Abstract Vector Spaces and Linear Transformations

(4/7/19)

Alex Nita

Abstract

We now step back from $\mathbb{R}^n$ and its standard coordinates $\sigma = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ to consider what happens when vector *properties alone* are considered, without explicit numerical realizations. This is the abstract vector space V. We then *re-introduce* coordinates, but this time *as an option*, as a *choice*. If we choose to work with real numbers $\mathbb{R}$, then the choice can be expressed as a linear isomorphism $\varphi_\beta : V \to \mathbb{R}^n$. If we choose the complex numbers then our choice would be expressed as $\varphi_\beta : V \to \mathbb{C}^n$. If we choose to work over a finite field $\mathbb{F}_p$, then our choice is $\varphi_\beta : V \to \mathbb{F}_p^n$.

This distinction between the vector space V and its *field of numbers* F ($F = \mathbb{R}$, $\mathbb{C}$, $\mathbb{F}_p$, etc.) is an important one, for in making it we both extend the range over which many theorems apply, and simultaneously clarify the *different roles* played by vectors as compared with numbers. Vectors and their linear transformations are *characterized by properties alone*, conceptually cordoned off from numerics. Numbers, on the other hand, are *chosen*, as clothing is chosen, to *present* the vectors and transformations in a certain light, for certain occasions and purposes. ***By this neat distinction we can understand which aspects depend on properties, and which on numbers***, and thereby *gain* in tools and techniques. *Rank* and *nullity*, for example, depend almost entirely on properties, while the various *canonical forms* of linear transformations depend to a much larger degree on the field of numbers.

To see how the range over which linear algebra extends beyond $\mathbb{R}^n$ by this process, take a look at function spaces. Examples such as the smooth functions $C^\infty(\mathbb{R}^n)$ or distributions $\mathcal{D}(\mathbb{R}^n)$ on $\mathbb{R}^n$ come readily to mind. Because certain of their properties are *manifestly vectorial*, e.g. additivity and scalar multiplication,

$$f, g \in C^\infty(\mathbb{R}^n), \quad c \in \mathbb{R} \quad \Longrightarrow \quad f + g,\ cf \in C^\infty(\mathbb{R}^n)$$

we can stop trying to 'solve equations,' like the *heat equation*,

$$\frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2} = \frac{\partial u}{\partial t}$$

and instead treat functions $u \in C^\infty(\mathbb{R}^n)$ as vectors and differentiation, e.g. $P = \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2} - \frac{\partial}{\partial t}$, as linear operators

$$P = \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2} - \frac{\partial}{\partial t} \in \mathcal{L}\big(C^\infty(\mathbb{R}^n)\big)$$

The heat equation then becomes

$$\boxed{Pu = 0}$$

Since the setting is similar to the familiar one of $\mathbb{R}^n$, we naturally seek an *analog* of the linear system $A\mathbf{x} = \mathbf{0}$: perhaps there is something *like* row-reduction for $Pu = 0$, or perhaps there are eigenvalues and eigenvectors, a coordinate system in which $Pu = \sum_{j=1}^{\infty} \lambda_j v_j$. Certainly, P has a certain *rank* and *nullity*, just like A. Such observations lead directly to a *vast reformulation* of linear differential equations in terms of linear algebra, called *functional analysis*, which has played a tremendous role in the mathematical physics of quantum mechanics.

Yet even the study of $\mathbb{R}^n$ itself, and its companion *matrix algebra*, benefit from this distinction. The Rank-Nullity Theorem is one example, which we can pluck right out of the conceptual structure developed here. But there is more: the distinction between coordinate-free $\mathbb{R}^n$ and its coordinatized realizations give us greater freedom of motion, even with concrete objects such as matrices A. We can *diagonalize* them or put them into other *canonical forms*, as the occasion requires. From here it is but a short step to *quadratic forms*, especially *symmetric* forms (aka *inner products* or *generalized dot products*), which allows us to understand, for example, things like the *second derivative test* in Calc 3.

In this section we lay down the basic definitions in terms of properties, then explore their first consequences. We'll illustrate the notions with examples of all sorts.

1 Definitions, Notation and Examples

1.1 Sequence Spaces

In order to intruduce certain key constructions and examples, let us begin with a reformulation of the n-tuple formulation of a vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

in $\mathbb{R}^n$ that allows for generalization to *infinite dimensions*! This is the idea of *infinite sequence spaces*:

Definition 1.1 We will not fuss over the abstract details, but let us agree that a **field of numbers** F (also called the **ground field**) is anything *like* the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, or the complex numbers $\mathbb{C}$, in that division of nonzero numbers is allowed, and *every* nonzero number x has an inverse x^{-1}. While it is *possible* to say everything in maximum abstractness in terms of a list of properties, in this class we remain content to denote by F either $\mathbb{R}$ or $\mathbb{C}$:

$$F = \mathbb{R} \text{ or } \mathbb{C}$$

For us, the gains in working with arbitrary fields F are negligible, since the real and complex cases are the motivation for the abstraction, in the end. We need to master these cases first. ■

Another way to look at a coordinatized vector $\mathbf{x} = \langle x_1, \ldots, x_n \rangle$ in $\mathbb{R}^n$ is ***as a function***: Let $I = \{1, \ldots, n\}$ be the **index set**, and

$$\boxed{\mathbf{x} \,:\, I \to \mathbb{R}}$$

the function taking *real values* for each $i \in I$, the **components** of $\mathbf{x}$,

$$\boxed{\mathbf{x}(i) \stackrel{\text{def}}{=} x_i}$$

From this point of view, the actual n**-tuple** $\langle x_1, \ldots, x_n \rangle$ is merely the list of values which $\mathbf{x}$ takes, each value $x_i \in \mathbb{R}$ associated uniquely to a specific index $i \in I$. This n-tuple can be interpreted as a finite sequence, too,

$$\mathbf{x} = \langle x_1, \ldots, x_n \rangle = (x_i)_{i=1}^n$$

Once we put the matter in these terms, we see how to generalize: by *freeing the index set I from being finite!*

Definition 1.2 Let us start with the **index set**, an *arbitrary* set in general,

$$\boxed{I \stackrel{\text{def}}{=} \{i \mid i \in I\}}$$

The most interesting examples for us are $I = \mathbb{N}$ or $\mathbb{R}$ or $\mathbb{R}^n$, as we will see. Once in possession of I, we can define a **vector** $\mathbf{x}$ to be a real-, complex-, or other F-valued function:

$$\boxed{\mathbf{x} : I \to F}$$

which is also denoted in terms of its values as

$$\boxed{\mathbf{x} \;=\; (x_i)_{i \in I}}$$

The **set of all I-tuple vectors** is denoted, with a nod to $\mathbb{R}^n$, by

$$\boxed{F^I = \{(x_i)_{i \in I} \mid x_i \in F,\ i \in I\}}$$

and called the **Cartesian product** or **set of all I-tuples**. When $I = \mathbb{N}$, we get the **space of infinite sequences**,

$$\boxed{\begin{aligned} \mathbb{R}^{\mathbb{N}} &= \{(x_1,\, x_2, \ldots) \mid x_n \in \mathbb{R},\ n \in \mathbb{N}\} \\ &= \{(x_n)_{n \in \mathbb{N}} \mid x_n \in \mathbb{R},\ n \in \mathbb{N}\} \\ &\qquad \text{and} \\ \mathbb{C}^{\mathbb{N}} &= \{(z_1, z_2, \ldots) \mid z_n \in \mathbb{C},\ n \in \mathbb{N}\} \\ &= \{(z_n)_{n \in \mathbb{N}} \mid z_n \in \mathbb{C},\ n \in \mathbb{N}\} \end{aligned}}$$

For purposes of complex analysis, where Laurent series indexed over all integers $\mathbb{Z}$ are important, we include the possibility of $\mathbb{Z}$-indexed sequences, too,

$$\boxed{\begin{aligned} \mathbb{R}^{\mathbb{Z}} &= \{(x_n)_{n \in \mathbb{N}} \mid x_n \in \mathbb{R},\ n \in \mathbb{Z}\} \\ \mathbb{C}^{\mathbb{Z}} &= \{(z_n)_{n \in \mathbb{N}} \mid z_n \in \mathbb{C},\ n \in \mathbb{Z}\} \end{aligned}}$$

We endow these sets with **pointwise addition** and **scalar multiplication**: for all $(x_i)_{i\in I}$ and $(y_i)_{i\in I} \in F^I$, and for all $c \in F$,

$$\boxed{\begin{aligned}(x_i)_{i\in I} + (y_i)_{i\in I} &\stackrel{\text{def}}{=} (x_i + y_i)_{i\in I}\\ c(x_i)_{i\in I} &\stackrel{\text{def}}{=} (cx_i)_{i\in I}\end{aligned}}$$

Stated in terms of functions, these are

$$\boxed{\begin{aligned}(\mathbf{x}+\mathbf{y})(i) &\stackrel{\text{def}}{=} \mathbf{x}(i) + \mathbf{y}(i)\\ \big(c\mathbf{x}\big)(i) &\stackrel{\text{def}}{=} c\mathbf{x}(i)\end{aligned}}$$

This completes our first generalization of $\mathbb{R}^n$ and $\mathbb{C}^n$. After we introduce the definition of an abstract vector space, we will see that our definition here gives $\mathbb{R}^I$ and $\mathbb{C}^I$ the character of a vector space (real and complex, respectively). ■

Example 1.3 Let

$$\mathbf{x} = (n)_{n\in\mathbb{N}} = (1, 2, 3, \dots), \quad \mathbf{y} = \big((-1)^n\big)_{n\in\mathbb{N}} = (-1, 1, -1, \dots) \in \mathbb{R}^{\mathbb{N}}$$

be two real sequences. Then, for example

$$\mathbf{x}(7) = x_7 = 7 \quad \text{and} \quad \mathbf{y}(7) = y_7 = (-1)^7 = -1$$

and

$$\begin{aligned}\mathbf{x}+\mathbf{y} &= (n)_{n\in\mathbb{N}} + \big((-1)^n\big)_{n\in\mathbb{N}}\\ &= \big(n + (-1)^n\big)_{n\in\mathbb{N}}\\ &= \boxed{(0, 3, 2, 5, \dots)}\end{aligned}$$

and

$$\begin{aligned}5\mathbf{x} &= 5(n)_{n\in\mathbb{N}}\\ &= (5n)_{n\in\mathbb{N}}\\ &= \boxed{(5, 10, 15, \dots)}\end{aligned}$$

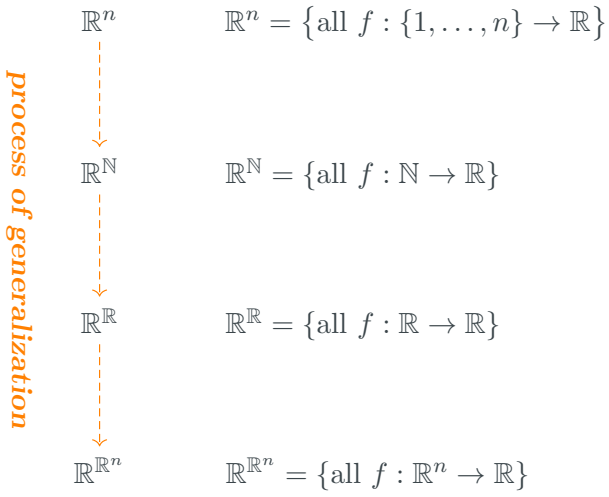
■

Exercise 1.4 Let $\mathbf{x} = (n)_{n\in\mathbb{N}}$, $\mathbf{y} = \big((-1)^n\big)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, as in the previous example. Find the general expression for the sequence $2\mathbf{x} - 3\mathbf{y}$, and use it to determine its 7th term, $2x_7 - 3y_7$. ■

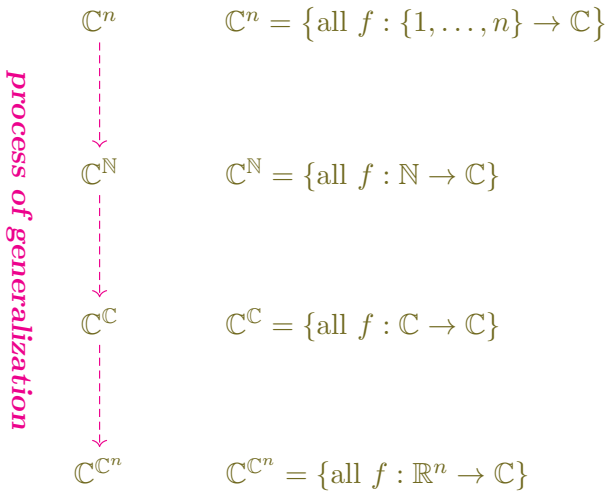
Exercise 1.5 Let $\mathbf{x} = \big(\frac{1}{n}\big)_{n\in\mathbb{N}}$ and $\mathbf{y} = \big(\frac{1}{n^2}\big)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}^{\mathbb{N}}$. Find the general expression for the sequence $2\mathbf{x} - 3\mathbf{y}$, and use it to determine its 7th term, $2x_7 - 3y_7$. ■

