Theorem 1.1 (Cancellation Law for Vector Addition). If x, y, and z are vectors in a vector space $\mathbf{V}$ such that $x + z = y + z$, then $x = y$.

Proof. There exists a vector v in $\mathbf{V}$ such that $z + v = \mathbf{0}$ (VS 4). Thus

$$x = x + \mathbf{0} = \mathrm{x} + (\mathrm{z} + \mathrm{v}) = (\mathrm{x} + \mathrm{z}) + \mathrm{v} = (\mathrm{y} + \mathrm{z}) + \mathrm{v} = \mathrm{y} + (\mathrm{z} + \mathrm{v}) = \mathrm{y} + \mathbf{0} = \mathrm{y}$$

by (VS 2) and (VS 3).

Corollary 1. The vector $\mathbf{0}$ described in (VS 3) is unique.

Corollary 2. The vector y described in (VS 4) is unique.

Theorem 1.2. In any vector space $\mathbf{V}$ the following statements are true:

(a) $0x = \mathbf{0}$ for each $x \in \mathbf{V}$.

(b) $(-a)x = -(ax) = a(-x)$ for each $a \in F$ and each $x \in \mathbf{V}$.

(c) $a\mathbf{0} = \mathbf{0}$ for each $a \in F$.

Proof. a. By (VS 8), (VS 3), and (VS 1) it follows that

$$0x + 0x = (0+0)x = 0x = 0x + \mathbf{0}.$$

Hence $0x = \mathbf{0}$ by Theorem 1.1.

b. The vector $-(ax)$ is the unique element of $\mathbf{V}$ such that $ax+[-(ax)] = \mathbf{0}$. Thus if $ax+(-a)x = \mathbf{0}$, Corollary 2 to Theorem 1.1 implies that $(-a)x = -(ax)$. But by (VS 8).

$$ax + (-a)x = [a + (-a)]x = 0x = \mathbf{0}$$

by (a). Consequently $(-a)x = -(ax)$. In particular $(-1)x = -x$. So by (VS 6),

$$a(-x) = a[(-1)x] = [a(-1)]x = (-a)x.$$

c. By (VS 3) and (VS 7) it follows that

$$a\mathbf{0} + \mathrm{a}\mathbf{0} = \mathrm{a}(\mathbf{0} + \mathbf{0}) = \mathrm{a}\mathbf{0} = \mathrm{a}\mathbf{0} + \mathbf{0}.$$

Hence $a\mathbf{0} = \mathbf{0}$ by Theorem 1.1.

Theorem 1.3. Let $\mathbf{V}$ be a vector space and $\mathbf{W}$ be a subset of $\mathbf{V}$. Then $\mathbf{W}$ is a subspace of $\mathbf{V}$ if and only if the following three conditions hold for the operations defined in $\mathbf{V}$:

(a) $\mathbf{0} \in \mathbf{W}$.

(b) $(x + y) \in \mathbf{W}$ whenever $x, y \in \mathbf{W}$.

(c) $cx \in \mathbf{W}$ whenever $c \in F$ and $x \in \mathbf{W}$.

Proof. We begin by assuming that $\mathbf{W}$ is a subspace of $\mathbf{V}$; then $\mathbf{W}$ is a vector space with the operations of addition and scalar multiplication defined on $\mathbf{V}$. Hence conditions (b) and (c) hold, and there exists a vector $\mathbf{0_w} \in \mathbf{W}$ such that $0_w + x = x$ for each $x \in \mathbf{W}$. But also $x + \mathbf{0} = x$, and thus $\mathbf{0_w} = \mathbf{0}$ by Theorem 1.1. So condition (a) holds.

Next suppose that conditions (a), (b), and (c) hold, then by the discussion on pg 17, $\mathbf{W}$ is a subspace of $\mathbf{V}$ if the additive inverse of each vector in $\mathbf{W}$ lies in $\mathbf{W}$. But if $x \in \mathbf{W}$, then $(-1)x \in \mathbf{W}$ by condition (c), and $-x = (-1)x$ by Theorem 1.2. Hence $\mathbf{W}$ is a subspace of $\mathbf{V}$.

Theorem 1.4. Any intersection of subspaces of a vector space $\mathbf{V}$ is a subspace of $\mathbf{V}$.

Proof. Let C be a collection of subspaces of $\mathbf{V}$ and let $\mathbf{W}$ denote the intersection of the subspaces of C. Since every subspace contains the zero vector, $\mathbf{0} \in \mathbf{W}$. Let $a \in F$ and $x, y \in \mathbf{W}$. Then x and y are contained in each subspace of C. Because each subspace is closed under addition and scalar multiplication, it follows that $x + y$ and ax are contained in each subspace of C. Hence $x + y$ and ax are also contained in $\mathbf{W}$, so that $\mathbf{W}$ is a subspace of $\mathbf{V}$ by Theorem 1.3.

Theorem 1.5. The span of any subset S of a vector space $\mathbf{V}$ is a subspace of $\mathbf{V}$. Moreover, any subspace of $\mathbf{V}$ that contains S must also contain the span of S.

Proof. The result is immediate if $S = \phi$ because $\text{span}(\phi) = \{\mathbf{0}\}$, which is a subspace contained in any subspace of $\mathbf{V}$.

If $S \neq \phi$, then S contains a vector z. So $0z = \mathbf{0}$ is in $\text{span}(S)$. Let $x, y \in \text{span}(S)$. Then there exists vectors $u_1, u_1, \ldots, u_m, v_1, v_2, \ldots, v_n$ in S and scalars $a_1, a_2, \ldots, a_m, b_1, b_2, \ldots, b_n$ such that

$$x = a_1u_1 + a_2u_2 + \cdots + a_mu_m \ , \qquad y = b_1v_1 + b_2v_2 + \cdots + b_nv_n.$$

Then

$$x + y = a_1u_1 + a_2u_2 + \cdots + a_mu_m + b_1v_1 + b_2v_2 + \cdots + b_nv_n$$

and for any scalar c,

$$cx = (ca_1)u_1 + (ca_2)u_2 + \cdots + (ca_m)u_m$$

are clearly linear combinations of vectors in S; so $x+y$ and cx are in $\text{span}(S)$. Thus $\text{span}(S)$ is a subspace of $\mathbf{V}$.

Now let $\mathbf{W}$ denote any subspace of $\mathbf{V}$ that contains S. If $w \in \text{span}(S)$, then w has the form $w = c_1w_1 + c_2w_2 + \cdots + c_kw_k$ for some vectors $w_1, w_2, \ldots, w_k$ in S and some scalars $c_1, c_2, \ldots, c_k$. Since $S \subseteq \mathbf{W}$, we have $w_1, w_2, \ldots, w_k \in \mathbf{W}$. Therefore $w = c_1w_1 + c_2w_2 + \cdots + c_kw_k$ is in $\mathbf{W}$ by exercise 20 of Section 1.3. Because w, an arbitrary vector in $\text{span}(S)$, belongs to $\mathbf{W}$, it follows that $\text{span}(S) \subseteq \mathbf{W}$.

Theorem 1.6. Let $\mathbf{V}$ be a vector space, and let $S_1 \subseteq S_2 \subseteq \mathbf{V}$. If S_1 is linearly dependent, then S_2 is linearly dependent.

Proof. If S_1 is linearly dependent, then there exist vectors $u_1, u_2, \ldots, u_m$ in S_1 and nonzero scalars $a_1, a_2, \ldots, a_m$ such that

$$a_1 u_1 + a_2 u_2 + \cdots + a_m u_m = \mathbf{0}.$$

Since $S_1 \subseteq S_2$, the vectors $u_1, u_2, \ldots, u_m$ are also elements of S_2, so that by definition, S_2 is linearly dependent.

Corollary. Let $\mathbf{V}$ be a vector space, and let $S_1 \subseteq S_2 \subseteq \mathbf{V}$. If S_2 is linearly independent, then S_1 is linearly independent.

Proof. Suppose to the contrary that S_1 is linearly dependent; by Theorem 1.6, this implies that S_2 is linearly dependent which is a contradiction. Hence S_1 is linearly independent.

Theorem 1.7. Let S be a linearly independent subset of a vector space $\mathbf{V}$, and let v be a vector in $\mathbf{V}$ that is not in S. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in \operatorname{span}(S)$.

Proof. We begin by assuming that $S \cup \{v\}$ is linearly dependent; then there are vectors $u_1, u_2, \ldots, u_m$ in $S \cup \{v\}$ such that

$$a_1u_1 + a_2u_2 + \cdots + a_mu_m = \mathbf{0}$$

for some nonzero scalars $a_1, a_2, \ldots, a_m$. Because S is linearly independent, one of the u_i's, say u_1, equals v. Thus

$$a_1v + a_2u_2 + \cdots + a_mu_m = \mathbf{0},$$

and so

$$v = a_1^{-1}(-a_2u_2 - \cdots - a_mu_m) = -(a_1^{-1}a_2)u_2 - \cdots - (a_1^{-1}a_m)u_m$$

Since v is a linear combination of $u_2, \ldots, u_n$, which are in S, we have $v \in \operatorname{span}(S)$.

Next assume that $v \in \operatorname{span}(S)$. Then there exists vectors $v_1, v_2, \ldots, v_m$ in S and scalars $b_1, b_2, \ldots, b_m$ such that

$$v = b_1v_1 + b_2v_2 + \cdots + b_mv_m.$$

Hence

$$\mathbf{0} = \mathrm{b_1v_1 + b_2v_2 + \cdots + b_mv_m + (-1)v}.$$

Since $v \neq v_i$ for $i = 1, 2, \ldots, m$, the coefficient of v in this linear combination is nonzero, and so the set $\{v_1, \ldots, v_m, v\}$ is linearly dependent. Therefore $S \cup \{v\}$ is linearly dependent by Theorem 1.6.

Theorem 1.8. Let $\mathbf{V}$ be a vector space and $\beta = \{u_1, u_2, \ldots, u_m\}$ be a subset of $\mathbf{V}$. Then β is a basis for $\mathbf{V}$ if and only if each $v \in \mathbf{V}$ can be uniquely expressed as a linear combination of the vectors of β, that is, can be expressed in the form

$$v = a_1 u_1 + a_2 u_2 + \cdots + a_n u_n$$

for unique scalars $a_1, a_2, \ldots, a_n$.

Proof. Let β be a basis for $\mathbf{V}$. If $v \in \mathbf{V}$, then $v \in \text{span}(\beta)$ because $\text{span}(\beta) = \mathbf{V}$. Thus v is a linear combination of the vectors of β. Suppose that

$$v = a_1 u_1 + a_2 u_2 + \cdots + a_n u_n \qquad\qquad v = b_1 u_1 + b_2 u_2 + \cdots + \ b_n u_n$$

are two such representations of v. Subtracting the second equation from the first gives

$$\mathbf{0} = (\mathrm{a}_1 - \mathrm{b}_1)\mathrm{u}_1 + (\mathrm{a}_2 - \mathrm{b}_2)\mathrm{u}_2 + \cdots + (\mathrm{a}_\mathrm{n} - \mathrm{b}_\mathrm{n})\mathrm{u}_\mathrm{n}.$$

Since β is linearly independent, it follows that

$$a_1 - b_1 = a_2 - b_2 = \cdots = a_n - b_n = 0.$$

Hence, $a_1 = b_1, a_2 = b_2, \cdots, a_n = b_n$, and so v is uniquely expressible as a linear combination of the vectors of β.

Next suppose that each $v \in \mathbf{V}$ can be uniquely expressed as a linear combination of the vectors of β. Clearly the vectors in β generate $\mathbf{V}$. For β to be a basis we must show that the vectors in β are linearly independent. Since the zero vector is in $\mathbf{V}$, it also has a unique representation of the form

$$\mathbf{0} = \mathrm{a}_1\mathrm{u}_1 + \mathrm{a}_2\mathrm{u}_2 + \cdots + \mathrm{a}_\mathrm{n}\mathrm{u}_\mathrm{n}.$$

It is obvious that we can write

$$\mathbf{0} = 0\mathrm{u}_1 + 0\mathrm{u}_2 + \cdots + 0\mathrm{u}_\mathrm{n}$$

and that this must therefore be the unique representation; hence by definition, the vectors of β are linearly independent and thus a basis of $\mathbf{V}$.

Theorem 1.9. If a vector space $\mathbf{V}$ is generated by a finite set S, then some subset of S is a basis for $\mathbf{V}$. Hence, $\mathbf{V}$ has a finite basis.

Proof. If $S = \phi$ or $S = \{\mathbf{0}\}$, then $\mathbf{V} = \{\mathbf{0}\}$ and $S = \phi$ is a subset of S that is a basis for $\mathbf{V}$. Otherwise S contains a nonzero vector u_1. By item 2 on page 37, $\{u_1\}$ is a linearly independent set. Continue if possible choosing vectors $u_2, \ldots, u_k$ in S such that $\{u_1, u_2, \ldots, u_k\}$ is linearly independent. Since S is a finite set, we must eventually reach a stage at which $\beta = \{u_1, u_2, \ldots, u_k\}$ is a linearly independent subset of S but adjoining to β any vector in S not in β produces a linearly dependent set. We claim that β is a basis for $\mathbf{V}$. Because β is linearly independent by construction it is sufficient to show that β spans $\mathbf{V}$. By Theorem 1.5 we need to show that $S \subseteq \text{span}(\beta)$. Let $v \in S$. If $v \in \beta$, then clearly $v \in \text{span}(\beta)$. Otherwise, if $v \notin \beta$, then the preceding construction shows that $\beta \cup \{v\}$ is linearly dependent. So $v \in \text{span}(\beta)$ by Theorem 1.7. Thus $S \subseteq \text{span}(\beta)$.

Theorem 1.10 (Replacement Theorem). Let $\mathbf{V}$ be a vector space that is generated by a set G containing exactly n vectors, and let L be a linearly independent subset of $\mathbf{V}$ containing exactly m vectors. Then $m \leq n$ and there exists a subset H of G containing exactly $n-m$ vectors such that $L \cup H$ generates $\mathbf{V}$.

Proof. The proof is by mathematical induction on m. The induction begins with $m = 0$; in this case $L = \phi$, and so taking $H = G$ gives the desired result.

Now suppose that the theorem is true for some integer $m \geq 0$. We prove that the theorem is true for $m+1$. Let $L = \{v_1, v_2, \ldots, v_{m+1}\}$ be a linearly independent set of $\mathbf{V}$ consisting of $m+1$ vectors. By the Corollary to Theorem 1.6, $\{v_1, v_2, \ldots, v_m\}$ is linearly independent, and so we may apply the induction hypothesis to conclude that $m \leq n$ and that there is a subset $\{u_1, u_2, \ldots, u_{n-m}\}$ of G such that $\{v_1, v_2, \ldots, v_m\} \cup \{u_1, u_2, \ldots, u_{n-m}\}$ generates $\mathbf{V}$. Thus there exists scalars $a_1, a_2, \ldots, a_m, b_1, b_2, \ldots, b_{n-m}$ such that

$$a_1v_1 + a_2v_2 + \cdots + a_mv_m + b_1u_1 + b_2u_2 + \cdots + b_{n-m}u_{n-m} = v_{m+1}.$$

Note that $n-m > 0$, lest v_{m+1} be a linear combination of $v_1, v_2, \ldots, v_m$ which by theorem 1.7 contradicts the assumption that L is linearly independent. Hence $n > m$; that is $n \geq m+1$. Moreover, some b_i, say b_1 is nonzero, for otherwise we obtain the same contradiction. Solving for u_1 gives

$$u_1 = (-b_1^{-1}a_1)v_1 + (-b_1^{-1}a_2)v_2 + \cdots + (-b_1^{-1}a_m)v_m + (b_1^{-1})v_{m+1} + (-b_1^{-1}b_2)u_2 + \cdots + (-b_1^{-1}b_{n-m})u_{n-m}.$$

Let $H = \{u_2, \ldots, u_{n-m}\}$. Then $u_1 \in \text{span}(L \cup H)$, and because $v_1, v_2, \ldots, v_m, u_2, \ldots, u_{n-m}$ are clearly in $\text{span}(L \cup H)$, it follows that

$$\{\mathrm{v_1, v_2, \ldots, v_m, u_1, u_2, \ldots, u_{n-m}}\} \subseteq \mathrm{span(L \cup H)}.$$

Because $\{v_1, v_2, \ldots, v_m, u_1, u_2, \ldots, u_{n-m}\}$ generates $\mathbf{V}$, Theorem 1.5 implies that $\text{span}(L \cup H) = \mathbf{V}$. Since H is a subset of G that contains $(n-m) - 1 = n - (m+1)$ vectors, the theorem is true for $m+1$. This completes the induction.

Corollary 1. Let **V** be a vector space having a finite basis. Then every basis for **V** contains the same number of vectors.

Proof. Suppose that β is a finite basis for **V** that contains exactly n vectors and let γ be any other basis for **V**. If γ contains more that n vectors, then we can select a subset $S$of γ containing exactly $n+1$ vectors. Since S is linearly independent and β generates **V**, the replacement theorem implies that $n+1 \leq n$, a contradiction. Therefore γ is finite, and the number m of vectors in γ satisfies $m \leq n$. Reversing the roles of β and γ and arguing as above, we obtain $n \leq m$. Hence $m = n$.

Corollary 2. Let **V** be a vector space with dimension n.
(a) Any finite generating set for **V** contains at least n vectors, and a generating set for **V** that contains exactly n vectors is a basis for **V**.
(b) Any linearly independent subset of **V** that contains exactly n vectors is a basis for **V**.
(c) Every linearly independent subset of **V** can be extended to a basis for **V**.

Proof. Let β be a basis for **V**.
(a) Let G be a finite generating set for **V**. By Theorem 1.9 some subset H of G is a basis for **V**. Corollary 1 implies that H contains exactly n vectors. Since a subset of G contains n vectors, G must contain at least n vectors. Moreover, if G contains exactly n vectors, then we must have $H = G$, so that G is a basis for **V**.

(b) Let L be a linearly independent subset of **V** containing exactly n vectors. It follows from the replacement theorem that there is a subset H of β containing $n - n = 0$ vectors such that $L \cup H$ generates **V**. Thus $H = \phi$, and L generates **V**. Since L is also linearly independent, L is a basis for **V**.

(c) If L is a linearly independent subset of **V** containing m vectors, then the replacement theorem asserts that there is a subset H of β containing exactly $n - m$ vectors such that $L \cup H$ generates **V**. Now $L \cup H$ contains at most n vectors; therefore (a) implies that $L \cup H$ contains exactly n vectors and that $L \cup H$ is a basis for **V**.

Theorem 1.11. Let **W** be a subspace of a finite-dimensional vector space **V.** Then **W** is finite-dimensional and $\dim(\mathbf{W}) \leq \dim(\mathbf{V})$. Moreover if $\dim(\mathbf{W}) = \dim(\mathbf{V})$, then $\mathbf{W} = \mathbf{V}$.

Proof. Let $\dim(\mathbf{V}) = n$. If **W**$= \{0\}$, then **W** is finite dimensional and $\dim(\mathbf{W}) = 0 \leq n$. Otherwise, **W** contains a nonzero vector x_1; so $\{x_1\}$ is a linearly independent set. Continue choosing vectors, $x_1, x_2, \ldots, x_k$ in **W** such that $\{x_1, x_2, \ldots, x_k\}$ is linearly independent. Since no linearly independent subset of **V** can contain more than n vectors, this process must stop at a stage where $k \leq n$ and $\{x_1, x_2, \ldots, x_k\}$ is linearly independent but adjoining any other vector from **W**produces a linearly dependent set. Theorem 1.7 implies that $\{x_1, x_2, \ldots, x_k\}$ generates **W**, and hence is a basis for **W**. Therefore $\dim(\mathbf{W}) = k \leq n$.

If $\dim(\mathbf{W}) = n$, then a basis for **W** is a linearly independent subset of **V** containing n vectors. But by Corollary 2 of the replacement theorem implies that this basis for **W** is also a basis for **V**; so **W**= **V**.

Corollary. If **W** is a subspace of a vector space **V**, then any basis for **W** can be extended to a basis for **V**.

Proof. Let S be a basis for **W**. Because S is a linearly independent subset of **V**, Corollary 2 of the replacement theorem guarantees that S can be extended to a basis for **V**.

Theorem 2.1. Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces and $\mathbf{T}\colon \mathbf{V} \to \mathbf{W}$ be linear. Then $\mathbf{N}(\mathbf{T})$ and $\mathbf{R}(\mathbf{T})$ are subspaces of $\mathbf{V}$ and $\mathbf{W}$, respectively.

Proof. To clarify the notation, we use the symbols $\mathbf{0_V}$ and $\mathbf{0_W}$ to denote the zero vectors of $\mathbf{V}$ and $\mathbf{W}$, respectively.

