Linear Algebra Notes

Apiros3

First Version : Jan 19, 2025
Last Update : Jan 19, 2025

Contents

1 Introduction

This note aims to provide an introduction to Linear Algebra.

This section aims to summarize the main concepts that follows in the notes, effectively acting as a synopsis. Section 2 first provides a basic set of definitions about vector spaces, then provides an explanation for why dimensions are well defined. The section then further shows how a basis can be extended to any subspace of a finite vector space. We then show some theorems that relate dimensions between these spaces. We finish the section with the rank nullity theorem. Section 3 then provides a basic set of definitions for matrices that will be used throughout the rest of the notes. Section 4 then introduces Reduced Row Echelon Forms (RREF), which can tell properties about rank and nullity of the matrix.

2 Vector Space

Definition 2.1 *A* ***field*** $\mathbb{F}$ *is a set with operations* $(+)$ *and* $(\times)$ *that satisfy the following properties for any* $a, b, c \in \mathbb{F}$*:*

- $(+)$ *and* $(\times)$ *are both associative and commutative*
- *Additive and multiplicative identity: there exist two distinct elements* $0, 1 \in \mathbb{F}$ *such that* $a + 0 = a$, $a \times 1 = a$
- *Additive inverse: there exists an element in* $\mathbb{F}$ *denoted* $(-a)$ *such that* $a + (-a) = 0$
- *Multiplicative inverse: there exists an element in* $\mathbb{F}$ *denoted* (a^{-1}) *or* $1/a$ *such that* $a \times a^{-1} = 1$
- *Distributivity:* $a \times (b + c) = a \times b + a \times c$

Example 2.2 *The following are some exaples of fields:*

- $\mathbb{Z}/p\mathbb{Z}$ *- the field of integers modulo a prime* p
- $\mathbb{Q}$ *- the field of rational numbers*
- $\mathbb{R}$ *- the field of real numbers*
- $\mathbb{C}$ *- the field of complex numbers*

Definition 2.3 *A* ***vector space*** *is a non-empty set* V *over a field* $\mathbb{F}$ *with a binary operation* $(+) : V \times V \to V$ *sending* $(u, v) \mapsto u + v$ *and a map* $\mathbb{F} \times V \to V$ *by* $(\lambda, v) \mapsto \lambda v$ *that satisfy the following rules:*

- $(+)$ *is associative and commutative*
- *Additive identity: there exists* $0_V \in V$ *such that for all* $v \in V$, $v + 0_V = v$
- *Additive inverse: for all* $v \in V$*, there exists* $w \in V$ *such that* $v + w = 0_V$
- *Distributivity: for all* $u, v \in V, \lambda \in \mathbb{F}$, $\lambda(u + v) = \lambda u + \lambda v$, $(\lambda + \mu)v = \lambda v + \mu v$, $(\lambda\mu)v = \lambda(\mu v)$
- *Identity on scalar multiplication: for all* $v \in V$, $1_{\mathbb{F}} v = v$

Example 2.4 *The following are some examples of vector spaces:*

- *A field* $\mathbb{F}$ *is a vector space over itself, where addition and scalar multiplication are inherited from the structure of* $\mathbb{F}$.
- *For any field* $\mathbb{F}$ *and* $m, n \geq 1$, $\mathcal{M}_{m \times n}(\mathbb{F})$ *is a vector space over* $\mathbb{F}$.
- $V = \mathbb{R}^{\mathbb{R}} = \{f : \mathbb{R} \to \mathbb{R}\}$ *with addition and scalar multiplication defined pointwise.*
- $V = \mathbb{R}^{\mathbb{N}} = \{(x_0, x_1, \dots) : x_i \in \mathbb{R}\}$ *with addition and scalar multiplication defined component-wise.*
- $V = \mathbb{R}^n$

Notation 2.5 *For any sets U and V with an operation $(+)$ that is defined between elements of the two,*

$$U + W = \{u + w \mid u \in U, w \in W\}$$

Lemma 2.6 *Let V be a vector space over $\mathbb{F}$. The additive identiy element of V is unique.*

Proof. Let 0 and 0' be two elements that satisfy the property of an additive identity. Then, $0 = 0 + 0' = 0'$. ∎

Notation 2.7 *We will write 0 to refer to the additive identity and 1 for the multiplicative identity when it is clear what object we are referring to, and write with a subscript when necessary.*

Lemma 2.8 *Let V be a vector space over $\mathbb{F}$. For any $v \in V$, there is a unique additive inverse of v. Specifically, if there exists $w_1, w_2 \in V$ such that $v + w_1 = v + w_2 = 0$, then $w_1 = w_2$.*

Proof.

$$\begin{aligned} w_1 &= 0 + w_1 \\ &= (w_2 + v) + w_1 \\ &= w_2 + (v + w_1) \\ &= w_2 + 0 \\ &= w_2 \end{aligned}$$

∎

Notation 2.9 *Where it is clear, we will write $(-v)$ to refer to the unique additive inverse of v.*

Proposition 2.10 (Basic Properties of vector spaces) *Let V be a vector space over a field $\mathbb{F}$. Take any $v \in V, \lambda \in \mathbb{F}$. Then,*

- $\lambda 0_V = 0_V$
- $0v = 0_V$
- $(-\lambda)v = -(\lambda v) = \lambda(-v)$
- *if $\lambda v = 0_V$, then $\lambda = 0$ or $v = 0_V$*
- $-v = (-1)v$

Proof. For the first case, note that

$$\lambda 0_V = \lambda(0_V + 0_V) = \lambda 0_V + \lambda 0_V$$

In the second case, note that

$$0v = (0 + 0)v = 0v + 0v$$

For the third case, we have

$$\lambda v + \lambda(-v) = \lambda(v + (-v)) = \lambda 0_V = 0_V$$

and

$$\lambda v + (-\lambda)v = (\lambda + (-\lambda))v = 0v = 0_V$$

For the fourth case, if $\lambda \neq 0$, as $\lambda^{1} \in \mathbb{F}$ and

$$\lambda^{-1}(\lambda v) = \lambda^{-1} 0_V = 0_V$$

So,

$$(\lambda^{-1}\lambda)v = 0_V$$

giving $v = 1v = 0_V$ Finally,

$$\begin{aligned} v + (-1)v &= 1v + (-1)v \\ &= (1 + (-1))v \\ &= 0v \\ &= 0_V \end{aligned}$$

■

Definition 2.11 *Let V be a vector space over $\mathbb{F}$. A* ***subspace*** *of V is a non-empty subset of V that is closed under addition and scalar multiplication. Specifically, a subset $U \subseteq V$ such that*

- $U \neq \emptyset$
- *for all $u, w \in U$, $u + w \in U$ $(U + U \subseteq U)$*
- *for all $u \in U, \lambda \in \mathbb{F}$, $\lambda u \in U$*

Remark 2.12 *The sets $\{0_V\}$ and V are subspaces of V.*

Example 2.13 *Given a fixed x, y, the set $\{(\alpha x, \alpha y) : \alpha \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$*

Notation 2.14 *We write $U \leq V$ to denote that U is a subspace of V.*

Proposition 2.15 (Subspace Test) *Let V be a vector space over $\mathbb{F}$ and $U \leq V$. Then, U is a subspace if and only if*
(i) $0_V \in U$
(ii) for all $u, w \in U$ and $\lambda \in \mathbb{F}$, $\lambda u + w \in U$