1.2 Function Spaces

If we go one step further, if extend the range of the index set I from $\mathbb{N}$ to $\mathbb{R}$, and by two more steps to $\mathbb{R}^n$ and $\mathbb{C}^n$, then we will have arrived at functions and function spaces as we know them from calculus[1] Generalizing from $\mathbb{R}^n$ to $\mathbb{R}^{\mathbb{N}}$ to $\mathbb{R}^{\mathbb{R}}$ to $\mathbb{R}^{\mathbb{R}^n}$,



and duplicating to the complex case,



We can also mix and match, for example:

$$\mathbb{C}^{\mathbb{R}} = \{\text{all } f : \mathbb{R} \to \mathbb{C}\}$$

and

$$\mathbb{R}^{\mathbb{C}} = \{\text{all } f : \mathbb{C} \to \mathbb{R}\}$$

The process outlined above was the process of progressively widening the domain of f, from $\{1, \ldots, n\}$ to $\mathbb{N}$ to $\mathbb{R}$ to $\mathbb{R}^n$, and similarly in the complex direction. We could

[1] Perhaps including also those with discontinuities.

also mess around with the range. We can widen it, first to $\mathbb{R}^n$, then to $M_{m,n}(\mathbb{R})$. Or we can widen it in other ways. For example

Example 1.6 Consider the set of all functions from $\mathbb{R}$ into $\mathbb{R}^n$,

$$(\mathbb{R}^n)^{\mathbb{R}} \stackrel{\text{def}}{=} \{\text{all } f : \mathbb{R} \to \mathbb{R}^n\}$$

or the set of all functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ (*not necessarily linear*!)

$$(\mathbb{R}^m)^{\mathbb{R}^n} \stackrel{\text{def}}{=} \{\text{all } f : \mathbb{R}^n \to \mathbb{R}^m\}$$

We can go further, and consider all $m \times n$ **matrix-valued functions**,

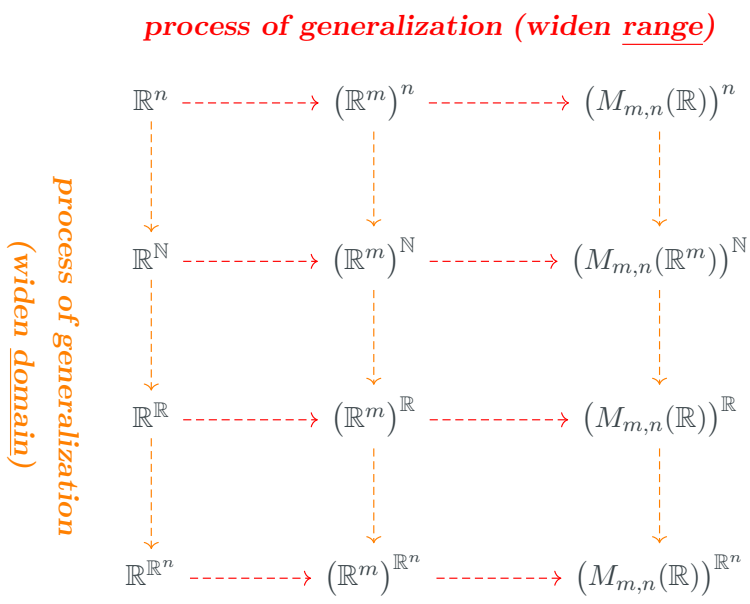
$$f : \mathbb{R} \to M_{m,n}(\mathbb{R}),$$
$$f(t) = \begin{pmatrix} f_{11}(t) & f_{12}(t) & \cdots & f_{1n}(t) \\ f_{21}(t) & f_{22}(t) & \cdots & f_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1}(t) & f_{m2}(t) & \cdots & f_{mn}(t) \end{pmatrix}$$

I suppose we could denote this accordingly, although the notation is becoming cluttered:

$$\big(M_{m,n}(\mathbb{R})\big)^{\mathbb{R}}$$

The examples could continue by now messing with the domain. But let us consider other ways of concocting function spaces. ■

Diagramatically, this is the 'horizontal' direction of the previous diagram:



Remark 1.7 Why did we choose *these* examples as ways of generalizing? **Answer:** We need things we can add and scale, and what's interesting about these examples is, in widening the domain we followed *intuition* more closely, enlarging what we started with, $\{1, \ldots, n\}$, progressively to $\mathbb{R}^n$ or $\mathbb{C}^n$. We could easily throw in $M_{m,n}(\mathbb{R})$ for the next installment. There was no particular reason to do it this way beyond following our nose. ***The index set I could be any set whatsoever*** **($\mathbb{N}$, $\mathbb{R}$, $\mathbb{C}^n$, even some weird probability sample space or something else).**

But the case with the *range* is very different. *We need the range to be something that allows adding and scaling*, because we defined those operations *componentwise!* This is why we chose $\mathbb{R}^n$ and $M_{m,n}(\mathbb{R})$, because *these* sets allow for precisely that. ■

There are other directions we could take this generalization, for example in the direction of *subsets of the above examples*. This is achieved by isolating some desirable property like *continuity* or *differentiability*.

Example 1.8 Consider the subset of $\mathbb{R}^{\mathbb{R}}$, the set of *all* real-valued functions of a real variable, consisting of *only the* ***continuous functions*** *(real-valued of one real variable)*:

$$C(\mathbb{R}) \stackrel{\text{def}}{=} \{ f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous} \}$$

We suppress the range or codomain of the continuous functions mainly for aesthetic purposes. Here, $C(\mathbb{R})$ means $C(\mathbb{R}, \mathbb{R})$, that is, the continuous functions are *real-valued*. In the analysis literature, $C(\mathbb{R})$ usually denotes *complex-valued* functions, i.e. $C(\mathbb{R}, \mathbb{C})$. In this course, however, we take real-valued functions as our default, and emphasize anything else accordingly. For example, $C(\mathbb{R}, \mathbb{C}^n)$ denotes continuous complex vector-valued functions $f : \mathbb{R} \to \mathbb{C}^n$.

Or consider the subset of $\mathbb{R}^{\mathbb{R}}$, the set of *all* real-valued functions of a real variable, consisting of *only the* ***differentiable functions*** *(real-valued of one real variable)*:

$$\mathcal{D}(\mathbb{R}) \stackrel{\text{def}}{=} \{ f : \mathbb{R} \to \mathbb{R} \mid f \text{ is differentiable} \}$$

Then there are the ***k-times continuously differentiable functions***, those for which all derivatives up to and including order k exist and are continuous,

$$C^k(\mathbb{R}) \stackrel{\text{def}}{=} \left\{ f : \mathbb{R} \to \mathbb{R} \;\middle|\; \frac{d^j f}{dx^j} \text{ exists and is continuous for all } 1 \leq j \leq k \right\}$$

Functions which have continuous derivatives of *all* orders are called **smooth functions** or **C^∞-functions**,

$$C^\infty(\mathbb{R}) \stackrel{\text{def}}{=} \left\{ f : \mathbb{R} \to \mathbb{R} \;\middle|\; \frac{d^j f}{dx^j} \text{ exists and is continuous for all } j \in \mathbb{N} \right\}$$

There are functions which are smooth, but which do not everywhere *equal* their own Taylor series, e.g.

$$f(x) = \begin{cases} e^{-1/x}, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0, \end{cases}$$

See Rudin, *Real and Complex Analysis*, exercise p. 418. This example stands in contrast to functions such as e^x and $\sin x$, which *do* equal their Taylor series everywhere (and in fact become *defined* in terms of their series centered at 0). Hence, we introduce the **(real) analytic functions**, at least in a neighborhood of a point $x = a$,

$$C^\omega(\mathbb{R}) \stackrel{\text{def}}{=} \left\{ f \in C^\infty(\mathbb{R}) \;\middle|\; f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n, \text{ for all } a \in \mathbb{R} \right\}$$

The most familiar functions which equal their own Taylor series are the **polynomials**, which our book denotes non-standardly as $\mathbb{P}$, but which we will denote also by $\mathbb{R}[x]$:

$$\mathbb{P} \text{ or } \mathbb{R}[x] \stackrel{\text{def}}{=} \left\{ p(x) \in C^\omega(\mathbb{R}) \;\middle|\; p(x) = \sum_{k=0}^{n} a_k x^k, \; n \in \mathbb{N} \right\}$$

In other words, ***polynomials are those analytic functions whose Taylor series are finite!*** For example, $p(x) = 2x^2 - 3x + 1$ satisfies $p'(x) = 4x - 3$ and $p''(x) = 4$, with higher derivatives $p^{(n)}(x) = 0$, so that, centering the Taylor series at $x = 0$ we have

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{p^{(n)}(0)}{n!}(x-0)^n &= p(0) + \frac{p'(0)}{1!}x + \frac{p''(0)}{2!}x^2 \\ &= 1 + \frac{-3}{1}x + \frac{4}{2}x^2 \\ &= 1 - 3x + 2x^2 \\ &= p(x) \end{aligned}$$

We may also filter the set of polynomials into subclasses, namely **polynomials of degree at most** n,

$$\mathbb{P}_n \text{ or } \mathbb{R}_n[x] \stackrel{\text{def}}{=} \{ p \in \mathbb{R}[x] \mid \deg(p) \leq n \}$$

It is an easy theorem to prove in general, that for any $p(x) = \sum_{k=0}^{n} a_k x^k$ we have all coefficients $a_k = \frac{p^{(k)}(0)}{k!}$, and we accordingly leave it as an exercise. ■

Exercise 1.9 Prove that all real polynomials are real analytic, i.e. that $\mathbb{R}[x] \subseteq C^\omega(\mathbb{R})$. This is most easily achieved by demonstrating that for any $p(x) = \sum_{k=0}^{n} a_k x^k \in \mathbb{R}[x]$, the coefficients satisfy $a_k = \frac{p^{(k)}(0)}{k!}$. ■

Remark 1.10 The above example deepened the left slot in the third row of the 'generalization' diagram above,

$$\mathbb{R}^{\mathbb{R}} \dashrightarrow \left(\mathbb{R}^m\right)^{\mathbb{R}} \dashrightarrow \left(M_{m,n}(\mathbb{R})\right)^{\mathbb{R}}$$

that of $\mathbb{R}^{\mathbb{R}}$, as follows (*the hooked arrows denote inclusion*):

$$\mathbb{P}_0 = \mathbb{R} \hookrightarrow \mathbb{P}_1 \hookrightarrow \cdots \hookrightarrow \mathbb{P}_n \hookrightarrow \cdots \hookrightarrow \mathbb{P} = \mathbb{R}[x] \hookrightarrow C^\omega(\mathbb{R}) \hookrightarrow C^\infty(\mathbb{R}) \hookrightarrow \cdots \hookrightarrow C^k(\mathbb{R}) \hookrightarrow \cdots \hookrightarrow C^2(\mathbb{R}) \hookrightarrow C^1(\mathbb{R}) \hookrightarrow \mathcal{D}(\mathbb{R}) \hookrightarrow C(\mathbb{R}) \hookrightarrow \mathbb{R}^{\mathbb{R}}$$

This is the 'third dimension' of the generalization diagram, behind the slot of $\mathbb{R}^{\mathbb{R}}$. *We could repeat this with each of the other slots in that row, by applying the above ideas to each of the component functions.* For example, a function $f \in (\mathbb{R}^m)^{\mathbb{R}}$ is just a **vector-valued function** $f : \mathbb{R} \to \mathbb{R}^m$,

$$f(t) = \begin{pmatrix} f_1(t) \\ \vdots \\ f_m(t) \end{pmatrix}$$

each component function f_i being a function in $\mathbb{R}^{\mathbb{R}}$, to which the above ideas apply. And, as noted above, a functoin $f \in \left(M_{m,n}(\mathbb{R})\right)^{\mathbb{R}}$ is just a **matrix-valued function** $f : \mathbb{R} \to M_{m,n}(\mathbb{R})$,

$$f(t) = \begin{pmatrix} f_{11}(t) & f_{12}(t) & \cdots & f_{1n}(t) \\ f_{21}(t) & f_{22}(t) & \cdots & f_{2n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1}(t) & f_{m2}(t) & \cdots & f_{mn}(t) \end{pmatrix}$$

each component function f_{ij} being a function in $\mathbb{R}^{\mathbb{R}}$. ***One could consider these paths, respectively in*** $\mathbb{R}^n$ ***and*** $M_{m,n}(\mathbb{R})$**.** In any case, we more or less automatically get $C^k(\mathbb{R}, \mathbb{R}^m)$, $C^\omega(\mathbb{R}, \mathbb{R}^m)$, $C^\infty\left(\mathbb{R}, M_{m,n}(\mathbb{R})\right)$, etc., in the same manner as for $\mathbb{R}^{\mathbb{R}}$ (modulo some easy theorems from real analysis). ■

Remark 1.11 In the real case (when the domain is $\mathbb{R}$ or $\mathbb{R}^n$), the filtration above is strict, in the sense that there are functions $f \in C^k(\mathbb{R})$ which are not in $C^{k+1}(\mathbb{R})$.

- $f(x) = |x|$ lies in $C(\mathbb{R})$ by not in $C^1(\mathbb{R})$.
- $f(x) = x|x|$ lies in $C^1(\mathbb{R})$ but not in $C^2(\mathbb{R})$.
- $f(x) = x^k|x|$ lies in $C^k(\mathbb{R})$ but not in $C^{k+1}(\mathbb{R})$.
- $f(x) = \begin{cases} e^{-1/x}, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0, \end{cases}$ lies in $C^\infty(\mathbb{R})$ but not in $C^\omega(\mathbb{R})$.

