**Question:** What is space?

**Answer:** $\mathbb{R}^n$

**(2.A)** Whatever "space" is, it is usually said to consist of "points". We will begin with Descartes' revolutionary idea (1637):

$$\textit{a point} \equiv \textit{an ordered list of numbers.}$$

Having made this identification, it becomes irresistible to try to do algebraic things with points, like *add* them:

$$(u_1, u_2, \ldots, u_n) + (v_1, v_2, \ldots, v_n) := (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n).$$

This has the unfortunate side effect of singling out one special point $\mathbf{0}$ with the property that $\mathbf{u}+\mathbf{0} = \mathbf{u}$ for all points $\mathbf{u}$. We probably don't *want* to single out a special point, but we can fix this later. In the meantime, we can use the special point $\mathbf{0}$ to make an identification between the point $\mathbf{u}$ and the directed line segment $\mathbf{0} \to \mathbf{u}$ (called a **vector**). This shows us that it was not really *points* we were trying to add before, but *vector displacements*, which makes a lot more sense. (We can think of vector addition as functional composition of translations.) Today we consider *vectors* as primary and *points* as secondary in the axiomatic development of space.

**(2.B)** The question remains: what kind of "numbers" should we use to define our "vectors"? We will temporarily retreat from the identification of numbers with $\mathbb{R}$, since the real number system has topological and analytic subtleties that are not always relevant. In building up an axiomatic definition of space we prefer to begin with the weakest concept of "number" that still yields interesting geometry. This is the concept of a field, which we denote by $K$ for Körper. Let $K^\times$ stand for the nonzero elements of $K$.

*A* **field** *is a structure* $(K, +, \times, 0, 1)$ *in which*

$(K, +, 0)$ *is an "abelian group":*
- $\forall\, a \in K, a + 0 = a$,
- $\forall\, a, b \in K, a + b = b + a$,
- $\forall\, a, b, c \in K, a + (b + c) = (a + b) + c$,
- $\forall\, a \in K, \exists\, b \in K, a + b = 0$.

$(K^\times, \times, 1)$ *is an "abelian group":*
- $\forall\, a \in K^\times, 1a = a$,
- $\forall\, a, b \in K^\times, ab = ba$,
- $\forall\, a, b, c \in K^\times, a(bc) = (ab)c$,
- $\forall\, a \in K^\times, \exists\, b \in K^\times, ab = 1$.

*And "multiplication distributes over addition":*
- $\forall\, a, b, c \in K, a(b + c) = ab + ac$.

Exercise: Should we also say that $0 \neq 1$? These nine axioms are clearly intended to model the properties of the rational numbers $\mathbb{Q}$, but they are not sufficient to **characterize** the rational numbers. Indeed, $\mathbb{R}$ also satisfies these properties, but $\mathbb{R} \neq \mathbb{Q}$ (Pythagoras).

We can also define number systems intermediate between $\mathbb{Q}$ and $\mathbb{R}$. For example, given any non-square $d \in \mathbb{N}$ we define

$$\mathbb{Q}[\sqrt{d}] := \{a + b\sqrt{d} : a, b \in \mathbb{Q}\},$$

which is a field because

$$\frac{1}{a+b\sqrt{d}} = \frac{1}{a+b\sqrt{d}}\frac{(a-b\sqrt{d})}{(a-b\sqrt{d})} = \frac{a-b\sqrt{d}}{a^2-b^2d} = \left(\frac{a}{a^2-b^2d}\right) + \left(\frac{-b}{a^2-b^2d}\right)\sqrt{d}.$$

It turns out that the field concept is rigid enough to support interesting geometry. When working over a general field, we will be doing "geometric algebra" in the spirit of Emil Artin (1957). After developing geometry at this level of generality, we can add in further (e.g., topological) structure as desired.

**(2.C)** We are now ready to define a notion of "space". Instead of **constructing** space, we will try to **characterize** space in terms of abstract properties (just as the field axioms characterize the concept of "number"). This was first attempted by Peano (1888) but his definition was premature—perhaps lacking sufficient examples—and was ignored for decades. Please consult Gregory Moore (1995) for the full story.

Following Peano, we define space as a set of **vectors** (the name is not important: you may call them beer mugs if you like), and we say how they are allowed to behave.