Since $\mathbf{T}(\mathbf{0_V}) = \mathbf{0_W}$, we have that $\mathbf{0_V} \in \mathbf{N}(\mathbf{T})$. Let $x, y \in \mathbf{N}(\mathbf{T})$ and $c \in F$. Then $\mathbf{T}(x+y) = \mathbf{T}(x) + \mathbf{T}(y) = \mathbf{0_W} + \mathbf{0_W} = \mathbf{0_W}$, and $\mathbf{T}(cx) = c\mathbf{T}(x) = c\mathbf{0_W} = \mathbf{0_W}$. Hence $x + y \in \mathbf{N}(\mathbf{T})$ and $cx \in \mathbf{N}(\mathbf{T})$, so that $\mathbf{N}(\mathbf{T})$ is a subspace of $\mathbf{V}$.

Because $\mathbf{T}(\mathbf{0_V}) = \mathbf{0_W}$, we have that $\mathbf{0_W} \in \mathbf{R}(\mathbf{T})$. Now let $x, y \in \mathbf{R}(\mathbf{T})$ and $c \in F$. Then there exist v and w in $\mathbf{V}$ such that $\mathbf{T}(v) = x$ and $\mathbf{T}(w) = y$. So $\mathbf{T}(v+w) = \mathbf{T}(v) + \mathbf{T}(w) = x + y$, and $\mathbf{T}(cv) = c\,\mathbf{T}(v) = cx$. Thus $x + y \in \mathbf{R}(\mathbf{T})$ and $cx \in \mathbf{R}(\mathbf{T})$, so $\mathbf{R}(\mathbf{T})$ is a subspace of $\mathbf{W}$.

Theorem 2.2. Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces, and $\mathbf{T}$: $\mathbf{V}\rightarrow$ $\mathbf{W}$ be linear. If $\beta = \{v_1, v_2, \ldots, v_n\}$ is a basis for $\mathbf{V}$, then $\mathbf{R}(\mathbf{T}) = \text{span}(\mathbf{T}(\beta)) = \text{span}(\{\mathbf{T}(v_1),\mathbf{T}(v_2),\ldots,\mathbf{T}(v_n)\})$.

Proof. Clearly $\mathbf{T}(v_i) \in \mathrm{R}(\mathbf{T})$ for each i. Because $\mathbf{R}(\mathbf{T})$ is a subspace, $\mathbf{R}(\mathbf{T})$ contains $\text{span}(\{\mathbf{T}(v_1),\mathbf{T}(v_2),\ldots,\mathbf{T}(v_n)\}) = \text{span}(\mathbf{T}(\beta))$ by Theorem 1.5.

Now suppose that $w \in \mathbf{R}(\mathbf{T})$. Then $w = \mathbf{T}(v)$ for some $v \in \mathbf{V}$. Because β is a basis for $\mathbf{V}$, we have

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

for some $a_1, a_2, \ldots, a_n \in F$. Since $\mathbf{T}$ is linear, it follows that

$$w = \mathbf{T}(v) = \textstyle\sum_{i=1}^{n} a_i \mathbf{T}(v_i) \in \text{span}(\mathbf{T}(\beta)).$$

So $\mathbf{R}(\mathbf{T})$ is contained in $\text{span}(\mathbf{T}(\beta))$.

Theorem 2.3 (Dimension Theorem). Let **V** and **W** be vector spaces, and **T**: **V**→ **W** be linear. If **V** is finite-dimensional then

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

Proof. Suppose that $\dim(\mathbf{V}) = n$, $\dim(\mathbf{N}(\mathbf{T})) = k$, and $\{v_1, v_2, \ldots, v_k\}$ is a basis for $\mathbf{N}(\mathbf{T})$. By the corollary to Theorem 1.11, we may extend $\{v_1, v_2, \ldots, v_k\}$ to a basis $\beta = \{v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ for **V**. We claim that $S = \{\mathbf{T}(v_{k+1}),\mathbf{T}(v_{k+2}),\ldots,\mathbf{T}(v_n)\}$ is a basis for $\mathbf{R}(\mathbf{T})$.

First we prove that S generates $\mathbf{R}(\mathbf{T})$. Using Theorem 2.2 and the fact that $\mathbf{T}(v_i) = \mathbf{0}$ for $1 \le i \le k$, we have

$$\mathbf{R}(\mathbf{T}) = \text{span}(\{\mathbf{T}(v_1),\mathbf{T}(v_2),\ldots,\mathbf{T}(v_n)\})$$

$$=\text{span}(\{\mathbf{T}(v_{k+1}),\mathbf{T}(v_{k+2}),\ldots,\mathbf{T}(v_n)\}) = \text{span}(S).$$

Now we prove that S is linearly independent. Suppose that

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

Using the fact that **T** is linear, we have

$$\mathbf{T}\Big(\sum_{\mathrm{i=k+1}}^{\mathrm{n}} \mathrm{b_i v_i}\Big) = \mathbf{0}.$$

So

$$\sum_{i=k+1}^{n} b_i v_i \in \mathbf{N}(\mathbf{T}).$$

Hence there exists $c_1, c_2, \ldots, c_k \in F$ such that

$$\sum_{i=k+1}^{n} b_i v_i = \sum_{i=1}^{k} c_i v_i \quad or \quad \sum_{i=1}^{k}(-c_i)v_i + \sum_{i=k+1}^{n} b_i v_i = \mathbf{0}.$$

Since β is a basis for **V**, we have $b_i = 0$ for all i. Hence S is linearly independent. Notice that this argument also shows that $\mathbf{T}(v_{k+1}),\mathbf{T}(v_{k+2}),\ldots,\mathbf{T}(v_n)$ are distinct; therefore $\text{rank}(\mathbf{T}) = n - k$.

Theorem 2.4. Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces, and $\mathbf{T}\colon \mathbf{V}\rightarrow \mathbf{W}$ be linear. Then $\mathbf{T}$ is one-to-one if and only if $\mathbf{N}(\mathbf{T}) = \{\mathbf{0}\}$.

Proof. Suppose that $\mathbf{T}$ is one-to-one and $x \in \mathbf{N}(\mathbf{T})$. Then $\mathbf{T}(\mathrm{x}) = \mathbf{0} = \mathbf{T}(\mathbf{0})$. Since $\mathbf{T}$ is one-to-one, we have $x = \mathbf{0}$. Hence $\mathbf{N}(\mathbf{T}) = \{\mathbf{0}\}$.

Now assume that $\mathbf{N}(\mathbf{T}) = \{\mathbf{0}\}$, and suppose that $\mathbf{T}(\mathrm{x}) = \mathbf{T}(\mathrm{y})$. Then $\mathbf{0} = \mathbf{T}(\mathrm{x}) - \mathbf{T}(\mathrm{y}) = \mathbf{T}(\mathrm{x}-\mathrm{y})$ since $\mathbf{T}$ is linear. Therefore $x - y \in \mathbf{N}(\mathbf{T}) = \{\mathbf{0}\}$. So $x - y = \mathbf{0}$ or $x = y$. This means that $\mathbf{T}$ is one-to-one.

Theorem 2.5. Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces of equal (finite) dimension, and $\mathbf{T}\colon \mathbf{V}\to \mathbf{W}$ be linear. Then the following are equivalent.
(a) $\mathbf{T}$ is one-to-one.
(b) $\mathbf{T}$ is onto.
(c) $\text{rank}(\mathbf{T}) = \dim(\mathbf{V})$.

Proof. From the dimension theorem, we have

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

Now with the use of Theorem 2.4, we have that $\mathbf{T}$ is one-to-one if and only if $\mathbf{N}(\mathbf{T}) = \{\mathbf{0}\}$, if and only if $\text{nullity}(\mathbf{T}) = 0$, if and only if $\text{rank}(\mathbf{T}) = \dim(\mathbf{V})$, if and only if $\text{rank}(\mathbf{T}) = \dim(\mathbf{W})$, and if and only if $\dim(\mathbf{R}(\mathbf{T})) = \dim(\mathbf{W})$. By Theorem 1.11, this equality is equivalent to $\mathbf{R}(\mathbf{T}) = \mathbf{W}$, the definition of $\mathbf{T}$ being onto.

Theorem 2.6. Let **V** and **W** be vector spaces over F, and suppose that $\{v_1, v_2, \ldots, v_n\}$ is a basis for **V**. For $w_1, w_2, \ldots, w_n$ in **W**, there exists exactly one linear transformation **T**: **V**→ **W** such that $\mathbf{T}(v_i) = w_i$ for $i = 1, 2, \ldots, n$.

Proof. Let $x \in \mathbf{V}$. Then

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

where $a_1, a_2, \ldots, a_n$ are unique scalars. Define

$$\mathbf{T} : \mathbf{V} \to \mathbf{W} \quad \text{by} \quad \mathbf{T}(\mathrm{x}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{a_i w_i}.$$

(a) **T** is linear: Suppose that $u, v \in \mathbf{V}$ and $d \in F$. Then we may write

$$\mathrm{u} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{b_i v_i} \quad \text{and} \quad \mathrm{v} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{c_i v_i}$$

for some scalars $b_1, b_2, \ldots, b_n, c_1, c_2, \ldots, c_n$. Thus

$$du + v = \sum_{i=1}^{n} (db_i + c_i) v_i.$$

So

$$\mathbf{T}(\mathrm{du + v}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{db_i + c_i}) \mathrm{w_i} = \mathrm{d} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{b_i w_i} + \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{c_i w_i} = \mathrm{d}\mathbf{T}(\mathrm{u}) + \mathbf{T}(\mathrm{v}).$$

(b) Clearly

$$\mathbf{T}(\mathrm{v_i}) = \mathrm{w_i} \quad \text{for} \quad \mathrm{i} = 1, 2, \ldots, \mathrm{n}.$$

(c) **T** is unique: Suppose that **U**: **V**→ **W** is linear and $\mathbf{U}(v_i) = w_i$ for $i = 1, 2, \ldots, n$. Then for $x \in \mathbf{V}$ with

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

we have

$$\mathbf{U}(\mathrm{x}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{a_i} \mathbf{U}(\mathrm{v_i}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{a_i w_i} = \mathbf{T}(\mathrm{x}).$$

Hence $\mathbf{U} = \mathbf{T}$.

Corollary. Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces, and suppose that $\mathbf{V}$ has a finite basis $\{v_1, v_2, \ldots, v_n\}$. If $\mathbf{U}$,$\mathbf{T}$: $\mathbf{V}\rightarrow$ $\mathbf{W}$ are linear and $\mathbf{U}(\mathrm{v_i}) = \mathbf{T}(\mathrm{v_i})$ for $i = 1, 2, \ldots, n$, then $\mathbf{U} = \mathbf{T}$.

Theorem 2.7. Let **V** and **W** be vector spaces over a field F, and **T**,**U**: **V**$\rightarrow$ **W** be linear.
(a) For all $a \in F$, $a\mathbf{T}+\mathbf{U}$ is linear.
(b) Using the operations of addition and scalar multiplication in the preceding definition, the collection of all linear transformations from **V** to **W** is a vector space over F.

Proof. (a) Let $x, y \in \mathbf{V}$and $c \in F$. Then

$$\begin{aligned}(a\mathbf{T}+\mathbf{U})(\mathrm{cx}+\mathrm{y}) &= \mathrm{a}\mathbf{T}(\mathrm{cx}+\mathrm{y})+\mathbf{U}(\mathrm{cx}+\mathrm{y}) \\ &= a[\mathbf{T}(\mathrm{cx}+\mathrm{y})]+\mathrm{c}\mathbf{U}(\mathrm{x})+\mathbf{U}(\mathrm{y}) \\ &= a[c\mathbf{T}(\mathrm{x})+\mathbf{T}(\mathrm{y})]+\mathrm{c}\mathbf{U}(\mathrm{x})+\mathbf{U}(\mathrm{y}) \\ &= ac\mathbf{T}(\mathrm{x})+\mathrm{c}\mathbf{U}(\mathrm{x})+\mathrm{a}\mathbf{T}(\mathrm{y})+\mathbf{U}(\mathrm{y}) \\ &= c(a\mathbf{T}+\mathbf{U})(\mathrm{x})+(\mathrm{a}\mathbf{T}+\mathbf{U})(\mathrm{y}).\end{aligned}$$

So $a\mathbf{T}+\mathbf{U}$ is linear.

(b) Noting that $\mathbf{T}_0$, the zero transformation, plays the role of the zero vector, it is easy to verify that the axioms of a vector space are satisfied, and hence that the collection of all linear transformations from **V** to **W** is a vector space over F.

Theorem 2.8. Let $\mathbf{V}$ and $\mathbf{W}$ be finite-dimensional vector spaces with ordered bases β and γ, respectively, and let $\mathbf{T}$,$\mathbf{U}$: $\mathbf{V}\rightarrow$ $\mathbf{W}$ be linear transformations. Then
(a) $[\mathbf{T}+\mathbf{U}]_\beta^\gamma = [\mathbf{T}]_\beta^\gamma + [\mathbf{U}]_\beta^\gamma$
(b) $[a\mathbf{T}]_\beta^\gamma = \mathrm{a}[\mathbf{T}]_\beta^\gamma$ for all scalars a.

Proof. (a) Let $\beta = \{v_1, v_2, \ldots, v_n\}$ and $\gamma = \{w_1, w_2, \ldots, w_m\}$. There exists unique scalars a_{ij} and b_{ij} $(1 \leq i \leq m,\ 1 \leq j \leq n)$ such that

$$\mathbf{T}(\mathrm{v_j}) = \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{a_{ij}w_i} \quad \text{and} \quad \mathbf{U}(\mathrm{v_j}) = \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{b_{ij}w_i} \quad \text{for} \ \ 1 \leq \mathrm{j} \leq \mathrm{n}.$$

Hence

$$(\mathbf{T}+\mathbf{U})(\mathrm{v_j}) = \sum_{\mathrm{i=1}}^{\mathrm{m}} (\mathrm{a_{ij}} + \mathrm{b_{ij}})\mathrm{w_i}.$$

Thus

$$([\mathbf{T}+\mathbf{U}]_\beta^\gamma)_{\mathrm{ij}} = \mathrm{a_{ij}} + \mathrm{b_{ij}} = ([\mathbf{T}]_\beta^\gamma + [\mathbf{U}]_\beta^\gamma)_{\mathrm{ij}}.$$

So (a) is proved, and the proof of (b) is similar.

Theorem 2.9. Let **V**, **W**, and **Z** be vector spaces over the same field F, and let **T**: **V**$\rightarrow$ **W** and **U**: **W**$\rightarrow$ **Z** be linear. Then **UT**: **V**$\rightarrow$ **Z** is linear.

Proof. Let $x, y \in \mathbf{V}$ and $z \in F$. Then

$$\mathbf{UT}(\mathrm{ax} + \mathrm{y}) = \mathbf{U}(\mathbf{T}(\mathrm{ax} + \mathrm{y})) = \mathbf{U}(\mathrm{a}\mathbf{T}(\mathrm{x}) + \mathbf{T}(\mathrm{y}))$$

$$= \mathbf{U}(\mathrm{a}\mathbf{T}(\mathrm{x})) + \mathbf{U}(\mathbf{T}(\mathrm{y})) = \mathrm{a}\mathbf{U}(\mathbf{T}(\mathrm{x})) + \mathbf{U}(\mathbf{T}(\mathrm{y})) = \mathrm{a}(\mathbf{UT})(\mathrm{x}) + (\mathbf{UT})(\mathrm{y}).$$

Theorem 2.10. Let $\mathbf{V}$ be a vector space. Let $\mathbf{T}, \mathbf{U_1}, \mathbf{U_2} \in \mathbf{L}(\mathbf{V})$. Then
(a) $\mathbf{T}(\mathbf{U_1} + \mathbf{U_2}) = \mathbf{TU_1} + \mathbf{TU_2}$ and $(\mathbf{U_1} + \mathbf{U_2})\mathbf{T} = \mathbf{U_1T} + \mathbf{U_2T}$
(b) $\mathbf{T}(\mathbf{U_1U_2}) = (\mathbf{TU_1})\mathbf{U_2}$
(c) $\mathbf{TI} = \mathbf{IT} = \mathbf{T}$
(d) $a(\mathbf{U_1U_2}) = (\mathrm{a}\mathbf{U_1})\mathbf{U_2} = \mathbf{U_1}(\mathrm{a}\mathbf{U_2})$ for all scalars a.

Proof. (a) Let $x \in \mathbf{V}$. Then

$$\mathbf{T}(\mathbf{U_1} + \mathbf{U_2})(\mathrm{x}) = \mathbf{T}((\mathbf{U_1} + \mathbf{U_2})(\mathrm{x})) = \mathbf{T}(\mathbf{U_1}(\mathrm{x}) + \mathbf{U_2}(\mathrm{x}))$$

$$= \mathbf{T}(\mathbf{U_1}(\mathrm{x})) + \mathbf{T}(\mathbf{U_2}(\mathrm{x})) = (\mathbf{TU_1})(\mathrm{x}) + (\mathbf{TU_2})(\mathrm{x}) = \mathbf{TU_1}(\mathrm{x}) + \mathbf{TU_2}(\mathrm{x})$$

so that $\mathbf{T}(\mathbf{U_1} + \mathbf{U_2}) = \mathbf{TU_1} + \mathbf{TU_2}$.

Likewise

$$(\mathbf{U_1} + \mathbf{U_2})\mathbf{T}(\mathrm{x}) = (\mathbf{U_1} + \mathbf{U_2})(\mathbf{T}(\mathrm{x})) = \mathbf{U_1}(\mathbf{T}(\mathrm{x})) + \mathbf{U_2}(\mathbf{T}(\mathrm{x}))$$

$$= (\mathbf{U_1T})(\mathrm{x}) + (\mathbf{U_2T})(\mathrm{x}) = \mathbf{U_1T}(\mathrm{x}) + \mathbf{U_2T}(\mathrm{x})$$

so that $(\mathbf{U_1} + \mathbf{U_2})\mathbf{T} = \mathbf{U_1T} + \mathbf{U_2T}$.

The proof of the other properties are developed in a similar manner.

Theorem 2.11. Let **V**, **W**, and **Z** be finite-dimensional vector spaces with ordered bases α, β, and γ, respectively. Let **T**: **V**$\to$ **W** and **U**: **W**$\to$ **Z** be linear transformations. Then

$$[\mathbf{UT}]_\alpha^\gamma = [\mathbf{U}]_\beta^\gamma [\mathbf{T}]_\alpha^\beta.$$

Proof. Let $A = [\mathbf{U}]_\beta^\gamma$, $B = [\mathbf{T}]_\alpha^\beta$, and $C = [\mathbf{UT}]_\alpha^\gamma$ where $\alpha = \{v_1, v_2, \ldots, v_n\}$, $\beta = \{w_1, w_2, \ldots, w_m\}$, and $\gamma = \{z_1, z_2, \ldots, z_p\}$ are ordered bases for **V**, **W**, and **Z**, respectively. We wish to determine the elements of the matrix $[\mathbf{UT}]_\alpha^\gamma$. For $1 \leq j \leq n$, we have

$$(\mathbf{UT})(\mathrm{v_j}) = \mathbf{U}(\mathbf{T}(\mathrm{v_j})) = \mathbf{U}\Big(\sum_{\mathrm{k=1}}^{\mathrm{m}} \mathrm{B_{kj} w_k}\Big) = \sum_{\mathrm{k=1}}^{\mathrm{m}} \mathrm{B_{kj}} \mathbf{U}(\mathrm{w_k})$$

$$= \sum_{k=1}^{m} B_{kj}\Big(\sum_{i=1}^{p} A_{ik} z_i\Big) = \sum_{k=1}^{m}\Big(\sum_{i=1}^{p} B_{kj} A_{ik} z_i\Big) = \sum_{k=1}^{m}\sum_{i=1}^{p} B_{kj} A_{ik} z_i$$

$$= \sum_{i=1}^{p}\sum_{k=1}^{m} B_{kj} A_{ik} z_i = \sum_{i=1}^{p}\Big(\sum_{k=1}^{m} A_{ik} B_{kj}\Big) z_i.$$

Hence $C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj} = (AB)_{ij}$ where the final equality results from the definition of matrix multiplication. Thus $C = AB$ or $[\mathbf{UT}]_\alpha^\gamma = [\mathbf{U}]_\beta^\gamma [\mathbf{T}]_\alpha^\beta$.

Corollary. Let **V** be a finite-dimensional vector space with an ordered basis β. Let $\mathbf{T}, \mathbf{U} \in \mathbf{L}(\mathbf{V})$. Then $[\mathbf{UT}]_\beta = [\mathbf{U}]_\beta [\mathbf{T}]_\beta$.

Theorem 2.12. Let Let A be a $m \times n$ matrix, B and C be $n \times p$ matrices, and D and E be $q \times m$ matrices. Then
(a) $A(B+C) = AB + AC$ and $(D+E)A = DA + EA$.
(b) $a(AB) = (aA)B = A(aB)$ for any scalar a.
(c) $I_m A = A = AI_n$.
(d) If $\mathbf{V}$ is an n-dimensional vector space with an ordered basis β, then $[I_V]_\beta = I_n$.

Proof. The proof of the first half of (a) and (c) are given.