- Certainly $e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$ lies in $C^\omega(\mathbb{R})$ but not in $\mathbb{P}$.

The situation with complex differentiability is different, because by Cauchy's theorem a complex-differentiable function is automatically analytic,

$$\mathcal{D}_{\mathbb{C}}(\mathbb{C}) = C^\omega(\mathbb{C})$$

Important theorems in topology, calculus/real analysis and complex analysis establish the basic facts about all these sets, and among them is one key one: the assurance that adding two functions in one of these sets, or scaling one, lands us *back* in the set. I.e. these sets are *closed under addition and scalar multiplication.* ■

Exercise 1.12 Prove that $f(x) = x^k|x|$ lies in $C^k(\mathbb{R})$ but not in $C^{k+1}(\mathbb{R})$. ■

Remark 1.13 We could duplicate everything we achieved in the third row of the 'generalization' diagram in its 4th row:

$$\mathbb{R}^{\mathbb{R}^n} \dashrightarrow \left(\mathbb{R}^m\right)^{\mathbb{R}^n} \dashrightarrow \left(M_{m,n}(\mathbb{R})\right)^{\mathbb{R}^n}$$

In the first slot, $\mathbb{R}^{\mathbb{R}^n}$, we could consider continuous functions $C(\mathbb{R}^n)$, smooth functions $C^\infty(\mathbb{R})$, and the rest, just like we did in the single-variable case. This is the content of the *second semester* of real analysis—Calc 3 with proofs. The only thing that changes is polynomials look a little different: instead of $\mathbb{P} = \mathbb{R}[x]$, with x the only variable, we now have polynomials of n-variables,

$$\mathbb{R}[\mathbf{x}] = \mathbb{R}[x_1, \ldots, x_n]$$

For example, $p(x, y, z) = x^5y^2z^3 - 4x^2y + 20z^8 \in \mathbb{R}[x, y, z]$. In fact, this p has degree 10 (the highest sum of powers in any of its monomial terms), so lies in $\mathbb{R}_{10}[x, y, z]$ in fact. With this in mind, the deepening process looks like this:

$$\mathbb{R} = \mathbb{R}_0[\mathbf{x}] \hookrightarrow \mathbb{R}_1[\mathbf{x}] \hookrightarrow \cdots \hookrightarrow \mathbb{R}_n[\mathbf{x}] \hookrightarrow \cdots \hookrightarrow \mathbb{R}[\mathbf{x}] \hookrightarrow C^\omega(\mathbb{R}^n) \hookrightarrow$$
$$C^\infty(\mathbb{R}^n) \hookrightarrow \cdots \hookrightarrow C^k(\mathbb{R}^n) \hookrightarrow \cdots \hookrightarrow C^2(\mathbb{R}^n) \hookrightarrow C^1(\mathbb{R}) \hookrightarrow \mathcal{D}(\mathbb{R}^n) \hookrightarrow$$
$$C(\mathbb{R}^n) \hookrightarrow \mathbb{R}^{\mathbb{R}^n}$$

And as we proceeded with the single variable case, we do here as well: the second and third slots of the fourth row may be deepened in exactly the same way. ■

There are many other examples of highly refined function spaces, but we merely content ourselves with mentioning some of them, letting the reader wait until he or she

takes functional analysis for the full monty. The point here is that **analysis**, the differential theory behind calculus and complex variables, is the branch of math dedicated to intricate methods of filtering function spaces, in ways that lead down to polynomials, the simplest types of functions—polynomials are constructed out of arithmetic operations. Fine tuning of this type provides a crucial bridge between algebra and analysis, whose traversal leads eventually to complete solutions to differential equations, and along the way gives us spectral analysis (the method of eigenvalues) and other techniques which are fundamental in quantum mechanics and its generalization to quantum field theory—a current field of intense research.

Example 1.14 The following are subsets of $(\mathbb{R})^{\mathbb{R}^n}$, functions $f : \mathbb{R}^n \to \mathbb{R}$. We encourage the reader to explore the functional analysis and patial differential equations literature on these:

- Schwartz functions $\mathcal{S}(\mathbb{R}^n)$.
- Bump or test functions $C_c(\mathbb{R}^n)$.
- Functions vanishing at infinity $C_0(\mathbb{R}^n)$.
- Bounded functions $\mathcal{B}(\mathbb{R}^n)$.
- Riemann integrable functions $\mathcal{R}(\mathbb{R})$.
- Lebesgue square-integrable functions $L^2(\mathbb{R})$.
- Sobolev spaces $H^p(\mathbb{R}^n)$ and $W^{k,p}(\mathbb{R}^n)$.

Good places to look at these spaces are Reed and Simon, *Functional Analysis*; Kadison and Ringrose, *Fundamentals of the Theory of Operator Algebras, Volume I: Elementary Theory*; Duistermaat and Kolk, *Distributions: Theory and Applications*; Leoni, *A First Course in Sobolev Spaces*. The background required is a graduate real analysis course, e.g. Rudin's *Real and Complex Analysis*. ■

What about the second row, can we deepen $\mathbb{R}^{\mathbb{N}}$*, and by analogy also* $(\mathbb{R}^m)^{\mathbb{N}}$ *and* $\big(M_{m,n}(\mathbb{R})\big)^{\mathbb{N}}$*?* ***These are sequences in*** $\mathbb{R}$***,*** $\mathbb{R}^n$ ***and*** $M_{m,n}(\mathbb{R})$***, respectively.***

1.3 Sequence Spaces

In the previous section we saw how to take one of the rows of the generalization diagram, namely the third

$$\textbf{Row 3} \quad \mathbb{R}^{\mathbb{R}} \dashrightarrow \big(\mathbb{R}^m\big)^{\mathbb{R}} \dashrightarrow \big(M_{m,n}(\mathbb{R})\big)^{\mathbb{R}}$$

and fourth

$$\textbf{Row 4} \quad \mathbb{R}^{\mathbb{R}^n} \dashrightarrow \big(\mathbb{R}^m\big)^{\mathbb{R}^n} \dashrightarrow \big(M_{m,n}(\mathbb{R})\big)^{\mathbb{R}^n}$$

and deepen them by considering nested sequences of *subspaces*, which are subsets closed under addition and scalar multiplication. Can this be duplicated in the second row?

$$\textbf{Row 2} \quad \mathbb{R}^{\mathbb{N}} \dashrightarrow \big(\mathbb{R}^m\big)^{\mathbb{N}} \dashrightarrow \big(M_{m,n}(\mathbb{R})\big)^{\mathbb{N}}$$

The answer is, of course, Yes! But we cannot proceed in the same way, since continuity and differentiability make sense when the domain is $\mathbb{R}$ or $\mathbb{R}^n$, but not when it is $\mathbb{N}$. The criterion must be various sorts of *convergence*.

Example 1.15 (**Space of Convergent Sequences** c) A sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^{\mathbb{N}}$ may *converge*,

$$\boxed{\lim_{n\to\infty} x_n = x \ \text{ exists in } \mathbb{R}}$$

and theorems from Calc 2 guarantee that when two sequences, $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$, converge, x_n to x and y_n to y, then their sum $(x_n + y_n)_{n\in\mathbb{N}}$ converges to $x + y$,

$$\boxed{\lim_{n\to\infty}(x_n + y_n) = x + y = \lim_{n\to\infty}(x_n) + \lim_{n\to\infty}(y_n)}$$

That is, when dealing with convergenct sequences, limits distribute over sums. They similarly distribute over scalar multiplication: if $k \in \mathbb{R}$,

$$\boxed{\lim_{n\to\infty} kx_n = kx = k \lim_{n\to\infty} x_n}$$

This calls for a name and a notation:

$$\boxed{c \overset{\text{def}}{=} \big\{(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \,\big|\, \lim_{n\to\infty} x_n = x \text{ exists in } \mathbb{R}\big\}}$$

denotes the **space of convergent sequences** in $\mathbb{R}^{\mathbb{N}}$. The boxed identities (theorems in Calc 2) show that

$$(x_n)_{n\in\mathbb{N}},\ (y_n)_{n\in\mathbb{N}} \in c,\quad k \in \mathbb{R} \implies k(x_n)_{n\in\mathbb{N}},\ (x_n)_{n\in\mathbb{N}} + (y_n)_{n\in\mathbb{N}} \in c$$

i.e. c is closed under addition and scalar multiplication, so is a subspace of $\mathbb{R}^{\mathbb{N}}$. ■

Example 1.16 (**Space of Sequences Vanishing at Infinity** c_0) Let c_0 denote the **sequences vanishing at infinity**, or **space of null sequences**, those sequences in c converging to 0,

$$c_0 \stackrel{\text{def}}{=} \{(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid \lim_{n\to\infty} x_n = 0\}$$

By the previous example it is also closed under addition and scalar multiplication. The sequence

$$\left(\frac{1}{n}\right)_{n\in\mathbb{N}} \in c_0$$

for example. The sequence $\left(\tan\frac{1}{n}\right)_{n\in\mathbb{N}}$ also lies in c_0. ■

Example 1.17 (**Space of Finite Sequences** c_{00}) If two sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^{\mathbb{N}}$ *have only finitely many nonzero terms*, then so does their sum and any scalar multiple of them. Thus,

$$c_{00} \stackrel{\text{def}}{=} \{(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid \text{only finitely many terms } x_n \neq 0\}$$ ■

Example 1.18 (**Bounded Sequences** b) Let b denote the **set of all bounded sequences** in $\mathbb{R}^{\mathbb{N}}$,

$$b \stackrel{\text{def}}{=} \{(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid |x_n| < M < \infty, \text{ for all } n \in \mathbb{N}, \text{ for some } M > 0\}$$

The number M is called the bound, and can be defined as $M = \sup_{n\in\mathbb{N}} |x_n|$, where 'sup' means *supremum*, or *least upper bound*. ■

Remark 1.19 If a sequence $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ has only finitely many nonzero terms, then $\lim_{n\to\infty} x_n = 0$, so we see that $c_{00} \subseteq c_0$.

If we think of $\mathbb{R}^n$ as sequences of length n, $\mathbf{x} = \langle x_1, \ldots, x_n\rangle = (x_k)_{k=1}^n$, then by letting $x_k = 0$ for all $k > n$ we may include $\mathbb{R}^n$ in c_{00}. Since all convergent sequences are bounded (this is a theorem of real analysis), we finally conclude

$$\mathbb{R}^n \subseteq c_{00} \subseteq c_0 \subseteq c \subseteq b \subseteq \mathbb{R}^{\mathbb{N}}$$

or, in the sense of an 'inclusion' diagram,

$$\mathbb{R}^n \hookrightarrow c_{00} \hookrightarrow c_0 \hookrightarrow c \hookrightarrow b \hookrightarrow \mathbb{R}^{\mathbb{N}}$$

Note the analogy between c_0 and $C_0(\mathbb{R})$, the space of functions 'vanishing at infinity,' which in the real case means those $f \in C(\mathbb{R})$ satisfying $\lim_{x\to\infty} f(x) = 0$! ■

This is one way to deepen $\mathbb{R}^n$. Let's see what other sequence spaces we can concoct. What about sequences involved in *infinite series*? Recall that an infinite series

$$\sum_{n=1}^{\infty} x_n$$

is really a sequence $(s_N)_{N\in\mathbb{N}}$ of partial sums

$$s_N = \sum_{n=1}^{N} x_n$$

so that the series $\sum_{n=0}^{\infty} x_n$ converges by definition if the sequence of partial sums converges

$$S = \sum_{n=1}^{\infty} x_n = \lim_{N\to\infty} s_N = \lim_{N\to\infty} \sum_{n=1}^{N} x_n$$

We can apply the logic of sequences developed above to the case of partial sums:

Example 1.20 (**Convergent Series**) Since two series $\sum_{n=1}^{\infty} x_n$ and $\sum_{n=1}^{\infty} y_n$ in $\mathbb{R}$ are convergent iff their corresponding sequences of partial sums $(S_N)_{N\in\mathbb{N}}$ and $(T_N)_{N\in\mathbb{N}}$ converge, say to S and T, respectively, and since sums of convergent sequences converge, we see that sums of convergent series also converge,

$$\sum_{n=1}^{\infty} x_n + \sum_{n=1}^{\infty} y_n \stackrel{\text{def}}{=} \sum_{n=1}^{\infty} (x_n + y_n) = S + T$$

and similarly with scalar multiplication,

$$c\sum_{n=1}^{\infty} x_n \stackrel{\text{def}}{=} \sum_{n=1}^{\infty} cx_n = cS$$

we see that the space of sequences $(x_n)_{n\in\mathbb{N}}$ whose *series* $\sum_{n=1}^{\infty} x_n$ converge is closed under vector operations, and must be a subspace of $\mathbb{R}^{\mathbb{N}}$, called the **space of summable sequences**, and denoted

$$\ell \stackrel{\text{def}}{=} \left\{ (x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \;\middle|\; \sum_{n=1}^{\infty} x_n = S \in \mathbb{R} \right\}$$

The notation ℓ is by analogy with L in the function space of **Lebesgue-integrable functions**,

$$L(\mathbb{R}) \stackrel{\text{def}}{=} \left\{ f \in \mathbb{R}^{\mathbb{R}} \mid \int_{\mathbb{R}} f\, dx \text{ exists in } \mathbb{R} \right\}$$

The integral here isn't the Riemann integral, but the Lebesgue integral of graduate real analysis (cf. Rudin, *Real and Complex Analysis*).

