LINEAR OPERATORS

BRANKO ĆURGUS

Throughout this note $\mathscr{V}$ is a vector space over a scalar field $\mathbb{F}$. $\mathbb{N}$ denotes the set of positive integers and $i, j, k, l, m, n, p \in \mathbb{N}$.

1. Functions

First we review formal definitions related to functions. In this section A and B are nonempty sets.

The formal definition of function identifies a function and its graph. A justification for this is the fact that if you know the graph of a function, then you know the function, and conversely, if you know a function you know its graph. Simply stated the definition below says that a function from a set A to a set B is a subset f of the Cartesian product $A \times B$ such that for each $x \in A$ there exists unique $y \in B$ such that $(x, y) \in F$.

A *function* from A into B is a subset f of the Cartesian product $A \times B$ such that

(a) $\forall\, x \in A \;\; \exists\, y \in B \quad (x, y) \in f$,

(b) $\forall\, x \in A \;\forall\, y \in B \;\forall\, z \in B \quad (x, y) \in f \wedge (x, z) \in f \;\Rightarrow\; y = z$.

The relationship $(x, y) \in f$ is commonly written as $y = f(x)$. The symbol $f : A \to B$ denotes a function from A to B.

The set A is the *domain* of $f : A \to B$. The set B is the *codomain* of $f : A \to B$. The set

$$\bigl\{ y \in B : \exists\, x \in A \;\; y = f(x) \bigr\}$$

is called the *range* of $f : A \to B$. It is denoted by $\operatorname{ran} f$.

A function $f : A \to B$ is a *surjection* if for every $y \in B$ there exists $x \in A$ such that $y = f(x)$.

A function $f : A \to B$ is an *injection* if for every $x_1, x_2 \in A$ the following implication holds: $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$.

A function $f : A \to B$ is a *bijection* if it is both: a surjection and an injection.

Next we give a formal definition of a composition of two functions. However, before giving a definition we need to prove a proposition.

Proposition 1.1. *Let $f : A \to B$ and $g : C \to D$ be functions. If $\operatorname{ran} f \subseteq C$, then*

$$\bigl\{ (x, z) \in A \times D : \exists\, y \in B \;\; (x, y) \in f \wedge (y, z) \in g \bigr\} \tag{1.1}$$

is a function from A to D.

Date: February 3, 2017 at09:57.

Proof. A proof is a nice exercise. $\square$

The function defined by (1.1) is called the *composition* of functions f and g. It is denoted by $f \circ g$.

The function

$$\bigl\{(x, x) \in A \times A : x \in A\bigr\}$$

is called the *identity function* on A. It is denoted by id_A. In the standard notation id_A is the function $\mathrm{id}_A : A \to A$ such that $\mathrm{id}_A(x) = x$ for all $x \in A$.

A function $f : A \to B$ is *invertible* if there exist functions $g : B \to A$ and $h : B \to A$ such that $f \circ g = \mathrm{id}_B$ and $h \circ f = \mathrm{id}_A$.

Theorem 1.2. *Let $f : A \to B$ be a function. The following statements are equivalent.*

(a) *The function f is invertible.*
(b) *The function f is a bijection.*
(c) *There exists a unique function $g : B \to A$ such that $f \circ g = \mathrm{id}_B$ and $g \circ f = \mathrm{id}_A$.*

If f is invertible, then the unique g whose existence is proved in Theorem 1.2 (c) is called the *inverse* of f; it is denoted by f^{-1}.

Let $f : A \to B$ be a function. It is common to extend the notation $f(x)$ for $x \in A$ to subsets of A. For $X \subseteq A$ we introduce the notation

$$f(X) = \bigl\{y \in B : \exists\, x \in X \ y = f(x)\bigr\}.$$

With this notation, the range of f is simply the set $f(A)$.

Below are few exercises about functions from my Math 312 notes.

Exercise 1.3. Let A, B and C be nonempty sets. Let $f : A \to B$ and $g : B \to C$ be injections. Prove that $g \circ f : A \to C$ is an injection.

Exercise 1.4. Let A, B and C be nonempty sets. Let $f : A \to B$ and $g : B \to C$ be surjections. Prove that $g \circ f : A \to C$ is a surjection.

Exercise 1.5. Let A, B and C be nonempty sets. Let $f : A \to B$ and $g : B \to C$ be bijections. Prove that $g \circ f : A \to C$ is a bijection. Prove that $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Exercise 1.6. Let A, B and C be nonempty sets. Let $f : A \to B$, $g : B \to C$. Prove that if $g \circ f$ is an injection, then f is an injection.

Exercise 1.7. Let A, B and C be nonempty sets and let $f : A \to B$, $g : B \to C$. Prove that if $g \circ f$ is a surjection, then g is a surjection.

Exercise 1.8. Let A, B and C be nonempty sets and let $f : A \to B$, $g : B \to C$ and $h : C \to A$ be three functions. Prove that if any two of the functions $h \circ g \circ f$, $g \circ f \circ h$, $f \circ h \circ g$ are injections and the third is a surjection, or if any two of them are surjections and the third is an injection, then f, g, and h are bijections.

2. Linear operators

In this section $\mathscr{U}$, $\mathscr{V}$ and $\mathscr{W}$ are vector spaces over a scalar field $\mathbb{F}$.