(a) We have

$$[A(B+C)]_{ij} = \sum_{k=1}^{n} A_{ik}(B+C)_{kj} = \sum_{k=1}^{n} A_{ik}(B_{kj} + C_{kj})$$

$$= \sum_{k=1}^{n}(A_{ik}B_{kj} + A_{ik}C_{kj}) = \sum_{k=1}^{n} A_{ik}B_{kj} + \sum_{k=1}^{n} A_{ik}C_{kj} = (AB)_{ij} + (AC)_{ij} = [AB+AC]_{ij}.$$

So $A(B+C) = AB + AC$.

(c) We have

$$(I_m A)_{ij} = \sum_{k=1}^{m} (I_m)_{ik} A_{kj} = \sum_{k=1}^{m} \delta_{ik} A_{kj} = A_{ij}.$$

Corollary. Let A be an $m \times n$ matrix, $B_1, B_2, \ldots, B_k$ be $n \times p$ matrices, $C_1, C_2, \ldots, C_k$ be $q \times m$ matrices, and $a_1, a_2, \ldots, a_k$ be scalars. Then

$$A\left(\sum_{i=1}^{k} a_i B_i\right) = \sum_{i=1}^{k} a_i A B_i \quad \text{and} \quad \left(\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{a_i C_i}\right)\mathrm{A} = \sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{a_i C_i A}.$$

Proof. Induction on k.

Theorem 2.13. Let A be a $m \times n$ matrix, B be $n \times p$ matrix. For each j $(1 \leq j \leq p)$ let u_j and v_j denote the jth columns of AB and B respectively. Then
(a) $u_j = Av_j$
(b) $v_j = Be_j$, where e_j is the jth standard vector of $\mathbf{F}^{\mathrm{p}}$.

Proof. The proof of the first half of (a) and (c) are given.

(a) We have

$$u_j = \begin{pmatrix} (AB)_{1j} \\ (AB)_{2j} \\ \vdots \\ (AB)_{mj} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} A_{1k}B_{kj} \\ \sum_{k=1}^{n} A_{2k}B_{kj} \\ \vdots \\ \sum_{k=1}^{n} A_{mk}B_{kj} \end{pmatrix} = A \begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{mj} \end{pmatrix} = Av_j.$$

Hence (a) is proved.

(b) The proof of (b) follows immediately if we substitute B for A and I_p for B into the result from (a).

Theorem 2.14. Let **V** and **W** be finite-dimensional vector spaces having ordered bases β and γ, respectively, and let **T**: **V**→ **W** be linear. Then, for each $u \in \mathbf{V}$, we have

$$[\mathbf{T}(\mathrm{u})]_\gamma = [\mathbf{T}]_\beta^\gamma [\mathrm{u}]_\beta.$$

Proof. Fix $u \in \mathbf{V}$, and define the linear transformations $f : \mathbf{F} \to \mathbf{V}$ by $f(a) = au$ and $g : \mathbf{F} \to \mathbf{W}$ by $g(a) = a(\mathbf{T}\mathrm{u})$ for all $a \in \mathbf{F}$. Let $\alpha = \{1\}$ be the standard ordered basis for **F**. Notice that $g = \mathbf{T}\mathrm{f}$. Identifying the column vectors as matrices and using Theorem 2.11, we obtain

$$[\mathbf{T}(\mathrm{u})]_\gamma = [\mathrm{g}(1)]_\gamma = [\mathrm{g}]_\alpha^\gamma = [\mathbf{T}\mathrm{f}]_\alpha^\gamma = [\mathbf{T}]_\beta^\gamma [\mathrm{f}]_\alpha^\beta = [\mathbf{T}]_\beta^\gamma [\mathrm{f}(1)]_\beta = [\mathbf{T}]_\beta^\gamma [\mathrm{u}]_\beta.$$

Alternative Proof. Let $\beta = \{v_1, v_2, \ldots, v_n\}$, $\gamma = \{w_1, w_2, \ldots, w_m\}$, and $u = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$. Then

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

$$\mathbf{T}(\mathrm{u}) = \mathrm{a}_1 \mathbf{T}(\mathrm{v}_1) + \mathrm{a}_2 \mathbf{T}(\mathrm{v}_2) + \cdots + \mathrm{a}_\mathrm{n} \mathbf{T}(\mathrm{v}_\mathrm{n}),$$

and

$$[\mathbf{T}(\mathrm{u})]_\gamma = [\mathrm{a}_1 \mathbf{T}(\mathrm{v}_1) + \mathrm{a}_2 \mathbf{T}(\mathrm{v}_2) + \cdots + \mathrm{a}_\mathrm{n} \mathbf{T}(\mathrm{v}_\mathrm{n})]_\gamma$$

$$= a_1 [\mathbf{T}(\mathrm{v}_1)]_\gamma + \mathrm{a}_2 [\mathbf{T}(\mathrm{v}_2)]_\gamma + \cdots + \mathrm{a}_\mathrm{n} [\mathbf{T}(\mathrm{v}_\mathrm{n})]_\gamma$$

$$= \Big([\mathbf{T}(\mathrm{v}_1)]_\gamma \ [\mathbf{T}(\mathrm{v}_2)]_\gamma \ \cdots \ [\mathbf{T}(\mathrm{v}_\mathrm{n})]_\gamma \Big) \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

$$= [\mathbf{T}]_\beta^\gamma [\mathrm{u}]_\beta.$$

Theorem 2.15. Let A be an $m \times n$ matrix with entries form $\mathbf{F}$. Then the left-multiplication transformation $\mathbf{L}_{\mathrm{A}} : \mathbf{F}^{\mathrm{n}} \to \mathbf{F}^{\mathrm{m}}$ is linear. Furthermore if B is any other $m \times n$ matrix (with entries from $\mathbf{F}$) and β and γ are the standard ordered bases for $\mathbf{F}^{\mathrm{n}}$ and $\mathbf{F}^{\mathrm{m}}$, respectively, then we have the following properties.
(a) $[\mathbf{L}_{\mathrm{A}}]_{\beta}^{\gamma} = \mathrm{A}$.
(b) $\mathbf{L}_{\mathrm{A}} = \mathbf{L}_{\mathrm{B}}$ if and only if $A = B$.
(c) $\mathbf{L}_{\mathrm{A+B}} = \mathbf{L}_{\mathrm{A}} + \mathbf{L}_{\mathrm{B}}$ and $\mathbf{L}_{\mathrm{aA}} = \mathrm{a}\mathbf{L}_{\mathrm{A}}$ for all $a \in \mathbf{F}$.
(d) If $\mathbf{T} : \mathbf{F}^{\mathrm{n}} \to \mathbf{F}^{\mathrm{m}}$ is linear, then there exists a unique $m \times n$ matrix C such that $\mathbf{T} = \mathbf{L}_{\mathrm{C}}$. In fact, $C = [\mathbf{T}]_{\beta}^{\gamma}$.
(e) If E is an $n \times p$ matrix, then $\mathbf{L}_{\mathrm{AE}} = \mathbf{L}_{\mathrm{A}}\mathbf{L}_{\mathrm{E}}$.
(f) If $m = n$, then $\mathbf{L}_{\mathrm{I_n}} = \mathrm{I}_{\mathbf{F}^{\mathrm{n}}}$.

Proof. The fact that $\mathbf{L}_{\mathrm{A}}$ is linear follows immediately from Theorem 2.12.
(a) The jth column of $[\mathbf{L}_{\mathrm{A}}]_{\beta}^{\gamma}$ is equal to $\mathbf{L}_{\mathrm{A}}(\mathrm{e_j})$. However, $\mathbf{L}_{\mathrm{A}}(\mathrm{e_j}) = \mathrm{Ae_j}$, which is also the jth column of A by Theorem 2.13(b). So $[\mathbf{L}_{\mathrm{A}}]_{\beta}^{\gamma} = \mathrm{A}$.
(b) If $\mathbf{L}_{\mathrm{A}} = \mathbf{L}_{\mathrm{B}}$, then we may use (a) to write $A = [\mathbf{L}_{\mathrm{A}}]_{\beta}^{\gamma} = [\mathbf{L}_{\mathrm{B}}]_{\beta}^{\gamma} = \mathrm{B}$. The proof of the converse is trivial.
(c) The proof is left as an exercise.
(d) Let $C = [\mathbf{T}]_{\beta}^{\gamma}$. By Theorem 2.14, we have $[\mathbf{T}(\mathrm{x})]_{\gamma} = [\mathbf{T}]_{\beta}^{\gamma}[\mathrm{x}]_{\beta}$, or
$\mathbf{T}(\mathrm{x}) = \mathrm{Cx} = \mathbf{L}_{\mathrm{C}}(\mathrm{x})$ for all $x \in \mathbf{F}^{\mathrm{n}}$. So $\mathbf{T} = \mathbf{L}_{\mathrm{C}}$. The uniqueness of C follows from (b).
(e) For any j $(1 \leq j \leq p)$, we may apply Theorem 2,13 several times to note that $(AE)e_j$ is the jth column of AE and that the jth column of AE is also equal to $A(Ee_j)$. So $(AE)e_j = A(Ee_j)$. Thus

$$\mathbf{L}_{\mathrm{AE}}(\mathrm{e_j}) = (\mathrm{AE})\mathrm{e_j} = \mathrm{A}(\mathrm{Ee_j}) = \mathbf{L}_{\mathrm{A}}(\mathrm{Ee_j}) = \mathbf{L}_{\mathrm{A}}(\mathbf{L}_{\mathrm{E}}(\mathrm{e_j})).$$

Hence $\mathbf{L}_{\mathrm{AE}} = \mathbf{L}_{\mathrm{A}}\mathbf{L}_{\mathrm{E}}$ by the corollary to Theorem 2.6.
(f) The proof is left as an exercise.

Theorem 2.16. Let A, B, and C be matrices such that $A(BC)$ is defined. Then $(AB)C$ is also defined and $A(BC) = (AB)C$; that is matrix multiplication is associative.

Proof. It is left to the reader to show that $(AB)C$ is defined. Using (e) of Theorem 2.15 and the associativity of functional composition, we have

$$\mathbf{L}_{\mathrm{A(BC)}} = \mathbf{L}_{\mathrm{A}}\mathbf{L}_{\mathrm{BC}} = \mathbf{L}_{\mathrm{A}}(\mathbf{L}_{\mathrm{B}}\mathbf{L}_{\mathrm{C}}) = (\mathbf{L}_{\mathrm{A}}\mathbf{L}_{\mathrm{B}})\mathbf{L}_{\mathrm{C}} = \mathbf{L}_{\mathrm{AB}}\mathbf{L}_{\mathrm{C}} = \mathbf{L}_{\mathrm{(AB)C}}.$$

So from (b) of Theorem 2.15, it follows that $A(BC) = (AB)C$.

Alternative Proof. Let A be an $m \times n$ matrix, B be an $n \times p$ matrix, and C be an $p \times q$ matrix, all with components from the same field F. Using the definition of matrix multiplication we have

$$\begin{aligned} (A(BC))_{ij} &= \sum_{s=1}^{n} A_{is}(BC)_{sj} = \sum_{s=1}^{n} A_{is}\Big(\sum_{t=1}^{p} B_{st}C_{tj}\Big) \\ &= \sum_{s=1}^{n}\sum_{t=1}^{p} A_{is}B_{st}C_{tj} = \sum_{t=1}^{p}\sum_{s=1}^{n} A_{is}B_{st}C_{tj} \\ &= \sum_{t=1}^{p}\Big(\sum_{s=1}^{n} A_{is}B_{st}\Big)C_{tj} = \sum_{t=1}^{p}(AB)_{it}C_{tj} = ((AB)C)_{ij}. \end{aligned}$$

Hence, $A(BC) = (AB)C$.

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

Proof. Let $y_1, y_2 \in \mathbf{W}$ and $c \in \mathbf{F}$. Since $\mathbf{T}$ is onto and one-to-one, there exists unique vectors x_1 and x_2 such that $\mathbf{T}(\mathrm{x}_1) = \mathrm{y}_1$ and $\mathbf{T}(\mathrm{x}_2) = \mathrm{y}_2$. Thus $x_1 = \mathbf{T^{-1}}(\mathrm{y}_1)$ and $x_2 = \mathbf{T^{-1}}(\mathrm{y}_2)$; so

$$\mathbf{T^{-1}}(\mathrm{cy}_1+\mathrm{y}_2) = \mathbf{T^{-1}}(\mathrm{c}\mathbf{T}(\mathrm{x}_1)+\mathbf{T}(\mathrm{x}_2)) = \mathbf{T^{-1}}(\mathbf{T}(\mathrm{cx}_1+\mathrm{x}_2)) = \mathrm{cx}_1+\mathrm{x}_2 = \mathrm{c}\mathbf{T^{-1}}(\mathrm{y}_1)+\mathbf{T^{-1}}(\mathrm{y}_2).$$

Lemma. Let $\mathbf{T}$ be an invertible transformation from $\mathbf{V}$ to $\mathbf{W}$. Then $\mathbf{V}$ is finite-dimensional if and only if $\mathbf{W}$ is finite-dimensional. In this case, $\dim(\mathbf{V}) = \dim(\mathbf{W})$.

Proof. Suppose $\mathbf{V}$ is finite dimensional. let $\beta = \{x_1, x_2, \ldots, x_n\}$ be a basis for $\mathbf{V}$. By Theorem 2.2, $\mathbf{T}(\beta)$ spans $\mathbf{R}(\mathbf{T}) = \mathbf{W}$; hence $\mathbf{W}$ is finite dimensional by Theorem 1.9. Conversely, if $\mathbf{W}$ is finite dimensional, then so is $\mathbf{V}$ by a similar argument, using $\mathbf{T}^{-1}$.

Now suppose that $\mathbf{V}$ and $\mathbf{W}$ are finite dimensional. Because $\mathbf{T}$ is one-to-one and onto, we have

$$\text{nullity}(\mathbf{T}) = 0 \quad \text{and} \quad \text{rank}(\mathbf{T}) = \dim(\mathbf{R}(\mathbf{T})) = \dim(\mathbf{W}).$$

So by the dimension theorem, if follows that $\dim(\mathbf{V}) = \dim(\mathbf{W})$.

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

Proof. Suppose that $\mathbf{T}$ is invertible. By the lemma, we have $\dim(\mathbf{V}) = \dim(\mathbf{W})$. Let $n = \dim(\mathbf{V})$. So $[\mathbf{T}]_\beta^\gamma$ is an $n \times n$ matrix. Now $\mathbf{T}^{-1} : \mathbf{W} \to \mathbf{V}$ satisfies $\mathbf{T}\mathbf{T}^{-1} = \mathbf{I_W}$ and $\mathbf{T}^{-1}\mathbf{T} = \mathbf{I_V}$. Thus

$$I_n = [\mathbf{I_V}]_\beta = [\mathbf{T}^{-1}\mathbf{T}]_\beta = [\mathbf{T}^{-1}]_\gamma^\beta [\mathbf{T}]_\beta^\gamma.$$

Similarly, $[\mathbf{T}]_\beta^\gamma [\mathbf{T}^{-1}]_\gamma^\beta = \mathrm{I_n}$. So $[\mathbf{T}]_\beta^\gamma$ is invertible and $([\mathbf{T}]_\beta^\gamma)^{-1} = [\mathbf{T}^{-1}]_\gamma^\beta$.

Now suppose that $A = [\mathbf{T}]_\beta^\gamma$ is invertible. Then there exists an $n \times n$ matrix B such that $AB = BA = I_n$. By Theorem 2.6, there exists $\mathbf{U} \in \mathbf{L}(\mathbf{W}, \mathbf{V})$ such that

$$\mathbf{U}(\mathrm{w_j}) = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{B_{ij} v_i} \quad \text{for } \mathrm{j} = 1, 2, \ldots, \mathrm{n}.$$

where $\gamma = \{w_1, w_2, \ldots, w_n\}$ and $\beta = \{v_1, v_2, \ldots, v_n\}$. It follows that $[\mathbf{U}]_\gamma^\beta = \mathrm{B}$. To show that $\mathbf{U} = \mathbf{T}^{-1}$, observe that

$$[\mathbf{UT}]_\beta = [\mathbf{U}]_\gamma^\beta [\mathbf{T}]_\beta^\gamma = \mathrm{BA} = \mathrm{I_n} = [\mathbf{I_V}]_\beta$$

by Theorem 2.11. So $\mathbf{UT} = \mathbf{I_V}$, and similarly $\mathbf{TU} = \mathbf{I_W}$.

Corollary 1. Let $\mathbf{V}$ be a finite-dimensional vector space with an ordered basis β, and let $\mathbf{T} : \mathbf{V} \to \mathbf{V}$ be linear. Then $\mathbf{T}$ is invertible if and only if $[\mathbf{T}]_\beta$ is invertible. Furthermore, $[\mathbf{T}^{-1}]_\beta = ([\mathbf{T}]_\beta)^{-1}$.

Corollary 2. Let A be an $n \times n$ matrix. Then A is invertible if and only if $\mathbf{L}_{\mathrm{A}}$ is invertible. Furthermore, $(\mathbf{L}_{\mathrm{A}})^{-1} = \mathbf{L}_{\mathrm{A}^{-1}}$.

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

Proof. Suppose that $\mathbf{V}$ is isomorphic to $\mathbf{W}$ and that $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ is an isomorphism from $\mathbf{V}$ to $\mathbf{W}$. By the lemma preceding Theorem 2.18, we have that $\dim(\mathbf{V}) = \dim(\mathbf{W})$.

Now suppose that $\dim(\mathbf{V}) = \dim(\mathbf{W})$, and let $\beta = \{v_1, v_2, \ldots, v_n\}$ and $\gamma = \{w_1, w_2, \ldots, w_n\}$ be bases for $\mathbf{V}$ and $\mathbf{W}$ respectively. By Theorem 2.6, there exists $\mathbf{T} : \mathbf{V} \to \mathbf{W}$ such that $\mathbf{T}$ is linear and $\mathbf{T}(\mathrm{v_i}) = \mathrm{w_i}$ for $i = 1, 2, \ldots, n$. Using Theorem 2.2, we have

$$\mathbf{R}(\mathbf{T}) = \mathrm{span}(\mathbf{T}(\beta)) = \mathrm{span}(\gamma) = \mathbf{W}.$$

So $\mathbf{T}$ is onto. From Theorem 2.5, we have $\mathbf{T}$ is also one-to-one. Hence $\mathbf{T}$ is an isomorphism.

Corollary. Let $\mathbf{V}$ be a vector space over F. Then $\mathbf{V}$ is isomorphic to F^n if and only if $\dim(\mathbf{V}) = n$.

Theorem 2.20. Let **V** and **W** be finite-dimensional vector spaces over **F** of dimensions n and m, respectively, and let β and γ be ordered bases for **V** and **W** respectively. Then the function $\mathbf{\Phi} : \mathbf{L}(\mathbf{V}, \mathbf{W}) \rightarrow \mathbf{M}_{\mathrm{m\times n}}(\mathbf{F})$, defined by $\mathbf{\Phi}(\mathbf{T}) = [\mathbf{T}]_\beta^\gamma$ for $\mathbf{T} \in \mathbf{L}(\mathbf{V}, \mathbf{W})$ is an isomorphism.

Proof. By Theorem 2.8, $\mathbf{\Phi}$ is linear. Hence we must show that $\mathbf{\Phi}$ is one-to-one and onto. This is accomplished if we show that for every $m \times n$ matrix A, there exists a unique linear transformation $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{W}$ such that $\mathbf{\Phi}(\mathbf{T}) = \mathrm{A}$. Let $\beta = \{v_1, v_2, \ldots, v_n\}$, $\gamma = \{w_1, w_2, \ldots, w_n\}$, and A be a given $m \times n$ matrix. By Theorem 2.6, there exists a unique linear transformation $\mathbf{T} : \mathbf{V} \rightarrow \mathbf{W}$ such that

$$\mathbf{T}(\mathrm{v_j}) = \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{A_{ij} w_i} \quad \text{for } 1 \leq \mathrm{j} \leq \mathrm{n}.$$

But this means that $[\mathbf{T}]_\beta^\gamma = \mathrm{A}$, or $\mathbf{\Phi}(\mathbf{T}) = \mathrm{A}$. Thus $\mathbf{\Phi}$ is an isomorphism.

Corollary. Let **V** and **W** be finite-dimensional vector spaces of dimensions n and m, respectively. Then $\mathbf{L}(\mathbf{V}, \mathbf{W})$ is finite-dimensional of dimension mn.

Proof. The proof follows from Theorems 2.20 and 2.19 and the fact that $\dim(\mathbf{M}_{\mathrm{m\times n}}(\mathbf{F})) = \mathrm{mn}$.

Theorem 2.21. For any finite-dimensional vector space $\mathbf{V}$ with ordered basis β, ϕ_β is an isomorphism.