Proof. $(\Rightarrow)$ Assume that $U \leq V$. For (i), as U is nonempty, there exists a $u \in U$. As U is closed under scalar multiplication, $0u = 0_V \in U$. For (ii), this follows from the fact $\lambda u \in U$ and $w \in U$ so $\lambda u + w \in U$ (by closure under scalar multiplication and addition).
$(\Leftarrow)$ Assume both (i) and (ii). Then,

- $U \neq \emptyset$: as $0_V \in U$ by (i)
- closure under addition: for any $u, w \in U$, $u = w = 1u + w \in U$ by (ii)
- closure under multiplication: for any $u \in U$ and $\lambda \in \mathbb{F}$, $\lambda u = \lambda u + 0_V \in U$ by (ii)

■

Proposition 2.16 *Let V be a vector space over $\mathbb{F}$. Then, the subspaces of V that are vector spaces over $\mathbb{F}$ with the inherited operations.*

Proof. A subset that is a vector space over $\mathbb{F}$ is clearly a subspace of V, as it is a nonempty set that has well-defined operators that are closed under addition and scalar multiplication by elements of $\mathbb{F}$. Any subspace U is a vector space over $\mathbb{F}$, as properties for it to be a vector space are inherited from V. The operations are well defined as the restriction of $(+)$ gives a map $U \times U \to U$ and scalar multiplication gives a map $\mathbb{F} \times U \to U$ due to closure properties of the two.

Proposition 2.17 *The subspace operator $(\leq)$ is transitive.*

Proof. Follows immediately from definitions. ■

Proposition 2.18 *Let V be a vector space and $U, W \leq V$. Then, $U + W \leq V$ and $U \cap W \leq V$.*

Proof. We will use the subspace test for both cases. For $U + W$, note that $0_V \in U$ and $0_V \in W$ so $0_V \in U + W$. Take any $v_1, v_2 \in U + W$ and $\lambda \in \mathbb{F}$. Then, there exists $u_1, u_2 \in U$ and $w_1, w_2 \in W$ such that $v_1 = u_1 + w_1$ and $v_2 = u_2 + w_2$. Then,

$$\lambda v_1 + v_2 = \lambda(u_1 + w_1) + u_2 + w_2 = (\lambda u_1 + u_2) + (\lambda w_1 + w_2) \in U + W$$

as $U \leq V$ and $W \leq V$. For $U \cap W$, first note that $0_V \in U$ and $0_V \in W$ so $0_V \in U \cap W$. Taking any $v_1, v_2 \in U \cap W$ and $\lambda \in \mathbb{F}$, we see that $\lambda v_1 + v_2 \in U$ as $v_1, v_2 \in U$, and $\lambda v_1 + v_2 \in W$ as $v_1, v_2 \in W$, using also the fact that $U \leq V$ and $W \leq V$. Therefore, $\lambda v_1 + v_2 \in U \cap W$. ■

Remark 2.19 *It follows that if V be a vector space with $U, W \leq V$, $U + W$ is the smallest subspace of V that contains U and W, while $U \cap W$ is the largest subspsace of V that is contained in both U and W.*

Definition 2.20 *Let V be a vector space over $\mathbb{F}$ and $u_1, \dots, u_n \in V$. A* ***Linear Combination*** *of $u_1, \dots, u_n$ is a vector $\alpha_1 u_1 + \cdots + \alpha_n u_n$ for some $\alpha_1, \dots, \alpha_n \in \mathbb{F}$.*

Definition 2.21 *Let V be a vector space over $\mathbb{F}$ and $S \subseteq V$, where S can be finite or infinite. The* ***span*** *of S is defined as*

$$\langle S \rangle := \{\alpha_1 s_1 + \cdots + \alpha_n s_n : n \geq 0, s_1, \dots, s_n \in S, \alpha_1, \dots, \alpha_n \in \mathbb{F}\}$$

By convention, $\langle \emptyset \rangle = \{0_V\}$.

Notation 2.22 *For a finite set $S = \{u_1, \dots, u_n\}$, we often write $\langle u_1, \dots u_n \rangle$ to represent $\langle S \rangle$.*

Example 2.23 *The span of S only ever involves finite sums of elements, even if S is infinite. For instance, consider $V = \mathbb{R}^{\mathbb{N}} = \{(x_0, x_1, \dots) : x_i \in \mathbb{R}\}$ and $S = \{\mathbf{e}_i : i \in \mathbb{N}\}$, where*

$$\mathbf{e}_i = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}$$

Then, $\langle S \rangle = \{(x_0, x_1, \dots) : \exists N \in \mathbb{N},\ \forall n \geq N,\ x_n = 0\}$. That is, the set of sequences that eventually become zero. In particular, note that $(1, 1, \dots) \notin \langle S \rangle$.

Lemma 2.24 *Let V be a vector space over a field $\mathbb{F}$, and take any possibly empty $S = \{u_1, u_2, \dots, u_n\} \subseteq V$. Take $U := \langle S \rangle$. Then, $U \leq V$.*

Proof. Follows from the subspace test, writing each element of U as a linear combination of elements in S. ■

Definition 2.25 *Let V be a vector space over a field $\mathbb{F}$. If $S \subseteq V$ and $V = \langle S \rangle$, we say that S* ***spans*** *V and that S is a* ***spanning set*** *of V.*

Definition 2.26 *Let V be a vector space over $\mathbb{F}$. We say that $v_1, \ldots, v_n \in V$ are* ***linearly independent*** *if*

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0_V \qquad \alpha_1, \ldots, \alpha_n \in \mathbb{F}$$

implies that

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

Else, $v_1, \ldots, v_n$ are said to be linearly dependent, meaning there is a non-trivial linear combination of $v_1, \ldots, v_n$ that adds to 0_V.

Given $S \subseteq V$, we say that S is linearly independent if every finite subset of S is linearly independent.

Proposition 2.27 *Let V be a vector space and $S = \{v_1, \ldots v_n\} \subseteq V$ be a linearly independent set. Then,*

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = \beta_1 v_1 + \cdots + \beta_n v_n$$

if and only if $\alpha_i = \beta_i$ for all $1 \leq i \leq m$.

Proof. The ($\Leftarrow$) direction is immedidate. For ($\Rightarrow$), by rewriting the given equation as

$$(\alpha_1 - \beta_1)v_1 + \cdots + (\alpha_n - \beta_n)v_n = 0_V$$

and noting that S is linearly independent, it follows that $\alpha_i - \beta_i = 0$ for all i. ■

Proposition 2.28 *Let $v_1, \ldots, v_n$ be linearly independent elements of a vector space V. Take, $v_{n+1} \in V$. Then, $v_1, \ldots, v_n, v_{n+1}$ are linearly independent if and only if*

$$v_{n+1} \notin \langle v_1, \ldots, v_n \rangle$$

Proof. ($\Rightarrow$) Suppose $v_1, \ldots, v_{n+1}$ are linearly independent. Assume for a contradiction that $v_{n+1} \in \langle v_1, \ldots, v_n \rangle$. So, there exists $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that

$$v_{n+1} = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

But then $\alpha_1 v_1 + \cdots + \alpha_n v_n - v_{n+1} = 0_V$, which contradicts the linear independence of $v_1, \ldots, v_{n+1}$.