Since the sequence of partial sums of any convergent series is convergent, it is bounded,

$$M = \sup_{N\in\mathbb{N}} |s_N| = \sup_{N\in\mathbb{N}} \left| \sum_{n=1}^{N} x_n \right| < \infty$$

Therefore, the original sequence $(x_n)_{n\in\mathbb{N}}$ must be bounded, meaning that each term in the *series* must satisfy

$$|x_n| < N < \infty$$

for some $0 < N < \infty$. This is Cauchy's criterion for convergent sequences in $\mathbb{R}$, and concerns precisely the issue of limits that seemed so technical in Calc 1 (another topic covered in undergrad real analysis). Well, this means

$$\boxed{\ell \subseteq b}$$

Indeed, by an elementary theorem from Calc 2, proved in undergraduate real analysis, we know that if a series $\sum_{n=1}^{\infty} x_n$ converges, then $\lim_{n\to\infty} x_n = 0$, so that in fact

$$\boxed{\ell \subseteq c_0}$$

We have extend our inclusion chain above by one link:

$$\mathbb{R}^n \hookrightarrow c_{00} \hookrightarrow \ell \hookrightarrow c_0 \hookrightarrow c \hookrightarrow b \hookrightarrow \mathbb{R}^{\mathbb{N}}$$

In fact, we can cram a few more things in there. ■

Example 1.21 (**p-Summable Sequences**) Let $p \in \mathbb{R}$ and recall that for any $a > 0$ exponentiating a by p means $a^p \overset{\text{def}}{=} e^{p \ln a}$. Now let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}^{\mathbb{N}}$, and consider the **set of all p-summable sequences**,

$$\ell^p \overset{\text{def}}{=} \left\{ (x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \;\middle|\; \left(\sum_{n=1}^{\infty} |x_n|^p \right)^{1/p} < \infty \right\}$$

For example, ℓ^1 is the **space of sequences with absolutely convergent series**,

$$\boxed{\ell^1 = \left\{ (x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \;\middle|\; \sum_{n=1}^{\infty} |x_n| < \infty \right\}}$$

Since a Calc 2 theorem (proved in undergrad analysis) tells us that any absolutely convergent series is convergent, we conclude that

$$\boxed{\ell_1 \subseteq \ell}$$

The **space of sequences with square-summable series**,

$$\boxed{\ell^2 = \left\{ (x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \;\middle|\; \left(\sum_{n=1}^{\infty} |x_n|^2 \right)^{1/2} < \infty \right\}}$$

should look very familiar. The length of a vector $\mathbf{x} = \langle x_1, \ldots, x_n \rangle \in \mathbb{R}^n$ has a similar form,

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$$

Thus, square-summable sequences are just 'infinite vectors' $\mathbf{x} = \langle x_1, x_2, \ldots \rangle \in \mathbb{R}^{\mathbb{N}}$ whose length is finite,

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \Big(\sum_{i=1}^{\infty} |x_i|^2 \Big)^{1/2} < \infty$$

As a result of Young's Inequality and Minkowski's Inequality (theorems in undergrad real analysis), sums and scalar multiples of p-summable sequences are again p-summable, which means ℓ^p spaces are subspaces of $\mathbb{R}^{\mathbb{N}}$.

Moreover, by a theorem in undergrad analysis, cf Theorem 17.21 in Yeh, *Real Analysis: The Theory of Measure and Integration*, 2nd Ed., we have

$$1 \leq p < q < \infty \implies \ell^q \subseteq \ell^p$$

Combining with $\ell^1 \subseteq \ell$, we conclude that

$$\boxed{1 \leq p < q < \infty \implies \ell^q \subseteq \ell^p \subseteq \ell^1 \subseteq \ell}$$

Lastly, consider the question of what happens to ℓ^p as $p \to \infty$, that is what happens when we consider the limit $\lim_{p\to\infty} \big(\sum_{n=1}^{\infty} |x_n|^p \big)^{1/p}$ for any sequence $(x_n)_{n\in\mathbb{N}} \in \ell^1$. By Theorem 16.50 in Yeh, *Real Analysis: The Theory of Measure and Integration*, 2nd Ed.,

$$\lim_{p\to\infty} \Big(\sum_{n=1}^{\infty} |x_n|^p \Big)^{1/p} = \sup_{n\in\mathbb{N}} |x_n|$$

and since $\ell^p \subseteq \ell^1 \subseteq \ell$, this must by finite, whence we see that

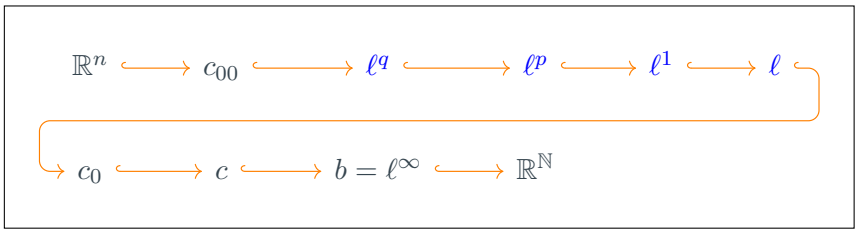
$$\boxed{b \cap c = \text{“} \lim_{p\to\infty} \ell^p \text{”} \subseteq \ell^1 \subseteq \ell}$$

and for this reason b is sometimes denoted ℓ^∞,

$$\boxed{b = \ell^\infty}$$

Note, however, that b is not contained in any ℓ^p or even in ℓ, because b consists of *merely bounded* sequences, not necessarily *convergent* sequences. ■

Remark 1.22 We have arrived at the following deepening of $\mathbb{R}^{\mathbb{N}}$: for all $1 \leq p < q < \infty$,



There is more to say, of course, but let us content ourselves with this overview. Needless to say, the procedure can be carried through on the other terms of that second row of the 'generalization' diagram. ■

1.4 Abstract Vector Spaces and Their Basic Properties

Let's return to the remark we made above concerning our process of generalizing $\mathbb{R}^n$, that it occurs in two directions:

(1) Widening the possibilities for the *domain* of any vector $\mathbf{x} : \{1, \ldots, n\} \to \mathbb{R}$ from $\{1, \ldots, n\}$ to *all conceivable* 'index sets' I.

(2) Widening the *codomain* (the target space containing the range) *in certain specific directions*, first from $\mathbb{R}$ to $\mathbb{R}^n$ and thence to $M_{m,n}(\mathbb{R})$. This is about as far as we went, though we could squeeze out a couple more of these types of codomain spaces if we really wanted to—we don't at the moment.

Why did we choose $\mathbb{R}^n$ and $M_{m,n}(\mathbb{R})$ as our extensions in (2)? **Answer:** *because what we need to have is things that add and scale like vectors in $\mathbb{R}^n$.* And if f *takes values* in $\mathbb{R}^n$ or $M_{m,n}(\mathbb{R})$, then by Theorems 1.6 and 1.11, 'Vectors in $\mathbb{R}^n$,' and their matrix analogs Theorems 2.1 and 2.2, 'Matrices and Linear Transformations,' these will add like vectors *in the sense of those eight properties.*

These eight properties form the unifying principle behind all our examples above, because it turns out that *all* proofs about the algebraic vectorial side of $\mathbb{R}^n$ depend on *just these eight properties*. This is the reason, it seems, that the eight properties *are* the definition of the generic, abstract vector space.

Definition 1.23 A **vector space** V, over a field of numbers F (which we take to be either $\mathbb{R}$ or $\mathbb{C}$ here), is a set V on which the operations **addition** $+$ and **scalar multiplication** $\cdot$ satisfy the following conditions: for all $u, v, w \in V$ and $a, b, 1 \in F$,

(1) $u + v = v + u$ (***commutativity of addition***)

(2) $(u + v) + w = u + (v + w)$ (***associativity of addition***)

(3) There is a **zero vector** $\vec{0} \in V$ satisfying $v + \vec{0} = v$ for all $v \in V$

(4) Every vector $v \in V$ possesses a **negative** $u \in V$, characterized by $u + v = 0$

(5) $1v = v$ for all $v \in V$ (here, $1 \in F = \mathbb{R}$ or $\mathbb{C}$)

(6) $(ab)v = a(bv)$ (***associativity of scalar multiplication***)

(7) $a(u + v) = au + av$ (***distributivity over vector addition***)

(8) $(a + b)v = av + bv$ (***distributivity over scalar addition***)

If $F = \mathbb{R}$, we call V a **real vector space**. If $F = \mathbb{C}$, we call V a **complex vector space**. ■

Lemma 1.24 (Cancellation Law) *Let u, v and w be any vectors in V. Then,*

$$u + w = v + w \implies u = v$$

That is, we may 'cancel w' or 'subtract w from both sides.'

Proof: Let z be the negative of w, so that $w + z = z + w = 0$. Then, by properties (2), (3)

$$u \overset{(3)}{=} u+0 \overset{\text{def z}}{=} u+(w+z) \overset{(2)}{=} (u+w)+z \overset{\text{hyp}}{=} (v+w)+z \overset{(2)}{=} v+(w+z) \overset{\text{def z}}{=} v+0 \overset{(3)}{=} v \quad ■$$

Example 1.25 Let V be a vector space over $F = \mathbb{R}$ or $\mathbb{C}$. Then for all $v \in V$ we have

$$0v = \vec{0}$$

where $0 \in F$, and $\vec{0} \in V$.

Solution: Use properties (3) and (8):

$$0v + 0v \overset{(8)}{=} (0+0)v = 0v \overset{(3)}{=} 0v + \vec{0}$$

Subtracting $0v$ from both sides shows that $0v = \vec{0}$. ■

Example 1.26 (a) Show that the negative of any vector v is unique. We accordingly denote it $-v$.

(b) Show that for any $v \in V$, we have $(-1)v = -v$.

Solution: (a) Suppose u and w are both negatives of v. Then $u + v = 0 = w + v$, so subtracting v from both sides gives $u = w$. (b) Use part (a):

$$(-1)v + v \overset{(5)}{=} (-1)v + 1v \overset{(8)}{=} ((-1)+1)v = 0v \overset{\text{prev ex}}{=} \vec{0}$$

Thus, since $(-1)v + v = \vec{0}$, part (a) kicks in to give us $(-1)v = -v$. ■

Example 1.27 Show that if any vector $v \in V$ satisfies $v = -v$, then $v = \vec{0}$. *Hint: Use the previous example.*

Solution: If $v = -v$, then by the previous exercise, part (b), $v = -v = (-1)v$, so subtracting $(-1)v$ from both sides gives

$$2v = (1+1)v \overset{(8)}{=} 1v+1v \overset{(5)}{=} 1v+v \overset{\text{hyp}}{=} 1v+(-v) \overset{(b)}{=} 1v+(-1)v \overset{(8)}{=} (1+(-1))v = 0v \overset{\text{prev ex}}{=} \vec{0}$$

But then dividing by 2,

$$v \overset{(5)}{=} 1v = \left(\frac{1}{2} \cdot 2\right)v \overset{(6)}{=} \frac{1}{2}(2v) \overset{\text{hyp}}{=} \frac{1}{2}\vec{0} \overset{\text{exercise}}{=} \vec{0}$$ ■

Exercise 1.28 Show that the zero vector $\vec{0}$ is *unique*. That is, suppose there were another zero vector $\vec{0}'$ satisfying (3) in the definition of V, and demonstrate, using only the properties of V, that $\vec{0} = \vec{0}'$. ■

Exercise 1.29 Show from the definition of V, the above lemma, and the previous exercise that for any scalar $a \in F$ we have $a\vec{0} = \vec{0}$. ■

Definition 1.30 Let V be a vector space over $F = \mathbb{R}$ or $\mathbb{C}$. A nonempty subset U of V is called a vector **subspace** if U is closed under addition and scalar multiplication:

(1) $u, v \in U \implies u + v \in U$.

(2) $u \in U$ and $c \in F \implies cu \in U$. ■

Remark 1.31 Since U is *nonempty* it will *have to* contain the zero vector 0, but this is now a consequence of a proposition. Don't get lost in these details. The bigger point is that we want to distinguish between a *subset* and a *subspace*: a subspace is supposed to be a *special type* of subset, one that is itself a *vector space*. This is the main idea. The notion that this is *due* to the closure properties is an innocent but slightly confused mixing of related ideas. Let me explain, maybe this will help you in reading math books.

You should always keep an eye on the cultural component of math in your learning. The professional mathematician gets quickly embroiled in the technical question, 'How best to reach back in the logical chain of things in order to *anchor* a desired concept to an axiom or some other "more basic" fact?' Once an anchor is hit upon (*such as the two closure properties in the definition above*), the mathematician proceeds to *prove* that the anchor indeed logically supports the desired conclusion, in this case that a subspace U of V (according to our closure-property definition) is a vector space itself.

But this sort of thing isn't always satisfying to the student, unless there is a *good reason for the anchoring itself*. The anchor is supposed to *support* a difficult or weird conclusion, to give a *reason* and an *explanation* for it. But this is rarely how it goes. Sometimes it happens, and it happens in modern math more than anywhere else, that the anchor is *logically equivalent* to the goal. In our case, the anchor, the definition of a subspace in terms of closure under $+$ and $\cdot$, is logically equivalent to the goal of viewing a subspace as a subset which is a vector space. When this sort of logical equivalence occurs, one will encounter in the literature a reversal of the order in some books. One book will take one statement for a goal and the other for an anchor, and another book will proceed in the reverse direction.