Proof. Let $\beta = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis for $\mathbf{V}$, $x, y \in \mathbf{V}$, and $c \in \mathbf{F}$ with

$$x = a_1v_1 + a_2v_2 + \cdots + a_nv_n, \quad y = b_1v_1 + b_2v_2 + \cdots + b_nv_n.$$

Then

$$\phi_\beta(cx + y) = \phi_\beta(c(a_1v_1 + a_2v_2 + \cdots + a_nv_n) + (b_1v_1 + b_2v_2 + \cdots + b_nv_n))$$

$$= \phi_\beta((ca_1 + b_1)v_1 + (ca_2 + b_2)v_2 + \cdots + (ca_n + b_n)v_n))$$

$$= \begin{pmatrix} ca_1 + b_1 \\ ca_2 + b_2 \\ \vdots \\ ca_n + b_n \end{pmatrix} = c\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = c\phi_\beta(x) + \phi_\beta(y)$$

so ϕ_β is linear.

Next consider the null space of ϕ_β; let x be given as above and suppose that $x \in \mathbf{N}(\phi_\beta)$, then

$$\phi_\beta(x) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

so that by the equalities of n-tuples, $a_1 = a_2 = \cdots = a_n = 0$. Thus $\mathbf{N}(\phi_\beta) = \{\mathbf{0_v}\}$ and nullity$(\mathbf{N}(\phi_\beta)) = 0$. By Theorem 2.4, ϕ_β is one-to-one and since $\dim(\mathbf{V}) = \dim(\mathbf{F}^{\mathrm{n}})$, ϕ_β is onto by Theorem 2.5. Hence ϕ_β is linear, one-to-one, and onto and therefore, by definition, an isomorphism.

Theorem 2.22. Let β and β' be two ordered bases for a finite-dimensional vector space $\mathbf{V}$, and let $Q = [\mathbf{I_V}]_{\beta'}^{\beta}$. Then
(a) Q is invertible.
(b) For any $v \in \mathbf{V}$, $[v]_\beta = Q[v]_{\beta'}$.

Proof. (a) Since $\mathbf{I_V}$ is invertible, Q is invertible by Theorem 2.18.
(b) For any $v \in \mathbf{V}$,

$$[v]_\beta = [\mathbf{I_V}(\mathrm{v})]_\beta = [\mathbf{I_V}]_{\beta'}^{\beta}[\mathrm{v}]_{\beta'} = \mathrm{Q}[\mathrm{v}]_{\beta'}$$

by Theorem 2.14.

Theorem 2.23. Let $\mathbf{T}$ be a linear operator on a finite-dimensional vector space $\mathbf{V}$, and let β and β' be ordered bases for $\mathbf{V}$. Suppose that Q is the change of coordinate matrix that changes β'-coordinates into β-coordinates. Then

$$[\mathbf{T}]_{\beta'} = \mathrm{Q}^{-1}[\mathbf{T}]_{\beta}\mathrm{Q}.$$

Proof. Let $\mathbf{I}$ be the identity transformation on $\mathbf{V}$. Then $\mathbf{T} = \mathbf{IT} = \mathbf{TI}$; hence by Theorem 2.11

$$Q[\mathbf{T}]_{\beta'} = [\mathbf{I}]_{\beta'}^{\beta}[\mathbf{T}]_{\beta'}^{\beta'} = [\mathbf{IT}]_{\beta'}^{\beta} = [\mathbf{TI}]_{\beta'}^{\beta} = [\mathbf{T}]_{\beta}^{\beta}[\mathbf{I}]_{\beta'}^{\beta} = [\mathbf{T}]_{\beta}\mathrm{Q}.$$

Therefore, $[\mathbf{T}]_{\beta'} = \mathrm{Q}^{-1}[\mathbf{T}]_{\beta}\mathrm{Q}$.

Corollary. Let Let $A \in \mathbf{M}_{\mathrm{n}\times\mathrm{n}}(\mathrm{F})$, and let γ be an ordered basis for $\mathbf{F}^{\mathrm{n}}$. Then $[\mathbf{L}_{\mathrm{A}}]_{\gamma} = \mathrm{Q}^{-1}\mathrm{AQ}$, where Q is the $n \times n$ matrix whose jth column is the jth column vector of γ.

Theorem 3.1. Let $A \in \mathbf{M}_{\mathrm{m\times n}}(\mathrm{F})$, and suppose that B is obtained from A by performing an elementary row [column] operation. Then there exists an $m \times m$ [$n \times n$] elementary matrix E such that $B = EA$ [$B = AE$]. In fact E is obtained from I_m [I_n] by performing the same elementary row [column] operation as that which was performed on A to obtain B. Conversely, if E is an elementary $m \times m$ [$n \times n$] matrix, then EA [AE] is the matrix obtained from A by performing the same elementary row [column] operation as that which produces E from I_m [I_n].

Proof. We will consider only the case where E is obtained from I_m by adding a scalar multiple a of row s to row t. The matrix E can be decomposed as $E = I_m + J$ where $(J)_{ij} = a$ for $i = t,\ j = s$ and the components of this matrix are equal to zero otherwise. Therefore

$$EA = A + JA.$$

By the definition of matrix multiplication

$$(JA)_{ij} = \sum_{k=1}^{m} J_{ik}A_{kj}$$

so that by the above development, $(JA)_{ij} = 0$ for $i \neq t$. Hence all the rows of JA are identically zero with the possible exception of the tth row. The elements in this row are of the form

$$(JA)_{tj} = \sum_{k=1}^{m} J_{tk}A_{kj} = J_{ts}A_{sj} = aA_{sj}$$

which implies that tth row of JA is a times the sth row of A. Thus performing the componentwise addition of the elements of the matrices A and JA we find that, with the exception of the tth row, the rows of the matrix $EA = A + JA$ are equal to the rows of the matrix A. The tth row of the product matrix is the sum of the tth row of A and a times the sth row of A.

Theorem 3.2. Elementary matrices are invertible, and the inverse of an elementary matrix is an elementary matrix of the same type.

Proof. Let E be an elementary $n \times n$ matrix. Then E can be obtained by an elementary row operation on I_n. By reversing the steps used to transform I_n into E, we can transform E back into I_n. The result is that I_n can be obtained from E by an elementary row operation of the same type. By Theorem 3.1, there is an elementary matrix $\bar{E}$ such that $\bar{E}E = I_n$. Therefore by Exercise 10 of section 2.4, E is invertible and $E^{-1} = \bar{E}$.

Lemma 1. Let $\mathbf{V}$, $\mathbf{W}$, and $\mathbf{Z}$ be finite dimensional vector spaces, with $\mathbf{V}$ isomorphic to $\mathbf{W}$, $\dim(\mathbf{V}) = \dim(\mathbf{W}) = n$, ϕ: $\mathbf{V} \rightarrow \mathbf{W}$ an isomorphism, and $\mathbf{U}$: $\mathbf{W} \rightarrow \mathbf{Z}$ a linear transformation, then $\text{rank}(\mathbf{U}\phi) = \text{rank}(\mathbf{U})$.

Proof. The Dimension Theorem gives

$$\text{nullity}(\mathbf{U}\phi) + \text{rank}(\mathbf{U}\phi) = \dim(\mathbf{V}) = \text{n}$$

$$\text{nullity}(\mathbf{U}) + \text{rank}(\mathbf{U}) = \dim(\mathbf{W}) = \text{n}.$$

Next suppose that δ is a basis for $\mathbf{N}(\mathbf{U})$, then $\phi^{-1}(\delta)$ is a set of linearly independent vectors in $\mathbf{V}$ and $\phi^{-1}(\delta) \subseteq \mathbf{N}(\mathbf{U}\phi)$ so that $\text{nullity}(\mathbf{U}\phi) \geq \text{nullity}(\mathbf{U})$. Likewise, if γ is a basis for $\mathbf{N}(\mathbf{U}\phi)$, then $\phi(\gamma)$ is a set of linearly independent vectors in $\mathbf{W}$ and $\phi(\gamma) \subseteq \mathbf{N}(\mathbf{U})$ so that $\text{nullity}(\mathbf{U}) \geq \text{nullity}(\mathbf{U}\phi)$. Hence $\text{nullity}(\mathbf{U}) = \text{nullity}(\mathbf{U}\phi)$ and rewriting the above equations yields

$$\text{rank}(\mathbf{U}\phi) = \text{n} - \text{nullity}(\mathbf{U}\phi) = \text{n} - \text{nullity}(\mathbf{U}) = \text{rank}(\mathbf{U}).$$

Lemma 2. Let $\mathbf{V}$, $\mathbf{W}$, and $\mathbf{Z}$ be finite dimensional vector spaces, with $\mathbf{W}$ isomorphic to $\mathbf{Z}$, $\dim(\mathbf{V}) = n$, ϕ: $\mathbf{W} \rightarrow \mathbf{Z}$ an isomorphism, and $\mathbf{U}$: $\mathbf{V} \rightarrow \mathbf{W}$ a linear transformation, then $\text{rank}(\phi\mathbf{U}) = \text{rank}(\mathbf{U})$.

Proof. The Dimension Theorem gives

$$\text{nullity}(\phi\mathbf{U}) + \text{rank}(\phi\mathbf{U}) = \dim(\mathbf{V}) = \text{n}$$

$$\text{nullity}(\mathbf{U}) + \text{rank}(\mathbf{U}) = \dim(\mathbf{V}) = \text{n}.$$

Next suppose that $v \in \mathbf{N}(\mathbf{U})$, then

$$\phi\mathbf{U}(\text{v}) = \phi(\mathbf{U}(\text{v})) = \phi(\mathbf{0_W}) = \mathbf{0_Z}$$

so that $v \in \mathbf{N}(\phi\mathbf{U})$, $\mathbf{N}(\mathbf{U}) \subseteq \mathbf{N}(\phi\mathbf{U})$, and $\text{nullity}(\mathbf{U}) \leq \text{nullity}(\phi\mathbf{U})$.

Now suppose that $v \in \mathbf{N}(\phi\mathbf{U})$, then $\mathbf{0_Z} = \phi\mathbf{U}(\text{v}) = \phi(\mathbf{U}(\text{v}))$ so that $\mathbf{U}(\text{v}) \in \mathbf{N}(\phi)$; since ϕ is an isomorphism, $\mathbf{N}(\phi) = \{\mathbf{0_W}\}$ and thus $v \in \mathbf{N}(\mathbf{U})$.

Therefore $\mathbf{N}(\phi\mathbf{U}) \subseteq \mathbf{N}(\mathbf{U})$ and $\text{nullity}(\phi\mathbf{U}) \leq \text{nullity}(\mathbf{U})$. Combining the above inequalites gives $\text{nullity}(\phi\mathbf{U}) = \text{nullity}(\mathbf{U})$ and rewriting the equations from the Dimension Theorem yields

$$\text{rank}(\phi\mathbf{U}) = \text{n} - \text{nullity}(\phi\mathbf{U}) = \text{n} - \text{nullity}(\mathbf{U}) = \text{rank}(\mathbf{U}).$$

Theorem 3.3. Let $\mathbf{T}$: $\mathbf{V} \rightarrow \mathbf{W}$ be a linear transformation between finite-dimensional vector spaces, and let β and γ be ordered bases for $\mathbf{V}$ and $\mathbf{W}$ respectively. Then $\text{rank}(\mathbf{T}) = \text{rank}([\mathbf{T}]_\beta^\gamma)$.

Proof. Let $\dim(\mathbf{V}) = n$ and $\dim(\mathbf{W}) = m$ and consider the following diagram.

$$\begin{array}{ccc} \mathbf{V} & \longrightarrow^{\mathbf{T}}\longrightarrow & \mathbf{W} \\ \phi_\beta \downarrow & & \downarrow \phi_\gamma \\ \mathbf{F^n} & \longrightarrow^{\mathbf{L_A}}\longrightarrow & \mathbf{F^m} \end{array}$$

where $A = [\mathbf{T}]_\beta^\gamma$.

We know that $\mathbf{L}_\text{A}\phi_\beta,\ \phi_\gamma\mathbf{T} : \mathbf{V} \rightarrow \mathbf{F}^\text{m}$ and that $\mathbf{L}_\text{A}\phi_\beta = \phi_\gamma\mathbf{T}$ so that

$$\text{rank}(\mathbf{L}_\text{A}\phi_\beta) = \text{rank}(\phi_\gamma\mathbf{T}).$$

Since ϕ_β is an isomorphism, Lemma 1 and the definition of the rank of a matrix gives

$$\text{rank}(\mathbf{L}_\text{A}\phi_\beta) = \text{rank}(\mathbf{L}_\text{A}) = \text{rank}([\mathbf{T}]_\beta^\gamma).$$

Since ϕ_γ is an isomorphism, Lemma 2 gives

$$\text{rank}(\phi_\gamma\mathbf{T}) = \text{rank}(\mathbf{T}).$$

Combining these results gives

$$\text{rank}(\mathbf{T}) = \text{rank}(\phi_\gamma\mathbf{T}) = \text{rank}(\mathbf{L}_\text{A}\phi_\beta) = \text{rank}(\mathbf{L}_\text{A}) = \text{rank}([\mathbf{T}]_\beta^\gamma).$$

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, respectively, then
(a) $\text{rank}(AQ) = \text{rank}(A)$,
(b) $\text{rank}(PA) = \text{rank}(A)$,
and therefore,
(c) $\text{rank}(PAQ) = \text{rank}(A)$.

Proof. First observe that

$$\mathrm{R}(\mathbf{L}_{\mathrm{AQ}}) = \mathrm{R}(\mathbf{L}_{\mathrm{A}}\mathbf{L}_{\mathrm{Q}}) = \mathbf{L}_{\mathrm{A}}\mathbf{L}_{\mathrm{Q}}(\mathbf{F}^{\mathrm{n}}) = \mathbf{L}_{\mathrm{A}}(\mathbf{L}_{\mathrm{Q}}(\mathbf{F}^{\mathrm{n}})) = \mathbf{L}_{\mathrm{A}}(\mathbf{F}^{\mathrm{n}}) = \mathrm{R}(\mathbf{L}_{\mathrm{A}})$$

since $\mathbf{L}_{\mathrm{Q}}$ is onto. Therefore

$$\mathrm{rank(AQ)} = \dim(\mathrm{R}(\mathbf{L}_{\mathrm{AQ}})) = \dim(\mathrm{R}(\mathbf{L}_{\mathrm{A}})) = \mathrm{rank(A)}.$$

This establishes (a). To establish (b), apply exercise 17 of section 2.4 to $\mathbf{T} = \mathbf{L}_{\mathrm{P}}$ and obtain

$$\mathrm{rank(PA)} = \mathrm{rank}(\mathbf{L}_{\mathrm{PA}}) = \dim(\mathrm{R}(\mathbf{L}_{\mathrm{PA}})) = \dim(\mathbf{L}_{\mathrm{PA}}(\mathbf{F}^{\mathrm{n}}))$$

$$= \dim(\mathbf{L}_{\mathrm{P}}(\mathbf{L}_{\mathrm{A}}(\mathbf{F}^{\mathrm{n}}))) = \dim(\mathbf{L}_{\mathrm{A}}(\mathbf{F}^{\mathrm{n}})) = \dim(\mathrm{R}(\mathbf{L}_{\mathrm{A}})) = \mathrm{rank}(\mathbf{L}_{\mathrm{A}}) = \mathrm{rank(A)}.$$

Finally applying (a) and (b), we have

$$\mathrm{rank(PAQ)} = \mathrm{rank(PA)} = \mathrm{rank(A)}.$$

Corollary. Elementary row and column operations on a matrix are rank-preserving.

Proof. If B is obtained from a matrix A by an elementary row operation, then there exists an elementary matrix E such that $B = EA$. By Theorem 3.2, E is invertible and hence $\text{rank}(B) = \text{rank}(A)$ by Theorem 3.4. The proof that elementary column operations are rank preserving is left as an exercise.

Theorem 3.5. The rank of any matrix equals the maximum number of its linearly independent columns; that is, the rank of a matrix is the dimension of the subspace generated by its columns.

Proof. For any $A \in M_{n\times n}(F)$,

$$\mathrm{rank(A)} = \mathrm{rank}(\mathbf{L}_\mathrm{A}) = \mathrm{dim(R}(\mathbf{L}_\mathrm{a})).$$

Let β be the standard ordered basis for $\mathbf{F}^\mathrm{n}$. Then β spans $\mathbf{F}^\mathrm{n}$ and hence, by Theorem 2.2,

$$\mathrm{R}(\mathbf{L}_\mathrm{A}) = \mathrm{span}(\mathbf{L}_\mathrm{A}(\beta)) = \mathrm{span}(\{\mathbf{L}_\mathrm{A}(\mathrm{e_1}), \mathbf{L}_\mathrm{A}(\mathrm{e_2}), \ldots, \mathbf{L}_\mathrm{A}(\mathrm{e_n})\}).$$

But, for any j, we have seen in Theorem 2.13(b) that $\mathbf{L}_\mathrm{A}(\mathrm{e_j}) = \mathrm{Ae_j} = a_j$, where a_j is the jth column of A. Hence

$$\mathrm{R}(\mathbf{L}_\mathrm{A}) = \mathrm{span}(\{\mathrm{a_1}, \mathrm{a_2}, \ldots, \mathrm{a_n}\}).$$

Thus

$$\mathrm{rank(A)} = \mathrm{dim(R}(\mathbf{L}_\mathrm{A})) = \mathrm{dim(span}(\{\mathrm{a_1}, \mathrm{a_2}, \ldots, \mathrm{a_n}\})).$$

Theorem 3.6. Let A be an $m \times n$ matrix of rank r. Then $r \leq m,\ r \leq n$, and by means of a finite number of elementary row and column operations, A can be transformed into the matrix

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

where $O_1,\ O_2$, and O_3 are zero matrices. Thus $D_{ii} = 1$ for $i \leq r$ and $D_{ij} = 0$ otherwise.

Proof. See text.

Corollary 1. Let A be an $m \times n$ matrix of rank r. Then there exist invertible matrices B and C of sizes $m \times m$ and $n \times n$, respectively, such that $D = BAC$, where

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

is the $m \times n$ matrix in which $O_1,\ O_2$, and O_3 are zero matrices.

Proof. By Theorem 3.6, A can be transformed by a means of a finite number of elementary row and column operations into the matrix D. We can appeal to Theorem 3.1 each time we perform an elementary operation. Thus there exist elementary $m \times m$ matrices $E_1, E_2, \ldots, E_p$ and elementary $n \times n$ matrices $G_1, G_2, \ldots, G_q$ such that

$$D = E_p E_{p-1} \cdots E_2 E_1 A G_1 G_2 \cdots G_q.$$

By Theorem 3.2, each E_j and G_j is invertible. Let $B = E_p E_{p-1} \cdots E_2 E_1$ and $C = G_1 G_2 \cdots G_q$. Then B and C are invertible by exercise 4 of section 2.4, and $D = BAC$.

Corollary 2. Let A be an $m \times n$ matrix. Then
(a) $\text{rank}(A^t) = \text{rank}(A)$.
(b) The rank of any matrix equals the maximum number of its linearly independent rows; that is, the rank of a matrix is the dimension of the subspace generated by its rows.
(c) The rows and columns of any matrix generate subspaces of the same dimension, numerically equal to the rank of the matrix.

Proof. (a) By Corollary 1, there exist invertible matrices B and C such that $D = BAC$, where D satisfies the stated conditions of the corollary. Taking transposes, we have

$$D^t = (BAC)^t = C^t A^t B^t.$$

Since B and C are invertible, so are B^t and C^t by exercise 5 of section 2.4. Hence by Theorem 3.4,

$$\mathrm{rank(A^t) = rank(C^t A^t B^t) = rank(D^t)}.$$

Suppose that $r = \text{rank}(A)$. Then D^t is an $n \times m$ matrix with the form of the matrix D in Corollary 1, and hence $\text{rank}(D^t) = r$ by Theorem 3.5. Thus

$$\mathrm{rank(A^t) = rank(D^t) = r = rank(A)}.$$

This establishes (a).

The proofs of (b) and (c) are left as exercises.

Corollary 3. Every invertible matrix is a product of elementary matrices.

Proof. If A is an invertible $n \times n$ matrix, then $\text{rank}(A) = n$. Hence the matrix D in Corollary 1 equals I_n, and there exist invertible matrices B and C such that $I_n = BAC$.

As in the proof of Corollary 1, note that $B = E_p E_{p-1} \cdots E_1$ and $C = G_1 G_2 \cdots G_q$, where the E_i's and G_i's are elementary matrices. Thus $A = B^{-1} I_n C^{-1} = B^{-1} C^{-1}$, so that

$$A = E_1^{-1} E_2^{-1} \cdots E_p^{-1} G_q^{-1} G_{q-1}^{-1} \cdots G_1^{-1}.$$

The inverses of elementary matrices are elementary matrices, however, and hence A is the product of elementary matrices.

Theorem 3.7. Let $\mathbf{T}$: $\mathbf{V} \rightarrow \mathbf{W}$ and $\mathbf{U}$: $\mathbf{W} \rightarrow \mathbf{Z}$ be linear transformations on finite-dimensional vector spaces $\mathbf{V}$,$\mathbf{W}$, and $\mathbf{Z}$, and let A and B be matrices such that the product AB is defined. Then
(a) $\text{rank}(\mathbf{UT}) \leq \text{rank}(\mathbf{U})$.
(b) $\text{rank}(\mathbf{UT}) \leq \text{rank}(\mathbf{T})$.
(c) $\text{rank}(AB) \leq \text{rank}(A)$.
(c) $\text{rank}(AB) \leq \text{rank}(B)$.