*A* **vector space** *is a structure* $(V, K, \cdot)$ *in which*

- $V = (V, +, \mathbf{0})$ *is an abelian group (of* **vectors***),*
- $K = (K, +, \times, 0, 1)$ *is a field (of* **scalars***),*
- *The field* $K$ **acts linearly** *on vectors* $V$ *by "scaling". That is, we have a function* $K \times V \to V$ *denoted* $(a, \mathbf{v}) \mapsto a \cdot \mathbf{v}$ *which satisfies*
  - $\forall\, \mathbf{v} \in V, 1 \cdot \mathbf{v} = \mathbf{v}$,
  - $\forall\, \mathbf{v} \in V, \forall\, a, b \in K, (ab) \cdot \mathbf{v} = a \cdot (b \cdot \mathbf{v})$,
  - $\forall\, \mathbf{v} \in V, \forall\, a, b \in K, (a+b) \cdot \mathbf{v} = a \cdot \mathbf{v} + b \cdot \mathbf{v}$.

Exercise: Show that $0 \cdot \mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \in V$. It is sloppy (but harmless) that we use the same symbol "+" for addition of vectors and addition of scalars. When the field is understood we will denote the vector space $(V, K)$ simply by $V$. One should think that each individual vector $\mathbf{v} \in V$ generates a "line" $K(\mathbf{v}) := \{a \cdot \mathbf{v} : a \in K\} \subseteq V$ isomorphic to $K$. Of course, this makes the most sense when $K = \mathbb{R}$, but the intuition is helpful in general.

Such an abstract definition should not be tolerated without motivation, so I will hurry to prove a fundamental theorem. First let me define the notion of the **linear closure** (a.k.a. **span**) of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$:

$$K(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) := \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n : a_1, a_2, \ldots, a_n \in K\} \subseteq V.$$

Exercise: Show that this is the smallest **subspace** of $V$ containing the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Next we will assume a "finiteness condition". We say that a vector space $V$ is **finitely generated** if there exists a finite set $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ such that $V = K(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$, and in this case we say that $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a **generating set** for $V$. [Highbrow Remark: I could have said "Noetherian" instead of "finitely generated".] Given a set of vectors $S \subseteq V$ we will say that a vector $\mathbf{u} \in S$ is **redundant** if $K(S) = K(S - \mathbf{u})$. This leads to the notion of linear independence.

*We say that the vectors* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \in V$ *form an* **independent set** *if none of the vectors is redundant, that is, if*

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_m\mathbf{u}_m = \mathbf{0} \in V$$

*necessarily implies that* $a_1 = a_2 = \cdots = a_m = 0 \in K$.

We now prove the fundamental lemma of Steinitz. This lemma is the irreducible core of linear algebra and we take it to motivate the definitions.

**Steinitz' Exchange Lemma (1910):** Let $V$ be a finitely generated vector space, let $\mathsf{IND} \subseteq V$ be *any* independent set, and let $\mathsf{GEN} \subseteq V$ be *any finite* generating set (we assume that $V$ is finitely generated). Then $\mathsf{IND}$ is finite, and moreover we have

$$|\mathsf{IND}| \leq |\mathsf{GEN}|.$$

**Proof:** Let $\mathsf{IND} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be an independent set (assume $m > 1$, so that $\mathbf{u}_i \neq \mathbf{0}$ for all $i$) and let $\mathsf{GEN} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a generating set for $V$. Assume for contradiction that $m > n$.

Since $\mathsf{GEN}$ is generating, we can write

$$\mathbf{u}_1 = r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n$$

for some scalars $r_1, \ldots, r_n \in K$. Since $\mathbf{u}_1 \neq \mathbf{0}$, not all of the coefficients are zero. Without loss, assume that $r_1 \neq 0$. Thus we can write

$$\mathbf{v}_1 = \frac{1}{r_1}\mathbf{u}_1 - \frac{r_2}{r_1}\mathbf{v}_2 - \cdots - \frac{r_n}{r_1}\mathbf{v}_n$$

and it follows that $\{\mathbf{u}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n\}$ is also a generating set for $V$. Now suppose (for induction) we have shown that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_n\}$ is a generating set for some $1 \leq i < n$. In this case we can write

$$\mathbf{u}_{i+1} = s_1\mathbf{u}_1 + s_2\mathbf{u}_2 + \cdots + s_i\mathbf{u}_i + s_{i+1}\mathbf{v}_{i+1} + \cdots + s_n\mathbf{v}_n$$