($\Leftarrow$) Suppose that $v_{n+1} \notin \langle v_1, \ldots, v_n \rangle$. Take any $\alpha_1, \ldots, \alpha_{n+1} \in \mathbb{F}$ such that

$$\alpha_1 v_1 + \cdots + \alpha_{n+1} v_{n+1} = 0_V$$

Suppose that $\alpha_{n+1} \neq 0$. Then,

$$v_{n+1} = -\frac{1}{a_{m+1}}(\alpha_1 v_1 + \cdots + \alpha_n v_n) \in \langle v_1, \ldots, v_n \rangle$$

which contradicts our assumption. So, $\alpha_{n+1} = 0$ and $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$. By linear independence of $v_1, \ldots v_n$, we conclude that $\alpha_1 = \cdots = \alpha_n = 0$. ■

Definition 2.29 *Let V be a vector space. A **basis** of V is a linearly independent, spanning set. If V has a finite basis, then we say that V is **finite dimensional**.*

Example 2.30 *Consider $V = \mathbb{F}[x]$ over $\mathbb{F}$, the set of $\mathbb{F}$ coefficient polynomials. Then, the set*

$$\{1, x, x^2, \dots\}$$

is a basis for V.

Example 2.31 *Taking $V = \mathbb{R}^n$, for $1 \leq i \leq n$, define $\mathbf{e}_i$ to be the row vector with coordinate 1 in the i-th entry and 0 otherwise. Then, $\mathbf{e}_1, \dots, \mathbf{e}_n$ forms a basis for $\mathbb{R}^n$. This is the **standard basis** or **canonical basis** for $\mathbb{R}^n$.*

Proposition 2.32 *Let V be a vector space over $\mathbb{F}$. Take $S = \{v_1, \dots, v_n\} \subseteq V$. Then S is a basis of V if and only if every vector in V can be written uniquely as a linear combination of elements in S.*

Proof. ($\Rightarrow$) Let S be a basis for V. Take any $v \in V$. Then, as S spans V, there exists $\alpha_1, \dots, \alpha_n \in \mathbb{F}$ such that $v = \alpha_1 v_1 + \cdots \alpha_n v_n$. Now, by Proposition 2.27, these scalars are unique.

($\Leftarrow$) Suppose every vector in V has a unique representation as a linear combination of elements of S. Then, S is spanning as for any $v \in V$ we may write it as a linear combination of elements in S. Taking any $\alpha_1, \dots, \alpha_n$ such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0_V = 0 v_1 + \cdots + 0 v_n$, by uniqueness we have $\alpha_i = 0$ for all i, giving linear independence. ∎

Definition 2.33 *Given a basis $\{v_1, \dots, v_n\}$ of V, then every $v \in V$ can be written uniquely as*

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

*with some scalars $\alpha_1, \dots \alpha_n$, which we call the **coordinates** of v with respect to the basis $v_1, \dots, v_n$.*

Remark 2.34 *Choosing a different basis gives different coordinates for the same vector, so it must be specified which basis we are referring to when talking about coordinates.*

Theorem 2.35 (Steinitz Exchange Lemma) *Let V be a vector space over a field $\mathbb{F}$. Take any finite $X = \{v_1, \dots, v_n\} \subseteq V$. Take any $u \in \langle X \rangle$ but $u \notin \langle X \backslash \{v_i\} \rangle$ for some i. Take*

$$Y = (X \backslash \{v_i\}) \cup \{u\}$$

Then, $\langle Y \rangle = \langle X \rangle$.

Proof. Let $u \in \langle X \rangle$. There are $\alpha_1, \dots, \alpha_n \in \mathbb{F}$ such that

$$u = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

By assumption, there is some $v_i \in X$ such that $u \notin \langle X \backslash \{v_i\} \rangle$. Without loss of generality, take $i = n$. As $u \notin \langle X \backslash \{v_n\} \rangle$, $\alpha_n \neq 0$. This gives

$$v_n = \frac{1}{\alpha_n}(u - \alpha_1 v_1 - \cdots - \alpha_{n-1} v_{n-1})$$

Taking any $w \in \langle Y \rangle$, then we can write w as a linear combination of elements in Y. Replacing u with $\alpha_1 v_1 + \cdots + \alpha_n v_n$, we can write w as linear combination of elements of X. This gives $\langle Y \rangle \subseteq \langle X \rangle$. Taking $w \in \langle X \rangle$, we can write w as a linear combination of elements of X. Replacing v_n by the equality above, we can write w as a linear combination of elements of Y. This gives $\langle X \rangle \subseteq \langle Y \rangle$. ∎

Theorem 2.36 *Let V be a vector space. Take U, W to be finite subsets of V. If U is linearly independent and W spans V, then $|U| \leq |W|$.*

Proof. Assume U is linearly independent and W spans V. Let $U = u_1, \ldots, u_m$ and $W = w_1, \ldots, w_n$. Take $T_0 = \{w_1, \ldots, w_n\}$. Noting $\langle T_0 \rangle = V$, take the smallest i such that $u_1 \in \langle w_1, \ldots, w_i \rangle$. Then, as $u_1 \notin \langle w_1, \ldots, w_{i-1} \rangle$, by the Steinitz Exchange Lemma,

$$\langle w_1, \ldots, w_i \rangle = \langle u_1, w_1, \ldots, w_{i-1} \rangle$$

It then follows that

$$\begin{aligned} V &= \langle w_1, \ldots, w_n \rangle \\ &= \langle w_1, \ldots, w_i \rangle + \langle w_{i+1}, \ldots, w_n \rangle \\ &= \langle u_1, w_1, \ldots, w_{i-1} \rangle + \langle w_{i+1}, \ldots, w_n \rangle \\ &= \langle u_1, w_1, \ldots, w_{i-1}, w_{i+1}, \ldots, w_n \rangle \end{aligned}$$

By relabelling elements of T, we can assume without loss of generality that u_1 has been exchanged for w_1, and we set

$$T_1 = \{u_1, w_2, \ldots, w_n\}$$

with $\langle T_1 \rangle = V$. Repeating this process inductively, we can obtain a T_k such that

$$T_k = \{u_1, \ldots, u_k, w_{k+1}, \ldots, w_n\}$$

with $\langle T_k \rangle = V$. Note that at any iteration $u_{k+1} \in \langle T_k \rangle$ but $u_{k+1} \notin \langle u_1, \ldots, u_k \rangle$ as U is independent. We can repeat this process until T_m, which gives $m \leq n$. ■

Corollary 2.37 *Let V be a finite dimensional vector space. Any basis for V is finite and are of the same size.*

Proof. As V is finite dimensional, it has a finite basis $\mathcal{B}$. By Theorem 2.36, any finite linearly independent subset of V has size at most $|\mathcal{B}|$. If $\mathcal{E}$ is another basis for V, every finite subset of $\mathcal{E}$ is linearly independent, meaning $\mathcal{E}$ is finite and $|\mathcal{E}| \leq |\mathcal{B}|$. But as $\mathcal{B}$ is linearly independent and $\mathcal{E}$ is a spanning set, by the same theorem $|\mathcal{B}| \leq |\mathcal{E}|$. ■

Definition 2.38 *Let V be a finite dimensional vector space. The **dimension** of V, written $\dim V$, is the size of any basis of V. We assure this value is well-defined by the previous corollary.*

Proposition 2.39 *Let V be a vector space over $\mathbb{F}$ and U be a finite spanning set. Then U contains a basis.*

Proof. Let U be a finite spanning set for V. Take any $W \subseteq U$ such that W is linearly independent and is maximal (no linearly independent subset of U strictly contains W). We assure such W exists, as U is finite. Suppose then for a contradiction that $\langle W \rangle \neq V$. Then, as $\langle U \rangle = V$, there exists a $v \in U \backslash \langle W \rangle$. Now, as $W \cup \{v\}$ is linearly independent with $W \cup \{v\} \subseteq U$, this contradicts the maximality of W. So W spans V and is linearly independent, thus is a basis. ■

Proposition 2.40 *Let V be a finite dimensional vector space with $U \leq V$. Then U is finite dimensional with $\dim U \leq \dim V$. If $\dim U = \dim V$ then $U = V$.*

Proof. Let $n = \dim V$. Take W to be the largest linearly independent set contained in U. It follows from 2.36 that $|W| \leq n$. We first show that $\langle W \rangle = U$, as else this contradicts the maximality of W. Thus, W is a basis for U. Therefore, $\dim U \leq \dim V$.