2.1. **The definition and the vector space of all linear operators.** A function $T : \mathscr{V} \to \mathscr{W}$ is said to be a *linear operator* if it satisfies the following conditions:

$$\forall\, u \in \mathscr{V} \;\; \forall\, v \in \mathscr{V} \qquad T(u+v) = T(u) + f(v), \tag{2.1}$$

$$\forall\, \alpha \in \mathbb{F} \;\; \forall\, v \in \mathscr{V} \qquad T(\alpha v) = \alpha\, T(v). \tag{2.2}$$

The property (2.1) is called *additivity*, while the property (2.2) is called *homogeneity*. Together additivity and homogeneity are called *linearity*.

Denote by $\mathscr{L}(\mathscr{V},\mathscr{W})$ the set of all linear operators from $\mathscr{V}$ to $\mathscr{W}$. Define the addition and scaling in $\mathscr{L}(\mathscr{V},\mathscr{W})$. For $S,T \in \mathscr{L}(\mathscr{V},\mathscr{W})$ and $\alpha \in \mathbb{F}$ we define

$$(S+T)(v) = S(v) + T(v), \qquad \forall\, v \in \mathscr{V}, \tag{2.3}$$

$$(\alpha T)(v) = \alpha\, T(v), \qquad \forall\, v \in \mathscr{V}. \tag{2.4}$$

$$\tag{2.5}$$

Notice that two plus signs which appear in (2.3) have different meanings. The plus sign on the left-hand side stands for the addition of linear operators that is just being defined, while the plus sign on the right-hand side stands for the addition in $\mathscr{W}$. Notice the analogous difference in empty spaces between α and T in (2.4). Define the zero mapping in $\mathscr{L}(\mathscr{V},\mathscr{W})$ to be

$$0_{\mathscr{L}(\mathscr{V},\mathscr{W})}(v) = 0_{\mathscr{W}}, \qquad \forall\, v \in \mathscr{V}.$$

For $T \in \mathscr{L}(\mathscr{V},\mathscr{W})$ we define its opposite operator by

$$(-T)(v) = -T(v), \qquad \forall\, v \in \mathscr{V}.$$

Proposition 2.1. *The set $\mathscr{L}(\mathscr{V},\mathscr{W})$ with the operations defined in* (2.3), *and* (2.4) *is a vector space over $\mathbb{F}$.*

For $T \in \mathscr{L}(\mathscr{V},\mathscr{W})$ and $v \in \mathscr{V}$ it is customary to write Tv instead of $T(v)$.

Example 2.2. Assume that a vector space $\mathscr{V}$ is a direct sum of its subspaces $\mathscr{U}$ and $\mathscr{W}$, that is $\mathscr{V} = \mathscr{U} \oplus \mathscr{W}$. Define the function $P : \mathscr{V} \to \mathscr{V}$ by

$$Pv = w \qquad \Leftrightarrow \qquad v = u + w, \quad u \in \mathscr{U}, \quad w \in \mathscr{W}.$$

Then P is a linear operator. It is called the *projection* of $\mathscr{V}$ onto $\mathscr{W}$ parallel to $\mathscr{U}$; it is denoted by $P_{\mathscr{W}\|\mathscr{U}}$.

2.2. **Composition, inverse, isomorphism.** In the next two propositions we prove that the linearity is preserved under composition of linear operators and under taking the inverse of a linear operator.

Proposition 2.3. *Let $S : \mathscr{U} \to \mathscr{V}$ and $T : \mathscr{V} \to \mathscr{W}$ be linear operators. The composition $T \circ S : \mathscr{U} \to \mathscr{W}$ is a linear operator.*

Proof. Prove this as an exercise. $\square$

When composing linear operators it is customary to write simply TS instead of $T \circ S$.

The identity function on $\mathscr{V}$ is denoted by $I_{\mathscr{V}}$. It is defined by $I_{\mathscr{V}}(v) = v$ for all $v \in \mathscr{V}$. It is clearly a linear operator.

Proposition 2.4. *Let $T : \mathscr{V} \to \mathscr{W}$ be a linear operator which is invertible. Then the inverse $T^{-1} : \mathscr{W} \to \mathscr{V}$ of T is a linear operator.*

Proof. Since T is invertible, by Theorem 1.2 there exists a function $S : \mathscr{W} \to \mathscr{V}$ such that $ST = I_{\mathscr{V}}$ and $TS = I_{\mathscr{W}}$. Since T is linear and $TS = I_{\mathscr{W}}$ we have

$$T\bigl(\alpha Sx + \beta Sy\bigr) = \alpha T(Sx) + \beta T(Sy) = \alpha (TS)x + \beta (TS)y = \alpha x + \beta y$$

for all $\alpha, \beta \in \mathbb{F}$ and all $x, y \in \mathscr{W}$. Applying S to both sides of

$$T\bigl(\alpha Sx + \beta Sy\bigr) = \alpha x + \beta y$$

we get

$$(ST)\bigl(\alpha Sx + \beta Sy\bigr) = S\bigl(\alpha x + \beta y\bigr) \qquad \forall\, \alpha, \beta \in \mathbb{F} \quad \forall x, x \in \mathscr{W}.$$

Since $ST = I_{\mathscr{V}}$, we get

$$\alpha Sx + \beta Sy = S\bigl(\alpha x + \beta y\bigr) \qquad \forall\, \alpha, \beta \in \mathbb{F} \quad \forall x, x \in \mathscr{W},$$

thus proving the linearity of S. Since by definition $S = T^{-1}$ the proposition is proved. $\square$

A linear operator $T : \mathscr{V} \to \mathscr{W}$ which is a bijection is called an *isomorphism* between vector spaces $\mathscr{V}$ and $\mathscr{W}$.