for some scalars $s_1, \ldots, s_n \in K$. Since $\mathsf{IND}$ is independent we can assume that the coefficients $s_{i+1}, s_{i+2}, \ldots, s_n$ are not all zero. Without loss, assume that $s_{i+1} \neq 0$. Thus we can write

$$\mathbf{v}_{i+1} = -\frac{s_1}{s_{i+1}}\mathbf{u}_1 - \cdots - \frac{s_i}{s_{i+1}}\mathbf{u}_i + \frac{1}{s_{i+1}}\mathbf{u}_{i+1} - \frac{s_{i+2}}{s_{i+1}}\mathbf{v}_{i+2} - \cdots - \frac{s_n}{s_{i+1}}\mathbf{v}_n$$

and it follows that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{i+1}, \mathbf{v}_{i+2}, \ldots, \mathbf{v}_n\}$ is a generating set. By induction we conclude that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a generating set for $V$. But, if so, we can write

$$\mathbf{u}_{n+1} = t_1\mathbf{u}_1 + t_2\mathbf{u}_2 + \cdots + t_n\mathbf{u}_n.$$

for some scalars $t_1, \ldots, t_n \in K$, which contradicts the fact that $\mathsf{IND}$ is independent. We conclude that $m \leq n$. □

Why do we care? Steinitz' Lemma is fundamental in that it allows us to define the concept of "dimension", a concept that any good "space" should have.

*Let* $V$ *be a finitely generated vector space. We say that* $\mathcal{B} \subseteq V$ *is a* **basis** *if*

- $\mathcal{B}$ *is independent,*
- $\mathcal{B}$ *is a generating set for* $V$.

First note that bases exist. Indeed, if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ is a generating set that is *not* independent then there exists a relation

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

with, say, $a_n \neq 0$. Thus we can express $\mathbf{v}_n$ as

$$\mathbf{v}_n = -\frac{a_1}{a_n}\mathbf{v}_1 - \cdots - \frac{a_{n-1}}{a_n}\mathbf{v}_{n-1},$$

which implies that $V = K(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = K(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1})$. Continue to throw away redundant generators until a basis is reached.

**Corollary to Steinitz:** All bases have the same size.

**Proof:** Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two bases. Applying Steinitz one way gives $|\mathcal{B}_1| \leq |\mathcal{B}_2|$ and applying it the other way gives $|\mathcal{B}_2| \leq |\mathcal{B}_1|$. Hence $|\mathcal{B}_1| = |\mathcal{B}_2|$. □

*We define the* **dimension** $\dim(V)$ *of a finitely generated vector space $V$ to be the common size of any basis.*

Exercise: Show that the dimension of a vector space $V$ coincides with the length of any unrefinable chain of subspaces:

$$\{\mathbf{0}\} < V_1 < V_2 < \cdots < V_{\dim(V)} = V.$$

This is the prototype for the concept of "dimension" in any branch of geometry, and many people assume that it is trivial. (After all, don't we teach this to freshmen?) As you see, it is *not* trivial, and relies on a subtle interplay of the vector space axioms. Hassler Whitney (1935) abstracted the Steinitz Exchange Property (even further) to define the notion of a **matroid** (a generalization of "matrix"), which is still an active topic of research.

**(2.D)** Since this is a book about classification, we will now attempt to classify finitely generated vector spaces. The first step is easy.

Given a field $K$ and a positive integer $n \in \mathbb{N}$, we construct the **Cartesian space**

$$K^n := \left\{ \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} : u_1, u_2, \ldots, u_n \in K \right\},$$

with "componentwise" addition and scalar multiplication:

$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} := \begin{pmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{pmatrix} \qquad \text{and} \qquad a \cdot \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} := \begin{pmatrix} au_1 \\ au_2 \\ \vdots \\ au_n \end{pmatrix}.$$

Exercise: Show that $(K^n, K)$ is a vector space, and observe that the "standard basis vectors"

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ \ldots, \ \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

are indeed a basis for $K^n$, hence $K^n$ is $n$-dimensional. It turns out that this is the only $n$-dimensional vector space over the field $K$.

**Classification Theorem:** If the vector space $(V, K)$ has dimension $n$, then

$$(V, K) \approx K^n.$$

**Proof (Choose Coordinates):** The symbol "$\approx$" of course means the existence of a bijection $V \longleftrightarrow K^n$ that preserves vector space structure. We will construct such a map by "choosing coordinates".

