array}$$

Application. We can create *augmented matrices* by merging matrices side by side into one large block. By creating an augmented matrix of some square matrix A and the identity matrix, we can row reduce to determine if A is invertible. If the left block of the augmented matrix reduces to the identity matrix, then A is invertible and A^{-1} can be found on the right block. If A is not invertible, we will not reach the identity matrix.

$$\left[\begin{array}{c|c} A & I_n \end{array}\right] \Rightarrow \left[\begin{array}{c|c} I_n & A^{-1} \end{array}\right]$$

Matrix Rank.

We have defined invertibility of a linear transform, then extended it to the matrix. Now we will extend the definition of rank to a matrix in a similar manner.

The *rank* of a matrix A, $\text{rank}(A)$ is defined as the rank of the left-multiplication transform by A; that is, for $L_A : \mathbb{F}^n \to \mathbb{F}^m$, $\text{rank}(A) = \text{rank}(L_A)$.

Remark. The rank of an invertible $n \times n$ square matrix must equal n, since an invertible linear transform $T : V \to W$ has rank $\text{rank}(T) = \dim(V)$, and $(L_A)^{-1} = L_{A^{-1}}$.

Theorem 3.3. For a linear transformation $T : V \to W$ where V and W are finite-dimensional vector spaces, with ordered bases β and γ respectively, $\text{rank}(T) = \text{rank}([T]_\beta^\gamma)$. In other words, for any linear transformation, the rank of the transformation is equal to the rank of its representative matrix.

In order to more easily determine the rank of a matrix, it would be useful to know what operations to not affect a matrix's rank. This next theorem helps to develop such methods.

Theorem 3.4. Let A be an $m \times n$ matrix. If P and Q are invertible $m \times m$ and $n \times n$ matrices, then

1. $\text{rank}(AQ) = \text{rank}(A)$ and
2. $\text{rank}(PA) = \text{rank}(A)$.

Corollary. It follows directly from this that elementary row and column operations on a matrix are rank-preserving.

Theorem 3.5. The rank of any matrix is equal to the maximum number of its linearly independent columns. Alternatively, the rank of a matrix is equal to the dimension of the *column space* (the subspace generated by the columns of the matrix). This is called the *column rank* of the matrix. So by this definition, we have put the rank of a matrix equal to its column rank.

Theorem 3.6. For an $m \times n$ matrix A of rank r, $r \leq m$ and $r \leq n$, and through elementary row and column operations, A can be reduced to the form

$$D = \begin{bmatrix} I_r & O_1 \\ O_2 & O_3 \end{bmatrix}$$

where O_i are zero matrices.

Corollary 1. The rank of any matrix is equal to the rank of its transpose: $\text{rank}(A) = \text{rank}(A^t)$. Additionally, the rank of any matrix is equal to the number of its linearly independent rows, as this forms an equivalence with columns. Thus, the rows and columns of any matrix generate a subspace of the same rank, equal to the rank of the matrix. So the rank of the *row space*, the subspace generated by the rows of the matrix, is equal to that of the column space. So the *row rank* of the matrix is equal to its column rank.

Corollary 2. The rank r of any invertible matrix must be its size, as noted above. As we have demonstrated here, then it must be reduceable to I_r. As such, then any invertible matrix can be expressed as a product of elementary matrices.

Homogenous Equations and the Nullity of a Matrix.

A system $Ax = b$ of m linear equations in n unknowns is said to be *homogenous* if $b = 0$ (or *non-homogenous* if $b \neq 0$).

We can see easily that all homogenous systems have at least one solution: $x = 0$.

Theorem 3.8. Let $Ax = 0$ be a homogenous system of m linear equations with n unknowns over a field $\mathbb{F}$. If K denotes the set of all solutions, then $K = N(L_A)$; so K is a subspace of $\mathbb{F}^n$, with $\dim(K) = n - \text{rank}(L_A) = n - \text{rank}(A) = \text{nullity}(A)$.

A word of caution! Though this may resemble a statement of the dimension theorem, it is NOT by any means. It is entirely incorrect to refer to the dimension of a matrix. $\dim(A)$ is not defined.

Now we can tie the definition of rank in with our knowledge of RREF.

Theorem 3.16. Let A be an $m \times n$ matrix of rank $r > 0$, and let $B = A_{\text{RREF}}$. Then the following properties hold:

1. B has exactly r non-zero rows.
2. For each i with $1 \leq i \leq r$, there is a column $b_{j_i} = e_i$.
3. These columns of A, numbered $j_1, \ldots, j_r$ are linearly independent.

Corollary. The RREF of a matrix is unique.

Remark. Comparing a matrix with its RREF, we can see that the column space differs, and the row space stays the same.

Proof. For our column space, consider $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

Clearly, the column space of this matrix and its RREF differ.

Now in considering the matrix's row space, we must look at the effects of elementary row operations on the span of the rows. Swapping rows and multiplying rows by a scalar constant certainly would not have any effect. And since adding a multiple of a row to another replaces a row with a linear combination from the last rowspace, this will not change the row space either. So the row space of a matrix and its RREF are the same.