Proof. We prove these items in the order: (a), (c), (d), and (b).
(a) Clearly, $\mathbf{R}(\mathbf{T}) \subseteq \mathbf{W}$. Hence

$$\mathbf{R}(\mathbf{UT}) = \mathbf{UT}(\mathbf{V}) = \mathbf{U}(\mathbf{T}(\mathbf{V})) = \mathbf{U}(\mathbf{R}(\mathbf{T})) \subseteq \mathbf{U}(\mathbf{W}) = \mathbf{R}(\mathbf{U}).$$

Thus

$$\text{rank}(\mathbf{UT}) = \dim(\text{R}(\mathbf{UT})) \leq \dim(\mathbf{R}(\mathbf{U})) = \text{rank}(\mathbf{U}).$$

(c) By (a),

$$\text{rank}(\text{AB}) = \text{rank}(\mathbf{L}_{\text{AB}}) = \text{rank}(\mathbf{L}_{\text{A}}\mathbf{L}_{\text{B}}) \leq \text{rank}(\mathbf{L}_{\text{A}}) = \text{rank}(\text{A}).$$

(d) By (c) and Corollary 2 to Theorem 3.6,

$$\text{rank}(\text{AB}) = \text{rank}((\text{AB})^{\text{t}}) = \text{rank}(\text{B}^{\text{t}}\text{A}^{\text{t}}) \leq \text{rank}(\text{B}^{\text{t}}) = \text{rank}(\text{B}).$$

(b) Let α, β, and γ be ordered bases for $\mathbf{V}$,$\mathbf{W}$, and $\mathbf{Z}$, respectively, and let $A' = [\mathbf{U}]_\beta^\gamma$ and $B' = [\mathbf{U}]_\alpha^\beta$. Then $A'B' = [\mathbf{UT}]_\alpha^\gamma$ by Theorem 2.11. Hence by Theorem 3.3 and (d),

$$\text{rank}(\mathbf{UT}) = \text{rank}(\text{A}'\text{B}') \leq \text{rank}(\text{B}') = \text{rank}(\mathbf{T}).$$

Theorem 3.8. Let $Ax = 0$ be a homogeneous system of m linear equations in n unknowns over a field $\mathbf{F}$. Let $\mathbf{K}$ denote the set of all solutions to $Ax = 0$. Then $\mathbf{K} = \mathbf{N}(\mathbf{L}_{\mathrm{A}})$; hence $\mathbf{K}$ is a subspace of $\mathbf{F}^{\mathrm{n}}$ of dimension $n-$ rank$(\mathbf{L}_{\mathrm{A}}) = n-$ rank(A).

Proof. Clearly, $\mathbf{K} = \{\mathrm{s} \in \mathbf{F}^{\mathrm{n}} : \mathrm{As} = 0\} = \mathbf{N}(\mathbf{L}_{\mathrm{A}})$. The second part now follows from the dimension theorem.

Corollary. If $m < n$, the system $Ax = 0$ has a nonzero solution.

Proof. Suppose that $m < n$. Then $\mathrm{rank}(A) = \mathrm{rank}(\mathbf{L}_{\mathrm{A}}) \leq m$. Hence,

$$\dim(\mathbf{K}) = \mathrm{n} - \mathrm{rank}(\mathbf{L}_{\mathrm{A}}) \geq \mathrm{n} - \mathrm{m} > 0,$$

where $\mathbf{K} = \mathbf{N}(\mathbf{L}_{\mathrm{A}})$. Since $\dim(\mathbf{K}) > 0$, $\mathbf{K} \neq \{\mathbf{0}\}$. Thus there exists a nonzero vector $s \in \mathbf{K}$; so s is a nonzero solution to $Ax = 0$.

Theorem 3.9. Let $\mathbf{K}$ be the solution set of a system of linear equations $Ax = b$, and let $\mathbf{K_H}$ be the solution set of the corresponding homogeneous system $Ax = \mathbf{0}$. Then for any solution s to $Ax = b$

$$\mathbf{K} = \mathrm{s} + \mathbf{K_H} = \{\mathrm{s} + \mathrm{k} : \mathrm{k} \in \mathbf{K_H}\}.$$

Proof. Let s be any solution to $Ax = b$. We must show that $\mathbf{K} = \{\mathrm{s}\} + \mathbf{K_H}$. If $w \in \mathbf{K}$, then $Aw = b$. Hence

$$A(w - s) = Aw - As = b - b = \mathbf{0}.$$

So $w - s \in \mathbf{K_H}$. Thus there exist $k \in \mathbf{K_H}$ such that $w - s = k$. It follows that $w = s + k \in \{s\} + \mathbf{K_H}$, and therefore

$$\mathbf{K} \subseteq \mathrm{s} + \mathbf{K_H}.$$

Conversely, suppose that $w \in \{s\} + \mathbf{K_H}$; then $w = s + k$ for some $k \in \mathbf{K_H}$. But then $Aw = A(s + k) = As + Ak = b + \mathbf{0} = \mathrm{b}$; so $w \in \mathbf{K}$. Therefore $s + \mathbf{K_H} \subseteq \mathbf{K}$, and thus $\mathbf{K} = \mathrm{s} + \mathbf{K_H}$.

Theorem 3.10. Let $Ax = b$ be a system of n linear equations in n unknowns. If A is invertible, then the system has exactly one solution, namely, $A^{-1}b$. Conversely, if the system has exactly one solution, then A is invertible.

Proof. Suppose A is invertible. Substituting $A^{-1}b$ into the system, we have $A(A^{-1}b) = (AA^{-1})b = b$. Thus $A^{-1}b$ is a solution. If s is an arbitrary solution then $As = b$. Multiplying both sides by A^{-1} gives $s = A^{-1}b$. Thus the system has one and only one solution, namely, $A^{-1}b$.

Conversely suppose that the system has exactly one solution s. Let $\mathbf{K_H}$ denote the solution set for the corresponding homogeneous system $Ax = \mathbf{0}$. By Theorem 3.9, $\{s\} = \{s\} + \mathbf{K_H}$. But this is so only if $\mathbf{K_H} = \mathbf{0}$. Thus $\mathbf{N}(\mathbf{L}_{\mathrm{A}}) = \{\mathbf{0}\}$, and hence A is invertible.

Theorem 3.11. Let $Ax = b$ be a system of linear equations. Then the system is consistent if and only if $\text{rank}(A) = \text{rank}(A|b)$.

Proof. To say that $Ax = b$ has a solution is equivalent to saying $b \in \mathbf{R}(\mathbf{L}_\mathrm{A})$. (Exercise 9) In the proof of Theorem 3.5, we saw that

$$\mathbf{R}(\mathbf{L}_\mathrm{A}) = \mathrm{span}(\{\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{n}\}),$$

the span of the columns of A. thus $Ax = b$ has a solution if and only if $b \in \text{span}(\{a_1, a_2, \ldots, a_n\})$. But $b \in \text{span}(\{a_1, a_2, \ldots, a_n\})$ if and only if $\text{span}(\{a_1, a_2, \ldots, a_n\}) = \text{span}(\{a_1, a_2, \ldots, a_n, b\})$. This last statement is equivalent to

$$\mathrm{dim}(\mathrm{span}(\{\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{n}\})) = \mathrm{dim}(\mathrm{span}(\{\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{n}, \mathrm{b}\})).$$

So by Theorem 3.5, the preceding equation reduces to

$$\mathrm{rank}(\mathrm{A}) = \mathrm{rank}(\mathrm{A}|\mathrm{b}).$$

Theorem 3.13. Let $Ax = b$ be a system of m linear equations in n unknowns, and let C be an invertible $m \times m$ matrix. Then the system $(CA)x = Cb$ is equivalent to $Ax = b$.

Proof. Let K be the solution set for $Ax = b$ and K' the solution set for $(CA)x = Cb$. If $w \in K$, then $Aw = b$. So $(CA)w = Cb$, and hence $w \in K'$. Thus $K \subseteq K'$.

Conversely, if $w \in K'$, then $(CA)w = Cb$. Hence

$$Aw = C^{-1}(CAw) = C^{-1}(Cb) = b;$$

so $w \in K$. Thus $K' \subseteq K$, and therefore, $K = K'$.

Corollary. Let $Ax = b$ be a system of m linear equations in n unknowns. If $(A'|b')$ is obtained from $(A|b)$ by a finite number of elementary row operations, then the system $A'x = b'$ is equivalent to the original system.

Proof. Suppose that $(A'|b')$ is obtained from $(A|b)$ by elementary row operations. These may be executed by multiplying $(A|b)$ by elementary $m \times m$ matrices $E_1, E_2, \ldots, E_p$. Let $C = E_p \cdots E_2 E_1$; then

$$(A'|b') = C(A|b) = (CA|Cb).$$

Since each E_i is invertible, so is C. Now $A' = CA$ and $b' = Cb$. Thus by Theorem 3.13, the system $A'x = b'$ is equivalent to the system $Ax = b$.

Theorem 3.14. Gaussian elimination transforms any matrix into its reduced row echelon form.

Theorem 3.15. Let $Ax = b$ be a system of r nonzero equations in n unknowns. Suppose that $\text{rank}(A) = \text{rank}(A|b)$ and that $(A|b)$ is in reduced echelon form. Then
(a) $\text{rank}(A) = r$.
(b) If the general solution obtained by the procedure above (Gaussian elimination) is of the form

$$s = s_0 + t_1 u_1 + t_2 u_2 + \cdots + t_{n-r} u_{n-r}$$

then $\{u_1, u_2, \ldots, u_{n-r}\}$ is a basis for the solution set of the corresponding homogeneous system, and s_0 is a solution to the original system.

Proof. Since $(A|b)$ is in reduced echelon form, $(A|b)$ must have r nonzero rows. Clearly these rows are linearly independent by the definition of the reduced row echelon form, and so $\text{rank}(A|b) = r$. Thus $\text{rank}(A) = r$.

Let K be the solution set for $Ax = b$, and let $\mathbf{K_H}$ be the solution set for $Ax = \mathbf{0}$. Setting $t_1 = t_2 = \cdots = t_{n-r} = 0$, we see that $s = s_0 \in K$. But by Theorem 3.9, $K = \{s_0\} + \mathbf{K_H}$. Hence

$$\mathbf{K_H} = -\{\mathrm{s_0}\} + \mathrm{K} = \mathrm{span}(\{\mathrm{u_1}, \mathrm{u_2}, \ldots, \mathrm{u_{n-r}}\}).$$

Because $\text{rank}(A) = r$, we have $\dim(\mathbf{K_H}) = n - r$. Thus since $\dim(\mathbf{K_H}) = n - r$ and $\mathbf{K_H}$ is generated by a set $\{u_1, u_2, \ldots, u_{n-r}\}$ containing at most $n - r$ vectors we conclude that this set is a basis for $\mathbf{K_H}$.

Theorem 3.16. Let Let A be an $m \times n$ matrix of rank r, where $r > 0$, and let B be the reduced row echelon form of A. Then
(a) The number of nonzero rows in B is r.
(b) For each $i = 1, 2, \ldots, r$, there is a column b_{j_i} of each B such that $b_{j_i} = e_i$.
(c) The columns of A numbered $j_1, j_2, \ldots, j_r$ are linearly independent.
(d) For each $k = 1, 2, \ldots, n$, if column k of B is $d_1e_1 + d_2e_2 + \cdots + d_re_r$, then column k of A is $d_1a_{j_1} + d_2a_{j_1} + \cdots + d_ra_{j_r}$.

Proof.
(a) Let A be a $m \times n$ matrix and let B be the reduced echelon form of A. If the rank of A is r, then the rank of B is also r by the corollary to Theorem 3.4. Because B is in reduced echelon form, no nonzero row of B can be a linear combination of the other rows of B. Hence B must have exactly r nonzero rows.

(b) Denote the columns of A by $a_1, a_2, \ldots, a_n$ and the columns of B by $b_1, b_2, \ldots, b_n$. Since B has r nonzero rows then for $r \geq 1$, the vectors $e_1, e_2, \ldots, e_r$ must occur among the columns of B. For $i = 1, 2, \ldots, r$, let j_i denote the column of B such that $b_{j_i} = e_i$.

(c) We claim the $a_{j_1}, a_{j_2}, \ldots, a_{j_r}$, the columns of A corresponding to these columns of B, are linearly independent. For suppose there are scalars $c_1, c_2, \ldots, c_r$ such that

$$c_1a_{j_1} + c_2a_{j_2} + \cdots + c_ra_{j_r} = \mathbf{0}.$$

Because B can be obtained from A by a sequence of elementary row operations, there exists an invertible $m \times m$ matrix M such that $MA = B$. Multiplying the preceding equation by M yields

$$c_1Ma_{j_1} + c_2Ma_{j_2} + \cdots + c_rMa_{j_r} = \mathbf{0}.$$

Since $Ma_{j_i} = b_{j_i} = e_i$, it follows that

$$c_1e_1 + c_2e_2 + \cdots c_re_r = \mathbf{0}.$$

Hence $c_1 = c_2 = \cdots = c_r = 0$, proving that the vectors $a_{j_1}, a_{j_2}, \ldots, a_{j_r}$ are linearly independent.

(d) Because B has only r nonzero rows, every column of B has the form

$$\begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

for scalars $d_1, d_2, \ldots, d_r$. the corresponding column of A must be

$$M^{-1}(d_1 e_1 + d_2 e_2 + \cdots + d_r e_r) = d_1 M^{-1} e_1 + d_2 M^{-1} e_2 + \cdots + d_r M^{-1} e_r$$

$$= d_1 M^{-1} b_{j_1} + d_2 M^{-1} b_{j_2} + \cdots + d_r M^{-1} b_{j_r} = d_1 a_{j_1} + d_2 a_{j_2} + \cdots + d_r a_{j_r}.$$

Corollary. The reduced row echelon form of a matrix is unique.

Theorem 4.1. The function det: $\mathbf{M}_{2\times2}(\mathrm{F}) \to \mathrm{F}$ is a linear function of each row of a 2×2 matrix when the other row is held fixed. That is, if u, v, and w are in $\mathbf{F}^2$ and k is a scalar, then

$$\det\begin{pmatrix} u+kv \\ w \end{pmatrix} = \det\begin{pmatrix} u \\ w \end{pmatrix} + \mathrm{k}\cdot\det\begin{pmatrix} v \\ w \end{pmatrix}$$

and

$$\det\begin{pmatrix} w \\ u+kv \end{pmatrix} = \det\begin{pmatrix} w \\ u \end{pmatrix} + \mathrm{k}\cdot\det\begin{pmatrix} w \\ v \end{pmatrix}.$$

Proof. Let $u = (a_1, a_2),\ v = (b_1, b_2)$, and $w = (c_1, c_2)$ be in $\mathbf{F}^2$ and k be a scalar. Then

$$\det\begin{pmatrix} u \\ w \end{pmatrix} + \mathrm{k}\cdot\det\begin{pmatrix} v \\ w \end{pmatrix} = \det\begin{pmatrix} a_1 & a_2 \\ c_1 & c_2 \end{pmatrix} + \mathrm{k}\cdot\det\begin{pmatrix} b_1 & b_2 \\ c_1 & c_2 \end{pmatrix}$$

$$= (a_1c_2 - a_2c_1) + k(b_1c_2 - b_2c_1) = (a_1 + kb_1)c_2 - (a_2 + kb_2)c_1$$

$$= \det\begin{pmatrix} a_1 + kb_1 & a_2 + kb_2 \\ c_1 & c_2 \end{pmatrix} = \det\begin{pmatrix} u+kv \\ w \end{pmatrix}.$$

A similar calculation shows that

$$\det\begin{pmatrix} w \\ u \end{pmatrix} + \mathrm{k}\cdot\det\begin{pmatrix} w \\ v \end{pmatrix} = \det\begin{pmatrix} w \\ u+kv \end{pmatrix}.$$

Theorem 4.2. Let $A \in \mathbf{M}_{2\times 2}(\mathrm{F})$. Then the determinant of A is nonzero if and only if A is invertible. Moreover, if A is invertible, then

$$\mathrm{A}^{-1} = (1/\det(\mathrm{A})) \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}.$$

Proof. If $\det(A) \neq 0$, then we can define a matrix

$$\mathrm{M} = (1/\det(\mathrm{A})) \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}.$$

A straightforward calculation shows that $AM = MA = I$, and so A is invertible and $M = A^{-1}$.

Conversely, suppose that A is invertible. A remark on page 152 shows that the rank of

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

must be 2. Hence $A_{11} \neq 0$ or $A_{21} \neq 0$. If $A_{11} \neq 0$, add $-A_{21}/A_{11}$ times row 1 of A to row 2 to obtain the matrix

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} - \frac{A_{12}A_{21}}{A_{11}} \end{pmatrix}.$$

Because elementary row operations are rank preserving by the corollary to Theorem 3.4, it follows that

$$A_{22} - \frac{A_{12}A_{21}}{A_{11}} \neq 0.$$

Therefore $\det(A) = A_{11}A_{22} - A_{12}A_{21} \neq 0$. On the other hand, if $A_{21} \neq 0$, we see that $\det(A) \neq 0$ by adding $-A_{11}/A_{21}$ times row 2 of A to row 1 and applying a similar argument. Thus in either case, $\det(A) \neq 0$.

Theorem 4.3. The determinant of an $n \times n$ matrix is a linear function of each row when the remaining rows are held fixed. That is, for $1 \leq r \leq n$, we have

$$\det\begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u+kv \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} = \det\begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} + \mathrm{k}\cdot\det\begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ v \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix}$$

whenever k is a scalar and u, v, and each a_i are row vectors in $\mathbf{F}^{\mathrm{n}}$.

Proof. The proof is by mathematical induction on n. The result is immediate if $n = 1$. Assume that for some integer $n \geq 2$ the determinant of any $(n-1) \times (n-1)$ matrix is a linear function of each row when the remaining rows are held fixed. Let A be an $n \times n$ matrix with rows $a_1, a_2, \ldots, a_n$ respectively, and suppose that for some r $(1 \leq r \leq n)$, we have $a_r = u + kv$ for some $u, v \in \mathbf{F}^{\mathrm{n}}$ and some scalar k. Let $u = (b_1, b_2, \ldots, b_n)$ and $v = (c_1, c_2, \ldots, c_n)$, and let B and C be the matrices obtained from A by replacing row r of A by u and v respectively. We must prove that $\det(A) = \det(B) + k\ \det(C)$. We leave the proof of this fact to the reader for the case $r = 1$. For $r > 1$ and $1 \leq j \leq n$, the rows of $\tilde{A}_{1j}$, $\tilde{B}_{1,j}$, and $\tilde{C}_{1j}$ are the same except for row $r-1$. Moreover, row $r-1$ of $\tilde{A}_{1j}$ is

$$(b_1 + kc_1, \ldots, b_{j-1} + kc_{j-1}, b_{j+1} + kc_{j+1}, \ldots, b_n + kc_n),$$

which is the sum of row $r-1$ of $\tilde{B}_{1j}$ and k times row $r-1$ of $\tilde{C}_{1j}$. Since $\tilde{B}_{1j}$ and $\tilde{C}_{1j}$ are $(n-1) \times (n-1)$ matrices, we have

$$\mathrm{det(\tilde{A}_{1j}) = det(\tilde{B}_{1j}) + k \cdot det(\tilde{C}_{1j})}$$

by the induction hypothesis. Thus since $A_{1j} = B_{1j} = C_{1j}$, we have

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

$$= \sum_{\mathrm{j=1}}^{\mathrm{n}} (-1)^{\mathrm{1+j}} \mathrm{A_{1j}} \cdot [\mathrm{det(\tilde{B}_{1j}) + k \cdot det(\tilde{C}_{1j})}]$$

$$= \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \cdot \det(\tilde{B}_{1j}) + k \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \cdot \det(\tilde{C}_{1j})$$

$$= \det(B) + k \cdot \det(C).$$

This shows that the theorem is true for $n \times n$ matrices, and so the theorem is true for all square matrices by mathematical induction.

Corollary. If $A \in \mathbf{M}_{n\times n}(F)$ has a row consisting entirely of zeros, then $\det(A) = 0$.

Proof. Exercise.

Lemma. Let $B \in \mathbf{M}_{\mathrm{n \times n}}(\mathrm{F})$, where $n \geq 2$. If row i of B equals e_k for some k $(1 \leq k \leq n)$, then $\det(B) = (-1)^{i+k} \det(\tilde{B}_{ik})$.