Let $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n \in V$ be *any* basis for $V$. We define a function $\varphi : V \to K^n$ on the basis by setting $\varphi(\mathbf{b}_i) := \mathbf{e}_i$ and we extend this ("linearly") to all of $V$ by setting

$$\begin{aligned}
\varphi(a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + \cdots + a_n\mathbf{b}_n) &:= a_1\varphi(\mathbf{b}_1) + a_2\varphi(\mathbf{b}_2) + \cdots + a_n\varphi(\mathbf{b}_n) \\
&= a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \cdots + a_n\mathbf{e}_n \\
&= a_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + a_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.
\end{aligned}$$

Clearly $\varphi$ is a bijection, with $\varphi^{-1}(\mathbf{e}_i) = \mathbf{b}_i$. It is also true (but too boring for a human to check) that $\varphi$ preserves vector space structure. □

Also note that there is a unique 0-dimensional vector space over $K$ consisting of a single vector, which we might as well call $\mathbf{0}$. We could extend notation by referring to this space as $K^0$ (consisting of the unique 0-tuple of numbers). [Highbrow Remark: The vector space $K^0$ is the zero object in the category of vector spaces over $K$.] This allows us to say the following:

*a finitely generated vector space ≡ (a field, a natural number).*

And if the field is understood, we can say:

*a finitely generated vector space ≡ a natural number (!)*

So, maybe the axiomatic definition of vector spaces was overkill? To complete the classification of finitely generated vector spaces, we must try to classify fields.

**(2.E)** Is it possible to classify fields? Not really. But I will at least shine a light on the terrain. The basic framework for this discussion was set down by Steinitz (1910). Unfortunately I will have to utter the word "ring". Just as a field is something like $\mathbb{Q}$, a ring is something like $\mathbb{Z}$:

*A (commutative)* **ring** *$R$ (with 1) satisfies all of the field axioms,* except

- $\forall\, 0 \neq a \in R, \exists\, b \in R, ab = 1.$

That is, in a ring we can add, subtract, and multiply; but **not** necessarily divide. As a word of warning: rings are infinitely more complicated/interesting than fields. The prototypical ring is $\mathbb{Z}$, and it satisfies a special property.

**Special Property of $\mathbb{Z}$:** *Let $R$ be any ring (commutative with 1, of course). Then there exists a unique ring map from $\mathbb{Z}$ to $R$. Indeed, if $\varphi : \mathbb{Z} \to R$ is any map preserving ring structure then we must have $\varphi(0_\mathbb{Z}) = 0_R$ and $\varphi(1_\mathbb{Z}) = 1_R$. It follows that*

$$\varphi(n) = \varphi(1_\mathbb{Z} + \cdots + 1_\mathbb{Z}) = \varphi(1_\mathbb{Z}) + \cdots + \varphi(1_\mathbb{Z}) = 1_R + \cdots + 1_R,$$

*and*

$$\varphi(-n) = -\varphi(n) = -(1_R + 1_R + \cdots + 1_R),$$

*and now the map is determined. [Highbrow Translation: $\mathbb{Z}$ is the initial object in the category of rings.]*

Now let $R = K$ be a field and consider the unique ring map $\varphi : \mathbb{Z} \to K$. The kernel of this map has the form $\ker \varphi = n\mathbb{Z} = \{na : a \in \mathbb{Z}\}$ for some $n \in \mathbb{N}$ (see Exercise **(2.3)**). By the Dedekind-Noether-First Isomorphism Theorem (see McLarty (2006)) we have

$$\mathbb{Z}/n\mathbb{Z} \approx \operatorname{im} \varphi \subseteq K.$$

Since $\operatorname{im} \varphi$ is a subring of a field it is a domain (has no zero divisors), hence $n\mathbb{Z}$ is a prime ideal of $\mathbb{Z}$. We conclude that $n = 0$ or $n = p$ for some prime $p \in \mathbb{N}$.

*We say that* $\operatorname{char}(K) := n$ *(either prime or zero) is the* **characteristic** *of the field $K$.*

Exercise: Does there exist a field of characteristic 1? If you didn't understand any of that, you can take $\operatorname{char}(K)$ to be the smallest positive integer $n$ such that $\varphi(n) = 0$ in $K$, and $\operatorname{char}(K) = 0$ if no such $n$ exists. (But that's much less elegant, don't you think?) Now here is the key definition of the present subsection:

*We say that a field is* **prime** *if it has no proper subfield.*

The following result of Steinitz (1910) is the foundation for any possible classification of fields.