This leaves the student with the impression that there is no *best* starting point when two or more things are logically equivalent. Of course, this is just the nature of 'logical equivalence,' it is bi-directional. 'Starting points' and 'ending points' appear, logically, to be a matter of perspective, and therefore convenience, only. It is the modern way, the way of convenience. But convenience itself is perspectival. What is convenient to a professional mathematician, intent upon organizing things towards a certain purpose—or, perhaps, towards no purpose at all, just personal taste—may not be convenient to a student learning the material for the first time, who needs *good reasons* for things, especially *unusual, counterintuitive things*. When one encounters a *more obscure* reason for something perfectly simple, one should inquire more deeply into the matter, for likely there is a hidden *cultural* reason for this. There might not be a satifying answer to culture, either, but this is how curiosity leads towards

answers, if there are any to be had. [2]

In this course, we will try to lay the formalism over an intuitively understandable progression from simple concepts to those more complex. To return to our case of a subspace U defined as closed under addition and scalar multiplication, and from this deducing the more satisfying statement that U must on that account be a vector space itself (and so containing 0), let us say this: the idea here is convenience, that 2 properties (as in the definition of a subspace) are easier to remember than 8 (as in the definition of a vector space). One therefore takes the 2-property definition and works once and for all to deduce the other 8. Thereafter we only have to check these 2, for we know they imply the other 8. ■

Proposition 1.32 *If U is a subspace of V according to our Definition 1.30, then U is a vector space.*

Proof: We need to check the list of 8 properties of Definition 1.23. From the fact that U is nonempty and closed under $+$ and $\cdot$ (scalar mult.), as per Definition 1.30, we can conclude that some vector u lies in U, and therefore, by Example 1.26, that its negative $-u = (-1)u$ also lies in U, being a scalar multiple of u. Thus U contains all negatives of its members, which is property (4) of a vector space. Also, $\vec{0} \in U$, since $\vec{0} = u + (-u)$ for any $u \in U$. This is property (3). Properties (1)-(2) and (5)-(8) hold automatically, since they hold in V and U is closed under $+$ and $\cdot$. ■

Definition 1.33 Let $U_1, U_2, \ldots, U_k$ be subspaces of V, and define their **sum** to be

$$U_1 + U_2 + \cdots + U_k \stackrel{\text{def}}{=} \{u_1 + u_2 + \cdots + u_k \in V \mid u_i \in U_i \text{ for all } i\}$$

The sum is called a **direct sum** if additionally any two subspaces intersect trivially, $U_i \cap U_j = \{0\}$, and in this case we write

$$U_1 \oplus U_2 \oplus \cdots \oplus U_k \equiv \bigoplus_{i=1}^{n} U_i$$

This idea works for infinitely many subspaces $\{U_i\}_{i \in I}$. We have the sum $\sum_{i \in I} U_i$ and the direct sum $\bigoplus_{i \in I} U_i$. The elements of each are *finite* sums $u_{i_1} + \cdots + u_{i_k}$. ■

Exercise 1.34 Show that the intersection $\bigcap_{i \in I} U_i$ and sum $\sum_{i \in I} U_i$ of any number of subspaces U_i of V are again subspaces. ■

[2] In fact, I think the above remarks constitute a major reason for the difficulty of math. It is today a bit of a forest of logical equivalences mixed with uni-directional implications, the beginning and end of which are lost in the haze. In principle, set theory and logic form the foundation of all math, but geometry sits somewhat uneasily with this way of thinking. In any case, to the student encountering this tangle for the first time it can be discouraging without some sort of Ariadne's thread, for Descartes' original notion of clarity is mostly local, the global proving somewhat elusive.

2 Bases and Subspaces

As with $\mathbb{R}^n$, so here we must consider the 'generation' of a subspace from a spanning set, and conversely find a spanning set to generate the subspace. We list the definitions here:

Definition 2.1 Suppose S is a subset of a vector space V. Take n vectors $v_1, \ldots, v_n$ in S and n scalars $a_1, \ldots, a_n$ in F, and call the sum of scalar multiples

$$a_1 v_1 + \cdots + a_n v_n$$

a **linear combination** of the vectors $v_1, \ldots, v_n$. The result of a linear combination of vectors is another vector in V. This is why we say S **generates** all the vectors in its **span**, which is the set of all linear combinations of vectors in S,

$$\operatorname{span}(S) \stackrel{\text{def}}{=} \{a_1 v_1 + \cdots + a_n v_n \mid a_i \in F,\ v_i \in S,\ n \in \mathbb{N}\}$$

The relationship between *arbitrary* vectors and linear combinations of *known* vectors is precisely the question of *coordinates*. Anyway, the scalars $a_i \in F$ are called the **coefficients** of the linear combination, or more generally the β-**coordinates**. ■

Definition 2.2 By Example 1.25 above we know that $0v = \vec{0}$ for all $v \in V$. Therefore, a linear combination of the form $0v_1 + \cdots + 0v_n$ equals $\vec{0} + \cdots + \vec{0} = \vec{0}$. We call this the **trivial representation of** $\vec{0}$ in the vectors $v_1, \ldots, v_n$. Since $\vec{0}$ thus always has the trivial representation, we shall be interested in vectors v_i which possess a **nontrivial representation of** $\vec{0}$, a linear combination equal to $\vec{0}$

$$a_1 v_1 + \cdots + a_n v_n = \vec{0}$$

but in which *not all* coefficients a_i equal 0. In this case we say the vectors $v_1, \ldots, v_n$ are **linearly dependent**. If the vectors only allow for the trivial representation of $\vec{0}$ then we say $v_1, \ldots, v_n$ are **linearly *in*dependent**. ■

We have the abstract analog of Theorem 1.26 in our 'Bases, Coordinates and Representations' notes:

Theorem 2.3 (Characterization of Linear Independence) *Let V be a real or complex vector space, and let β be a nonempty collection of nonzero vectors in V, possibly infinite. Then, β is linearly independent iff none of its vectors is a linear combination of the remainder, i.e.*

$$\beta \text{ is linearly independent} \iff \text{no } v \in \beta \text{ lies in } \operatorname{span}\big(\beta \backslash \{v\}\big)$$

■

We also have the analog of Theorem 3.7 in the 'Bases, Coordinates and Representations' notes, but leave the proof to the appendix. The proof, too, is entirely analogous, which is why I'm relegating it to the back.

Theorem 2.4 (Equivalent Characterizations of Bases) *Let β be a subset of a vector space V over $F = \mathbb{R}$ or $\mathbb{C}$. The following are logically equivalent statements about β:*

(1) *β is a **basis** for V.*

(2) *Every nonzero vector $v \in V$ has a **unique representation** in β, meaning*

$$v = a_1 v_1 + \cdots + a_k v_k$$

*for unique $a_i \in \mathbb{R}$ (called β-**coordinates**) and $v_i \in \beta$.*

(3) *$\operatorname{span}(\beta) = V$ but no $v \in \beta$ lies in the span of $\beta \backslash \{v\}$.*

(4) *β is a **minimal spanning set**, meaning $\operatorname{span}(\beta) = V$ but no proper subset $\tilde{\beta}$ of β spans V.*

(5) *β is a **maximal linearly independent set**, meaning β is linearly independent, but no strictly larger extension γ of β, i.e. $\beta \subset \gamma$, is linearly independent.* ■

Corollary 2.5 *If V is a vector space, then a collection of vectors $\beta = \{v_1, \ldots, v_n\}$ is a basis for V iff*

$$\boxed{V = \operatorname{span}(v_1) \oplus \cdots \oplus \operatorname{span}(v_n) = \bigoplus_{i=1}^{n} \operatorname{span}(v_i)}$$ ■

Theorem 2.6 (Existence of a Basis) *Let V be a nontrivial vector space and I a linearly independent set of vectors in V which is contained in a spanning set S,*

$$I \subseteq S \subseteq V, \quad I \text{ linearly independent and } V = \operatorname{span}(S)$$

then there exists a basis β for V 'between I and S,'

$$\boxed{I \subseteq \beta \subseteq S}$$

From this we conclude that:

(1) *Any nonzero vector space has a basis.*

(2) *Any linearly independent set in V is contained in a basis.*

(3) *Any spanning set in V contains a basis.* ■

Lemma 2.7 *If V is a vector space and S and T are subsets of V, then the following hold:*

(1) $S \subseteq T \subseteq V \implies \operatorname{span}(S) \subseteq \operatorname{span}(T)$

(2) $S \subseteq T \subseteq V$ *and* $\operatorname{span}(S) = V \implies \operatorname{span}(T) = V$

(3) $\operatorname{span}(S \cup T) = \operatorname{span}(S) + \operatorname{span}(T)$

(4) $\operatorname{span}(S \cap T) \subseteq \operatorname{span}(S) \cap \operatorname{span}(T)$ ■

Corollary 2.8 (All Subpaces Have Complements) *If U is a nontrivial subspace of a vector space V, then there exists a subspace T of V such that*

$$\boxed{V = U \oplus T}$$ ■

This corollary is also true if $U = \{\vec{0}\}$, but why would we ever bother complementing $\{\vec{0}\}$ with V? We already know that $V = \{\vec{0}\} \oplus V$.

Theorem 2.9 (Replacement Theorem) *If V is a vector space such that $V = \text{span}(S)$ for some subset S of V with $|S| = n$, and if B is a linearly independent subset of V with $|B| = m$, then*

(1) $m \leq n$

(2) *There exists a subset C of S with $|C| = n - m$ such that $V = \text{span}(B \cup C)$* ■

Corollary 2.10 (Dimension) *If V is a vector space with a finite spanning set, then every basis for V contains the same number of vectors, and this number is called the* **dimension** *of V.* ■

The number may be an infinite cardinal number, but it's still a measure of the size of V, because even infinite cardinals come in different sizes. Let's make this explicit, and along the way see what happens to finite-dimensional vector spaces

Corollary 2.11 *If V is a finite-dimensional vector space with $\dim(V) = n$, then the following hold:*

(1) *Any finite spanning set for V contains at least n vectors, and a spanning set for V that contains exactly n vectors is a basis for V.*

(2) *Any linearly independent subset of V that contains exactly n vectors is a basis for V.*

(3) *Every linearly independent subset of V can be extended to a basis for V.* ■

Theorem 2.12 *Any two bases for a vector space V have the same cardinality, even infinite spaces. This cardinality is called the* **dimension** *of V.* ■

Corollary 2.13 *If V is a vector space and U is a subspace of V:*

(1) $\dim(U) \leq \dim(V)$

(2) $\dim(U) = \dim(V) < \infty \implies S = V$

(3) *V is infinite-dimensional iff it contains an infinite linearly independent subset.*

■

Theorem 2.14 *Let V be a vector space.*

(1) *If B is a basis for V and $B = B_1 \cup B_2$, where $B_1 \cap B_2 = \varnothing$, then*

$$V = \operatorname{span}(B_1) \oplus \operatorname{span}(B_2)$$

(2) *If*

$$V = S \oplus T$$

and we have bases B_1 for S and B_2 for T, then $B_1 \cap B_2 = \varnothing$ and $B = B_1 \cup B_2$ is a basis for V. ■

Theorem 2.15 *If S and T are subspaces of a vector space V, then*

$$\boxed{\dim(S) + \dim(T) = \dim(S + T) + \dim(S \cap T)}$$

As a consequence,

$$\boxed{V = S \oplus T \iff \dim(V) = \dim(S) + \dim(T)}$$ ■

Corollary 2.16 *If S and T are subspaces of a vector space V, then*

(1) $\dim(S + T) \leq \dim(S) + \dim(T)$

(2) $\dim(S + T) \leq \max\{\dim(S), \dim(T)\}$*, if S and T are finite-dimensional.* ■

Theorem 2.17 (Direct Sums) *If $\mathcal{F} = \{S_i \mid i \in I\}$ is a family of subspaces of a vector space V such that $V = \sum_{i \in I} S_i$, then the following statements are equivalent:*

(1) $V = \bigoplus_{i \in I} S_i$.

(2) $\vec{0} \in V$ *has a* **unique expression** *as a sum of vectors each from different S_i, namely as a sum of zeros: for any distinct $j_1, \ldots, j_n \in I$, we have*

$$v_{j_1} + v_{j_2} + \cdots + v_{j_n} = \vec{0} \text{ and } v_{j_i} \in S_{j_i} \text{ for each } i \implies v_{j_1} = \cdots = v_{j_n} = \vec{0}$$

(3) *Each $v \in V$ has a* **unique expression** *as a sum of distinct $v_{j_i} \in S_{j_i} \backslash \{\vec{0}\}$,*

$$v = v_{j_1} + v_{j_2} + \cdots + v_{j_n}$$

(4) *If γ_i is a basis for S_i, then $\gamma = \bigcup_{i \in I} \gamma_i$ is a basis for V. If V is finite-dimensional, then this may be restated in terms of* ordered *bases γ_i and γ.* ■

3 Appendix 1: Proofs of the Basis and Subspace Theorems

Proof of Theorem 2.3:

Proof: The proof is essentially the same. Suppose first β is linearly independent. If there were some $v \in \beta$ lying in $\text{span}(\beta\backslash\{v\})$, so that $v = \sum_{j=1}^{k} a_j v_j$ for some other $v_j \in \beta$, then by subtracting v from both sides we would have a nontrivial representation of $\vec{0}$ in β,

$$\vec{0} = \sum_{j=1}^{k} a_j v_j + (-1)v$$

an impossibility. Note that we used here $-v = (-1)v$, which was proved in Example 1.26.