□

§4. The Determinant.

Determinants of Order 1 and 2.

The *determinant* of a 1×1 matrix A is denoted $\det(A) = |A|$ and is defined as $\det(A) = A_{11}$.

The determinant of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is denoted

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

and is defined as the scalar $\det(A) = ad - bc$.

Example.

$$\det\left(\begin{bmatrix} 4 & 6 \\ 1 & 2 \end{bmatrix}\right) = \begin{vmatrix} 4 & 6 \\ 1 & 2 \end{vmatrix} = 4 \cdot 2 - 6 \cdot 1 = 2.$$

Theorem 4.1. The function $\det : M_{2\times 2}(\mathbb{F}) \to \mathbb{F}$ is a linear function of each row when the other row is held constant. That is to say,

$$\det\left(\begin{bmatrix} u + kv \\ w \end{bmatrix}\right) = \det\left(\begin{bmatrix} u \\ w \end{bmatrix}\right) + k\det\left(\begin{bmatrix} v \\ w \end{bmatrix}\right)$$

Theorem 4.2. The determinant of $A \in M_{2\times 2}(\mathbb{F})$ is non-zero if and only if A is invertible.

Determinants of Order n.

Cofactor Expansion: The recursive definition of the order n determinant.

For $A \in M_{n\times n}(\mathbb{F})$. For $n = 1$, put $\det(A) = (A_{11})$, and for $n \geq 2$, we define $\det(A)$ recursively as

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \cdot \det(\tilde{A}_{1j}),$$

where $\tilde{A}_{1j}$ is defined as the modified A matrix with the 1st row and jth column deleted.

Example.

$$\begin{aligned} \det\left(\begin{bmatrix} 5 & 2 & 1 \\ 3 & 3 & 3 \\ 0 & 4 & 1 \end{bmatrix}\right) &= (-1)^{1+1}(5) \cdot \begin{vmatrix} 3 & 3 \\ 4 & 1 \end{vmatrix} + (-1)^{1+2}(2) \cdot \begin{vmatrix} 3 & 3 \\ 0 & 1 \end{vmatrix} + (-1)^{1+3}(1) \cdot \begin{vmatrix} 3 & 3 \\ 0 & 4 \end{vmatrix} \\ &= -45 - 6 + 12 \\ &= -39 \end{aligned}$$

Theorem 4.3. For an $n \times n$ matrix A, the determinant function is linear on a row when all other rows are held constant. This is a generalization of Theorem 4.1.

Corollary. If a matrix A has a row consisting of all zeros, then $\det(A) = 0$.

Theorem 4.4. Cofactor expanasion is possible along any row. That is, for a matrix $A \in M_{n\times n}(\mathbb{F})$, for any i with $1 \leq i \leq n$,

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij}).$$

Corollary. If a matrix A has two identical rows, then $\det(A) = 0$.

These next three theorems will give us a way to find the determinant from a matrix to which we applied elementary row operations. Among other benefits, this allows us to find the determinant of a matrix A from its RREF.

These are all corollaries to Theorem 4.3, as they are based in the linearity of the determinant when rows are held constant.

Theorem 4.5. For a matrix $A \in M_{n\times n}(\mathbb{F})$, if B is obtained from A by swapping any two rows of A, then $\det(B) = -\det(A)$.

Theorem 4.6. For a matrix $A \in M_{n\times n}(\mathbb{F})$, if B is obtained from A by adding a multiple of one row to another, then $\det(B) = \det(A)$.

Theorem. For a matrix $A \in M_{n\times n}(\mathbb{F})$, if B is obtained from A by multiplying one row of A by a scalar factor λ, then $\det(B) = \lambda \det(A)$.

Corollary. If a matrix $A \in M_{n\times n}(\mathbb{F})$ has rank less than n, then $\det(A) = 0$. Of course, this is also clear from the fact that only invertible matrices, with $\operatorname{rank}(A) = n$, have non-zero determinants (a generalization of Theorem 4.2 which follows directly from the recursive definition).

Properties of Determinants.

Theorem 4.7. The determinant function is multiplicative. That is, for any two matrices $A, B \in M_{n\times n}(\mathbb{F})$, $\det(AB) = \det(A)\det(B)$.

Corollary. As we know, a matrix A is invertible if and only if $\det(A) \neq 0$. If A is invertible, then $\det(A^{-1}) = 1/\det(A)$.

Theorem 4.8. For any $A \in M_{n\times n}(\mathbb{F})$, $\det(A^t) = \det(A)$.

Corollaries. With this theorem, we can now extend what we know about the determinant and its relationship to the rows of a matrix to columns:

1. With Theorem 4.4, we can extend cofactor expansion to follow along columns rather that rows.
2. What we know about the effects of elementary row operations on the determinant can be extended to elementary column operations.
3. If any two columns of a matrix are equal, the determinant is 0.