**Classification of Prime Fields:** Every prime field is isomorphic to

$$\mathbb{Q} \quad \text{or} \quad \mathbb{F}_p := \mathbb{Z}/p\mathbb{Z} \text{ for some prime } p \in \mathbb{N}.$$

**Proof:** Let $K$ be a prime field and consider the unique ring map $\varphi : \mathbb{Z} \to K$. If $\operatorname{char}(K) = p > 0$ then we have

$$\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z} \approx \operatorname{im} \varphi \subseteq K.$$

Since $\mathbb{F}_p$ is a field (Euclid) we conclude that $\mathbb{F}_p \approx \operatorname{im} \varphi$ is a subfield of $K$. But $K$ has no proper subfields, hence $\mathbb{F}_p \approx \operatorname{im} \varphi = K$. If $\operatorname{char}(K) = 0$ then we have

$$\mathbb{Z} \approx \mathbb{Z}/0\mathbb{Z} \approx \operatorname{im} \varphi \subseteq K.$$

Then the smallest subfield of $K$ containing $\operatorname{im} \varphi \, (\approx \mathbb{Z})$ is isomorphic to $\mathbb{Q}$. But $K$ has no proper subfields, hence $\mathbb{Q} \approx K$. □

Thus "prime" is a reasonable name for these fields. Having classified the prime fields, we will try to reduce the study of general fields to the study of primes. Let $K$ be a general field.

*The intersection of all subfields of $K$ is called the* **prime subfield** *of $K$.*

Clearly the prime subfield of $K$ is prime, hence it is isomorphic to $\mathbb{Q}$ or $\mathbb{F}_p$, depending on the characteristic of $K$. We will thus be done if we can classify all field extensions of $\mathbb{Q}$ and $\mathbb{F}_p$. This problem is far too difficult, so I will merely offer a definition. This definition is my own since I couldn't find any comparable idea on the internet. Feel free to use it.

*We say that a field is* **numeric** *if it is algebraic over a topological completion of its prime subfield.*

To say that a field $K$ is "algebraic" over a subfield $k$ means that each element of $K$ satisfies a polynomial equation with coefficients in $k$, i.e., each *element* of $K$ is "algebraic" over $k$. Examples of numeric fields are: $\mathbb{F}_p$, $\mathbb{Q}$, $\mathbb{Q}[\sqrt{d}]$, $\mathbb{Q}_p$, $\mathbb{R}$, $\mathbb{C}$. An example of a non-numeric field is the field of rational functions in one variable:

$$K(x) := \left\{ \frac{f(x)}{g(x)} : f(x), g(x) \in K[x], g(x) \not\equiv 0 \right\}.$$

The non-algebraic element here is the "variable" $x \in K(x)$; it is called "transcendental". Non-numeric fields (also called **function fields**) lie in the domain of "algebraic geometry", which is quite different from "geometric algebra". Algebraic geometry is also quite interesting, but I will have nothing to say about it in this book. Please consult Jean Dieudonné (1972).

## Exercises

I should note that, while the classification of general fields is too difficult, *finite* fields **can** be classified. This classification was known to Galois (1830) and was one of the earliest motivations for the abstract field concept.

**(2.1)** Let $k$ be a subfield of $K$. Show that the action of $k$ on $K$ by multiplication defines a vector space structure $(K, k)$. Let $[K : k]$ denote the dimension of this vector space.

**(2.2)** Let $K$ be a finite field, so that its prime subfield is $\mathbb{F}_p$ for some prime $p > 0$. Show that $K$ is finitely generated as a vector space over $\mathbb{F}_p$, and hence $K \approx \mathbb{F}_p^n$ for some $n \in \mathbb{N}$. Conclude that

$$|K| = |\mathbb{F}_p^n| = |\mathbb{F}_p|^n = p^n.$$

Thus every finite field has size the power of a prime. Conversely, for each prime $p$ and for each $n \in \mathbb{N}$ you will show that there exists a field of size $p^n$.