For the latter case, suppose $\dim U = \dim V$ with $U \neq V$. Then, there exists a $v \in V \backslash U$ which can be added to U to create a linearly independent subset of V that is greater than $\dim V$, which is a contradiction. ∎

Proposition 2.41 (Extending a Linearly Independent Set to a Basis) *Let V be a finite dimensional vector space over $\mathbb{F}$ and let U be a linearly independent set. Then, there exists a basis $\mathcal{B}$ such that $U \subseteq \mathcal{B}$.*

Proof. If $\langle U \rangle = V$, then we are done. Else, we can extend U to $U_1 = U \cup \{u_1\}$ where $u_1 \in U \backslash \langle U \rangle$ to create a larger linearly independent set. We repeat this process until $\langle S_k \rangle = V$, which must terminate as V is finite dimensional. ∎

Corollary 2.42 *A maximal linearly independent subset of a finite dimensional vector space is a basis.*

Proof. We note that adding an element that is not contained in the span of a linearly independent set maintains linear independence. Therefore, the maximal such set must span the entire vector space. ∎

Theorem 2.43 (The Dimension Formula) *Let V be a finite dimensional vector space over $\mathbb{F}$ and $U, W \leq V$. Then,*

$$\dim(U + W) + \dim(U \cap W) = \dim U + \dim W$$

Proof. Take a basis $\{v_1, \ldots, v_k\}$ for $U \cap W$. By Proposition 2.41, there exists an extension from this set to a basis $\{v_1, \ldots, v_k, u_1, \ldots, u_m\}$ for U and $\{v_1, \ldots, v_k, w_1, \ldots, w_n\}$ for of W. Then, we see that

$$\dim(U \cap W) = k \qquad \dim U = k + m \qquad \dim W = k + n$$

It is then sufficient to show that

$$S = \{v_1, \ldots, v_k, u_1, \ldots, u_m, w_1, \ldots, w_n\}$$

is a basis for $U + W$.

$\langle S \rangle = U + W$: Take any $v \in U + W$ such that $v = u + w$ for some $u \in U$ and $w \in W$. Then, there exists $\alpha_i, \beta_i \in \mathbb{F}$ such that

$$u = \alpha_1 v_1 + \cdots + \alpha_k v_k + \alpha_{k+1} u_1 + \alpha_{k+m} u_m$$
$$w = \beta_1 v_1 + \cdots + \beta_k v_k + \beta_{k+1} w_1 + \beta_{k+n} w_n$$

Then,

$$v = u + w = (\alpha_1 + \beta_1) v_1 + \cdots + (\alpha_k + \beta_k) v_k + \alpha_{k+1} u_1 + \cdots + \alpha_{k+m} u_m + \beta_{k+1} w_1 + \cdots + \beta_{k+n} w_n \in \langle S \rangle$$

S is linearly independent: Take $\alpha_i, \beta_i, \gamma_i \in \mathbb{F}$ such that

$$\alpha_1 v_1 + \cdots + \alpha_k v_k + \beta_1 u_1 + \cdots + \beta_m u_m + \gamma_1 w_1 + \cdots + \gamma_1 w_n = 0$$

Then,

$$\alpha_1 v_1 + \cdots + \alpha_k v_k + \beta_1 u_1 + \cdots + \beta_m u_m = -(\gamma_1 w_1 + \cdots + \gamma_1 w_n)$$

where the left is in U and the right is in W. Therefore, they are both in $U \cap W$. As $\{v_1, \ldots, v_k\}$ form a basis for $U \cap W$, there exists $\lambda_i \in \mathbb{F}$ such that

$$-(\gamma_1 w_1 + \cdots + \gamma_1 w_n) = \lambda_1 v_1 + \cdots + \lambda_k v_k$$

which rearranges to

$$\gamma_1 w_1 + \cdots + \gamma_1 w_n + \lambda_1 v_1 + \cdots + \lambda_k v_k = 0$$

As $\{v_1, \ldots, v_k, w_1, \ldots, w_n\}$ is linearly independent, it follows that for all i, $\gamma_i = 0$. Then,

$$\alpha_1 v_1 + \cdots + \alpha_k v_k + \beta_1 u_1 + \cdots + \beta_m u_m = 0$$

As $\{v_1, \ldots, v_k, u_1, \ldots, u_m\}$ is linearly independent, every $\alpha_i, \beta_i = 0$ ■

Definition 2.44 *Let V be a vector space and $U, W \leq V$. If $U \cap W = \{0_V\}$ and $V = U + W$, then V is the **direct sum** of U and W, writing $V = U \oplus W$.*

Proposition 2.45 *Let V be a finite dimensional vector space V. The following are equivalent:*

1. $V = U \oplus W$
2. *every $v \in V$ has a unique expression as $u + w$ with $u \in U$ and $w \in W$*
3. $\dim V = \dim U + \dim W$ *and* $V = U + W$
4. $\dim V = \dim U + \dim W$ *and* $U \cap W = \{0_V\}$
5. *if $\{u_1, \ldots, u_m\}$ is a basis for U and $\{w_1, \ldots, w_n\}$ is a basis for W, $\{u_1, \ldots, u_m, w_1, \ldots, w_n\}$ is a basis for V.*

Proof. Follows from definitions and using the dimension formula. ■

Definition 2.46 *A vector space V is said to be the (internal) direct sum of subspaces $X_1, \ldots, X_n \leq V$ if every $v \in V$ can be written as*

$$v = x_1 + \cdots + x_n$$

where $x_i \in X_i$ for all i, writing $X_1 \oplus \cdots \oplus X_n$. Given vector spaces $V_1, \ldots, V_n$, the (external) direct sum

$$V_1 \oplus \cdots \oplus V_n$$

is the set $V_1 \times \cdots \times V_n$ with addition and scalar multiplication defined componentwise.

Definition 2.47 *Let V, W be vector spaces over $\mathbb{F}$. A map $T : V \to W$ is **linear** if*

- $T(v_1 + v_2) = T(v_1) + T(v_2)$ *for all* $v_1, v_2 \in V$
- $T(\lambda v) = \lambda T(v)$ *for all* $v \in V, \lambda \in \mathbb{F}$

*Then T is a **linear transformation** or a **linear map**.*

Example 2.48 *Some examples of linear transformations include:*

- *Let V be a vector space. The* ***identity map*** $id_V : V \to V$ *taking* $v \mapsto v$.
- *Let V, W be vector spaces. The* ***zero map*** *from* $V \to W$ *that takes* $v \mapsto 0_W$
- *Let V be a vector space over $\mathbb{F}$ with subspaces U, W such that $V = U \oplus W$. For $v \in V$ there exists unique $u \in U, w \in W$ such that $v = u + w$. Define the* ***projection of V onto W along U*** *as the map $P : V \to V$ by $P(v) = w$. One can easily check that this defines a linear map.*
- *Let $\mathbb{R}_n[x]$ be the vector space of real polynomials with degree at most n. The map $\mathbb{R}_n[x] \to \mathbb{R}_n[x]$ by $p(x) \mapsto p'(x)$ is linear.*
- *Let X be any set and $V = \mathbb{R}^X$. For any $a \in X$, the* ***evaluation map*** *$E_a : V \to \mathbb{R}$ with $f \mapsto f(a)$ is a linear map.*