By Theorem 1.2 and Proposition 2.4 each isomorphism is invertible and its inverse is also an isomorphism.

In the next proposition we introduce the most important isomorphism $C_{\mathscr{B}}$.

Proposition 2.5. *Let $\mathscr{V}$ be a finite dimensional vector space over $\mathbb{F}$ and $n = \dim \mathscr{V}$. Let $\mathscr{B} = \{v_1, \ldots, v_n\}$ be a basis for $\mathscr{V}$. There exists a function $C_{\mathscr{B}} : \mathscr{V} \to \mathbb{F}^n$ such that*

$$C_{\mathscr{B}}(v) = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \qquad \Leftrightarrow \qquad v = \sum_{j=1}^{n} \alpha_j v_j.$$

The function $C_{\mathscr{B}}$ is an isomorphism between $\mathscr{V}$ and $\mathbb{F}^n$.

Proof. Since $\mathscr{B}$ spans $\mathscr{V}$, for every $v \in \mathscr{V}$ there exist $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $v = \sum_{j=1}^{n} \alpha_j v_j$. Thus $C_{\mathscr{B}}$ is defined for every $v \in \mathscr{V}$. To prove that $C_{\mathscr{B}}$ is a function assume

$$C_{\mathscr{B}}(v) = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \qquad \text{and} \qquad C_{\mathscr{B}}(v) = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}.$$

Then

$$v = \sum_{j=1}^{n} \alpha_j v_j \qquad \text{and} \qquad v = \sum_{j=1}^{n} \beta_j v_j.$$

Therefore $0_{\mathscr{V}} = \sum_{j=1}^{n}(\alpha_j - \beta_j)v_j$. Since $\mathscr{B}$ is linearly independent, $\alpha_j = \beta_j$ for all $j \in \{1, \ldots, n\}$. Thus

$$C_{\mathscr{B}}(v) = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \quad \text{and} \quad C_{\mathscr{B}}(v) = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \qquad \Rightarrow \qquad \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}.$$

This proves that $C_{\mathscr{B}}$ is a function.

The linearity of $C_{\mathscr{B}}$ is easy to verify.

The injectivity of $C_{\mathscr{B}}$ follows from the linear independence of $\mathscr{B}$.

The surjectivity of $C_{\mathscr{B}}$ follows from the fact that for arbitrary $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ we have $v = \sum_{j=1}^{n} \alpha_j v_j \in \mathscr{V}$ and therefore

$$C_{\mathscr{B}}(v) = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}. \qquad \square$$

In the last part of the proof of Proposition 2.5 we showed that the formula for the inverse $(C_{\mathscr{B}})^{-1} : \mathbb{F}^n \to \mathscr{V}$ of $C_{\mathscr{B}}$ is given by

$$(C_{\mathscr{B}})^{-1} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \sum_{j=1}^{n} \alpha_j v_j, \qquad \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{F}^n. \tag{2.6}$$

Notice that (2.6) defines a function from $\mathbb{F}^n$ to $\mathscr{V}$ even if $\mathscr{B}$ is not a basis of $\mathscr{V}$.

Example 2.6. Inspired by the definition of $C_{\mathscr{B}}$ and (2.6) we define a general operator of this kind. Let $\mathscr{V}$ and $\mathscr{W}$ be vector spaces over $\mathbb{F}$. Let $\mathscr{V}$ be finite dimensional, $n = \dim \mathscr{V}$ and let $\mathscr{B}$ be a basis for $\mathscr{V}$. Let $\mathscr{C} = (w_1, \ldots, w_n)$ be any n-tuple of vectors in $\mathscr{W}$. The entries of an n-tuple can be repeated, they can all be equal, for example to $0_{\mathscr{V}}$. We define the linear operator $L_{\mathscr{C}}^{\mathscr{B}} : \mathscr{V} \to \mathscr{W}$ by

$$L_{\mathscr{C}}^{\mathscr{B}}(v) = \sum_{j=1}^{n} \alpha_j w_j \qquad \text{where} \qquad \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = C_{\mathscr{B}}(v). \tag{2.7}$$

In fact, $L_{\mathscr{C}}^{\mathscr{B}} : \mathscr{V} \to \mathscr{W}$ is a composition of $C_{\mathscr{B}} : \mathscr{V} \to \mathbb{F}^n$ and the operator $\mathbb{F}^n \to \mathscr{W}$ defined by

$$\begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} \mapsto \sum_{j=1}^{n} \xi_j w_j \qquad \text{for arbitrary} \qquad \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} \in \mathbb{F}^n. \tag{2.8}$$

It is easy to verify that (2.8) defines a linear operator.

Denote by $\mathscr{E}$ the standard basis of $\mathbb{F}^n$, that is the basis which consists of the columns of the identity matrix. Then $C_{\mathscr{B}} = L_{\mathscr{E}}^{\mathscr{B}}$ and $(C_{\mathscr{B}})^{-1} = L_{\mathscr{B}}^{\mathscr{E}}$.

Exercise 2.7. Let $\mathscr{V}$ and $\mathscr{W}$ be vector spaces over $\mathbb{F}$. Let $\mathscr{V}$ be finite dimensional, $n = \dim \mathscr{V}$ and let $\mathscr{B}$ be a basis for $\mathscr{V}$. Let $\mathscr{C} = (w_1, \ldots, w_n)$ be a list of vectors in $\mathscr{W}$ with n entries.