Conversely, if no v lies in the span of the remaining vectors in β, then any β-representation of $\vec{0}$ would *have* to be trivial, otherwise we contradict our spanning assumption: if, say, $\sum_{j=1}^{k} a_j v_j = \vec{0}$ for some $v_j \in \beta$ but not all $a_j = 0$, then one of them, say $a_i \neq 0$. We can solve for v_i and divide through by a_i to get the contradiction,

$$v_i = \sum_{j \neq i} \Big(-\frac{a_j}{a_i}\Big) v_j$$

which exhibits v_i, a member of β, as a linear combination of other vectors $v_j \in \beta$: a contradiction. We therefore have to conclude that all $a_j = 0$, and so β is linearly independent. ■

Proof of Theorem 2.4

Proof: (1) $\Longrightarrow$ (2): If β is a basis, it is linearly independent, and if $v \in V$ is a nonzero vector which potentially may have two different β-representations,

$$v = a_1 u_1 + \cdots + a_n u_n = b_1 v_1 + \cdots + b_m v_m$$

where the a_i and b_i are scalars in $F = \mathbb{R}$ or $\mathbb{C}$, and the u_i and v_i are vectors in β, then grouping any vectors u_{i_j} and v_{i_j} that are equal, we have

$$\begin{aligned} 0 = (a_{i_1} - b_{i_1})u_{i_1} + \cdots + (a_{i_k} - b_{i_k})s_{i_k} \qquad & (u_{i_j} = v_{i_j}) \\ + a_{i_{k+1}} u_{i_{k+1}} + \cdots + a_{i_n} u_{i_n} \qquad & (\textit{only the scalar multiples of the } u_i) \\ + b_{i_{k+1}} v_{i_{k+1}} + \cdots + b_{i_m} v_{i_m} \qquad & (\textit{only the the scalar multiples of the } v_i) \end{aligned}$$

implies $a_{i_{k+1}} = \cdots = a_{i_n} = b_{i_{k+1}} = \cdots = b_{i_m} = 0$, so $n = m = k$, and for $j = 1, \ldots, k$ we have $a_{i_j} = b_{i_j}$, so that $v = \sum_{j=1}^{k} a_i v_i$ in fact has only one β-representation (after relabeling and tyding up), so it is uniquely represented in β.

(2) $\Longrightarrow$ (1): If every vector $v \in V$ has a unique β-representation $v = \sum_{j=1}^{k} a_j v_j$, then $\vec{0} \in V$ does, too. But $\vec{0}$ already has the trivial representation, $\vec{0} = \sum_{j=1}^{k} 0 v_j$ according to Example 1.25, and this is the only one allowed, which means β is linearly independent. Since every vector is represented in β, it also spans V, and β is a basis.

(1) $\Longleftrightarrow$ (3): Suppose (1), that β is a basis. Then β spans V and is linearly independent, which by Theorem 2.3 means no $v \in \beta$ lies in the span of $\beta\backslash\{v\}$, so that β satisfies (3). Conversely, if β satisfies (3), then $\text{span}(\beta) = V$ and no $v \in \beta$ lies in the span of $\beta\backslash\{v\}$, which again by Theorem 2.3 means β is linearly independent.

(1) $\Longleftrightarrow$ (4): If β is a linearly independent spanning set and T is a proper subset of β that also spans V, then any vectors in $\beta\backslash T$ would have to be linear combinations of

the vectors in T, violating (3) for β, since (3) is equivalent to (1). Thus β is a minimal spanning set. Conversely, if β is a minimal spanning set, then it must be linearly independent, for otherwise there would be some $v \in \beta$ that is a linear combination of other vectors in β, which would mean $\beta \backslash \{v\}$ also spans V, contradicting minimality. Thus (4) implies (1).

(1) $\iff$ (5): If (1) holds, so that β is linearly independent and spans V, but we erroneously suppose β is not maximal in terms of linear independence, then we can count on there being a vect r $v \in V$ which is not in β with the property that $\beta \cup \{v\}$ is also linearly independent. But $v \notin \text{span}(\beta)$ if we remember that (3) is equivalent to (1), a contradiction. Therefore β is maximal and (1) implies (5). Conversely, if β is a maximal linearly independent set, then β must span V, for otherwise there is a $v \in V \backslash \beta$ that isn't a linear combination of vectors in β, implying $\beta \cup \{v\}$ is linearly independent proper superset, violating maximality. Therefore $\text{span}(\beta) = V$, and (5) implies (1). ■

Proof of Theorem 2.6

Proof: Let $\mathcal{A}$ be the set of all linearly independent sets L such that $I \subseteq L \subseteq S$,

$$\mathcal{A} \stackrel{\text{def}}{=} \{L \subseteq V \mid I \subseteq L \subseteq S,\ L \text{ is linearly independent}\}$$

Then $\mathcal{A}$ is non-empty because $I \subseteq I \subseteq S$, so $I \in \mathcal{A}$. Now, if

$$\mathcal{C} = \{I_k \mid k \in K\} \subseteq \mathcal{A}$$

is a chain in $\mathcal{A}$, that is a totally ordered (under set inclusion) subset of $\mathcal{A}$ ($I_k \subseteq I_{k+1} \subseteq \cdots$), then the union

$$U = \bigcup \mathcal{C} = \bigcup_{i \in K} I_i$$

is linearly independent and satisfies $I \subseteq U \subseteq S$, that is $U \in \mathcal{A}$. But by Zorn's lemma every chain has a maximal element β, so that we have $\beta \in \mathcal{A}$, a maximal element which is linearly independent. But of course such a β is a basis for $V = \text{span}(S)$, for if any $s \in S$ is not a linear combination of elements in β, then $\beta \cup \{s\}$ is linearly independent and β is contained in a strictly larger set $\beta \cup \{s\}$, contradicting the maximality of β. Therefore $S \subseteq \text{span}(\beta)$, and so $V = \text{span}(S) \subseteq \text{span}(\beta) \subseteq V$, or $V = \text{span}(\beta)$. This shows (1) that there is a basis β for V, (2) any linearly independent set I has is contained in some β, and (3) any spanning set S contains some basis β. ■

Proof of Lemma 2.7

Proof: (1) and (2) are immediate, so we only need to prove 3 and 4:

(3) If $v \in \text{span}(S \cup T)$, then there exist vectors $v_1, \ldots, v_m \in S$, $u_1, \ldots, u_n \in T$ and scalars $a_1, \ldots, a_m$ and $b_1, \ldots, b_n \in F$ such that

$$v = v_1 a_1 + \cdots + v_m a_m + b_1 u_1 + \cdots + b_n u_n$$

Note, however, $v_1 a_1 + \cdots + v_m a_m \in \text{span}(S)$ and $b_1 u_1 + \cdots + b_n u_n \in \text{span}(T)$, so that $v \in \text{span}(S) + \text{span}(T)$. Thus $\text{span}(S \cup T) \subseteq \text{span}(S) + \text{span}(T)$. Conversely, if $v = s + t \in \text{span}(S) + \text{span}(T)$, then $s \in \text{span}(S)$ and $t \in \text{span}(T)$, so that by (1), since $S \subseteq S \cup T$ and $T \subseteq S \cup T$, we must have $\text{span}(S) \subseteq \text{span}(S \cup T)$ and $\text{span}(T) \subseteq \text{span}(S \cup T)$. Consequently, $s, t \in \text{span}(S \cup T)$, and since $\text{span}(S \cup T)$ is a subspace, $v = s + t \in \text{span}(S \cup T)$. That is $\text{span}(S) + \text{span}(T) \subseteq \text{span}(S \cup T)$. Thus, $\text{span}(S) + \text{span}(T) = \text{span}(S \cup T)$.

(4) First, $\text{span}(S \cap T)$, $\text{span}(S)$ and $\text{span}(T)$ are subspaces, while by Exercise 5.2 we know that $\text{span}(S) \cap \text{span}(T)$ is also a subspace. Now, consider $x \in \text{span}(S \cap T)$.

There exist vectors $v_1, \ldots, v_n \in S \cap T$ and scalars $a_1, \ldots, a_n \in F$ such that $x = a_1v_1 + \cdots + a_nv_n$. But since $v_1, \ldots, v_n$ belong to both S and T, $x \in \text{span}(S)$ and $x \in \text{span}(T)$, so that $x \in \text{span}(S) \cap \text{span}(T)$. It is not in general true, however, that $\text{span}(S) \cap \text{span}(T) \subseteq \text{span}(S \cap T)$. For example, if $S = \{\mathbf{e}_1, \mathbf{e}_2\} \subseteq \mathbb{R}^2$ and $T = \{\mathbf{e}_1,\ (1,1)\}$, then $\text{span}(S) \cap \text{span}(T) = \mathbb{R}^2 \cap \mathbb{R}^2 = \mathbb{R}^2$, but $\text{span}(S \cap T) = \text{span}(\{e_1\}) = \mathbb{R}$, and $\mathbb{R}^2 \not\subseteq \mathbb{R}$. ■

Proof of Corollary 2.8

Proof: If $U = V$, then let $T = \{0\}$. Otherwise, if $V \backslash U \neq \varnothing$, then since V has a basis β and β is maximal, we must have $U = \text{span}(\beta \cap U)$ and $\beta \cap (V \backslash U) \neq \varnothing$. That is, there is a nonzero subspace $T = \text{span}(\beta \cap (V \backslash U))$, and by Lemma 2.7

$$V = \text{span}(\beta) = \text{span}((\beta \cap U) \cup (\beta \cap (V \backslash U))) = \text{span}(\beta \cap U) + \text{span}(\beta \cap (V \backslash U)) = U + T$$

But $U \cap T = \{0\}$ because of the essentially unique representation of vectors in V as linear combinations of vectors in β. Hence, $V = U \oplus T$, and T is a complement of U. ■

Proof of Theorem 2.9

Proof: The proof is by induction on m, beginning with $m = 0$. For this m, $B = \varnothing$, so taking $C = S$, we have $B \cup C = \varnothing \cup S = S$, which generates V. Now suppose the theorem holds for some $m \geq 0$. For $m+1$, let $B = \{v_1, \ldots, v_{m+1}\} \subseteq V$ be linearly independent. By removing v_{m+1} we still have a linearly independent set $B' = \{v_1, \ldots, v_m\}$, and so by the induction hypothesis $m \leq n$ and there is a $C' = \{u_1, \ldots, u_{n-m}\} \subseteq S$ such that $\{v_1, \ldots, v_n\} \cup C'$ generates V. This means there are scalars $a_1, \ldots, a_m$, and $b_1, \ldots, b_{n-m} \in F$ satisfying

$$v_{m+1} = a_1v_1 + \cdots + a_mv_m + b_1u_1 + \cdots + b_{n-m}u_{n-m} \tag{3.1}$$

Note that if $n = m$, $C' = \varnothing$, so that $v_{m+1} \in \text{span}(B')$, contradicting the assumption that B is linearly independent. Therefore, $n > m$, or $n \geq m + 1$. Moreover, some b_i, say b_1, is nonzero, for otherwise we have $v_{m+1} = a_1v_1 + \cdots + a_mv_m$, leading again to B being linearly dependent in contradiction to assumption. Solving (3.1) for u_1, we get

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

Let $C = \{u_2, \ldots, u_{n-m}\}$. Then $u_1 \in \text{span}(B \cup C)$, and since $v_1, \ldots, v_m, u_2, \ldots, u_{n-m} \in B \cup C$, they are also in $\text{span}(B \cup C)$, whence

$$B' \cup C' \subseteq \text{span}(B \cup C)$$

Now, since $\text{span}(B' \cup C') = V$ by the induction hypothesis, we have also $\text{span}(B \cup C) = V$. So we have B linearly independent with $|B| = m + 1$ vectors, $m + 1 \leq n$, $\text{span}(B \cup C) = V$, and $C \subseteq S$ with $|C| = (n - m) - 1 = n - (m + 1)$ vectors, so the theorem holds for $m + 1$. Therefore the theorem is true for all $m \in \mathbb{N}$ by induction. ■

Proof of Corollary 2.10

Proof: Suppose S is a finite spanning set for V, and let β and γ be two bases for V with $|\gamma| > |\beta| = k$, so that some subset $T \subseteq \gamma$ contains exactly $k + 1$ vectors. Since T is linearly independent and β generates V, the replacement theorem implies that $k + 1 \leq k$, a contradiction. Therefore γ is finite and $|\gamma| \leq k$. Reversing the roles of β and γ shows that $k \leq |\gamma|$, so that $|\gamma| = |\beta| = k$. ■

Proof of Corollary 2.11

Proof: Let B be a basis for V. (1) Let $S \subseteq V$ be a finite spanning set for V. By Theorem 2.6, S contains a basis B for V, and by Corollary 2.10 $|B| = n$ vectors. Therefore $|S| \geq n$, and $|S| = n \implies S = B$, so that S is a basis for V. (2) Let $I \subseteq V$ be linearly independent with $|I| = n$. By the Replacement Theorem there exists a subset T of B with $|T| = n - n = 0$ vectors such that $V = \text{span}(I \cup T) = \text{span}(I \cup \varnothing) = \text{span}(I)$. Since I is also linearly independent, it is a basis for V. (3) If $I \subseteq V$ is linearly independent with $|I| = m$ vectors, then the replacement theorem asserts that there exists a subset H of B containing exactly $n - m$ vectors such that $V = \text{span}(I \cup H)$. Now, $|I \cup H| \leq n$, so that by 1 $|I \cup H| = n$, and so $I \cup H$ is a basis for V extended from I. ■