Proposition 2.49 *Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$. Then $T(0_V) = 0_W$.*

Proof. Follows from the fact that

$$T(0_V) = T(0_V + 0_V) = T(0_V) + T(0_V)$$

■

Proposition 2.50 *Let V, W be vector spaces over $\mathbb{F}$ and $T : V \to W$. The following are equivalent:*

- *T is linear*
- *for all $v_1, v_2 \in V$ and $\alpha, \beta \in \mathbb{F}$, $T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2)$*
- *for any $n \geq 1$, if $v_1, \dots, v_n \in V$ and $\alpha_1, \dots, \alpha_n \in \mathbb{F}$, then*

$$T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$$

Proof. Is immedidate after following definitions. ■

Proposition 2.51 *Let V, W be vector spaces over a field $\mathbb{F}$. If $S, T : V \to W$ are linear maps and $\lambda \in \mathbb{F}$, then the maps $S + T : V \to W$ by $v \mapsto S(v) + T(v)$ and $\lambda S : V \to W$ by $v \to \lambda S(v)$ are also linear maps.*

Proof. Follows from using the definition of linear maps and expanding out and rearranging. ■

Proposition 2.52 *Let U, V, W be vector spaces over $\mathbb{F}$. Let $S : U \to V$, $T : V \to W$ be linear maps. Then $T \circ S : U \to W$ is linear.*

Proof. Follows from definitions. ■

Notation 2.53 *We will often write TS to represent $T \circ S$ where it is clear.*

Notation 2.54 *We write* $\mathrm{Hom}(V, W)$ *to be the set of linear transformations from V to W. Note that this forms a vector space.*

Definition 2.55 *Let V, W be vector spaces and $T : V \to W$ be linear. We say that T is invertible if there exists a linear map S such that $ST = id_V$ and $TS = id_W$. An invertible linear map is called an isomorphism.*

Notation 2.56 *We often write T^{-1} to represent the inverse of T. Note this is not ambiguous as T is a function, meaning that inverses are unique.*

Proposition 2.57 *Let V, W be vector spaces and $T : V \to W$ be linear. Then, T is invertible if and only if it is bijective.*

Proof.

3 Basic Properties of Matrices

Definition 3.1 (Matrix) *A matrix is a rectangular array of numbers arranged over rows and columns. For instance,*

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

is a 2 by 3 matrix. Generally, an $m \times n$ matrix is an array of values (over some set) arranged into m rows and n columns. We will often write $A = (a_{ij})$ and use these subscripts to refer to the i-th row j-th column entry of a matrix A. Row vectors in $\mathbb{R}^n$ are $1 \times n$ matrices and column vectors in $\mathbb{R}^n_{\text{col}}$ are $n \times 1$ matrices.

Definition 3.2 *A matrix is called **square** if it has the same number of rows as columns.*

Notation 3.3 *We will write $\mathcal{M}_{m\times n}(\mathbb{F})$ to refer to the set of $m \times n$ matrices over some $\mathbb{F}$, so $\mathcal{M}_{m\times n}(\mathbb{R})$ reers to the set of $m \times n$ matrices with entries in $\mathbb{R}$. For simplicity, we will also write $\mathcal{M}_{m\times n}$ to denote $\mathcal{M}_{m\times n}(\mathbb{R})$.*

For the rest of this section, we will only care about real matrices.

Definition 3.4 (Addition) *Addition on matrices is an inline binary function that is defined on $\mathcal{M}_{m\times n}$. Specifically, $+ : \mathcal{M}_{m\times n} \times \mathcal{M}_{m\times n} \to \mathcal{M}_{m\times n}$, where given $A = (a_{ij})$ and $B = (b_{ij})$, $C = A + B = (c_{ij})$, where $c_{ij} = a_{ij} + b_{ij}$.*

Remark 3.5 *Matrix addition is commutative, as addition in $\mathbb{R}$ is commutative. For the same reason, addition of matrices is also associative. The additive identity on $\mathcal{M}_{m\times n}$ is the $m \times n$ zero matrix, written 0_{mn}, which is a matrix with m rows and n columns where all entries in it are 0.*

Notation 3.6 *When m and n are clear, we often write 0 to refer to 0_{mn}.*

Definition 3.7 (Scalar Multiplication) *Let $A = (a_{ij}) \in \mathcal{M}_{m\times n}$, and $k \in \mathbb{R}$. Then, kA is the $m \times n$ matrix with the (i, j)th entry equal to ka_{ij}.*

Remark 3.8 *Scalar multiplication distributes over matrices with the following identities, where $A, B \in \mathcal{M}_{m\times n}$ and $\lambda, \mu \in \mathbb{R}$:*

- $(\lambda + \mu)A = \lambda A + \mu A$
- $\lambda(A + B) = \lambda A + \lambda B$
- $\lambda(\mu A) = (\lambda\mu)A$

Definition 3.9 (Matrix Multiplication) *We define matrix multiplication on a matrix $A = (a_{ij})$ with $B = (b_{ij})$, if A is a p by q matrix and B is a q by r matrix. Specifically, $\times : \mathcal{M}_{p\times q} \times \mathcal{M}_{q\times r} \to \mathcal{M}_{p\times r}$. Then, $C = AB = c_{ij}$, where*

$$c_{ij} = \sum_{k=1}^{q} a_{ik}b_{kj} \qquad \text{for } 1 \leq i \leq p \text{ and } 1 \leq j \leq r.$$

Remark 3.10 *Given a matrix $A \in \mathcal{M}_{m\times n}$, Matrix multiplication defines a map L_A from $\mathbb{R}^n_{\text{col}}$ to $\mathbb{R}^m_{\text{col}}$ by*

$$L_A(v) = Av$$

where v is a $n \times 1$ matrix (alternatively, a n dimensional column vector). Then, it can be shown that $L_{AB} = L_A \circ L_B$, which follows from associativity of matrix multiplication.

Example 3.11 *Let $V = \mathcal{M}_{m\times n}(\mathbb{F})$. Then, the standard basis for V is the set*

$$\{E_{ij} \mid 1 \le i \le m, 1 \le j \le n\}$$

where E_{ij} is the matrix with entry 1 on the (i,j)-th entry and 0 elsewhere. Then, we may also write $A = (a_{ij}) \in V$ as

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}E_{ij}$$

Definition 3.12 (Identity Matrix) *The $n \times n$ **identity matrix** I_n is the $n \times n$ matrix with entries*

$$\delta_{ij} = \begin{cases} 1 \; if \; i = j \\ 0 \; if \; i \neq j \end{cases}$$

Remark 3.13 *The identity matrix is the multiplicative identity. In general,*

- *Matrix multiplication is not commutative. That is, $\exists A, B$ such that $AB \neq BA$.*
- *Matrix multiplication is associative.*
- *Distributive laws: $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$*
- *$MN = 0$ does not imply that either M or N is 0.*

Notation 3.14 *We write A^2 to denote AA, and A^n for $\underbrace{AA\cdots A}_{n\ times}$. We also define $A^0 = I$. Given a polynomial $p(x) = a_kx^k + a_{k-1}x^{k-1} + \cdots + a_1x + a_0$, we define*

$$p(A) = a_kA^k + a_{k-1}A^{k-1} + \cdots + a_1A + a_0I.$$

Example 3.15 *Given a square matrix B, here may be infinite or no solutions to the equation $A^2 = B$, with some matrix A.*

Proof. For the former, let

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}$$

Then, $A^2 = I_2$ for any α. For the latter, take

$$B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Take any $a, b, c, d \in \mathbb{C}$ such that

$$B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 = \begin{pmatrix} a^2 + bc & b(a+d) \\ c(a+d) & bc + d^2 \end{pmatrix}$$

Using that $c(a + d) = 0$, if $a + d = 0$, we get a contradiction with the top right entry. If $c = 0$, we see $a = 0$ and $d = 0$, which again leads to a contradiction.