Proof. The proof is by mathematical induction on n. The lemma is easily proved for $n = 2$. Assume that for some integer $n \geq 3$, the lemma is true for $(n-1) \times (n-1)$ matrices and let B be a $n \times n$ matrix in which row i of B equals e_k for some k $(1 \leq k \leq n)$. The result follows immediately from the definition of the determinant if $i = 1$. Suppose therefore that $1 < i \leq n$. For each $j \neq k$ $(1 \leq j \leq n)$, let C_{ij} denote the $(n-2) \times (n-2)$ matrix obtained from B by deleting rows 1 and i and columns j and k. For each j, row $i - 1$ of $\tilde{B}_{1j}$ is the following vector in $\mathbf{F}^{\mathrm{n}-1}$:

$$\begin{cases} e_{k-1} & \text{if j < k} \\ 0 & \text{if j = k} \\ e_k & \text{if j > k.} \end{cases}$$

Hence by the induction hypothesis and the corollary to Theorem 4.3, we have

$$\det(\tilde{\mathrm{B}}_{1\mathrm{j}}) = \begin{cases} (-1)^{(\mathrm{i}-1)+(\mathrm{k}-1)}\det(\mathrm{C}_{\mathrm{ij}}) & \text{if j < k} \\ 0 & \text{if j = k} \\ (-1)^{(\mathrm{i}-1)+\mathrm{k}}\det(\mathrm{C}_{\mathrm{ij}}) & \text{if j > k.} \end{cases}$$

Therefore

$$\det(\mathrm{B}) = \sum_{\mathrm{j}=1}^{\mathrm{n}} (-1)^{1+\mathrm{j}}\mathrm{B}_{1\mathrm{j}} \cdot \det(\tilde{\mathrm{B}}_{1\mathrm{j}})$$

$$= \sum_{\mathrm{j}<\mathrm{k}} (-1)^{1+\mathrm{j}}\mathrm{B}_{1\mathrm{j}} \cdot \det(\tilde{\mathrm{B}}_{1\mathrm{j}}) + \sum_{\mathrm{j}>\mathrm{k}} (-1)^{1+\mathrm{j}}\mathrm{B}_{1\mathrm{j}} \cdot \det(\tilde{\mathrm{B}}_{1\mathrm{j}})$$

$$= \sum_{\mathrm{j}<\mathrm{k}} (-1)^{1+\mathrm{j}}\mathrm{B}_{1\mathrm{j}} \cdot [(-1)^{(\mathrm{i}-1)+(\mathrm{k}-1)}\det(\mathrm{C}_{\mathrm{ij}})] + \sum_{\mathrm{j}>\mathrm{k}} (-1)^{1+\mathrm{j}}\mathrm{B}_{1\mathrm{j}} \cdot [(-1)^{(\mathrm{i}-1)+\mathrm{k}}\det(\mathrm{C}_{\mathrm{ij}})]$$

$$= (-1)^{\mathrm{i}+\mathrm{k}} \Big[\sum_{\mathrm{j}<\mathrm{k}} (-1)^{1+\mathrm{j}}\mathrm{B}_{1\mathrm{j}} \cdot \det(\mathrm{C}_{\mathrm{ij}}) + \sum_{\mathrm{j}>\mathrm{k}} (-1)^{1+(\mathrm{j}-1)}\mathrm{B}_{1\mathrm{j}} \cdot \det(\mathrm{C}_{\mathrm{ij}}) \Big].$$

Because the expression inside the preceding bracket is the cofactor expansion of $\tilde{B}_{ik}$ along the first row, it follows that

$$\det(\mathrm{B}) = (-1)^{\mathrm{i}+\mathrm{k}}\det(\mathrm{B}_{\mathrm{ik}}).$$

This shows that the lemma is true for all $n \times n$ matrices, and so the lemma is true for all square matrices by induction.

Consider the matrix B below

$$B = \begin{pmatrix} B_{11} & B_{12} & B_{13} & B_{14} \\ B_{21} & B_{22} & B_{23} & B_{24} \\ 0 & 1 & 0 & 0 \\ B_{41} & B_{42} & B_{43} & B_{44} \end{pmatrix}$$

and note that $\hat{e}_2$ occurs in row three so that $k = 2$ and $i = 3$.

We define the matrix C_{ij} as the matrix obtained from B by deleting the 1st and ith rows
($i = 3$ in this example) and the jth and kth column where $j \neq k$ ($k = 2$ in this example).

Thus for the current example the allowed matrices are of the form C_{3j} where $j \in \{1, 3, 4\}$ Thus each of these matrices will result from B by deleting the 1st and 3rd rows and 2nd column.

$$B = \begin{pmatrix} * & B_{11} & B_{12} & B_{13} & B_{14} \\ & B_{21} & B_{22} & B_{23} & B_{24} \\ * & 0 & 1 & 0 & 0 \\ & B_{41} & B_{42} & B_{43} & B_{44} \\ & & * & & \end{pmatrix}.$$

For example, the matrix C_{31} is obtained from B by deleting the 1st and 3rd rows and 1st and 2nd columns.

$$C_{31} = \begin{pmatrix} B_{23} & B_{24} \\ B_{43} & B_{44} \end{pmatrix}.$$

The matrix C_{33} is obtained from B by deleting the 1st and 3rd rows and 2nd and 3rd columns.

$$C_{33} = \begin{pmatrix} B_{21} & B_{24} \\ B_{41} & B_{44} \end{pmatrix}.$$

The matrix C_{34} is obtained from B by deleting the 1st and 3rd rows and 2nd and 4th columns.

$$C_{34} = \begin{pmatrix} B_{21} & B_{23} \\ B_{41} & B_{43} \end{pmatrix}.$$

Now we consider the matrices $\tilde{B}_{1j}$ and list the four possibilities for the present example:

$$j=1, \quad 2=i-1, \quad \tilde{B}_{11}=\begin{pmatrix} B_{22} & B_{23} & B_{24} \\ 1 & 0 & 0 \\ B_{42} & B_{43} & B_{44} \end{pmatrix}$$

$$j=2, \quad 2=i-1, \quad \tilde{B}_{12}=\begin{pmatrix} B_{21} & B_{23} & B_{24} \\ 0 & 0 & 0 \\ B_{41} & B_{43} & B_{44} \end{pmatrix}$$

$$j=3, \quad 2=i-1, \quad \tilde{B}_{13}=\begin{pmatrix} B_{21} & B_{22} & B_{24} \\ 0 & 1 & 0 \\ B_{41} & B_{42} & B_{44} \end{pmatrix}$$

$$j=4, \quad 2=i-1, \quad \tilde{B}_{14}=\begin{pmatrix} B_{21} & B_{22} & B_{23} \\ 0 & 1 & 0 \\ B_{41} & B_{42} & B_{43} \end{pmatrix}.$$

Note that the $(i-1)$th row of $\tilde{B}_{1j}$ are given as follows:

$$j=1 \quad \rightarrow \quad \hat{e}_1$$

$$j=2 \quad \rightarrow \quad 0$$

$$j=3 \quad \rightarrow \quad \hat{e}_2$$

$$j=4 \quad \rightarrow \quad \hat{e}_2.$$

These results are summarized below:

$$\text{(i}-\text{1)th row of } \tilde{\mathrm{B}}_{\mathrm{1j}}=\begin{cases} \hat{e}_1 & j<k \quad (j<2) \\ 0_3 & j=k \quad (j=2) \\ \hat{e}_2 & j>k \quad (j>2) \end{cases}$$

Induction Hypothesis: Let $B \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$, where $n \geq 2$. If row s of B equals e_t for some k $(1 \leq t \leq n)$, then $\det(B) = (-1)^{s+t} \det(\tilde{B}_{st})$.

We use the induction hypothesis to calculate the determinants of the matrices $\tilde{B}_{1j}$; recall that for the original matrix B that $i = 3$ and $k = 2$:

$$\mathrm{i} = 3; \quad \mathrm{k} = 2; \quad \mathrm{j} = 1; \quad \mathrm{s} = 2; \quad \mathrm{t} = 1; \quad \tilde{\mathrm{B}}_{11} = \begin{pmatrix} B_{22} & B_{23} & B_{24} \\ 1 & 0 & 0 \\ B_{42} & B_{43} & B_{44} \end{pmatrix}$$

$$\det(\tilde{\mathrm{B}}_{11}) = (-1)^{2+1}\det\begin{pmatrix} B_{23} & B_{24} \\ B_{43} & B_{44} \end{pmatrix} = (-1)^{2+1}\det(\mathrm{C}_{31}) = (-1)^{(\mathrm{i}-1)+(\mathrm{k}-1)}\det(\mathrm{C}_{\mathrm{ij}})$$

$$\mathrm{i} = 3; \quad \mathrm{k} = 2; \quad \mathrm{j} = 2; \qquad \tilde{\mathrm{B}}_{12} = \begin{pmatrix} B_{21} & B_{23} & B_{24} \\ 0 & 0 & 0 \\ B_{41} & B_{43} & B_{44} \end{pmatrix}$$

$$\det(\tilde{\mathrm{B}}_{12}) = 0$$

$$\mathrm{i} = 3; \quad \mathrm{k} = 2; \quad \mathrm{j} = 3; \quad \mathrm{s} = 2; \quad \mathrm{t} = 2; \quad \tilde{\mathrm{B}}_{13} = \begin{pmatrix} B_{21} & B_{22} & B_{24} \\ 0 & 1 & 0 \\ B_{41} & B_{42} & B_{44} \end{pmatrix}$$

$$\det(\tilde{\mathrm{B}}_{13}) = (-1)^{2+2}\det\begin{pmatrix} B_{21} & B_{24} \\ B_{41} & B_{44} \end{pmatrix} = (-1)^{2+2}\det(\mathrm{C}_{33}) = (-1)^{(\mathrm{i}-1)+\mathrm{k}}\det(\mathrm{C}_{\mathrm{ij}})$$

$$\mathrm{i} = 3; \quad \mathrm{k} = 2; \quad \mathrm{j} = 4; \quad \mathrm{s} = 2; \quad \mathrm{t} = 2; \quad \tilde{\mathrm{B}}_{14} = \begin{pmatrix} B_{21} & B_{22} & B_{23} \\ 0 & 1 & 0 \\ B_{41} & B_{42} & B_{43} \end{pmatrix}$$

$$\det(\tilde{\mathrm{B}}_{14}) = (-1)^{2+2}\det\begin{pmatrix} B_{21} & B_{23} \\ B_{41} & B_{43} \end{pmatrix} = (-1)^{2+2}\det(\mathrm{C}_{34}) = (-1)^{(\mathrm{i}-1)+\mathrm{k}}\det(\mathrm{C}_{\mathrm{ij}}).$$

Summarizing these results:

$$\det(\tilde{\mathrm{B}}_{1\mathrm{j}}) = \begin{cases} (-1)^{(i-1)+(k-1)} det(C_{ij}) & j < k \quad (j < 2) \\ 0 & j = k \quad (j = 2) \\ (-1)^{(i-1)+k} det(C_{ij}) & j > k \quad (j > 2) \end{cases}.$$

As a final computation, we calculate a scalar multiple of the determinant of $\tilde{B}_{32}$ by performing a cofactor expansion along the first row; recall that

$$\tilde{B}_{32} = \begin{pmatrix} B_{11} & B_{13} & B_{14} \\ B_{21} & B_{23} & B_{24} \\ B_{41} & B_{43} & B_{44} \end{pmatrix}$$

so that

$$(-1)^{3+2}\det(\tilde{\mathrm{B}}_{32}) =$$

$$(-1)^{3+2}[(-1)^{1+1}\mathrm{B}_{11}\det\begin{pmatrix} B_{23} & B_{24} \\ B_{43} & B_{44} \end{pmatrix} + (-1)^{1+2}\mathrm{B}_{13}\det\begin{pmatrix} B_{21} & B_{24} \\ B_{41} & B_{44} \end{pmatrix} + (-1)^{1+3}\mathrm{B}_{14}\det\begin{pmatrix} B_{21} & B_{23} \\ B_{41} & B_{43} \end{pmatrix}]$$

$$= (-1)^{3+2}[(-1)^{1+1}\mathrm{B}_{11}\det(\mathrm{C}_{31}) + (-1)^{1+(3-1)}\mathrm{B}_{13}\det(\mathrm{C}_{33}) + (-1)^{1+(4-1)}\mathrm{B}_{14}\det(\mathrm{C}_{34})]$$

$$= (-1)^{3+2}[\sum_{\mathrm{j}<2}(-1)^{1+\mathrm{j}}\mathrm{B}_{1\mathrm{j}}\det(\mathrm{C}_{3\mathrm{j}}) + \sum_{\mathrm{j}>2}(-1)^{1+(\mathrm{j}-1)}\mathrm{B}_{1\mathrm{j}}\det(\mathrm{C}_{3\mathrm{j}})]$$

$$= \det(\mathrm{B}).$$

Theorem 4.4. The determinant of a square matrix can be evaluated by cofactor expansion along any row. That is, if $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$, then for any integer i $(1 \le i \le n)$,

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

Proof. Cofactor expansion along the first row of A gives the determinant of A by definition. So the result is true if $i = 1$. Fix $i > 1$. Row i of A can be written as $\sum_{j=1}^{n} A_{ij}e_j$. For $1 \le j \le n$, let B_j denote the matrix obtained from A by replacing row i of A by e_j. Then by Theorem 4.3 and the lemma, we have

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

Corollary. If $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$ has two identical rows, then $\det(A) = 0$.

Proof. The proof is by mathematical induction on n. We leave the proof of the result to the reader in the case $n = 2$. Assume that for some integer $n \ge 3$, it is true that for $(n-1) \times (n-1)$ matrices, and let rows r and s of $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$ be identical for $r \ne s$. Because $n \ge 3$, we can choose an integer i $(1 \le i \le n)$ other than r and s. Now

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

by Theorem 4.4. Since each $\tilde{A}_{ij}$ is an $(n-1) \times (n-1)$ matrix with two identical rows, the induction hypothesis implies that each $\det(\tilde{A}_{ij}) = 0$, and hence $\det(A) = 0$. This completes the proof for $n \times n$ matrices, and so the corollary is true for all square matrices by mathematical induction.

Theorem 4.5. If $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$ and B is a matrix obtained from A by interchanging any two rows of A, then $\det(B) = -\det(A)$.

Proof. Let the rows of $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$ be $a_1, a_2, \ldots, a_n$, and let B be the matrix obtained from A by interchanging rows r and s, where $r < s$. Thus

$$\mathrm{A} = \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix} \quad \text{and} \quad \mathrm{B} = \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix}.$$

Consider the matrix obtained from A by replacing rows r and s by $a_r + a_s$. By the corollary to Theorem 4.4 and Theorem 4.3, we have

$$0 = \det \begin{pmatrix} a_1 \\ \vdots \\ a_r + a_s \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix}$$

$$= \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix}$$

$$= 0 + \det(\mathrm{A}) + \det(\mathrm{B}) + 0.$$

Therefore $\det(B) = -\det(A)$.

Theorem 4.6. Let $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$, and let B be a matrix obtained by adding a multiple of one row of A to another row of A. Then $\det(B) = \det(A)$.

Proof. Suppose that B is the $n \times n$ matrix obtained from A by adding k times row r to row s, where $r \neq s$. Let the rows of A be $a_1, a_2, \ldots, a_n$, and the rows of B be $b_1, b_2, \ldots, b_n$. Then $b_i = a_i$ for $i \neq s$ and $b_s = a_s + ka_r$. Let C be the matrix obtained from A by replacing row s with a_r. Applying Theorem 4.3 to row s of B we obtain

$$\mathrm{det(B) = det(A) + k \cdot det(C) = det(A)}$$

because $\det(C) = 0$ by the corollary to Theorem 4.4.

Corollary. If $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$ has rank less than n, then $\det(A) = 0$.

Proof. If the rank of A is less than n, then the rows $a_1, a_2, \ldots, a_n$, of A are linearly dependent. By exercise 14 of section 1.5, some row of A, say row r, is a linear combination of the other rows. So there exists scalars c_i such that

$$a_r = c_1a_1 + \cdots + c_{r-1}a_{r-1} + c_{r+1}a_{r+1} + \cdots + c_na_n.$$

Let B the matrix obtained from A by adding $-c_i$ times row i to row r for each $i \neq r$. then row r of B consists entirely of zeros, and so $\det(B) = 0$. But by Theorem 4.6, $\det(B) = \det(A)$. Hence $\det(A) = 0$.

Theorem 4.7. For any $A, B \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$, $\det(AB) = \det(A)\cdot \det(B)$.

Proof. We begin by establishing the result when A is an elementary matrix. If A is an elementary matrix obtained by interchanging two rows of I, then $\det(A) = -1$. But by Theorem 3.1, AB is a matrix obtained by interchanging two row of B. Hence by Theorem 4.5, $\det(AB) = -\det(B) = \det(A)\cdot\det(B)$. Similar arguments establish the result when A is an elementary matrix of type 2 or 3.

If A is an $n \times n$ matrix with rank less than n, then $\det(A) = 0$ by the corollary to Theorem 4.6. Since $\mathrm{rank}(AB) \leq \mathrm{rank}(A) < n$ by Theorem 3.7, we have $\det(AB) = 0$. Thus $\det(AB) = \det(A)\cdot\det(B)$ in this case.

On the other hand, if A has rank n, then A is invertible and hence is a product of elementary matrices (Corollary 3 to Theorem 3.6), say, $A = E_m \cdots E_2E_1$. The first paragraph of this proof shows that

$$\mathrm{det(AB) = det(E_m \cdots E_2E_1B)}$$

$$\mathrm{= det(E_m) \cdot det(E_{m-1} \cdots E_2E_1B)}$$

$$\vdots$$

$$\mathrm{= det(E_m) \cdots det(E_2) \cdot det(E_1) \cdot det(B)}$$

$$\mathrm{= det(E_m \cdots E_2E_1) \cdot det(B)}$$

$$\mathrm{= det(A) \cdot det(B).}$$

Corollary. A matrix $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$ is invertible if and only if $\det(A) \neq 0$. Furthermore, if A is invertible, then $\det(A^{-1}) = 1/\det(A)$.

Proof. If $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$ is not invertible, then the rank of A is less than n. So $\det(A) = 0$ by the corollary to Theorem 4.6. On the other hand, if $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$ is invertible, then

$$\mathrm{det(A) \cdot det(A^{-1}) = det(AA^{-1}) = det(I) = 1}$$

by Theorem 4.7. Hence $\det(A) \neq 0$ and $\det(A^{-1}) = 1/\det(A)$.

Theorem 4.8. For any $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$, $\det(A^t) = \det(A)$.

Proof. If A is not invertible, then $\operatorname{rank}(A) < n$. But $\operatorname{rank}(A^t) = \operatorname{rank}(A)$ by Corollary 2 to Theorem 3.6, and so A^t is not invertible. Thus $\det(A^t) = 0 = \det(A)$ in this case.

On the other hand, if A is invertible, then A is the product of elementary matrices, say $A = E_m \dots E_2 E_1$. Since $\det(E_i) = \det(E_i^t)$ for every i by exercise 29 of section 4.2, by Theorem 4.7 we have

$$\det(\mathrm{A^t}) = \det(\mathrm{E_1^t E_2^t \cdots E_m^t}))$$

$$= \det(\mathrm{E_1^t}) \cdot \det(\mathrm{E_2^t}) \cdot \dots \cdot \det(\mathrm{E_m^t})$$

$$= \det(\mathrm{E_1}) \cdot \det(\mathrm{E_2}) \cdot \dots \cdot \det(\mathrm{E_m})$$

$$= \det(\mathrm{E_m}) \cdot \dots \cdot \det(\mathrm{E_2}) \cdot \det(\mathrm{E_1})$$

$$= \det(\mathrm{E_m \cdots E_2 E_1})$$

$$= \det(\mathrm{A}).$$

Thus in either case, $\det(A^t) = \det(A)$.

Theorem 4.9 (Cramer's Rule). Let $Ax = b$ be the matrix form of a system of n linear equations in n unknows, where $x = (x_1, x_2, \ldots, x_n)^t$. If $\det(A) \neq 0$, then this system has a unique solution, and for each k $(k = 1, 2, \ldots, n)$,

$$x_k = \mathrm{det(M_k)/det(A)}$$

where M_k is the $n \times n$ matrix obtained from A by replacing column k of A by b.

Proof. If $\det(A) \neq 0$, then the system $Ax = b$ has a unique solution by the corollary to Theorem 4.7 and Theorem 3.10. For each integer k $(1 \leq k \leq n)$, let a_k denote the kth column of A and X_k denote the matrix obtained from the $n \times n$ identity matrix by replacing column k by x. Then by Theorem 2.13, AX_k is the $n \times n$ matrix whose ith column is

$$\mathrm{Ae_i \quad if \quad i \neq k \quad and \quad Ax = b \quad if \quad i = k.}$$

Thus $AX_k = M_k$. Evaluating X_k by cofactor expansion along row k produces

$$\mathrm{det(X_k) = x_k \cdot det(I_{n-1}) = x_k.}$$

Hence by Theorem 4.7,

$$\mathrm{det(M_k) = det(AX_k) = det(A) \cdot det(X_k) = det(A) \cdot x_k.}$$

Therefore

$$\mathrm{x_k = [det(A)]^{-1} \cdot det(M_k).}$$

Theorem 5.1. A linear operator $\mathbf{T}$ on a finite-dimensional vector space $\mathbf{V}$ is diagonalizable if and only if there exists an ordered basis β for $\mathbf{V}$ consisting of eigenvectors of $\mathbf{T}$. Furthermore, if $\mathbf{T}$ is diagonalizable, $\beta = \{v_1, v_2, \ldots, v_n\}$ is an ordered basis of eigenvectors of $\mathbf{T}$, and $D = [\mathbf{T}]_\beta$, then D is a diagonal matrix and D_{jj} is the eigenvalue corresponding to v_j for $1 \leq j \leq n$.