(a) Characterize the injectivity of $L_{\mathscr{C}}^{\mathscr{B}} : \mathscr{V} \to \mathscr{W}$.
(b) Characterize the surjectivity of $L_{\mathscr{C}}^{\mathscr{B}} : \mathscr{V} \to \mathscr{W}$.
(c) Characterize the bijectivity of $L_{\mathscr{C}}^{\mathscr{B}} : \mathscr{V} \to \mathscr{W}$.
(d) If $L_{\mathscr{C}}^{\mathscr{B}} : \mathscr{V} \to \mathscr{W}$ is an isomorphism, find a simple formula for $(L_{\mathscr{C}}^{\mathscr{B}})^{-1}$.

2.3. **The nullity-rank theorem.** Let $T : \mathscr{V} \to \mathscr{W}$ is be a linear operator. The linearity of T implies that the set

$$\operatorname{nul} T = \bigl\{ v \in \mathscr{V} : Tv = 0_{\mathscr{W}} \bigr\}$$

is a subspace of $\mathscr{V}$. This subspace is called the *null space* of T. Similarly, the linearity of T implies that the range of T is a subspace of $\mathscr{W}$. Recall that

$$\operatorname{ran} T = \bigl\{ w \in \mathscr{W} : \exists\, v \in \mathscr{V} \quad w = Tv \bigr\}.$$

Proposition 2.8. *A linear operator $T : \mathscr{V} \to \mathscr{W}$ is an injection if and only if* $\operatorname{nul} T = \{0_{\mathscr{V}}\}$.

Proof. We first prove the "if" part of the proposition. Assume that $\operatorname{nul} T = \{0_{\mathscr{V}}\}$. Let $u, v \in \mathscr{V}$ be arbitrary and assume that $Tu = Tv$. Since T is linear, $Tu = Tv$ implies $T(u-v) = 0_{\mathscr{W}}$. Consequently $u-v \in \operatorname{nul} T = \{0_{\mathscr{V}}\}$. Hence, $u - v = 0_{\mathscr{V}}$, that is $u = v$. This proves that T is an injection.

To prove the "only if" part assume that $T : \mathscr{V} \to \mathscr{W}$ is an injection. Let $v \in \operatorname{nul} T$ be arbitrary. Then $Tv = 0_{\mathscr{W}} = T0_{\mathscr{V}}$. Since T is injective, $Tv = T0_{\mathscr{V}}$ implies $v = 0_{\mathscr{V}}$. Thus we have proved that $\operatorname{nul} T \subseteq \{0_{\mathscr{V}}\}$. Since the converse inclusion is trivial, we have $\operatorname{nul} T = \{0_{\mathscr{V}}\}$. □

Theorem 2.9 (Nullity-Rank Theorem)**.** *Let $\mathscr{V}$ and $\mathscr{W}$ be vector spaces over a scalar field $\mathbb{F}$ and let $T : \mathscr{V} \to \mathscr{W}$ be a linear operator. If $\mathscr{V}$ is finite dimensional, then* $\operatorname{nul} T$ *and* $\operatorname{ran} T$ *are finite dimensional and*

$$\dim(\operatorname{nul} T) + \dim(\operatorname{ran} T) = \dim \mathscr{V}. \tag{2.9}$$

Proof. Assume that $\mathscr{V}$ is finite dimensional. We proved earlier that for an arbitrary subspace $\mathscr{U}$ of $\mathscr{V}$ there exists a subspace $\mathscr{X}$ of $\mathscr{V}$ such that

$$\mathscr{U} \oplus \mathscr{X} = \mathscr{V} \qquad \text{and} \qquad \dim \mathscr{U} + \dim \mathscr{X} = \dim \mathscr{V}.$$

Thus, there exists a subspace $\mathscr{X}$ of $\mathscr{V}$ such that

$$(\operatorname{nul} T) \oplus \mathscr{X} = \mathscr{V} \qquad \text{and} \qquad \dim(\operatorname{nul} T) + \dim \mathscr{X} = \dim \mathscr{V}. \tag{2.10}$$

Since $\dim(\operatorname{nul} T) + \dim \mathscr{X} = \dim \mathscr{V}$, to prove the theorem we only need to prove that $\dim \mathscr{X} = \dim(\operatorname{ran} T)$. To this end, let $m = \dim \mathscr{X}$ and let

$x_1, \dots, x_m$ be a basis for $\mathscr{X}$. We will prove that vectors $Tx_1, \dots, Tx_m$ form a basis for $\operatorname{ran} T$. We first prove

$$\operatorname{span}\bigl\{Tx_1, \dots, Tx_m\bigr\} = \operatorname{ran} T. \tag{2.11}$$

Clearly $\bigl\{Tx_1, \dots, Tx_m\bigr\} \subseteq \operatorname{ran} T$. Consequently, since $\operatorname{ran} T$ is a subspace of $\mathscr{W}$, we have $\operatorname{span}\bigl\{Tx_1, \dots, Tx_m\bigr\} \subseteq \operatorname{ran} T$. To prove the converse inclusion, let $w \in \operatorname{ran} T$ be arbitrary. Then, there exists $v \in \mathscr{V}$ such that $Tv = w$. Since $\mathscr{V} = (\operatorname{nul} T) + \mathscr{X}$, there exist $u \in \operatorname{nul} T$ and $x \in \mathscr{X}$ such that $v = u + x$. Then $Tv = T(u+x) = Tu + Tx = Tx$. As $x \in \mathscr{X}$, there exist $\xi_1, \dots, \xi_m \in \mathbb{F}$ such that $x = \sum_{j=1}^{m} \xi_j x_j$. Now we use linearity of T to deduce

$$w = Tv = Tx = \sum_{j=1}^{m} \xi_j T x_j.$$

This proves that $w \in \operatorname{span}\bigl\{Tx_1, \dots, Tx_m\bigr\}$. Since w was arbitrary in $\operatorname{ran} T$ this completes a proof of (2.11).