Definition 3.16 *Let A be a square matrix. Then, B is the **inverse matrix** of A if $BA = AB = I$. A matrix with an inverse is invertible, and else is called singular.*

Lemma 3.17 (Properties of Inverses) *In general, for any square matrices A, B of the same size,*

- *If A has an inverse, it is unique. We write A^{-1} for such inverse.*
- *If A, B are both invertible, then AB is invertible with $(AB)^{-1} = B^{-1}A^{-1}$*
- *If A is invertible, so is A^{-1}, with $(A^{-1})^{-1} = A$.*

Proof. If B and C are both inverses for $A \in \mathcal{M}_{n\times n}$, then

$$C = I_nC = (BA)C = B(AC) = BI_n = B$$

For the second case, simply note that $(AB)(B^{-1}A^{-1}) = AIA^{-1} = I$. Finally, noting that

$$(A^{-1})A = A(A^{-1}) = I$$

$(A^{-1})^{-1} = A$ by uniqueness. ■

Lemma 3.18

Definition 3.19 *A **diagonal matrix** is a square matrix whose non-diagonal entries are all zero.*

Definition 3.20 *The transpose matrix of $A \in \mathcal{M}_{m\times n}$ written $A^T \in \mathcal{M}_{n\times m}$ is the matrix such that the (i, j)-th entry of A is the (j, i)-th entry of A^T.*

Proposition 3.21 (Basic Properties of Tranposes) *Generally,*

- $(A + B)^T = A^T + B^T$
- $(\lambda A)^T = \lambda A^T$
- $(AB)^T = B^T A^T$
- $(A^T)^T = A$
- *A square matrix A is invertible if and only if A^T is invertible. Then, $(A^T)^{-1} = (A^{-1})^T$.*

Proof. For the first four cases, simply observe what the (i, j)-th entry of both sides are. For the final case, note that

$$A^T(A^T)^{-1} = I = I^T = (A^{-1}A)^T = A^T(A^{-1})^T$$

and as inverses are unique, if they exist, $(A^T)^{-1} = (A^{-1})^T$. The last part of the same equality shows that if A is invertible, then A^T is also invertible. Finally, if A^T is invertible, $(A^T)^T = A$ is invertible by the previous sentence. ■

Definition 3.22 *A square matrix $A = (a_{ij})$ is*

- ***symmetric** if $A^T = A$*
- ***skew-symmetric** (or **antisymmetric**) if $A^T = -A$*

- ***upper triangular*** *if* $i > j$ *implies* $a_{ij} = 0$
- ***strictly upper triangular*** *if it* $i \geq j$ *implies* $a_{ij} = 0$
- ***lower triangular*** *if* $i < j$ *implies* $a_{ij} = 0$
- ***strictly lower triangular*** *if it* $i \leq j$ *implies* $a_{ij} = 0$
- ***triangular*** *if it is either upper or lower triangular*

Definition 3.23 *A* $n \times n$ *matrix in* ***orthogonal*** *if* $A^T = A^{-1}$.

Proposition 3.24 *Let* A *and* B *be orthogonal* $n \times n$ *matrices. Then, both* AB *and* A^{-1} *are orthogonal.*

Proof. Follows from the fact that

$$(AB)^T = B^T A^T = B^{-1} A^{-1} = (AB)^{-1}$$

and

$$(A^{-1})^T = (A^T)^{-1} = (A^{-1})^{-1}$$

■

Notation 3.25 *Let* $A = (a_{ij})$ *be a* $m \times n$ *matrix. We may write*

$$A = \begin{pmatrix} \mathbf{a_1} \\ \vdots \\ \mathbf{a_m} \end{pmatrix}$$

where $\mathbf{a_1}, \dots \mathbf{a_m}$ *are* n *dimensional row vectors. Similarly, we may write*

$$A = \begin{pmatrix} \mathbf{a_1'} & \cdots & \mathbf{a_n'} \end{pmatrix}$$

using m *dimensional column vectors.*

Definition 3.26 *Given a* $m \times n$ *matrix* A, *the* ***row space*** *of* A *is the span of its rows and its* ***column space*** *is the span of its columns. We write* $\mathrm{Row}(A) \leq \mathbb{R}^n$ *for its row space and* $\mathrm{Col}(A) \leq \mathbb{R}^m_{\mathrm{col}}$ *for its column space.*

4 Reduced Row Echelon Form

Definition 4.1 (Elementary Row Operations) *Elementary row operation (ERO) is an operation that is one of the following :*

- *Swapping of row i and j (which we will write S_{ij})*
- *Multiplying row i by some $\lambda \neq 0$ (which we will write $M_i(\lambda)$)*
- *For some $i \neq j$, adding λ times row i to row j (which we will write $A_{ij}(\lambda)$).*

Proposition 4.2 (Elementary Matrices) *Let A be any $m \times n$ matrix. Then, any ERO is equivalent to pre-multiplying A with certain matrices (also denoted S_{ij}, $M_i(\lambda)$, and $A_{ij}(\lambda)$). Specifically, these matrices are exactly the matrices one gets after applying the EROs to I_n.*

Proof. Follows from writing out each matrix and calculating the product with an arbitrary matrix explicity. ■

Proposition 4.3 *Elementary matrices are invertible.*

Proof. Follows from noting that

$$(S_{ij})^{-1} = S_{ji} \qquad (M_i(\lambda))^{-1} = M_i(\lambda^{-1}) \qquad (A_{ij}(\lambda))^{-1} = A_{ij}(-\lambda)$$

. ■

Definition 4.4 *Let $A = (a_{ij})$ be a $m \times n$ matrix, and $\mathbf{b} = (b_i)$ be a m dimensional column vector. Then, we call $\mathbf{x} = (x_i)$ to be the solution of $(A|\mathbf{b})$ if $A\mathbf{x} = \mathbf{b}$, where $(A|\mathbf{b})$ is a matrix with $\mathbf{b}$ augmented onto A. Specifically, such $\mathbf{x}$ satisfies*

$$a_{i1}x_1 + \cdots + a_{in}x_n = b_i$$

for all i. Alternatively, $\mathbf{x}$ is the solution to the linear system of m equations represented by $(A|\mathbf{b})$. If such $\mathbf{x}$ exists, we say that this system is consistent.