Proof. Let $\mathbf{T}$ be a linear operator on a finite-dimensional vector space $\mathbf{V}$.

First suppose that $\mathbf{T}$ is diagonalizable. By definition, there exists an ordered basis $\beta = \{v_1, v_2, \ldots, v_n\}$ such that $D = [\mathbf{T}]_\beta$ is diagonal. For any $v_j \in \beta$, the image of v_j under $\mathbf{T}$ is given by

$$\mathbf{T}(\mathrm{v_j}) = \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{D_{ij} v_i} = \mathrm{D_{jj} v_j} = \lambda_{\mathrm{j}} \mathrm{v_j}$$

where $\lambda_j = D_{jj}$. Hence by definition, v_j is an eigenvector of $\mathbf{T}$ and λ_j is the corresponding eigenvalue; β therefore is an ordered basis consisting of eigenvectors of $\mathbf{T}$ and the diagonal entries D_{jj} of the diagonal matrix $D = [\mathbf{T}]_\beta$ are the eigenvalues corresponding to v_j.

Next suppose that there exists an ordered basis $\beta = \{v_1, v_2, \ldots, v_n\}$ for $\mathbf{V}$ consisting of eigenvectors of $\mathbf{T}$. Then by definition there exists scalars $\lambda_1, \lambda_2, \ldots, \lambda_n$ (the eigenvalues for each of the eigenvectors) such that $\mathbf{T}(\mathrm{v_1}) = \lambda_1 \mathrm{v_1}, \mathbf{T}(\mathrm{v_2}) = \lambda_2 \mathrm{v_2}, \ldots, \mathbf{T}(\mathrm{v_n}) = \lambda_{\mathrm{n}} \mathrm{v_n}$. Thus, constructing $[\mathbf{T}]_\beta$ we find

$$[\mathbf{T}]_\beta = \mathrm{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

so that by definition $\mathbf{T}$ is diagonalizable and the diagonal entries D_{jj} of the diagonal matrix $D = [\mathbf{T}]_\beta$ are the eigenvalues corresponding to v_j.

Theorem 5.2. Let $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$. Then a scalar λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Proof. A scalar λ is an eigenvalue of A if and only if there exists a nonzero vector $v \in \mathbf{F}^{\mathrm{n}}$ such that $Av = \lambda v$, that is, $(A - \lambda I_n)(v) = \mathbf{0}$. By Theorem 2.5, this is true if and only if $A - \lambda I_n$ is not invertible. However, this result is equivalent to the statement that $\det(A - \lambda I_n) = 0$.

Lemma. If $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathbf{P}_1(\mathrm{R}))$, then $\det(A) \in \mathbf{P}_{\mathrm{n}}(\mathrm{R})$.

Proof. The lemma will be established using induction on n. For the case $n = 1$, $A = (a_{11})$ with $a_{11} \in \mathbf{P}_1(\mathrm{R})$; thus $\det(A) = a_{11} \in \mathbf{P}_1(\mathrm{R})$.

Next assume the result holds for $n = k$ and consider $n = k + 1$. In this case, using a cofactor expansion along the first row of A, $\det(A)$ can be written as

$$\det(\mathrm{A}) = \sum_{\mathrm{j=1}}^{\mathrm{k+1}} (-1)^{\mathrm{1+j}}\ \mathrm{a_{1j}} \cdot \det(\tilde{\mathrm{A}}_{\mathrm{1j}})$$

where $\tilde{A}_{1j}$ is the matrix obtained by deleting the 1st row and jth column from A so that $\tilde{A}_{1j} \in \mathbf{M}_{\mathrm{k\times k}}(\mathbf{P}_1(\mathrm{R}))$. Using the induction hypothesis, it is clear that each term in the cofactor expansion is the product of an element in $\mathbf{P}_1(\mathrm{R})$ and an element in $\mathbf{P}_{\mathrm{k}}(\mathrm{R})$ so that expansion terms are elements in $\mathbf{P}_{\mathrm{k+1}}(\mathrm{R})$; since $\mathbf{P}_{\mathrm{k+1}}(\mathrm{R})$ is a vector space and is closed under addition, $\det(A) \in \mathbf{P}_{\mathrm{k+1}}(\mathrm{R})$ so that the lemma is established by induction.

Theorem 5.3. Let $A \in \mathbf{M}_{\mathrm{n\times n}}(\mathrm{F})$.
(a) The characteristic polynomial of A is a polynomial of degree n with leading coefficient $(-1)^n$.
(b) A has at most n distinct eigenvalues.

Proof. (a) The first assertion will be proved using induction on n. For $n = 1$, the matrix A is a scalar, a_{11}, so that the characteristic polynomial is given by

$$f(t) = \det(\mathrm{A} - \mathrm{tI_1}) = \mathrm{a_{11}} - \mathrm{t} = (-1)\mathrm{t} + \mathrm{a_{11}}$$

which is a polynomial of degree one with a leading coefficient of $(-1) = (-1)^1$. Thus the assertion is true for this first case.

Next assume the assertion holds for $n = k - 1$ and consider the case $n = k$. For this value of n the characteristic polynomial is given by

$$f(t) = \det(\mathrm{A} - \mathrm{tI_k}) = \det \begin{pmatrix} a_{11} - t & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} - t & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} - t \end{pmatrix}$$

$$= (\mathrm{a_{11}}-\mathrm{t})\cdot\det([\mathrm{A}-\mathrm{tI_k}]_{11})-\mathrm{a_{12}}\cdot\det([\mathrm{A}-\mathrm{tI_k}]_{12})+\cdots+(-1)^{1+\mathrm{k}}\,\mathrm{a_{1k}}\cdot\det([\mathrm{A}-\mathrm{tI_k}]_{1\mathrm{k}})$$

where a cofactor expansion was used along the first row and $[A - tI_k]_{1j}$ is the matrix obtained by deleting the 1st row and jth column from $(A - tI_k)$. Since $[A - tI_k]_{1j} \in \mathbf{M}_{\mathrm{k-1\times k-1}}(\mathbf{P}_1(\mathrm{R}))$, the lemma provides
$\det([A - tI_k]_{1j}) \in \mathbf{P}_{\mathrm{k-1}}(\mathrm{R})$. The matrix $[A - tI_k]_{11}$ has a special form given by

$$\begin{pmatrix} a_{22}-t & a_{23} & \cdots & a_{2k} \\ a_{32} & a_{33}-t & \cdots & a_{3k} \\ \vdots & \vdots & & \vdots \\ a_{k2} & a_{k3} & \cdots & a_{kk}-t \end{pmatrix}$$

from which it is deduced that $[A - tI_k]_{11} = \tilde{A}_{11} - tI_{k-1}$. The induction hypothesis is used for this matrix with the result that
$\det([A - tI_k]_{11}) = \det(\tilde{A}_{11} - tI_{k-1})$ is a polynomial of degree $k - 1$ with a leading coefficient of $(-1)^{k-1}$.

Referring to the cofactor expansion of $\det(A - tI_k)$, the above results suggest that for $2 \le j \le k$, $(-1)^{1+j}\ a_{1j}\cdot\det(\ [A - tI_k]_{1j}) \in \mathbf{P}_{\mathrm{k-1}}(\mathrm{R})$ while $(a_{11} - t)\cdot\det([A - tI_k]_{11})$ is a kth degree polynomial with a leading coefficient of $(-1)^{k-1} \cdot (-1) = (-1)^k$. Summing the elements in the cofactor expansion establishes that the characteristic polynomial of A is a polynomial of degree k with leading coefficient $(-1)^k$. The hypothesis has thus been established for $n = k$ so that the result is proved by induction.

(b) This result follows directly from Theorem 5.2 and Corollary 2 of Theorem E.1 (pg 564).

Theorem 5.4. Let $\mathbf{T}$ be a linear operator on a vector space $\mathbf{V}$, and let λ be an eigenvalue of $\mathbf{T}$. A vector $v \in \mathbf{V}$ is an eigenvector of $\mathbf{T}$ corresponding to λ if and only if $v \neq \mathbf{0}$ and $v \in \mathbf{N}(\mathbf{T} - \lambda \mathrm{I})$.

Proof. Suppose first that $v \in \mathbf{V}$ is an eigenvector of $\mathbf{T}$ corresponding to λ. By definition, $v \neq \mathbf{0}$ and $\mathbf{T}(\mathrm{v}) = \lambda \mathrm{v}$. The last equation can be rewritten as

$$\mathbf{0} = \mathbf{T}(\mathrm{v}) - \lambda \mathrm{v} = \mathbf{T}(\mathrm{v}) - \lambda \mathbf{I}(\mathrm{v}) = (\mathbf{T} - \lambda \mathbf{I})(\mathrm{v})$$

so that $v \in \mathbf{N}(\mathbf{T} - \lambda \mathrm{I})$ by definition.

Next suppose that $v \in \mathbf{V}$, $v \neq \mathbf{0}$, and $v \in \mathbf{N}(\mathbf{T} - \lambda \mathrm{I})$. Reversing the above argument it is clear that $v \in \mathbf{N}(\mathbf{T} - \lambda \mathrm{I})$ implies that $\mathbf{T}(\mathrm{v}) = \lambda \mathrm{v}$, so by definition v is an eigenvector of $\mathbf{T}$ corresponding to λ.

Theorem 5.5. Let $\mathbf{T}$ be a linear operator on a vector space $\mathbf{V}$, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of $\mathbf{T}$. If $v_1, v_2, \ldots, v_k$ are eigenvectors of $\mathbf{T}$ such that λ_i corresponds to v_i $(1 \leq i \leq k)$, then $\{v_1, v_2, \ldots, v_k\}$ is linearly independent.

Proof. The proof is on mathematical induction on k. Suppose that $k = 1$. Then $v_1 \neq \mathbf{0}$ since v_1 is an eigenvector, and hence $\{v_1\}$ is linearly independent. Now assume that the theorem holds for $k-1$ distinct eigenvalues, where $k - 1 \geq 1$, and that we have k eigenvectors $v_1, v_2, \ldots, v_k$ corresponding to the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. We wish to show that $\{v_1, v_2, \ldots, v_k\}$ is linearly independent. Suppose that $a_1, a_2, \ldots, a_k$ are scalars such that

$$a_1v_1 + a_2v_2 + \cdots + a_kv_k = \mathbf{0}. \qquad (1)$$

Applying $\mathbf{T} - \lambda_{\mathbf{k}}\mathbf{I}$ to both sides of the equation, we obtain

$$(a_1)(\lambda_1 - \lambda_k)v_1 + (a_2)(\lambda_2 - \lambda_k)v_2 + \cdots + (a_{k-1})(\lambda_{k-1} - \lambda_k)v_{k-1} = \mathbf{0}.$$

By the induction hypothesis $\{v_1, v_2, \ldots, v_{k-1}\}$ is linearly independent, and hence

$$(a_1)(\lambda_1 - \lambda_k) = (a_2)(\lambda_2 - \lambda_k) = \cdots = (a_{k-1})(\lambda_{k-1} - \lambda_k) = 0.$$

Since $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct, it follows that $\lambda_i - \lambda_k \neq 0$ for $1 \leq i \leq k-1$. So $a_1 = a_2 = \cdots = a_{k-1} = 0$, and (1) therefore reduces to $a_kv_k = \mathbf{0}$. But $v_k \neq \mathbf{0}$, and therefore $a_k = 0$. Consequently $a_1 = a_2 = \cdots = a_k = 0$, and it follows that $\{v_1, v_2, \ldots, v_k\}$ is linearly independent.

Corollary. Let $\mathbf{T}$ be a linear operator on an n-dimensional vector space $\mathbf{V}$. If $\mathbf{T}$ has n distinct eigenvalues, then $\mathbf{T}$ is diagonalizable.

Proof. Suppose that $\mathbf{T}$ has n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. For each i choose an eigenvector v_i corresponding to λ_i. by Theorem 5.5, $\{v_1, v_2, \ldots, v_n\}$ is linearly independent, and since $\dim(\mathbf{V}) = n$, this set is a basis for $\mathbf{V}$. Thus by Theorem 5.1, $\mathbf{T}$ is diagonalizable.

Theorem 5.6. The characteristic polynomial of any diagonalizable linear operator splits.

Proof. Let **T** be a diagonalizable linear operator on the n-dimensional vector space **V**, and let β be an ordered basis for **V** such that $[\mathbf{T}]_\beta = \mathrm{D}$ is a diagonal matrix. Suppose that

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

and let $f(t)$ be the characteristic polynomial of **T**. Then

$$f(t) = \det(\mathrm{D} - \mathrm{tI}) = \det \begin{pmatrix} \lambda_1 - t & 0 & \cdots & 0 \\ 0 & \lambda_2 - t & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n - t \end{pmatrix}$$

$$= (\lambda_1 - t)(\lambda_2 - t)\cdots(\lambda_n - t) = (-1)^n (t - \lambda_1)(t - \lambda_2)\cdots(t - \lambda_n).$$

Theorem 5.7. Let $\mathbf{T}$ be a linear operator on a finite-dimensional vector space $\mathbf{V}$, and let λ be an eigenvalue of $\mathbf{T}$ having multiplicity m. Then $1 \leq \dim(\mathbf{E}_\lambda) \leq m$.

Proof. Choose an ordered basis $\{v_1, v_2, \ldots, v_p\}$ for $\mathbf{E}_\lambda$, extend it to an ordered basis $\beta = \{v_1, v_2, \ldots, v_p, v_{p+1}, \ldots, v_n\}$ for $\mathbf{V}$, and let $A = [\mathbf{T}]_\beta$. Observe that v_i $(1 \leq i \leq p)$ is an eigenvector of $\mathbf{T}$ corresponding to λ, and therefore

$$A = \begin{pmatrix} \lambda I_p & B \\ O & C \end{pmatrix}.$$

By exercise 21 of section 4.3, the characteristic polynomial of $\mathbf{T}$ is

$$f(t) = \det(\mathrm{A} - \mathrm{tI}) = \det \begin{pmatrix} (\lambda - t)I_p & B \\ O & C - tI_{n-p} \end{pmatrix}$$

$$= \det((\lambda - \mathrm{t})\mathrm{I_p})\det(\mathrm{C} - \mathrm{tI_{n-p}})$$

$$(\lambda - t)^p g(t),$$

where $g(t)$ is a polynomial. Thus $(\lambda - t)^p$ is a factor of $f(t)$, and hence the multiplicity of λ is at least p. But $\dim(\mathbf{E}_\lambda) = p$, and so $\dim(\mathbf{E}_\lambda) \leq m$.

Lemma. Let $\mathbf{T}$ be a linear operator, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of $\mathbf{T}$. For each $i = 1, 2, \ldots, k$, let $v_i \in \mathbf{E}_{\lambda_i}$, the eigenspace corresponding to λ_i. If

$$v_1 + v_1 + \cdots + v_k = \mathbf{0},$$

then $v_i = \mathbf{0}$ for all i.

Proof. Suppose otherwise. By renumbering if necessary, suppose that, for $1 \leq m \leq k$, we have $v_i \neq \mathbf{0}$ for $1 \leq i \leq m$, and $v_i = \mathbf{0}$ for $i > m$. Then for each $i \leq m$, v_i is an eigenvector of $\mathbf{T}$ corresponding to λ_i and

$$v_1 + v_1 + \cdots + v_m = \mathbf{0}.$$

But this contradicts Theorem 5.5, which states that these v_i's are linearly independent. We conclude, therefore, that $v_i = \mathbf{0}$ for all i.

Theorem 5.8. Let $\mathbf{T}$ be a linear operator on a vector space $\mathbf{V}$, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be distinct eigenvalues of $\mathbf{T}$. For each $i = 1, 2, \ldots, k$, let S_i be a finite linearly independent subset of the eigenspace $\mathbf{E}_{\lambda_i}$. Then
$S = S_1 \cup S_2 \cup \cdots \cup S_k$ is a linearly independent subset of $\mathbf{V}$.

Proof. Suppose that for each i

$$S_i = \{v_{i1}, v_{i2}, \ldots, v_{in_i}\}.$$

Then $S = \{v_{ij} : 1 \leq j \leq n_i, \text{ and } 1 \leq \text{i} \leq \text{k}\}$. Consider any scalars $\{a_{ij}\}$ such that

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} a_{ij} v_{ij} = \mathbf{0}.$$

For each i, let

$$w_i = \sum_{j=1}^{n_i} a_{ij} v_{ij}.$$

Then $w_i \in \mathbf{E}_{\lambda_i}$ for each i, and $w_1 + w_2 + \cdots + w_k = \mathbf{0}$. Therefore by the lemma, $w_i = \mathbf{0}$ for all i. But each S_i is linearly independent, and hence $a_{ij} = 0$ for all j. We conclude that S is linearly independent.

Theorem 5.9. Let $\mathbf{T}$ be a linear operator on a finite-dimensional vector space $\mathbf{V}$ such that the characteristic polynomial of $\mathbf{T}$ splits. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the distinct eigenvalues of $\mathbf{T}$. Then
(a) $\mathbf{T}$ is diagonalizable if and only if the multiplicity of λ_i is equal to $\dim(\mathbf{E}_{\lambda_i})$ for all i.
(b) If $\mathbf{T}$ is diagonalizable and β_i is an ordered basis for $\mathbf{E}_{\lambda_i}$ for each i, then $\beta = \beta_1 \cup \beta_1 \cup \cdots \cup \beta_k$ is an ordered basis for $\mathbf{V}$ consisting of the eigenvectors of $\mathbf{T}$.

Proof. For each i, let m_i denote the multiplicity of λ_i, $d_i = \dim(\mathbf{E}_{\lambda_i})$, and $n = \dim(\mathbf{V})$.

First suppose that $\mathbf{T}$ is diagonalizable. Let β be a basis for $\mathbf{V}$ consisting of eigenvectors of $\mathbf{T}$. For each i, let $\beta_i = \beta \cap \mathbf{E}_{\lambda_i}$, the set of vectors in β that are eigenvectors corresponding to λ_i, and let n_i denote the number of vectors in β_i. Then $n_i \leq d_i$ for each i because β_i is a linearly independent subset of a subspace of dimension d_i and $d_i \leq m_i$ by Theorem 5.7. The n_i's sum to n because β contains n vectors. The m_i's also sum to n because the degree of the characteristic polynomial of $\mathbf{T}$ is equal to the sum of the multiplicities of the eigenvalues. Thus

$$n = \sum_{i=1}^{k} n_i \leq \sum_{i=1}^{k} d_i \leq \sum_{i=1}^{k} m_i = n.$$

It follows that

$$\sum_{i=1}^{k} (m_i - d_i) = 0.$$

Since $(m_i - d_i) \geq 0$ for all i, we conclude that $m_i = d_i$ for all i.

Conversely suppose that $m_i = d_i$ for all i. We simultaneously show that $\mathbf{T}$ is diagonalizable and prove (b). For each i, let β_i be an ordered basis for $\mathbf{E}_{\lambda_i}$, and let $\beta = \beta_1 \cup \beta_1 \cup \cdots \cup \beta_k$. By Theorem 5.8, β is linearly independent. Furthermore, since $m_i = d_i$ for all i, β contains

$$\sum_{i=1}^{k} d_i = \sum_{i=1}^{k} m_i = n.$$

vectors. Therefore β is an ordered basis for $\mathbf{V}$ consisting of eigenvectors of $\mathbf{V}$, and we conclude that $\mathbf{T}$ is diagonalizable.

Theorem 5.21. Let $\mathbf{T}$ be a linear operator on a finite-dimensional vector space $\mathbf{V}$, and let $\mathbf{W}$ be a $\mathbf{T}$-invariant subspace of $\mathbf{V}$. Then the characteristic polynomial of $\mathbf{T_w}$ divides the characteristic polynomial of $\mathbf{T}$.

Proof. Choose an ordered basis $\gamma = \{v_1, v_2, \ldots, v_k\}$ for $\mathbf{W}$, and extend it to an ordered basis $\beta = \{v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ for $\mathbf{V}$. Let $A = [\mathbf{T}]_\beta$ and $B_1 = [\mathbf{T_W}]_\gamma$. Then by exercise 12, A can be written in the form

$$A = \begin{pmatrix} B_1 & B_2 \\ O & B_3 \end{pmatrix}.$$

Let $f(t)$ be the characteristic polynomial of $\mathbf{T}$ and $g(t)$ the characteristic polynomial of $\mathbf{T_W}$. Then

$$f(t) = \mathrm{det(A - tI_n)} = \det \begin{pmatrix} B_1 - tI_k & B_2 \\ O & B_3 - tI_{n-k} \end{pmatrix} = \mathrm{g(t) \cdot det(B_3 - tI_{n-k})}$$

by exercise 21 of section 4.3. Thus $g(t)$ divides $f(t)$.