Next we prove that the vectors $Tx_1, \dots, Tx_m$ are linearly independent. Let $\alpha_1, \dots, \alpha_m \in \mathbb{F}$ be arbitrary and assume that

$$\alpha_1 T x_1 + \cdots + \alpha_m T x_m = 0_{\mathscr{W}}. \tag{2.12}$$

Since T is linear (2.12) implies that

$$\alpha_1 x_1 + \cdots + \alpha_m x_m \in \operatorname{nul} T. \tag{2.13}$$

Recall that $x_1, \dots, x_m \in cX$ and $\mathscr{X}$ is a subspace of $\mathscr{V}$, so

$$\alpha_1 x_1 + \cdots + \alpha_m x_m \in \mathscr{X}. \tag{2.14}$$

Now (2.13), (2.14) and the fact that $(\operatorname{nul} T) \cap \mathscr{X} = \{0_{\mathscr{V}}\}$ imply

$$\alpha_1 x_1 + \cdots + \alpha_m x_m = 0_{\mathscr{V}}. \tag{2.15}$$

Since $x_1, \dots, x_m$ are linearly independent (2.15) yields $\alpha_1 = \cdots = \alpha_m = 0$. This completes a proof of the linear independence of $Tx_1, \dots, Tx_m$.

Thus $\bigl\{Tx_1, \dots, Tx_m\bigr\}$ is a basis for $\operatorname{ran} T$. Consequently $\dim(\operatorname{ran} T) = m$. Since $m = \dim \mathscr{X}$, (2.10) implies (2.9). This completes the proof. $\square$

A direct proof of NRT Theorem is as follows:

Proof. Since $\operatorname{nul} T$ is a subspace of $\mathscr{V}$ it is finite dimensional. Set $k = \dim\bigl(\operatorname{nul} T\bigr)$ and let $\mathscr{C} = \bigl\{u_1, \dots, u_k\bigr\}$ be a basis for $\operatorname{nul} T$.

Since $\mathscr{V}$ is finite dimensional there exists a finite set $\mathscr{F} \subset \mathscr{V}$ such that $\operatorname{span}(\mathscr{F}) = \mathscr{V}$. Then the set $T\mathscr{F}$ is a finite subset of $\mathscr{W}$ and $\operatorname{ran} T = \operatorname{span}\bigl(T\mathscr{F}\bigr)$. Thus $\operatorname{ran} T$ is finite dimensional. Let $\dim\bigl(\operatorname{ran} T\bigr) = m$ and let $\mathscr{E} = \bigl\{w_1, \dots, w_m\bigr\}$ be a basis of $\operatorname{ran} T$.

Since clearly for every $j \in \bigl\{1, \dots, m\bigr\}$, $w_j \in \operatorname{ran} T$, we have that for every $j \in \bigl\{1, \dots, m\bigr\}$ there exists $v_j \in \mathscr{V}$ such that $Tv_j = w_j$. Set $\mathscr{D} = \bigl\{v_1, \dots, v_m\bigr\}$.

Further set $\mathscr{B} = \mathscr{C} \cup \mathscr{D}$.

We will prove the following three facts:

(I) $\mathscr{C} \cap \mathscr{D} = \emptyset$,
(II) $\operatorname{span} \mathscr{B} = \mathscr{V}$,
(III) $\mathscr{B}$ is a linearly independent set.

To prove (I), notice that the vectors in $\mathscr{E}$ are nonzero, since $\mathscr{E}$ is linearly independent. Therefore, for every $v \in \mathscr{D}$ we have that $Tv \neq 0_{\mathscr{W}}$. Since for every $u \in \mathscr{C}$ we have $Tu = 0_{\mathscr{W}}$ we conclude that $u \in \mathscr{C}$ implies $u \notin \mathscr{D}$. This proves (I).

To prove (II), first notice that by the definition of $\mathscr{B} \subset \mathscr{V}$. Since $\mathscr{V}$ is a vector space, we have $\operatorname{span} \mathscr{B} \subseteq \mathscr{V}$.

To prove the converse inclusion, let $v \in \mathscr{V}$ be arbitrary. Then $Tv \in \operatorname{ran} T$. Since $\mathscr{E}$ spans $\operatorname{ran} T$, there exist $\beta_1, \ldots, \beta_m \in \mathbb{F}$ such that

$$Tv = \sum_{j=1}^{m} \beta_j w_j.$$

Set

$$v' = \sum_{j=1}^{m} \beta_j v_j.$$

Then, by linearity of T we have

$$Tv' = \sum_{j=1}^{m} \beta_j T v_j = \sum_{j=1}^{m} \beta_j w_j = Tv.$$

The last equality yields and the linearity of T yield $T(v - v') = 0_{\mathscr{W}}$. Consequently, $v - v' \in \operatorname{nul} T$. Since $\mathscr{C}$ spans $\operatorname{nul} T$, there exist $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ such that

$$v - v' = \sum_{j=1}^{k} \alpha_i u_i.$$

Consequently,

$$v = v' + \sum_{j=1}^{k} \alpha_i u_i = \sum_{j=1}^{k} \alpha_i u_i + \sum_{j=1}^{m} \beta_j v_j.$$

This proves that for arbitrary $v \in \mathscr{V}$ we have $v \in \operatorname{span} \mathscr{B}$. Thus $\mathscr{V} \subseteq \operatorname{span} \mathscr{B}$ and (II) is proved.