Lemma 4.5 *Let $(A|\mathbf{b})$ be a linear system of m equations and E be an elementary $m \times m$ matrix. Then $\mathbf{x}$ is a solution of $(A|\mathbf{b})$ iff $\mathbf{x}$ is a solution of $(EA|E\mathbf{b})$.*

Proof. If $A\mathbf{x} = \mathbf{b}$, then $EA\mathbf{x} = E\mathbf{b}$. If $EA\mathbf{x} = E\mathbf{b}$, by premultiplying both sides by E^{-1}, this implies $A\mathbf{x} = \mathbf{b}$. ■

Definition 4.6 (Reduced Row Echelon Form (RREF)) *A matrix A is in RREF if*

- *The first non-zero column of any non-zero row is 1*
- *In a column with a leading 1 (viewed by row), all other entries of that column are zero*
- *The leading 1 of a non-zero row appears to the right of the leading 1s of the rows above it*
- *zero rows are below the non-zero rows.*

Definition 4.7 *The reduction of applying EROs to transform a matrix into RREF is also referred to as Gauss-Jordan elimination.*

Lemma 4.8 (Solutions to RREF) *Let $(A|\mathbf{b})$ be a matrix in RREF representing a linear system $A\mathbf{x} = \mathbf{b}$ of m equations in n variables. Then*

- *The system has no solutions iff the last non-zero row is*

$$(0 \ \ 0 \ \ \cdots \ \ 0 \mid 1)$$

- *The system has a unique solution iff the non-zero rows of A form I_n and $(A|\mathbf{b})$ has as many non-zero rows as A.*

- *The system has infinitely many solutions if there are as many non-zero rows in $(A|\mathbf{b})$ as A, and there exists a column that is all zeros. Then, the solutions can be represented by a $n-k$ variable equation where k is the number of columns that has leading 1s.*

Proof. The first case is clear. For the second and third case, first note that if $(A|\mathbf{b})$ and A don't have the same number of non-zero rows, we are in the first case, so the system has no solutions. Otherwise, suppose that A has k non-zero rows, meaning there are $n-k$ columns without leading 1s. Without loss of generality (by a reordering of the variable numbers), we can assume that the leading 1s appear in the first r columns of the first r rows. In particular,

$$A = \left(\begin{array}{cccccc|c} & I_k & & & A' & & \mathbf{b}' \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \end{array}\right)$$

for some A'. Then, we can see that,

$$x_i + a_{i(k+1)}x_{k+1} + \cdots + a_{in}x_n = b_i \qquad \text{for } 1 \le i \le k$$

Therefore, by assigning parameter values $y_{k+1}, \ldots, y_n$ onto $x_{k+1}, \ldots, x_n$, we can find values for $x_1, \ldots, x_k$ that solve the system. Conversely, if we are given a solution $\mathbf{x} = (x_1, \ldots, x_n)$, then it is the solution that is given when we assign $y_{k+1} = x_{k+1}, \ldots, y_n = x_n$ to the parameters. This gives an infinite set of solutions associated with $n-k$ independent parameters when $n > k$ and a unique solution when $n = k$. ■

Theorem 4.9 (Existence of RREF) *Every $m \times n$ matrix A can be reduced by EROs to a matrix in RREF.*

Proof. We prove by an induction on m, the number of rows of the matrix. In the case $m = 1$, a $1 \times n$ matrix is either a 0 matrix, in which case we are done, or has a leading entry. In the latter case, we can put the matrix into RREF by dividing the row by the leading entry (equivalently, by a premultiplication of $M_1(1/\lambda)$), where λ is the leading entry of A. For the inductive case, if A is the 0 matrix, we are done. Otherwise, there exists a first column j with a non-zero element α. By EROs we can swap the row with this element with the first row and divide through by $\alpha \neq 0$. Now, the matrix takes the form :

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & a'_{1(j+1)} & \cdots & a'_{1n} \\ 0 & \cdots & 0 & a'_{2j} & \vdots & \vdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & a'_{mj} & a'_{m(j+1)} & \cdots & a'_{mn} \end{pmatrix}$$

for some new entries a'. By applying $A_{12}(-a'_{2j}), \ldots, A_{1m}(-a'_{mj})$, the matrix becomes,

$$A' = \left(\begin{array}{cccc|ccc} 0 & \cdots & 0 & 1 & a'_{1(j+1)} & \cdots & a'_{1n} \\ \hline 0 & \cdots & 0 & 0 & & & \\ \vdots & \cdots & \vdots & \vdots & & B & \\ 0 & \cdots & 0 & 0 & & & \end{array}\right)$$

for some block matrix B (with lines added for reference on it's dimension). By the inductive hypothesis there exists an E such that EB is in RREF, where $E = E_k \cdots E_1$ for some elementary matrices E_i. Noting that,

$$\begin{pmatrix} 1 & 0 \\ 0 & E \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & E_1 \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ 0 & E_k \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & 0 \\ 0 & E \end{pmatrix} A' = \left(\begin{array}{cccc|ccc} 0 & \cdots & 0 & 1 & a'_{1(j+1)} & \cdots & a'_{1n} \\ \hline 0 & \cdots & 0 & 0 & & & \\ \vdots & \cdots & \vdots & \vdots & & EB & \\ 0 & \cdots & 0 & 0 & & & \end{array}\right)$$

we can transform this into RREF by turning any of a'_{1k} into zeros if they are above a leading 1 in EB by premultiplying by $A_{r1}(-a'_{1k})$ where r is the line with this leading 1. ■

Lemma 4.10 *Let $A \in \mathcal{M}_{n\times n}$ and $E = E_k \cdots E_1$ where E_i are EROs and E reduces A into RREF. Then, $EA = I_n$ if and only if A is invertible. If it is invertible, $E = A^{-1}$.*

Proof. Consider the augmented matrix $(A|I_n) \in \mathcal{M}_{n\times 2n}$. Then, $E(A|I_n) = (EA|E)$. In the forward case,

$$I_n = EA = (E_k \cdots E_1)A \qquad \Longrightarrow \qquad A^{-1} = E_k \cdots E_1$$

as elementary matrices are invertible. Alternatively, we can prove this by a premultiplication of $E_1^{-1} \cdots E_k^{-1}$ on EA, showing that $A = E_1^{-1} \cdots E_k^{-1}$. Conversely, if $EA \neq I_n$ and EA is in RREF, the last row of EA must be the zero-row. Thus, $(0, \ldots, 0, 1)(EA) = \mathbf{0}$. If A is invertible, as E is also invertible, by a postmultiplication of $A^{-1}E^{-1}$, we get $(0, \ldots, 0, 1) = \mathbf{0}$, a contradiction. ■

Remark 4.11 *The proof for Lemma 4.10 gives an algorithmic approach for finding the inverse of a matrix, as we can algorithmicly convert a matrix to RREF.*

Theorem 4.12 (Uniqueness of RREF) *The reduced row echelon form of a $m \times n$ matrix A is unique.*

Proof. Follows from an induction on n. For the base case, note that the only $m \times 1$ matrices in RREF are $\mathbf{0}$ and $\mathbf{e}_1^T$. The zero matrix reduces to the former while non-zero matrices reduce to the latter. This proves uniquness for $n = 1$.