**(2.3)** Let $K$ be a field and consider the ring of polynomials $K[x]$. We say that $I \subseteq K[x]$ is an **ideal** if

- $(I, +, 0)$ is a subgroup of $(K[x], +, 0)$,
- $\forall\, f(x) \in I, \forall\, g(x) \in K[x], f(x)g(x) \in I$.

Prove that every ideal $I$ of $K[x]$ is **principal** in the sense that

$$I = (f(x)) := \{f(x)g(x) : g(x) \in K[x]\}$$

for some $f(x) \in K[x]$. [Hint: Choose $0 \not\equiv f(x) \in I$ with *minimum* degree and consider any $g(x) \in I$. Divide $g(x)$ by $f(x)$ to obtain $q(x), r(x) \in K[x]$ such that $g(x) = q(x)f(x) + r(x)$ where $r(x)$ is either zero or $\deg(r) < \deg(f)$. Show that $\deg(r) < \deg(f)$ leads to a contradiction because $r(x) \in I$.]

We say that a polynomial $f(x) \in K[x]$ is **irreducible** if $f(x) = g(x)h(x)$ implies that at least one of $g(x)$ or $h(x)$ is an invertible element of $K[x]$ (i.e., a nonzero constant).

**(2.4)** Let $f(x) \in K[x]$ be an irreducible polynomial and show that the principal ideal $(f(x))$ is **maximal** in the sense that $(f(x)) < I \leq K[x]$ implies $I = K[x]$. [Hint: By **(2.3)** we know that $I = (g(x))$ for some $g(x) \in K[x]$. If $(f(x)) < (g(x)) < K[x]$, then we find that $g(x)$ is a **nontrivial divisor** of $f(x)$.]

**(2.5)** Let $\mathbf{m}$ be a maximal ideal in a ring $R$ and define the **quotient ring**

$$R/\mathbf{m} := \{a + \mathbf{m} : a \in R\}$$

with addition $(a+\mathbf{m})+(b+\mathbf{m}) := (a+b)+\mathbf{m}$, multiplication $(a+\mathbf{m})(b+\mathbf{m}) := ab+\mathbf{m}$, and equality $a+\mathbf{m} = b+\mathbf{m} \Leftrightarrow a-b \in \mathbf{m}$. Prove that $R/\mathbf{m}$ is a **field**. [Hint: Consider a nonzero element $a+\mathbf{m} \neq 0+\mathbf{m}$. Since $a \notin \mathbf{m}$ we get a strictly larger ideal $\mathbf{m} < (a)+\mathbf{m}$. Since $\mathbf{m}$ is maximal we have $1 \in R = (a)+\mathbf{m}$, hence $1 = ab+u$ for some $b \in R$ and $u \in \mathbf{m}$. But then $(a+\mathbf{m})(b+\mathbf{m}) = 1+\mathbf{m}$.]

**(2.6)** Now let $f(x) \in \mathbb{F}_p[x]$ be an irreducible polynomial of degree $n$ with coefficients in $\mathbb{F}_p$. By **(2.4)** and **(2.5)** we know that $\mathbb{F}_p[x]/(f(x))$ is a field. Consider the natural ring map $\varphi : \mathbb{F}_p[x] \to \mathbb{F}_p[x]/(f(x))$ defined by $g(x) \mapsto g(x) + (f(x))$. If we identify $\mathbb{F}_p \subseteq \mathbb{F}_p[x]$ with the constant polynomials, show that $\varphi(\mathbb{F}_p)$ is a subfield of $\mathbb{F}_p[x]/(f(x))$ isomorphic to $\mathbb{F}_p$. Furthermore, show that $\varphi(1), \varphi(x), \ldots, \varphi(x^{n-1})$ is a basis for $\mathbb{F}_p[x]/(f(x))$ as a vector space over $\varphi(\mathbb{F}_p) \approx \mathbb{F}_p$. Conclude that

$$\left| \frac{\mathbb{F}_p[x]}{(f(x))} \right| = |\mathbb{F}_p^n| = |\mathbb{F}_p|^n = p^n.$$

It remains only to show that irreducible polynomials of all degrees exist in $\mathbb{F}_p[x]$. This was first proved by Gauss (1889), pages 602–629, using generating functions. You will follow a less combinatorial and more algebraic method.