To prove (III) let $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$ and $\beta_1, \ldots, \beta_m \in \mathbb{F}$ be arbitrary and assume that

$$\sum_{j=1}^{k} \alpha_i u_i + \sum_{j=1}^{m} \beta_j v_j = 0_{\mathscr{V}}. \tag{2.16}$$

Applying T to both sides of the last equality, and using the fact that $u_i \in \operatorname{nul} T$ and the definition of v_j we get

$$\sum_{j=1}^{m} \beta_j w_j = 0_{\mathscr{W}}.$$

Since $\mathscr{E}$ is a linearly independent set the last equality implies that $\beta_j = 0$ for all $j \in \{1, \ldots, m\}$. Now substitute these equalities in (2.16) to get

$$\sum_{j=1}^{k} \alpha_i u_i = 0_{\mathscr{V}}.$$

Since $\mathscr{C}$ is a linearly independent set the last equality implies that $\alpha_i = 0$ for all $i \in \{1, \ldots, k\}$. This proves the linear independence of $\mathscr{B}$.

It follows from (II) and (III) that $\mathscr{B}$ is a basis for $\mathscr{V}$. By (I) we have that $|\mathscr{B}| = |\mathscr{C}| + |\mathscr{D}| = k + m$. This completes the proof of the theorem. $\square$

The nonnegative integer $\dim(\operatorname{nul} T)$ is called the *nullity* of T; the nonnegative integer $\dim(\operatorname{ran} T)$ is called the *rank* of T.

The nullity-rank theorem in English reads: If a linear operator is defined on a finite dimensional vector space, then its nullity and its rank are finite and they add up to the dimension of the domain.

Proposition 2.10. *Let $\mathscr{V}$ and $\mathscr{W}$ be vector spaces over $\mathbb{F}$. Assume that $\mathscr{V}$ is finite dimensional. The following statements are equivalent*

(a) *There exists a surjection $T \in \mathscr{L}(\mathscr{V}, \mathscr{W})$.*
(b) *$\mathscr{W}$ is finite dimensional and* $\dim \mathscr{V} \geq \dim \mathscr{W}$.

Proposition 2.11. *Let $\mathscr{V}$ and $\mathscr{W}$ be vector spaces over $\mathbb{F}$. Assume that $\mathscr{V}$ is finite dimensional. The following statements are equivalent*

(a) *There exists an injection $T \in \mathscr{L}(\mathscr{V}, \mathscr{W})$.*
(b) *Either $\mathscr{W}$ is infinite dimensional or* $\dim \mathscr{V} \leq \dim \mathscr{W}$.

Proposition 2.12. *Let $\mathscr{V}$ and $\mathscr{W}$ be vector spaces over $\mathbb{F}$. Assume that $\mathscr{V}$ is finite dimensional. The following statements are equivalent*

(a) *There exists an isomorphism $T : \mathscr{V} \to \mathscr{W}$.*
(b) *$\mathscr{W}$ is finite dimensional and* $\dim \mathscr{W} = \dim \mathscr{V}$.

2.4. **Isomorphism between $\mathscr{L}(\mathscr{V}, \mathscr{W})$ and $\mathbb{F}^{n\times m}$.** Let $\mathscr{V}$ and $\mathscr{W}$ be finite dimensional vector spaces over $\mathbb{F}$, $m = \dim \mathscr{V}$, $n = \dim \mathscr{W}$, let $\mathscr{B} = \{v_1, \ldots, v_n\}$ be a basis for $\mathscr{V}$ and let $\mathscr{C} = \{w_1, \ldots, w_n\}$ be a basis for $\mathscr{W}$. The mapping $C_{\mathscr{B}}$ provides an isomorphism between $\mathscr{V}$ and $\mathbb{F}^m$ and $C_{\mathscr{C}}$ provides an isomorphism between $\mathscr{W}$ and $\mathbb{F}^n$.

Recall that the simplest way to define a linear operator from $\mathbb{F}^m$ to $\mathbb{F}^n$ is to use an $n \times m$ matrix B. It is convenient to consider an $n \times m$ matrix to be an m-tuple of its columns, which are vectors in $\mathbb{F}^n$. For example, let $\mathbf{b}_1, \ldots, \mathbf{b}_m \in \mathbb{F}^n$ be columns of an $n \times m$ matrix B. Then we write

$$B = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_m \end{bmatrix}.$$

This notation is convenient since it allows us to write a multiplication of a vector $\mathbf{x} \in \mathbb{F}^m$ by a matrix B as

$$B\mathbf{x} = \sum_{j=1}^{m} \xi_j \mathbf{b}_j \qquad \text{where} \qquad \mathbf{x} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}. \tag{2.17}$$

Notice the similarity of the definition in (2.17) to the definition (2.7) of the operator $L_{\mathscr{C}}^{\mathscr{B}}$ in Example 2.6. Taking $\mathscr{B}$ to be the standard basis of $\mathbb{F}^m$ and taking $\mathscr{C}$ to me the m-tuple given by B, we have $L_{\mathscr{C}}^{\mathscr{B}}(\mathbf{x}) = B\mathbf{x}$.