The determinant of an *upper triangular matrix*, a matrix in which all entries below the diagonal entries are 0, is equivalent to the product of all the diagonal entries. In particular, the determinant of the identity matrix, $\det(I) = 1$.

Extras: The Trace Function and Similar Matrices.

The *trace* of a square matrix is the sum of all the diagonal entries of the matrix. That is, for an $n \times n$ matrix A,

$$\operatorname{trace}(A) = \sum_{i=1}^{n} A_{ii}.$$

Theorem. For two $n \times n$ matrices A and B, $\operatorname{trace}(AB) = \operatorname{trace}(BA)$.

Let A and B be square matrices of the same size. We say that A is *similar* to B if there exists some invertible matrix Q such that $A = Q^{-1}BQ$.

Theorem. Similar matrices share the following properties:

1. Rank.
2. Determinant.
3. Trace.
4. Eigenvalues/Characteristic polynomial (as we will reveal in the next section).

§5. Eigenvalues and Eigenvectors.

Definitions.

Let T be a linear operator on a vector space V. A vector $v \in V$, $v \neq 0$ is called an *eigenvector* of T if there exists a scalar $\lambda \in \mathbb{F}$ such that $T(v) = \lambda v$. The scalar λ is called an *eigenvalue* corresponding to the vector v.

This definition can be applied to a matrix $A \in M_{n\times n}(\mathbb{F})$. A non-zero vector $v \in \mathbb{F}^n$ is called an *eigenvector of* A if v is an eigenvector of L_A such that $Av = \lambda v$ for some scalar λ. This λ is referred to as the *eigenvalue* of A corresponding to the eigenvector v.

The set of all eigenvectors of a linear operator T corresponding to an eigenvalue λ is called the *eigenspace* of T with respect to λ and is denoted $E_\lambda = \{x \in V : T(x) = \lambda x\}$.

Example.

Let $A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$, $v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $v_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

Since

$$L_A(v_1) = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = (-2)\begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2v_1$$

So v_1 is an eigenvector of L_A, and thus an eigenvector of A. $\lambda_1 = -2$ is the eigenvalue corresponding to v_1. Additionally,

$$L_A(v_2) = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 15 \\ 20 \end{bmatrix} = (5)\begin{bmatrix} 3 \\ 4 \end{bmatrix} = 5v_2$$

So v_2 is also an eigenvector of A with eigenvalue $\lambda_2 = 5$.

It would be nice to be able to calculate such values with a more elegant method than brute force. We will now explore this concept.

Theorem 5.2. Let $A \in M_{n\times n}(\mathbb{F})$. A scalar λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Proof. To find eigenvalues of A, we need to solve $Av = \lambda v$. We can manipulate this equation to the form $(A - \lambda I_n)(v) = 0$. This equation has a solution if and only if $A - \lambda I_n$ is not invertible. This is true if and only if $\det(A - \lambda I_n) = 0$. $\square$

The Characteristic Polynomial.

For $A \in M_{n\times n}(\mathbb{F})$, we define the *characteristic polynomial of* A as $f(\lambda) = \det(A - \lambda I_n)$.

It should be clear that the roots of the characteristic polynomial correspond to the eigenvalues of A.

Example.

Let $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$. The characteristic polynomial of A is

$$\begin{vmatrix} 1-\lambda & 1 & 0 \\ 0 & 2-\lambda & 2 \\ 0 & 0 & 3-\lambda \end{vmatrix} = (1-\lambda)(2-\lambda)(3-\lambda)$$

since this matrix is in upper-triangular form. The roots of the characteristic polynomial are $\lambda = 1, 2, 3$, so λ is an eigenvalue for A if and only if it is one of these values.

§6. Supplementary Content.

Definitions.

- The *dot product* between two vectors $x = (x_i)_{i=1}^n$ and $y = (y_i)_{i=1}^n$ of $\mathbb{R}^n$ is defined as
$$x \cdot y := \sum_{i=1}^{n} x_i y_i \, ,$$
the sum of the pairwise products of the elements of the two vectors.

- The *Euclidian norm* (length) of a vector is defined as
$$\|x\| := \sqrt{\sum_{i=1}^{n} {x_i}^2} \, ,$$
the square root of the sum of every element of the vector squared.

- The *Euclidian distance* between two vectors $x, \, y \in \mathbb{R}^n$ is defined by
$$d(x, \, y) := \|x - y\|.$$

- We call two vectors x and y *orthogonal* if $x \cdot y = 0$.

- A *unit* or *normal vector* is any vector with $\|x\| = 1$. We denote such a vector by $\hat{x}$.

- For any non-zero vector, we can find a unit vector in relation to this vector with $\frac{1}{\|x\|}x$. This action is referred to as *normalizing* x.

- We define the *orthogonal projection* of $\mathbb{R}^n$ onto the line spanned by a unit vector $\hat{x}$ is the mapping given by
$$\mathrm{Proj}_{\hat{x}}(y) := (y \cdot \hat{x}) \, \hat{x}$$