**(2.7)** Let $p$ be prime and let $n \in \mathbb{N}$. Consider the special polynomial $x^{p^n} - x \in \mathbb{F}_p[x]$. If $f(x) \in \mathbb{F}_p[x]$ is irreducible of degree $d$, show that

$$f(x) \text{ divides } (x^{p^n} - x) \quad \Longleftrightarrow \quad d \text{ divides } n.$$

[Hint: The multiplicative group of the finite field $\mathbb{F}_p[x]/(f(x))$ has size $p^d - 1$, hence we have $c^{p^d-1} = 1$ (and $c^{p^d-1} = c$) for all $c \in \mathbb{F}_p[x]/(f(x))$. If moreover we let $n = dk$, then raising to the $p^d$-th power $k$ times gives

$$c = c^{p^d} = c^{p^{2d}} = \cdots = c^{p^{kd}} = c^{p^n}$$

for all $c \in \mathbb{F}_p[x]/(f(x))$. In particular we have $x^{p^n} + (f(x)) = x + (f(x))$. Conversely, suppose that $x^{p^n} - x \in (f(x))$ and write $n = qd + r$ with $0 \leq r < d$. Since $d$ divides $qd$ we already know that $x^{p^{qd}} = x \bmod f(x)$. Hence

$$x = x^{p^n} = (x^{p^{qd}})^{p^r} = x^{p^r} \bmod f(x).$$

Recall the Freshman's Binomial Theorem, which say that $(a+b)^p = a^p + b^p \bmod p$ for $a, b$ in any ring. It follows that $g(x)^{p^r} = g(x) \bmod f(x)$ for any polynomial $g(x) \in \mathbb{F}_p[x]$. Thus every element of the field $\mathbb{F}_p[x]/(f(x))$ is a root of the polynomial $T^{p^r} - T \in \mathbb{F}_p[x]/(f(x))[T]$. If $r \neq 0$ then by **(1.2)** and **(2.6)** we conclude that $p^d \leq p^r$, hence $d \leq r$. Contradiction.]

**(2.8)** We have seen that the irreducible factors of $x^{p^n} - x$ in $\mathbb{F}_p[x]$ are precisely those irreducible polynomials with degree $d$ dividing $n$. Show that each of these factors occurs with multiplicity 1. [Hint: If $x^{p^n} - x$ has a repeated factor then it must share this factor with its formal derivative. But the formal derivative is $p^n x^{p^n-1} - 1 = 0x^{p^n-1} - 1 = -1$, which has no factors. Recall that $\mathbb{F}_p[x]$ has unique prime factorization (Euclid).]

Finally, let $N_p(n)$ be the number of irreducible polynomials in $\mathbb{F}_p[x]$ with degree $n$ and leading coefficient 1 (i.e., monic). Dividing through by leading coefficients and using **(2.7)** and **(2.8)**, we can write $x^{p^n} - x$ as the product of all monic irreducibles with degrees dividing $n$. Comparing degrees on both sides gives Gauss' formula:

$$p^n = \sum_{d|n} dN_p(d). \qquad (\star)$$

**(2.9)** Let $a, b, c \in \mathbb{N}$ be prime. Use $(\star)$ to show that for all $\alpha, \beta, \gamma \in \mathbb{N}$ we have:

- $a^\alpha N_p(a^\alpha) = p^{a^\alpha} - p^{a^{\alpha-1}}$
- $a^\alpha b^\beta N_p(a^\alpha b^\beta) = p^{a^\alpha b^\beta} - p^{a^{\alpha-1}b^\beta} - p^{a^\alpha b^{\beta-1}} + p^{a^{\alpha-1}b^{\beta-1}}$
- $a^\alpha b^\beta c^\gamma N_p(a^\alpha b^\beta c^\gamma) = \left\{ \begin{array}{ccc} & p^{a^\alpha b^\beta c^\gamma} & \\ -p^{a^{\alpha-1}b^\beta c^\gamma} & -p^{a^\alpha b^{\beta-1}c^\gamma} & -p^{a^\alpha b^\beta c^{\gamma-1}} \\ +p^{a^{\alpha-1}b^{\beta-1}c^\gamma} & +p^{a^{\alpha-1}b^\beta c^{\gamma-1}} & +p^{a^\alpha b^{\beta-1}c^{\gamma-1}} \\ & -p^{a^{\alpha-1}b^{\beta-1}c^{\gamma-1}} & \end{array} \right\}$