Let $T : \mathscr{V} \to \mathscr{W}$ be a linear operator. Our next goal is to connect T in a natural way to a certain $n \times m$ matrix B. That "natural way" is suggested by following diagram:

$$\begin{tikzcd}
\mathscr{V} \arrow[r, "T"] \arrow[d, "C_{\mathscr{B}}"'] & \mathscr{W} \arrow[d, "C_{\mathscr{C}}"] \\
\mathbb{F}^m \arrow[r, dashed, "B"'] & \mathbb{F}^n
\end{tikzcd}$$

We seek an $n \times m$ matrix B such that the action of T between $\mathscr{V}$ and $\mathscr{W}$ is in some sense replicated by the action of B between $\mathbb{F}^m$ and $\mathbb{F}^n$. Precisely, we seek B such that

$$C_{\mathscr{C}}(Tv) = B\bigl(C_{\mathscr{B}}(v)\bigr) \qquad \forall\, v \in \mathscr{V}. \tag{2.18}$$

In English: multiplying the vector of coordinates of v by B we get exactly the coordinates of Tv.

Using the basis vectors $v_1, \ldots, v_n \in \mathscr{B}$ in (2.18) we see that the matrix

$$B = \bigl[C_{\mathscr{C}}(Tv_1) \ \cdots \ C_{\mathscr{C}}(Tv_m)\bigr] \tag{2.19}$$

has the desired property (2.18).

For an arbitrary $T \in \mathscr{L}(\mathscr{V}, \mathscr{W})$ the formula (2.19) associates the matrix $B \in \mathbb{F}^{n\times m}$ with T. In other words (2.19) defines a function from $\mathscr{L}(\mathscr{V}, \mathscr{W})$ to $\mathbb{F}^{n\times m}$.

Theorem 2.13. *Let $\mathscr{V}$ and $\mathscr{W}$ be finite dimensional vector spaces over $\mathbb{F}$, $m = \dim \mathscr{V}$, $n = \dim \mathscr{W}$, let $\mathscr{B} = \{v_1, \ldots, v_m\}$ be a basis for $\mathscr{V}$ and let $\mathscr{C} = \{w_1, \ldots, w_n\}$ be a basis for $\mathscr{W}$. The function*

$$M_{\mathscr{C}}^{\mathscr{B}} : \mathscr{L}(\mathscr{V}, \mathscr{W}) \to \mathbb{F}^{n\times m}$$

defined by

$$M_{\mathscr{C}}^{\mathscr{B}}(T) = \bigl[C_{\mathscr{C}}(Tv_1) \ \cdots \ C_{\mathscr{C}}(Tv_m)\bigr], \qquad T \in \mathscr{L}(\mathscr{V}, \mathscr{W}) \tag{2.20}$$

is an isomorphism.

Proof. It is easy to verify that $M_{\mathscr{C}}^{\mathscr{B}}$ is a linear operator.

Since the definition of $M_{\mathscr{C}}^{\mathscr{B}}(T)$ coincides with (2.19), equality (2.18) yields

$$C_{\mathscr{C}}(Tv) = \bigl(M_{\mathscr{C}}^{\mathscr{B}}(T)\bigr) C_{\mathscr{B}}(v). \tag{2.21}$$

The most direct way to prove that $M_{\mathscr{C}}^{\mathscr{B}}$ is an isomorphism is to construct its inverse. The inverse is suggested by the diagram (2.22).

$$\begin{tikzcd} \mathscr{V} \arrow[r, dashed, "T"] \arrow[d, "C_{\mathscr{B}}"'] & \mathscr{W} \\ \mathbb{F}^m \arrow[r, "B"'] & \mathbb{F}^n \arrow[u, "(C_{\mathscr{C}})^{-1}"'] \end{tikzcd} \tag{2.22}$$

Define

$$N_{\mathscr{C}}^{\mathscr{B}} : \mathbb{F}^{n\times m} \to \mathscr{L}(\mathscr{V}, \mathscr{W})$$

by

$$\big(N_{\mathscr{C}}^{\mathscr{B}}(B)\big)(v) = (C_{\mathscr{C}})^{-1}\big(B(C_{\mathscr{B}}(v))\big), \qquad B \in \mathbb{F}^{n\times m}. \tag{2.23}$$

Next we prove that

$$N_{\mathscr{C}}^{\mathscr{B}} \circ M_{\mathscr{C}}^{\mathscr{B}} = I_{\mathscr{L}(\mathscr{V},\mathscr{W})} \quad \text{and} \quad M_{\mathscr{C}}^{\mathscr{B}} \circ N_{\mathscr{C}}^{\mathscr{B}} = I_{\mathbb{F}^{n\times m}}.$$

First for arbitrary $T \in \mathscr{L}(\mathscr{V}, \mathscr{W})$ and arbitrary $v \in \mathscr{V}$ we calculate

$$\begin{aligned} \Big(\big(N_{\mathscr{C}}^{\mathscr{B}} \circ M_{\mathscr{C}}^{\mathscr{B}}\big)(T)\Big)(v) &= (C_{\mathscr{C}})^{-1}\big(\big(M_{\mathscr{C}}^{\mathscr{B}}(T)\big)(C_{\mathscr{B}}(v))\big) && \text{by (2.23)} \\ &= (C_{\mathscr{C}})^{-1}\big(C_{\mathscr{C}}(Tv)\big) && \text{by (2.21)} \\ &= Tv. \end{aligned}$$

Thus $\big(N_{\mathscr{C}}^{\mathscr{B}} \circ M_{\mathscr{C}}^{\mathscr{B}}\big)(T) = T$ and thus, since $T \in \mathscr{L}(\mathscr{V}, \mathscr{W})$ was arbitrary, $N_{\mathscr{C}}^{\mathscr{B}} \circ M_{\mathscr{C}}^{\mathscr{B}} = I_{\mathscr{L}(\mathscr{V},\mathscr{W})}$.