Theorem 5.22. Let $\mathbf{T}$ be a linear operator on a finite-dimensional vector space $\mathbf{V}$, and let $\mathbf{W}$ denote the $\mathbf{T}$-cyclic subspace of $\mathbf{V}$ generated by a nonzero vector $v \in \mathbf{V}$. Let $k = \dim(\mathbf{W})$. Then
(a) $\{\mathrm{v}, \mathbf{T}(\mathrm{v}), \mathbf{T}^2(\mathrm{v}), \ldots, \mathbf{T}^{\mathrm{k}-1}(\mathrm{v})\}$ is a basis for $\mathbf{W}$.
(b) If $a_0 v + a_1 \mathbf{T}(\mathrm{v}) + \cdots + \mathrm{a}_{\mathrm{k}-1}\mathbf{T}^{\mathrm{k}-1}(\mathrm{v}) + \mathbf{T}^{\mathrm{k}}(\mathrm{v}) = \mathbf{0}$, then the characteristic polynomial of $\mathbf{T_w}$ is $f(t) = (-1)^k(a_0 + a_1 t + \cdots + a_{k-1}t^{k-1} + t^k)$.

Proof. (a) Since $v \neq \mathbf{0}$, the set $\{v\}$ is linearly independent. Let j be the largest positive integer for which

$$\beta = \{\mathrm{v}, \mathbf{T}(\mathrm{v}), \mathbf{T}^2(\mathrm{v}), \ldots, \mathbf{T}^{\mathrm{j}-1}(\mathrm{v})\}$$

is linearly independent. Such a j must exist because $\mathbf{V}$ is finite-dimensional. Let $\mathbf{Z} = \mathrm{span}(\beta)$. Then β is a basis for $\mathbf{Z}$. Furthermore, $\mathbf{T}^{\mathrm{j}}(\mathrm{v}) \in \mathbf{Z}$ by Theorem 1.7. We use this information to show that $\mathbf{Z}$ is a $\mathbf{T}$ invariant subspace of $\mathbf{V}$. Let $w \in \mathbf{Z}$. Since w is a linear combination of the vectors of β, there exists scalars $b_0, b_1, \ldots, b_{j-1}$ such that

$$w = \mathrm{b}_0\mathrm{v} + \mathrm{b}_1\mathbf{T}(\mathrm{v}) + \cdots + \mathrm{b}_{\mathrm{j}-1}\mathbf{T}^{\mathrm{j}-1}(\mathrm{v}),$$

and hence

$$\mathbf{T}(\mathrm{w}) = \mathrm{b}_0\mathbf{T}(\mathrm{v}) + \mathrm{b}_1\mathbf{T}^2(\mathrm{v}) + \cdots + \mathrm{b}_{\mathrm{j}-1}\mathbf{T}^{\mathrm{j}}(\mathrm{v}).$$

Thus $\mathbf{T}(w)$ is a linear combination of vectors in $\mathbf{Z}$, and hence belongs to $\mathbf{Z}$. So $\mathbf{Z}$ is $\mathbf{T}$-invariant. Furthermore, $v \in \mathbf{Z}$. By exercise 11, $\mathbf{W}$ is the smallest $\mathbf{T}$-invariant subspace of $\mathbf{V}$ that contains v, so that $\mathbf{W} \subseteq \mathbf{Z}$. Clearly, $\mathbf{Z} \subseteq \mathbf{W}$ and so we conclude that $\mathbf{Z} = \mathbf{W}$. It follows that β is a basis for $\mathbf{W}$, and therefore $\dim(\mathbf{W}) = j$. Thus $j = k$. This proves (a).

(b) Now view β (from (a)) as an ordered basis for $\mathbf{W}$. Let $a_0, a_1, \ldots, a_{k-1}$ be scalars such that

$$a_0 v + a_1\mathbf{T}(\mathrm{v}) + \cdots + \mathrm{a}_{\mathrm{k}-1}\mathbf{T}^{\mathrm{k}-1}(\mathrm{v}) + \mathbf{T}^{\mathrm{k}}(\mathrm{v}) = \mathbf{0}.$$

Observe that

$$[\mathbf{T_W}]_\beta = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{pmatrix},$$

which has the characteristic polynomial

$$f(t) = (-1)^k(a_0 + a_1 t + \cdots + a_{k-1}t^{k-1} + t^k)$$

by exercise 19. Thus $f(t)$ is the characteristic polynomial of $\mathbf{T_W}$, proving (b).

Theorem 5.23 (Cayley-Hamilton). Let $\mathbf{T}$ be a linear operator on a finite-dimensional vector space $\mathbf{V}$, and let $f(t)$ be the characteristic polynomial of $\mathbf{T}$. Then $f(\mathbf{T}) = \mathbf{T}_0$, the zero transformation. That is, $\mathbf{T}$ "satisfies" its characteristic polynomial.

Proof. We will show that $f(\mathbf{T})(\mathrm{v}) = \mathbf{0}$ for all $v \in \mathbf{V}$. This is obvious if $v = 0$ because $f(\mathbf{T})$ is linear; so suppose that $v \neq 0$. Let $\mathbf{W}$ be the $\mathbf{T}$-cyclic subspace generated by v, and suppose that $\dim(\mathbf{W}) = k$. By Theorem 5.22(a), there exists scalars $a_0, a_1, \ldots, a_{k-1}$ such that

$$\mathrm{a_0 v} + \mathrm{a_1}\mathbf{T}(\mathrm{v}) + \cdots + \mathrm{a_{k-1}}\mathbf{T}^{\mathrm{k-1}}(\mathrm{v}) + \mathbf{T}^{\mathrm{k}}(\mathrm{v}) = \mathbf{0}.$$

Hence Theorem 5.22(b) implies that

$$g(t) = (-1)^k(a_0 + a_1 t + \cdots + a_{k-1}t^{k-1} + t^k).$$

By Theorem 5.21, $g(t)$ divides $f(t)$; hence there exists a polynomial $q(t)$ such that $f(t) = q(t)g(t)$. So

$$f(\mathbf{T})(\mathrm{v}) = \mathrm{q}(\mathbf{T})\mathrm{g}(\mathbf{T})(\mathrm{v}) = \mathrm{q}(\mathbf{T})(\mathrm{g}(\mathbf{T})(\mathrm{v})) = \mathrm{q}(\mathbf{T})(\mathbf{0}) = \mathbf{0}.$$

Corollary (Cayley-Hamilton Theorem for Matrices). Let A be an $n \times n$ matrix, and let $f(t)$ be the characterisitc polynomial of A. Then $f(A) = O$, the $n \times n$ zero matrix.

Proof. Let $\mathbf{L_A}$ the left multiplication transformation associated with the matrix A. By definition, the characterisitc polynomial, $f(t)$, of $\mathbf{L_A}$ is the characteristic polynomial of A. Hence by the Cayley-Hamilton Theorem, $f(\mathbf{L_A}) = \mathbf{T_0} = \mathbf{L_O}$. Thus for each $v \in \mathbf{F^n}$,

$$f(A)v = f(\mathbf{L_A})(\mathrm{v}) = \mathbf{L_O}(\mathrm{v}) = \mathrm{Ov}.$$

Since this equality holds for each $v \in \mathbf{F^n}$, $f(A) = O$.

Theorem 6.1. Let $\mathbf{V}$ be an inner product space. Then for $x, y, z \in \mathbf{V}$ and $c \in \mathbf{F}$, the following statements are true.
(a) $\langle x, y+z \rangle = \langle x, y \rangle + \langle x, z \rangle$.
(b) $\langle x, cy \rangle = \bar{c} \langle x, y \rangle$.
(c) $\langle x, \mathbf{0} \rangle = \langle \mathbf{0}, \mathrm{x} \rangle = 0$.
(d) $\langle x, x \rangle = 0$ if and only if $x = \mathbf{0}$.
(e) If $\langle x, y \rangle = \langle x, z \rangle$ for all $x \in \mathbf{V}$, then $y = z$.

Proof. (a) We have

$$\langle x, y+z \rangle = \overline{\langle y+z, x \rangle} = \overline{\langle y, x \rangle + \langle z, x \rangle}$$

$$= \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} = \langle x, y \rangle + \langle x, z \rangle .$$

Theorem 6.2. Let $\mathbf{V}$ be an inner product space over $\mathbf{F}$. Then for all $x, y \in \mathbf{V}$ and $c \in \mathbf{F}$, the following statements are true.
(a) $\|cx\| = |c| \cdot \|x\|$.
(b) $\|x\| = 0$ if and only if $x = 0$. In any case, $\|x\| \geq 0$.
(c) (Cauchy-Schwartz Inequality) $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$.
(d) (Triangle Inequality) $\|x + y\| \leq \|x\| + \|y\|$.

Proof. We leave the proofs of (a) and (b) as exercises.

(c) If $y = \mathbf{0}$, the the result is immediate. So assume that $y \neq \mathbf{0}$. For any $c \in \mathbf{F}$, we have

$$0 \leq \|x - cy\|^2 = \langle x - cy, x - cy\rangle = \langle x, x - cy\rangle - c\langle y, x - cy\rangle$$

$$= \langle x, x\rangle - \bar{c}\langle x, y\rangle - c\langle y, x\rangle + c\bar{c}\langle y, y\rangle .$$

In particular, if we set

$$c = \langle x, y\rangle / \langle y, y\rangle ,$$

the inequality becomes

$$0 \leq \langle x, x\rangle - |\langle x, y\rangle|^2 / \langle y, y\rangle = \|x\|^2 - |\langle x, y\rangle|^2 / \|y\|^2 ,$$

from which (c) follows.

(d) We have

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + \langle y, x\rangle + \langle x, y\rangle + \langle y, y\rangle$$

$$= \|x\|^2 + 2\Re\langle x, y\rangle + \|x\|^2$$

$$\leq \|x\|^2 + 2|\langle x, y\rangle| + \|x\|^2$$

$$\leq \|x\|^2 + 2\|x\|\,\|y\| + \|x\|^2$$

$$= (\|x\| + \|y\|)^2,$$

where $\Re\langle x, y\rangle$ denotes the real part of the complex number $\langle x, y\rangle$. Note that we used (c) to prove (d).

Theorem 6.3. Let $\mathbf{V}$ be an inner product space and $S = \{v_1, v_2, \dots, v_k\}$ be an orthogonal subset of $\mathbf{V}$ consisting of nonzero vectors. If $y \in \text{span}(S)$, then

$$y = \sum_{i=1}^{k} \frac{\langle y, v_i \rangle}{\|v_i\|^2} v_i.$$

Proof. Write $y = \sum_{i=1}^{k} a_i v_i$, where $a_1, a_2, \dots, \in \mathbf{F}$. Then for $1 \leq j \leq k$, we have

$$\langle y, v_j \rangle = \left\langle \sum_{i=1}^{k} a_i v_i, v_j \right\rangle = \sum_{i=1}^{k} a_i \langle v_i, v_j \rangle = a_j \langle v_j, v_j \rangle = a_j \|v_j\|^2 .$$

So $a_j = \langle y, v_j \rangle / \|v_j\|^2$, and the result follows immediately from Theorem 6.3.

Corollary 1. If, in addition to the hypotheses of Theorem 6.3, S is orthonormal and $y \in$span(S), then

$$y = \sum_{i=1}^{k} \langle y, v_i \rangle v_i.$$

Corollary 2. Let $\mathbf{V}$ be an inner product space and S be an orthogonal subset of $\mathbf{V}$ consisting of nonzero vectors. Then S is linearly independent.

Proof. Suppose that $v_1, v_2, \dots, v_k \in S$ and

$$\sum_{i=1}^{k} a_i v_i = 0.$$

As in the proof of Theorem 6.3 with $y = 0$, we have $a_j = \langle 0, v_j \rangle / \|v_j\|^2 = 0$ for all j. So S is linearly independent.

Theorem 6.4. Let $\mathbf{V}$ be an inner product space and $S = \{w_1, w_2, \ldots, w_n\}$ be linearly independent subset of $\mathbf{V}$. Define $S' = \{v_1, v_2, \ldots, v_n\}$ where $v_1 = w_1$ and

$$v_k = w_k - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} v_j \quad \text{for} \quad 2 \leq \text{k} \leq \text{n}. \qquad (1)$$

Then S' is an orthogonal set of nonzero vectors such that $\text{span}(S') = \text{span}(S)$.

Proof. The proof is by mathematical induction on n, the number of vectors in S. For $k = 1, 2, \ldots, n$, let $S_k = \{w_1, w_2, \ldots, w_k\}$. If $n = 1$, then the theorem is proved by taking $S'_1 = S_1$; i.e., $v_1 = w_1 \neq 0$. Assume then that the set $S'_{k-1} = \{v_1, v_2, \ldots, v_{k-1}\}$ with the desired properties has been constructed by the repeated use of (1). We will show that the set $S'_k = \{v_1, v_2, \ldots, v_{k-1}, v_k\}$ also has the desired properties, where v_k is obtained from S'_{k-1} by (1). If $v_k = 0$, then (1) implies that $w_k \in \text{span}(S'_{k-1})$ $= \text{span}(S_{k-1})$, which contradicts the assumption that S_k is linearly independent. For $1 \leq i \leq k-1$, it follows from (1) that

$$\langle v_k, v_i \rangle = \langle w_k, v_i \rangle - \sum_{j=1}^{k-1} \frac{\langle w_k, v_j \rangle}{\|v_j\|^2} \langle v_j, v_i \rangle = \langle w_k, v_i \rangle - \frac{\langle w_k, v_i \rangle}{\|v_i\|^2} \|v_i\|^2 = 0,$$

since $\langle v_j, v_i \rangle = 0$ if $i \neq j$ by the induction assumption that S'_{k-1} is orthogonal. Hence S'_k is an orthogonal set of nonzero vectors. Now, by (1), we have that $\text{span}(S'_k) \subset \text{span}(S_k)$. But by Corollary 2 to Theorem 6.3, S'_k is linearly independent; so $\dim(\text{span}(S_k)) = \dim(\text{span}(S'_k)) = k$. Therefore $\text{span}(S'_k) = \text{span}(S_k)$.

Theorem 6.5. Let $\mathbf{V}$ be a nonzero finite-dimensional inner product space. Then $\mathbf{V}$ has an orthonormal basis β. Furthermore, if $\beta = \{v_1, v_2, \ldots, v_n\}$ and $x \in \mathbf{V}$, then

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle v_i.$$

Proof. Let β_0 be an ordered basis for $\mathbf{V}$. Apply Theorem 6.4 to obtain an orthogonal set β' of nonzero vectors with $\text{span}(\beta') = \text{span}(\beta_0) = \mathbf{V}$. By normalizing each vector in β', we obtain an orthonormal set β that generates $\mathbf{V}$. By Corollary 2 to Theorem 6.3, β is linearly independent; therefore β is an orthonormal basis for $\mathbf{V}$. The remainder of the theorem follows from Corollary 1 to Theorem 6.3.

Corollary. Let $\mathbf{V}$ be a finite-dimensional inner product space with an orthonormal basis $\beta = \{v_1, v_2, \ldots, v_n\}$. Let $\mathbf{T}$ be a linear operator on $\mathbf{V}$, and let $A = [\mathbf{T}]_\beta$. Then for any i and j, $A_{ij} = \langle \mathbf{T}(\mathrm{v_j}), \mathrm{v_i} \rangle$.

Proof. From Theorem 6.5, we have

$$\mathbf{T}(\mathrm{v_j}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \langle \mathbf{T}(\mathrm{v_j}), \mathrm{v_i} \rangle \, \mathrm{v_i}.$$

Hence $A_{ij} = \langle \mathbf{T}(\mathrm{v_j}), \mathrm{v_i} \rangle$.

Theorem 6.6. Let $\mathbf{W}$ be a finite-dimensional subspace of an inner product space $\mathbf{V}$, and let $y \in \mathbf{V}$. Then there exists unique vectors $u \in \mathbf{W}$ and $z \in \mathbf{W}^{\perp}$ such that $y = u + z$. Furthermore, if $\{v_1, v_2, \ldots, v_k\}$ is an orthonormal basis for $\mathbf{W}$, then

$$u = \sum_{i=1}^{k} \langle y, v_i \rangle v_i.$$

Proof. Let $\{v_1, v_2, \ldots, v_k\}$ be an orthonormal basis for $\mathbf{W}$, let u be as defined in the preceding equation, and let $z = y - u$. Clearly $u \in \mathbf{W}$ and $y = u + z$.

To show that $z \in \mathbf{W}^{\perp}$, it suffices to show by exercise 7, that z is orthogonal to each v_j. For any j, we have

$$\langle z, v_j \rangle = \left\langle \left(y - \sum_{i=1}^{k} \langle y, v_i \rangle v_i \right), v_j \right\rangle = \langle y, v_j \rangle - \sum_{i=1}^{k} \langle y, v_i \rangle \langle v_i, v_j \rangle = \langle y, v_j \rangle - \langle y, v_j \rangle = 0.$$

To show uniqueness of u and z, suppose that $y = u + z = u' + z'$, where $u' \in \mathbf{W}$ and $z' \in \mathbf{W}^{\perp}$. Then then $u - u' = z - z' \in \mathbf{W} \cap \mathbf{W}^{\perp} = \{\mathbf{0}\}$. Therefore, $u = u'$ and $z = z'$.

Corollary. In the notation of Theorem 6.6, the vector u is the unique vector in $\mathbf{W}$ that is " closest " to y; that is, for any $x \in \mathbf{W}$, $\|y - x\| \geq \|y - u\|$ and this inequality is an equality if and only if $x = u$.

Proof. As in Theorem 6.6, we have that $y - u = z$, where $z \in \mathbf{W}^{\perp}$. Let $x \in \mathbf{W}$. Then $u - x$ is orthogonal to z, so, by exercise 10 of section 6.1,we have

$$\|y - x\|^2 = \|u + z - x\|^2 = \|(u - x) + z\|^2 = \|u - x\|^2 + \|z\|^2 \geq \|z\|^2 = \|y - u\|^2.$$

Now suppose that $\|y - x\| = \|y - u\|$. Then the inequality above becomes an equality, and therefore $\|u - x\|^2 + \|z\|^2 = \|z\|^2$. It follows that $\|u - x\| = 0$. and hence $x = u$. The proof of the converse is obvious.

Theorem 6.7. Suppose that $S = \{v_1, v_2, \ldots, v_k\}$ is an orthonormal set in an n-dimensional inner product space $\mathbf{V}$. Then
(a) S can be extended to an orthonormal basis $\{v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n\}$ for $\mathbf{V}$.
(b) If $\mathbf{W} = \text{span}(S)$, then $S_1 = \{v_{k+1}, v_{k+2}, \ldots, v_n\}$ is an orthonormal basis for $\mathbf{W}^{\perp}$.
(c) If $\mathbf{W}$ is any subspace of $\mathbf{V}$, then $\dim(\mathbf{V}) = \dim(\mathbf{W}) + \dim(\mathbf{W}^{\perp})$.

Proof. (a) By Corollary 2 to the replacement theorem, S can be extended to an ordered basis $S' = \{v_1, v_2, \ldots, v_k, w_{k+1}, \ldots, w_n\}$ for $\mathbf{V}$. Now apply the Gram-Schmidt process to S'. The first k vectors resulting from this process are the vectors in S by exercise 8, and this new set spans $\mathbf{V}$. Normalizing the last $k - n$ vectors of this set produces an orthonormal set that spans $\mathbf{V}$. The result now follows.
(b) Because S_1 is a subset of a basis, it is linearly independent. Since S_1 is clearly a subset of $\mathbf{W}^{\perp}$, we need only show that it spans $\mathbf{W}^{\perp}$. Note that, for any $x \in \mathbf{V}$, we have

$$x = \sum_{i=1}^{n} \langle x, v_i \rangle \, v_i.$$

If $x \in \mathbf{W}^{\perp}$, then $\langle x, v_i \rangle = 0$ for $1 \leq i \leq k$. Therefore,

$$x = \sum_{i=k+1}^{n} \langle x, v_i \rangle \, v_i \in \text{span}(\mathrm{S}_1).$$

(c) Let $\mathbf{W}$ be a subspace of $\mathbf{V}$. It is a finite-dimensional inner product space because $\mathbf{V}$ is, and so it has an orthonormal basis $\{v_1, v_2, \ldots, v_k\}$. By (a) and (b), we have

$$\dim(\mathbf{V}) = \mathrm{k} + (\mathrm{n} - \mathrm{k}) = \dim(\mathbf{W}) + \dim(\mathbf{W}^{\perp}).$$

A cofactor expansion requires $n!$ multiplications; for a 25×25 matrix that is approximately 1.5×10^{25} claculations.

If a computer performs one trillion (1×10^{12}) multiplications per second, it would run for over 500,000 years to calculate the determinant of a 25×25 matrix using cofactor expansion.