For the inductive case, assume any $m \times (n-1)$ matrices M have a unique RREF, written $\mathrm{RRE}(M)$. Let the first $m-1$ columns of A be written as A'. Then, the RREF of A is of one of the following forms :

$$\begin{pmatrix} \begin{matrix}\text{non-zero} \\ \mathrm{RRE}(A') \text{ rows}\end{matrix} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ 0 \;\cdots\; 0 & 1 \\ 0_{(m-k-1)m} & \end{pmatrix}, \qquad \begin{pmatrix} \begin{matrix}\text{non-zero} \\ \mathrm{RRE}(A') \text{ rows}\end{matrix} & \begin{matrix} * \\ \vdots \\ * \end{matrix} \\ 0_{(m-k)m} & \end{pmatrix} = \begin{pmatrix} \mathbf{e}_1(R) \\ \vdots \\ \mathbf{e}_k(R) \\ 0_{(m-k)m} \end{pmatrix}$$

for some k. This is the case distinction between whether $\mathbf{e}_{k+1}^T$ is in the rowspace of A. In the first case where $\mathbf{e}_{k+1}^T$ is in the rowspace, the RREF of A is uniquely determined as $\mathrm{RRE}(A')$ is unique. In the second case, suppose that R_1 and R_2 are RREFs for A. By Corollary 4.16,

$$\mathrm{Row}(R_1) = \mathrm{Row}(A) = \mathrm{Row}(R_2)$$

As the rowspaces agree, for any $1 \leq i \leq k$, there exist $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ such that,

$$\mathbf{e}_i(R_1) = \sum_{l=1}^{k} \alpha_k \mathbf{e}_k(R_2)$$

Then by observing the values on the first $n-1$ columns,

$$\mathbf{e}_i(\mathrm{RRE}(A')) = \sum_{l=1}^{k} \alpha_k \mathbf{e}_k(\mathrm{RRE}(A'))$$

This forces $\lambda_i = 1$ and $\lambda_j = 0$ for $i \neq j$. Therefore, $\mathbf{e}_i(R_1) = \mathbf{e}_i(R_2)$ for each i, giving $R_1 = R_2$. ■

Notation 4.13 *For any matrix A, we will often write* $\mathrm{RRE}(A)$ *to refer to the unique matrix given by the RREF of A. We note that this matrix exists and is unique as shown in Theorems 4.9 and 4.12.*

Proposition 4.14 *Let R be a matrix in RREF. Then, the non-zero rows of R are independent.*

Proof. Let $\mathbf{r}_1, \ldots \mathbf{r}_n$ represent the non-zero row vectors of R, and suppose that $\alpha_1 \mathbf{r}_1 + \cdots \alpha_n \mathbf{r}_n = \mathbf{0}$. Suppose the leading 1 of $\mathbf{r}_1$ appears in the j-th column. Then,

$$\alpha_1 + \alpha_2 r_{2j} + \alpha_3 r_{3j} + \cdots + \alpha_n r_{nj} = 0$$

As R is in RREF, $r_{2j}, \ldots, r_{nj}$ are all zero, giving $\alpha_1 = 0$. Inductively, we can repeat this to get $\alpha_i = 0$ for all i. ■

Lemma 4.15 *Rowspaces are invariant under premultiplication by invertible matrices. That is, for any matrix $A \in \mathcal{M}_{m \times n}(\mathbb{F})$, if $B \in \mathcal{M}_{m \times m}(\mathbb{F})$ is an invertible matrix, then* $\mathrm{Row}(A) = \mathrm{Row}(BA)$. *In particular, rowspaces are invariant under EROs.*

Proof. We first prove that for any $B \in \mathcal{M}_{k \times m}(\mathbb{F})$, $\mathrm{Row}(BA) \subseteq \mathrm{Row}(A)$. Given $A = (a_{ij})$ and $B = (b_{ij})$, we note that the i-th row of BA is the row vector

$$\left(\sum_{l=1}^{m} b_{il} a_{l1}, \ldots, \sum_{l=1}^{m} b_{il} a_{ln} \right) = \sum_{l=1}^{m} b_{il} \underbrace{(a_{l1}, \ldots, a_{ln})}_{l\text{-th row of } A}$$

which is a linear combination of rows of A. So every row in BA is in $\mathrm{Row}(A)$. As any element in $\mathrm{Row}(BA)$ is a linear combination of rows in BA, this is a linear combination of rows in A. This establishes $\mathrm{Row}(BA) \subseteq \mathrm{Row}(A)$. When B is invertible, the rest of the lemma follows from the fact that $\mathrm{Row}(A) = \mathrm{Row}(B^{-1}(BA)) \subseteq \mathrm{Row}(BA)$. ■

Corollary 4.16 *Let A be a matrix and R be the matrix obtained by turing A into RREF. Then,* $\mathrm{Row}(A) = \mathrm{Row}(R)$.

Proof. As A can be turned into R by EROs, there exists matrices $E_1, \ldots, E_k$ such that $E_k \cdots E_1 A = R$. By Lemma 4.15, as rowspaces are invariant under EROs,

$$\text{Row}(A) = \text{Row}(E_1 A) = \cdots = \text{Row}(E_k \cdots E_1 A) = \text{Row}(R)$$

■

Corollary 4.17 *Let A be a $m \times n$ matrix. Then* $\text{RRE}(A)$ *contains a zero row if and only if the rows of A are dependent.*

Proof. Let $\text{RRE}(A) = EA$, where E is the product of EROs.

($\Rightarrow$) Suppose the i-th row of EA is $\mathbf{0}$. Then,

$$\mathbf{0} = \sum_{l=1}^{m} e_{il}(l\text{-th row of } A)$$

As E is invertible, not all entries of the i-th row of E is zero. Therefore, the above shows the rows of A are linearly dependent.

($\Leftarrow$) Conversely, suppose that the rows of A are linearly dependent. Let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ denote the rows vectors of A. Then, without loss of generality (swapping rows if necessary) there exists $\alpha_1, \ldots, \alpha_{m-1}$ such that

$$\mathbf{a}_m = \alpha_1 \mathbf{a}_1 + \cdots + \alpha_{m-1} \mathbf{a}_{m-1}$$

By performing EROs $A_{1m}(-c_1), \ldots, A_{(m-1)m}(-c_{m-1})$, we get a matrix with m-th row zero. By performing EROs on the top $m-1$ rows, we get a matrix in RREF. Therefore, $\text{RRE}(A)$ has a zero row. ■

Corollary 4.18 *Let A be a $m \times n$ matrix. Then the rows of A span $\mathbb{R}^n$ if and only if*

$$\text{RRE}(A) = \begin{pmatrix} I_n \\ 0_{(m-n)n} \end{pmatrix}$$

Proof. ($\Rightarrow$) For any i, $\mathbf{e}_i$ is in the rowspace of $\text{RRE}(A)$ as the rows span $\mathbb{R}^n$. Therefore every column contains a leading 1, and the proof follows.

($\Leftarrow$) Follows immediately from the fact that

$$\mathbb{R}^n = \text{Row}(\text{RRE}(A)) = \text{Row}(A)$$

where the latter equality comes from Corollary 4.16. ■

Remark 4.19 *It follows from the above corollaries that the rows of A form a basis for $\mathbb{R}^n$ if and only if the RREF of A is I_n. In particular, if there is a set of k vectors that are linearly independent in $\mathbb{R}^n$, then $k \leq n$. Further, if k vectors span $\mathbb{R}^n$, then $k \geq n$.*

Definition 4.20 *The **row rank** or **rank** of a matrix A is the number of non-zero rows in* $\text{RRE}(A)$. *We write* $\text{rank}(A)$ *for this value. Uniqueness of RREF ensures this number is well defined. Alternatively, it is the dimension of the row space.*

Remark 4.21 *The non-zero rows of the matrix are linearly independent, and as the rowspace is invariant under EROs, the non-zero rows of a matrix in RREF form the basis for the rowspace.*

Proposition 4.22 *Let A be a $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m_{\text{col}}$. Then,*

- *the system $A\mathbf{x} = \mathbf{b}$ has no solutions if and only if $\begin{pmatrix} 0 & 0 & \cdots & 0 & \mid & 1 \end{pmatrix}$ is in $\mathrm{Row}(A|\mathbf{b})$.*

If the system $A\mathbf{x} = \mathbf{b}$ is consistent,

- *there is a unique solution if and only if $\mathrm{rank}(A) = n$*
- *there are infinitely many solutions if $\mathrm{rank}(A) < n$. The set of solutions can be written using $n - rank(A)$ variables.*

Proof. Follows from Lemma 4.8. ∎