Let now $B \in \mathbb{F}^{n\times m}$ be arbitrary and calculate

$$\begin{aligned} \big(M_{\mathscr{C}}^{\mathscr{B}} \circ N_{\mathscr{C}}^{\mathscr{B}}\big)(B) &= M_{\mathscr{C}}^{\mathscr{B}}\big(N_{\mathscr{C}}^{\mathscr{B}}(B)\big) \\ &= \Big[C_{\mathscr{C}}\big(\big(N_{\mathscr{C}}^{\mathscr{B}}(B)\big)(v_1)\big) \ \cdots \ C_{\mathscr{C}}\big(\big(N_{\mathscr{C}}^{\mathscr{B}}(B)\big)(v_m)\big)\Big] && \text{by (2.20)} \\ &= \Big[B(C_{\mathscr{B}}(v_1)) \ \cdots \ B(C_{\mathscr{B}}(v_m))\Big] && \text{by (2.23)} \\ &= B\Big[C_{\mathscr{B}}(v_1) \ \cdots \ C_{\mathscr{B}}(v_m)\Big] && \text{matrix mult.} \\ &= B\, I_m && \text{def. of } C_{\mathscr{B}} \\ &= B. \end{aligned}$$

Thus $\big(M_{\mathscr{C}}^{\mathscr{B}} \circ N_{\mathscr{C}}^{\mathscr{B}}\big)(B) = B$ for all $B \in \mathbb{F}^{n\times m}$, proving that $M_{\mathscr{C}}^{\mathscr{B}} \circ N_{\mathscr{C}}^{\mathscr{B}} = I_{\mathbb{F}^{n\times m}}$.

This completes the proof that $M_{\mathscr{C}}^{\mathscr{B}}$ is a bijection. Since it is linear, $M_{\mathscr{C}}^{\mathscr{B}}$ is an isomorphism. □

Theorem 2.14. *Let $\mathscr{U}$, $\mathscr{V}$ and $\mathscr{W}$ be finite dimensional vector spaces over $\mathbb{F}$, $k = \dim \mathscr{U}$, $m = \dim \mathscr{V}$, $n = \dim \mathscr{W}$, let $\mathscr{A}$ be a basis for $\mathscr{U}$, let $\mathscr{B}$ be a basis for $\mathscr{V}$, and let $\mathscr{C}$ be a basis for $\mathscr{W}$. Let $S \in \mathscr{L}(\mathscr{U}, \mathscr{V})$ and $T \in \mathscr{L}(\mathscr{V}, \mathscr{W})$. Let $M_{\mathscr{B}}^{\mathscr{A}}(S) \in \mathbb{F}^{m\times k}$, $M_{\mathscr{C}}^{\mathscr{B}}(T) \in \mathbb{F}^{n\times m}$ and $M_{\mathscr{C}}^{\mathscr{A}}(TS) \in \mathbb{F}^{n\times k}$ be as defined in Theorem 2.13. Then*

$$M_{\mathscr{C}}^{\mathscr{A}}(TS) = M_{\mathscr{C}}^{\mathscr{B}}(T) M_{\mathscr{B}}^{\mathscr{A}}(S).$$

Proof. Let $\mathscr{A} = \{u_{,} \ldots, u_k\}$ and calculate

$$\begin{aligned}
M_{\mathscr{C}}^{\mathscr{A}}(TS) &= \Big[C_{\mathscr{C}}\big(TSu_1\big) \ \cdots \ C_{\mathscr{C}}\big(TSu_k\big) \Big] && \text{by (2.20)} \\
&= \Big[M_{\mathscr{C}}^{\mathscr{B}}(T)\big(C_{\mathscr{B}}(Su_1)\big) \ \cdots \ M_{\mathscr{C}}^{\mathscr{B}}(T)\big(C_{\mathscr{B}}(Su_k)\big) \Big] && \text{by (2.21)} \\
&= M_{\mathscr{C}}^{\mathscr{B}}(T) \Big[C_{\mathscr{B}}(Su_1) \ \cdots \ C_{\mathscr{B}}(Su_k) \Big] && \text{matrix mult.} \\
&= M_{\mathscr{C}}^{\mathscr{B}}(T) M_{\mathscr{B}}^{\mathscr{A}}(S). && \text{by (2.20)}
\end{aligned}$$

□

The following diagram illustrates the content of Theorem 2.14.

$$\begin{tikzcd}
 & \mathscr{V} \arrow[dr, "T"] \arrow[dd, "C_{\mathscr{B}}"] & \\
\mathscr{U} \arrow[ur, "S"] \arrow[rr, "TS"] \arrow[ddd, "C_{\mathscr{A}}"] & & \mathscr{W} \arrow[ddd, "C_{\mathscr{C}}"] \\
 & \mathbb{F}^m \arrow[ddr, "M_{\mathscr{C}}^{\mathscr{B}}(T)"] & \\
\mathbb{F}^k \arrow[ur, "M_{\mathscr{B}}^{\mathscr{A}}(S)"] \arrow[rr, "M_{\mathscr{C}}^{\mathscr{A}}(TS) = M_{\mathscr{C}}^{\mathscr{B}}(T) M_{\mathscr{B}}^{\mathscr{A}}(S)"'] & & \mathbb{F}^n
\end{tikzcd}$$