**(2.10)** Use **(2.9)** to show that $N_p(n) > 0$ for all $n \in \mathbb{N}$, and hence by **(2.6)** that finite fields *exist* of all sizes $p^n$. [Hint: We have

$$nN_p(n) = p^n + \sum \text{ distinct smaller powers of } p \text{ with coefficients from } \{+1, -1, 0\}.$$

The sum on the right is no less than

$$\sum_{0 \le i < n} -p^i = -1 - p - p^2 - \cdots - p^{n-1} = -\left(\frac{p^n - 1}{p - 1}\right) > -(p^n - 1) > -p^n.]$$

It turns out that the field of size $p^n$ is unique up to isomorphism (and we will denote it by $\mathbb{F}_{p^n}$). We won't prove this because you're probably tired of exercises. Google "existence and uniqueness of splitting fields" if you like.

**(2.11)*** A consequence of **(2.10)** is that there exist infinitely many irreducible polynomials in $\mathbb{F}_p[x]$. We may wish to examine their distribution. As $n \to \infty$ we have

$$N_p(n) \sim \frac{p^n}{n},$$

and substituting $X = p^n$ gives

$$N_p(n) \sim \frac{X}{\log_p X}.$$

Does this have a precise relationship to the Prime Number Theorem?

## Notes

Following the pioneering work of Évariste Galois and Richard Dedekind, the axiomatic definition of a field was first stated by Heinrich Weber (1893). The study of fields for their own sake began with the highly influential work of Ernst Steinitz (1910). Perhaps you think that the field concept has too many axioms? David Hilbert (1899) showed that it can be done with four axioms. Namely, he used a construction of von Staudt to show that

*a field ≡ a projective plane in which Pappus' Theorem holds.*

Warning: He attributed Pappus' Theorem incorrectly to Pascal (you know, the one with the "mystic hexagram").

As I mentioned earlier, an axiomatization of vector space was first given by Giuseppe Peano (1888). However, its level of abstraction was premature and it was not very influential. The axiomatic theory of vector spaces did not really take off until Stefan Banach's work in the 1920s and 1930s on infinite dimensional normed spaces (perhaps finite dimensional spaces were too simple to require an axiomatization). For the full story see Gregory Moore (1995).

I have proposed "vector space" as the mathematically most fundamental version of "space", but this is just my opinion. Pierre Cartier (2001) presents a fascinating discussion on the evolution of the concept of "space" and its role in the history of mathematics.

## Bibliography


Artin, Emil (1957). *Geometric Algebra*. Interscience, New York.

Descartes, Renée (1637). La Géométrie, Appendix in *Discours de la méthode pour bien conduire sa raison, et chercher la vérité dans les sciences.*

Dieudonné, Jean (1972). The Historical Development of Algebraic Geometry. *The American Mathematical Monthly*, **79**, 827–866.

Cartier, Pierre (2001). A mad day's work: from Grothendieck to Connes and Kontsevich The evolution of concepts of space and symmetry. *Bulletin of the American Mathematical Society*, **38**, 389–408.

Galois, Évariste (1830). Sur la théorie des nombres. *Bulletin des Sciences mathématiques*, **13**, 428.

Gauss, Carl Friedrich (1889). *Untersuchungen Über Höhere Arithmetik*. Springer, Berlin.

Hilbert, David (1899). *Grundlagen der Geometrie*. Teubner, Leipzig.

McLarty, Colin (2006). Emmy Noether's 'Set Theoretic' Topology: From Dedekind to the Rise of Functors. In *The Architecture of Modern Mathematics: Essays in history and philosopy (edited by Jeremy Grey and José Ferreirós)*. Oxford University Press.

Moore, Gregory H. (1995). The axiomatization of linear algebra: 1875–1940. *Historia Mathematica*, **22**, 262–303.

Peano, Giuseppe (1888). *Calcolo geometrico secondo l'Ausdehnungslehre di H. Grassmann, preceduto dalle operazioni della Iogica deduttiva*. Bocca, Turin.

Steinitz, Ernst (1910). Algebraische Theorie der Körper. *Journal für die reine und angewandte Mathematik*, **137**, 167–309.

Weber, Heinrich (1893). Die allgemeinen Grundlagen der Galois'schen Gleichungstheorie. *Mathematische Annalen*, **43**, 521–549.

Whitney, Hassler (1935). On the Abstract Properties of Linear Dependence. *American Journal of Mathematics*, **57**, 509–533.