Proof of Theorem 2.12

Proof: If V is finite-dimensional, then the previous corollary applies, so we need only consider bases that are infinite sets. Let $B = \{b_i \mid i \in I\}$ ba a basis for V and let C be any other basis for V. Then all $c \in C$ can be written as finite linear combinations of vectors in B, where all the coefficients are nonzero. That is, if $U_c = \{1, \ldots, n_c\}$, we have

$$c = \sum_{i=1}^{n_c} a_i b_i = \sum_{i \in U_c} a_i b_i$$

for unique $a_1, \ldots, a_{n_c} \in F$. But because C is a basis, we must have $I = \bigcup_{c \in C} U_c$, for if $\bigcup_{c \in C} \subsetneq I$, then all $c \in C$ can be expressed by a *proper* subset B' of B, so that $V = \text{span}(B')$, which is contradiction of the minimality of B as a spanning set. Now, for all $c \in C$ we have $|U_c| < \aleph_0$, which implies that

$$|B| = |I| = \left| \bigcup_{c \in C} U_c \right| \leq \aleph_0 |C| = |C|$$

Reversing the roles of B and C gives $|C| \leq |B|$, so by the Schröder-Bernstein theorem we have $|B| = |C|$. ■

Proof of Theorem 2.14

Proof: (1) If $B_1 \cap B_2 = \varnothing$ and $B = B_1 \cup B_2$ is a basis for V, then $0 \notin B_1 \cup B_2$. But, if a nonzero vector $v \in \text{span}(B_1) \cap \text{span}(B_2)$, then $B_1 \cap B_2 \neq \varnothing$, a contradiction. Hence $\{0\} = \text{span}(B_1) \cap \text{span}(B_2)$. Moreover, since $B_1 \cup B_2$ is a basis for V, and since it is also a basis for $\text{span}(B_1) + \text{span}(B_2)$, we must have $V = \text{span}(B_1) + \text{span}(B_2)$, and hence $V = \text{span}(B_1) \oplus \text{span}(B_2)$. (2) If $V = S \oplus T$, then $S \cap T = \{0\}$, and since $0 \notin B_1 \cup B_2$, we have $B_1 \cap B_2 = \varnothing$. Also, since all $v \in V = S \oplus T$ have the form $a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_n v_n$ for $u_1, \ldots, u_m \in B_1$ and $v_1, \ldots, v_n \in B_2$, $v \in \text{span}(B_1 \cup B_2)$, so $B_1 \cup B_2$ is a basis for V by Theorem 2.3. ■

Proof of Theorem 2.15

Proof: If $B = \{b_i \mid i \in I\}$ is a basis for $S \cap T$, then we can extend this to a basis $A \cup B$ for S and to another basis $B \cup C$ for T, where $A = \{a_j \mid j \in J\}$, $C = \{c_k \mid k \in K\}$ and $A \cap B = \varnothing$ and $B \cap C = \varnothing$. Then $A \cup B \cup C$ is a basis for $S + T$: clearly $\text{span}(A \cup B \cup C) = S + T$, so we need only verify that $A \cup B \cup C$ is linearly independent. To that end, suppose not, suppose there are nonzero scalars $c_1, \ldots, c_n \in F \backslash \{0\}$ and $v_1, \ldots, v_n \in A \cup B \cup C$ such that

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = 0$$

Then some $v_i \in A \cap C$ since by our construction $A \cup B$ and $B \cup C$ are linearly independent. Isolating the vectors in A, say $v_1, \ldots, v_k$, on one side of the equality shows that there is a nonzero vector,

$$x = \overbrace{a_1 v_1 + \cdots + a_k v_k}^{\in \text{span}(A)} = \underbrace{a_{k+1} v_{k+1} + \cdots + a_n v_n}_{\in \text{span}(B \cup C)} \implies x \in \text{span}(A) \cap \text{span}(B \cup C) = \text{span}(A) \cap T \subseteq S \cap T$$

since $\text{span}(A) \subseteq S$. Consequently $x = \vec{0}$, and therefore $a_1 = \cdots = a_n = 0$ since A and $B \cup C$ are linearly independent sets, a contradiction. Thus $A \cup B \cup C$ is linearly independent and hence a basis for $S + T$. Moreover,

$$\begin{aligned} \dim(S) + \dim(T) &= |A \cup B| + |B \cup C| \\ &= |A| + |B| + |B| + |C| \\ &= |A \cup B \cup C| + |B| \\ &= \dim(S + T) + \dim(S \cap T) \end{aligned}$$

Of course if $S \cap T = \{0\}$, then $\dim(S \cap T) = 0$, and

$$\dim(S) + \dim(T) = \dim(S \oplus T) = \dim(V)$$

while if $\dim(V) = \dim(S) + \dim(T)$, then $\dim(S \cap T) = 0$, so $S \cap T = \{0\}$, and so $V = S \oplus T$. ■

Proof of Theorem 2.17

Proof: (1) $\implies$ (2): Suppose (1) is true, that is $V = \bigoplus_{i \in I} S_i$, and choose $v_1, \ldots, v_n \in V$ such that $v_i \in S_{k_i}$ and $v_1 + \cdots + v_k = 0$. Then for any j,

$$-v_j = \sum_{i \neq j} v_i \in \sum_{i \neq j} S_i$$

However, $-v_j \in S_j$, whence

$$-v_j \in W_j \cap \sum_{i \neq j} S_i = \{\vec{0}\}$$

so that $v_j = \vec{0}$. This is true for all $j = 1, \ldots, k$, so $v_1 = \cdots = v_k = \vec{0}$, proving (2).

(2) $\implies$ (3): Suppose (2) is true and let $v \in V = \sum_{i \in I} S_i$ be given by

$$\begin{aligned} v &= u_1 + \cdots + u_n \\ &= w_1 + \cdots + w_m \end{aligned}$$

for some $u_i \in S_{k_i}$ and $w_j \in S_{k_j}$. Then, grouping terms from the same subspaces

$$\vec{0} = v - v = (u_{i_1} - w_{i_1}) + \cdots + (u_{i_k} - w_{i_k}) + u_{i_{k+1}} + \cdots + u_{i_n} - w_{i_{k+1}} - \cdots - w_{i_m}$$

$$\implies \begin{cases} u_{i_1} = w_{i_1}, \cdots, u_{i_k} = w_{i_k} \\ \text{and } \; u_{i_{k+1}} = \cdots = u_{i_n} = w_{i_{k+1}} = \cdots = w_{i_m} = \vec{0} \end{cases}$$

proving uniqueness.

(3) $\implies$ (4): Suppose each vector $v \in V = \sum_{i \in I}^{n} S_i$ can be uniquely written as $v = v_1 + \cdots + v_k$, for $v_i \in S_{j_i}$. For each $i \in I$ let γ_i be an ordered basis for S_i, so that since $V = \sum_{i \in I}^{n} S_i$, we must have that $V = \text{span}(\gamma) = \text{span}\left(\bigcup_{i=1}^{k} \gamma_i\right)$, so we only need to show that γ is linearly independent. To that end, let

$$\begin{gathered} v_{11}, v_{12}, \ldots, v_{1n_1} \in \gamma_{i_1} \\ \vdots \\ v_{m1}, v_{m2}, \ldots, v_{mn_m} \in \gamma_{i_m} \end{gathered}$$

for any bases γ_{i_j} for $S_{i_j} \in \mathcal{F}$, then let $a_{ij} \in F$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n_i$, and let $w_i = \sum_{j=1}^{n_i} a_{ij} v_{ij}$. Then, suppose

$$w_1 + \cdots + w_m = \sum_{i,j} a_{ij} v_{ij} = 0$$

since $\vec{0} \in \bigcap_{i=1}^{m} S_{j_i}$ and $\vec{0} = \vec{0} + \cdots + \vec{0}$, by the assumed uniqueness of expression of $v \in V$ by $w_i \in S_i$ we must have $w_i = \vec{0}$ for all i, and since all the γ_i are linearly independent, we must have $a_{ij} = 0$ for all i, j.

(4) $\implies$ (1): if (4) is true, then there exist ordered bases γ_i the S_i such that $\gamma = \bigcup_{i \in I} \gamma_i$ is an ordered basis for V. By Lemma 2.7 and Theorem 2.14,

$$V = \operatorname{span}(\gamma) = \operatorname{span}(\bigcup_{i \in I} \gamma_i) = \sum_{i \in I} \operatorname{span}(\gamma_i) = \sum_{i \in I} S_i$$

Choose any $j \in I$ and suppose for a nonzero vector $v \in V$ we have $v \in S_j \cap \sum_{i \neq j} S_i$. Then,

$$v \in S_j = \operatorname{span}(\gamma_j) \cap \operatorname{span}(\bigcup_{i \neq j} \gamma_i) = \sum_{i \neq j} S_i$$

which means v is a nontrivial linear combination of elements in γ_i and elements in $\bigcup_{i \neq j} \gamma_i$, so that v can be expressed as a linear combination of $\bigcup_{i \in I} \gamma_i$ in more than one way, which contradicts uniqueness of representation of vectors in V in terms of a basis for V, Theorem 2.3. Consequently, any such v must be 0. That is,

$$S_j \cap \sum_{i \neq j} S_i = \{\vec{0}\}$$

and the sum is direct, i.e. $V = \bigoplus_{i \in I} S_i$, proving 1. ■

4 Linear Transformations

As with linear transformations between $\mathbb{R}^n$ and $\mathbb{R}^m$, so with linear transformations $\mathcal{L}(V, W)$ between vector spaces V and W *over the same field* $F = \mathbb{R}$ or $\mathbb{C}$:

Definition 4.1 If V and W are vector spaces over the same field $F = \mathbb{R}$ or $\mathbb{C}$, the type of function $T : V \to W$ which *preserves* the algebraic vector properties of V in W is called a **linear transformation**. Namely, T is linear if for all vectors $u, v \in V$ and all scalars $c \in F$ we have

$$\begin{aligned} T(u+v) &= T(u) + T(v) \\ T(cv) &= cT(v) \end{aligned}$$

The **set of all linear transformations** is denoted

$$\mathcal{L}(V, W)$$

and acquires itslef the structure of a vector space under *pointwise addition and scalar multiplication*,

$$\begin{aligned} (S+T)(v) &\stackrel{\text{def}}{=} S(v) + T(v) \\ (cT)(v) &\stackrel{\text{def}}{=} cT(v) \end{aligned}$$

Prove this as an exercise! The linear transformations from V to itself are called **linear operators** and are denoted

$$\mathcal{L}(V)$$

Linear transformations come in different shapes and sizes. There are the **one-to-one** or **injective** linear transformations, which will be characterized by having their null space trivial, there are the **onto** or **surjective** ones, and there are the **invertible** or **bijective** ones, which we will call **isomorphisms** and which are both one-to-one and onto. The set of all isomorphisms between vector spaces V and W is denoted here

$$\text{Isom}(V, W) = \{T \in \mathcal{L}(V, W) \mid T \text{ is invertible}\}$$

Whenver there exists an isomorphism between two different vector spaces V and W, we say that the vector spaces are **isomorphic**, and sometimes denote this by

$$V \cong W \quad \stackrel{\text{def}}{\iff} \quad \text{there exists some } T \in \text{Isom}(V, W)$$

The set of all isomorphisms from V to itself is a group, called the **general linear group** of V, and denoted

$$\text{GL}(V) = \{T \in \mathcal{L}(V) \mid T \text{ is invertible}\}$$ ■

Definition 4.2 There are two subspaces associated to any linear transformation T in $\mathcal{L}(V,W)$, its **kernel** or **null space**

$$\ker T \quad \text{or} \quad N(T) \stackrel{\text{def}}{=} \{v \in V \mid T(v) = \vec{0} \in W\}$$

a subspace of V, and its **image** or **range**, the set of *achieved* w-values,

$$\text{im}\, T \quad \text{or} \quad R(T) \stackrel{\text{def}}{=} \{w \in W \mid w = T(v) \text{ for some } v \in V\}$$

a subspace of W. The *dimensions* of each have names, and become important measures of the behavior of T: the **nullity** is the dimension of the kernel/null space, and the **rank** is the dimension of the image/range:

$$\begin{aligned} \text{null}(T) &\stackrel{\text{def}}{=} \dim(\ker T) \\ \text{rank}(T) &\stackrel{\text{def}}{=} \dim(\text{im}\, T) \end{aligned}$$

■

5 Basic Properties of Linear Transformations

Theorem 5.1 *Let V and W be vector spaces over the same field $F = \mathbb{R}$ or $\mathbb{C}$, and let $T \in \mathcal{L}(V,W)$. Then $\ker T$ is a subspace of V and $\text{im}\, T$ is a subspace of W.* ■

Exercise 5.2 Prove this theorem. ■

Theorem 5.3 *If V and W are vector spaces and $\beta = \{v_i \mid i \in I\}$ is a basis for V, then for any $T \in \mathcal{L}(V,W)$ we have $\text{im}(T) = \text{span}(T(\beta))$.* ■

Exercise 5.4 Prove this theorem. ■

Theorem 5.5 (Any Linear Transformation Is Defined by Its Action on a Basis) *Let V and W be vector spaces over the same field F. If $\beta = \{v_i \mid i \in I\}$ is a basis for V, then we can define a unique linear transformation $T \in \mathcal{L}(V,W)$ by arbitrarily specifying the w-values on β, $w_i = T(v_i) \in W$, and extending T by linearity, i.e. specifying that for all $a_1, \ldots, a_n \in F$ our T satisfy*

$$T(a_1 v_1 + \cdots + a_n v_n) \stackrel{\text{def}}{=} a_1 T(v_1) + \cdots + a_n T(v_n) \stackrel{\text{def}}{=} a_1 w_1 + \cdots + a_n w_n$$ ■

Exercise 5.6 Prove this theorem. ■

Theorem 5.7 *Let V, W, Z be vector spaces over the same field F. If $T \in \mathcal{L}(V,W)$ and $U \in \mathcal{L}(W,Z)$, then $U \circ T \in \mathcal{L}(V,Z)$.* ■

Exercise 5.8 Prove this theorem. ■

Theorem 5.9 *If V and W are vector spaces over F, then $\mathcal{L}(V, W)$ is a vector space over F under pointwise addition and scalar multiplication of functions (see Definition 4.1).* ■

Exercise 5.10 Prove this theorem. ■

Theorem 5.11 (Invertibility) *Let V, W, Z be vector spaces over the same field F. If $T \in \text{Isom}(V, W)$ and $U \in \text{Isom}(W, Z)$ are isomorphisms, then*

(1) *T^{-1} is linear, $T^{-1} \in \mathcal{L}(W, V)$.*
(2) *$(T^{-1})^{-1} = T$, and hence $T^{-1} \in \text{Isom}(W, V)$.*
(3) *$(U \circ T)^{-1} = T^{-1} \circ U^{-1} \in \text{Isom}(Z, V)$.* ■

Exercise 5.12 Prove this theorem. ■

Theorem 5.13 (Injectivity 1) *If V and W are vector spaces over the same field F and $T \in \mathcal{L}(V, W)$, then T is injective iff $\ker(T) = \{0\}$.* ■

Exercise 5.14 Prove this theorem. ■

Theorem 5.15 (Injectivity 2) *Let V and W be vector spaces over the same field and $S \subseteq V$. Then for any $T \in \mathcal{L}(V, W)$,*

(1) *T is one-to-one iff it carries linearly independent sets into linearly independent sets.*
(2) *If T is one-to-one and $S \subseteq V$, then S is linearly independent in V iff $T(S)$ is linearly independent in W.* ■

Exercise 5.16 Prove this theorem. ■

Theorem 5.17 *If V and W are vector spaces over the same field F and $T \in \text{Isom}(V, W)$, then for any subset S of V we have*

(1) *$V = \text{span}(S)$ iff $W = \text{span}(T(S))$.*
(2) *S is linearly independent in V iff $T(S)$ is linearly independent in W.*
(3) *S is a basis for V iff $T(S)$ is a basis for W.* ■

Exercise 5.18 Prove this theorem. ■

Theorem 5.19 (Isomorphisms Preserve Bases) *If V and W are vector spaces over the same field F and β is a basis for V, then*

$$\boxed{T \in \mathrm{Isom}(V,W) \iff T(\beta) \text{ is a basis for } W \text{ wherever } \beta \text{ is a basis for } V}$$

■

Exercise 5.20 Prove this theorem. ■

Theorem 5.21 (Isomorphisms Preserve Dimension) *If V and W are vector spaces over the same field F, then*

$$\boxed{V \cong W \iff \dim V = \dim W}$$

Proof: If $V \cong W$, then there is some $T \in \mathrm{Isom}(V,W)$, so if β is a basis for V, then by the last theorem $T(\beta)$ is a basis for W, and $\dim(V) = |\beta| = |T(\beta)| = \dim(W)$. Conversely, if $\dim(V) = |\beta| = |\gamma| = \dim(W)$, where γ is a basis for W, then we can take any bijection $T : \beta \to \gamma$ and extend it to V by linearity, thereby defining a unique linear transformation $T \in \mathcal{L}(V,W)$ by Theorem 5.5. Moreover, T is an *isomorphism* because it is surjective, on account of $\mathrm{im}\, T = \mathrm{span}(T(\beta)) = W$, and injective because $\ker(T) = \{\vec{0}\}$ (because $T(\beta)$ is linearly independent, so if $v \in \ker T$ then $\vec{0} = T(v) = T(\sum_j a_j v_j) = \sum_j a_j T(v_j)$ implies all $a_j = 0$, and so $v = \vec{0}$), therefore $V \cong W$. ■

Corollary 5.22 *If V and W are vector spaces over the same field and $T \in \mathcal{L}(V,W)$, then*

$$\boxed{\begin{aligned} \dim V < \dim W &\implies T \text{ cannot be onto} \\ \dim V > \dim W &\implies T \text{ cannot be one-to-one} \end{aligned}}$$

■

Corollary 5.23 *If V is an n-dimensional vector space over F, then*

$$\boxed{V \cong F^n}$$

If κ is any cardinal number, β a set of cardinality κ and V is a κ-dimensional vector space over F, then

$$\boxed{V \cong (F^\beta)_0}$$

where $(F^\beta)_0 \stackrel{\text{def}}{=} \{\text{all } f : \beta \to F \text{ taking on only finitely many nonzero values}\}$.

Proof: This is a result of the fact that $\dim(F^n) = n$ and $\dim\big((F^\beta)_0\big) = \kappa$, since a basis for $(F^\beta)_0$ is $B = \{\delta_i : \beta \to F \mid \delta_i(v_j) = \delta_{ij},\ i \in I\}$, along with the previous theorem. Here, δ_{ij} is the *Kronecker delta*, which is defined by $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. ■

The last corollary, 5.23, brings us to the link between $F^n = \mathbb{R}^n$ or $\mathbb{C}^n$ and any abstract or other finite-dimensional vector space V **of the same dimension**.

Definition 5.24 Denote the **set of all bases** for the n-dimensional space V by

$$V_n(V)$$

This is sometimes called the **Stiefel manifold** of bases or *frames*, and is indeed a manifold in the sense of differential geometry. Importantly, it is in one-to-one correspondence with the set of isomorphisms $\mathrm{Isom}(V, F^n)$, a theorem to be proved below.

Definition 5.25 Let V be an n-dimensional vector space over $F = \mathbb{R}$ or $\mathbb{C}$. For each basis $\beta = (b_1, \ldots, b_n) \in V_n(V)$ we define an isomorphism by means of Theorems 5.5 and 5.19, the β-**coordinate map**

$$\varphi_\beta \in \mathrm{Isom}(V, F^n)$$

by defining it on β,

$$\varphi_\beta(b_i) \stackrel{\text{def}}{=} \mathbf{e}_i$$

then extending it by linearity. Then, for any $v = \sum_{i=1}^n a_i b_i \in V$, with coordinates $a_i \in F$,

$$\varphi_\beta(v) = \varphi_\beta\Big(\sum_{i=1}^n a_i b_i\Big) \stackrel{\text{def}}{=} \sum_{i=1}^n a_i \varphi(b_i) = \sum_{i=1}^n a_i \mathbf{e}_i = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \stackrel{\text{def}}{=} [v]_\beta \qquad \blacksquare$$

We now wish to prove the following statement:

Every basis for V defines a unique isomorphism between V and F^n, and conversely every isomorphism between V and F^n corresponds to a unique basis for V.

but we'll do it rigorously, with the notation we set up above.

Theorem 5.26 *Let V be an n-dimensional vector space over $F = \mathbb{R}$ or $\mathbb{C}$. Then there is a bijective function (a 'one-to-one correspondence')*

$$\varphi : V_n(V) \to \mathrm{Isom}(V, F^n)$$

given by

$$\varphi(\beta) = \varphi_\beta$$

That is, φ gives to each basis β its β-coordinate map, and by this means describes all isomorphisms to F^n.

Proof: Since each φ_β lies in $\mathrm{Isom}(V, F^n)$, we see that the range of φ is contained in $\mathrm{Isom}(V, F^n)$. It remains to show that φ is one-to-one and onto, which mean, respectively,

(1) $\varphi_\beta = \varphi_\gamma \implies \beta = \gamma$

(2) Every isomorphism $T \in \mathrm{Isom}(V, F^n)$ is of the form $T = \varphi_\beta$ for some $\beta \in V_n(V)$

Let us begin with (2): Choose $T \in \mathrm{Isom}(V, F^n)$, and use Theorem 5.11 to observe that $T^{-1} \in \mathrm{Isom}(F^n, V)$, which by Theorem 5.19 means that

$$\boxed{\beta \stackrel{\text{def}}{=} T^{-1}(\sigma), \quad \text{that is each} \quad b_i \stackrel{\text{def}}{=} T^{-1}(\mathbf{e}_i)}$$

is a basis for V, $\beta \in V_n(V)$. We then note with gladness that

$$T(b_i) = \mathbf{e}_i = \varphi_\beta(b_i)$$

which means T and φ_β agree on a basis $\beta \in V_n(V)$. Theorem 5.7 then assures us that $\boxed{T = \varphi_\beta = \varphi(\beta)}$, and thus φ is onto.

Let us now show (1): Suppose that $\varphi(\beta) = \varphi(\gamma)$, i.e. that $\varphi_\beta = \varphi_\gamma$ for $\beta, \gamma \in V_n(V)$. Then, since they *are the same function*, they will agree on each basis vector,

$$\varphi_\beta(b_i) = \mathbf{e}_i = \varphi_\gamma(b_i) \quad \text{and} \quad \varphi_\beta(c_i) = \mathbf{e}_i = \varphi_\gamma(c_i)$$

Using their invertibility we conclude that

$$c_i = \varphi_\gamma^{-1}(\mathbf{e}_i) = \varphi_\beta^{-1}(\mathbf{e}_i) = b_i$$

so $\boxed{\beta = \gamma}$. ■

6 Rank and Nullity

Lemma 6.1 *If V and W are vector spaces over the same field F and $T \in \mathcal{L}(V, W)$, then any complement of the $\ker T$ in V is isomorphic to $\mathrm{im}\, T$ in W,*

$$\boxed{V = \ker T \oplus (\ker T)^c \implies (\ker T)^c \cong \mathrm{im}\, T}$$

where $(\ker T)^c$ is any complement of $\ker T$ in V.

Proof: By Theorem 2.15 $\dim V = \dim \ker T + \dim (\ker T)^c$. Let

$$T^c = T|_{(\ker T)^c} : (\ker T)^c \to \mathrm{im}\, T$$

denote the *restriction* of T to $(\ker T)^c$ and note that T is injective by Theorem 5.1 because by definition of a direct sum

$$\ker T^c = \ker T \cap \ker T^c = \{0\}$$

Moreover T^c is surjective because $T^c(V) = T(V) = \operatorname{im} T$: first we obviously have $T^c(V) \subseteq \operatorname{im} T$, while for the reverse inclusion suppose $v \in \operatorname{im} T$. Then by Theorem 2.17 there are unique $s \in \ker T$ and $t \in (\ker T)^c$ such that $v = s + t$, which means

$$T(v) = T(s+t) = T(s) + \cancelto{0}{T(t)} = T(s) = T^c(s) \in T^c(V)$$

This concludes the demonstration that $\operatorname{im} T = T^c(V)$ and thus that T^c is surjective, so that T^c is an *isomorphism*, whence

$$(\ker T)^c \cong \operatorname{im} T$$ ■

Theorem 6.2 (Rank-Nullity Theorem) *If V and W are vector spaces over the same field F and $T \in \mathcal{L}(V, W)$, then*

$$\boxed{\dim V = \operatorname{rank} T + \operatorname{null} T}$$

Proof: By the lemma and we have $(\ker T)^c \cong \operatorname{im} T$, which by the previous theorem implies $\dim(\ker T)^c = \dim \operatorname{im} T = \operatorname{rank} T$, while by Theorem 2.15 $V = \ker(T) \oplus \ker(T)^c$ implies

$$\begin{aligned} \dim(V) &= \dim \ker T + \dim(\ker T)^c \\ &= \dim \ker T + \dim \operatorname{im} T \\ &= \operatorname{null} T + \operatorname{rank} T \end{aligned}$$

which completes the proof. ■

Corollary 6.3 *Let V and W are vector spaces over the same field F and consider any linear transformation in $T \in \mathcal{L}(V, W)$. If $\dim V = \dim W$, then the following are equivalent:*

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

Proof: By the Rank-Nullity Theorem, $\operatorname{rank} T + \operatorname{null} T = \dim(V)$ and some other recent theorems we have

$$\begin{aligned} T \text{ is 1-1} &\overset{\text{Thm 2.17}}{\iff} \ker T = \{0\} \\ &\overset{\text{Thm 5.13}}{\iff} \operatorname{null} T = 0 \\ &\overset{\text{RN}}{\iff} \dim \operatorname{im} T = \operatorname{rank} T = \dim V \overset{\text{hyp}}{=} \dim W \\ &\overset{\text{Thm 5.21}}{\iff} \operatorname{im} T = W. \\ &\iff T \text{ is onto} \end{aligned}$$

which completes the proof. ■