Introduction to Fields

Four of the most basic structures of modern algebra are vector spaces, rings, groups and fields. The theory of these structures is covered in Math 2250, 3500, 4510 and 4520 respectively. (The theory of vector spaces is *linear algebra*, while rings, groups and fields are the three most basic of structures whose theory constitutes *abstract algebra*. The tradition of separating vector spaces from rings, groups and fields in this way is rather artificial—all four algebraic structures are defined abstractly, leading to rich and beautiful theory; and all four have equally concrete realizations and practical applications.) While fields were already used in Math 2250 (every vector space is defined over a particular choice of field of scalars) and fields were defined and many examples constructed in Math 3500 (after all, a field is a particular type of ring), a more thorough investigation of fields has been postponed until this final course[1] of our algebra sequence.

Fields in a Nutshell

Essentially, a field is a number system with addition, subtraction, multiplication and division. By 'division', we insist that we can divide by any nonzero field element. These operations are required to satisfy some basic 'niceness' properties (in particular the same laws of commutativity, associativity and distributivity as one is familiar with from college algebra classes). A few pages later, we will list axioms that clarify precisely what is meant by a 'field'. Logically, it makes sense to start witht the precise definition; but pedagogically, I feel it is better to give examples so the student can see what it is that is being abstracted, before giving the abstraction. Some important examples of fields are

- the field $\mathbb{R}$ of real numbers;
- the field $\mathbb{Q}$ of rational numbers; and
- the field $\mathbb{C}$ of complex numbers. Note that $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$. Also we have
- finite fields $\mathbb{F}_q$, i.e. fields with a finite number of elements $q = p^r$ where p is prime and $r \geqslant 1$.

[1] The course carries the more generic official title 'Topics in Algebra'. While the official course description allows some flexibility, it does highlight field theory; and I think it best to limit ourselves to the theory of fields throughout the course.

There are many more fields besides these. Note that $\mathbb{Z}$ (the ring of integers) is not a field, since it is not closed under division. Every field is a ring, but not conversely. We will define more precisely what is meant by a 'field', or a 'ring', and proceed to prove general results about fields; but first, it will help to discuss the specific examples of fields listed above, looking for common features as well as unique properties of each example of a field.

Before going further, let's make clear that if F is *any* field, then for each positive integer n, the set

$$F^n = \{(a_1, a_2, \ldots, a_n) : a_1, a_2, \ldots, a_n \in F\}$$

is a vector space over the field F. This means that, for example, we can use Gaussian elimination to solve systems of linear equations with coefficients in F, as taught in Math 2250 (Elementary Linear Algebra). This does not work over a general ring, even over $\mathbb{Z}$, where a/b is in general undefined. One reason for prefering fields to rings, then, is to allow us to solve linear systems.

The Real Numbers

The real number system $\mathbb{R}$ includes all the integers, all the rational numbers, and many other numbers besides. For example,

$$3,\ -11,\ 0,\ \tfrac{31}{6},\ -7.1556,\ \sqrt{2},\ \pi,\ \frac{17-\sin(14.2)}{1+\sqrt{6}}\ \in \mathbb{R}.$$

It does *not* include the complex number $i = \sqrt{-1}$; and it does not include ∞ or any infinitesimal values.

What exactly is meant by a 'real number'? One goal of our course will be to give a satisfactory definition of $\mathbb{R}$; but let's not tackle this challenge yet. First let's simply point out some important features of $\mathbb{R}$, including what makes it different from many other fields.

The field $\mathbb{R}$ is *totally ordered.* This means that given any two real numbers $a, b \in \mathbb{R}$, one and only one of the following three relations holds:

$$a < b, \quad a = b, \quad \text{or} \quad a > b.$$

(We call this the *trichotomy property*.) The relation '$<$' satisfies many important properties; for example, if $a < b$ and $c > 0$, then $ca < cb$. Most fields do not share this property; in particular, $\mathbb{C}$ and $\mathbb{F}_q$ are not ordered. (We cannot have $i > 0$ since this would force $-1 = i^2 < 0$, a contradiction; and if $-i < 0$ then we may obtain a similar contradiction.) The field $\mathbb{Q}$ is ordered, but $\mathbb{Q}$ differs from $\mathbb{R}$ in other important ways as we'll soon observe.

The field $\mathbb{R}$ is *complete.* Without giving too many details yet on this important feature of the real number system, we point out that this property is what gives us the Intermediate Value Theorem of Calculus. For example, the function $f(x) = x^2 - 2$ is continuous (since it is a polynomial). Since $f(1) < 0$ and $f(2) > 0$, there exists a real number c between 0 and 1 such that $f(c) = 0$. This statement is no longer true if we replace 'real' by 'rational'; and the important point is that $\mathbb{R}$ is complete, whereas $\mathbb{Q}$ is not. Details later...

The field $\mathbb{R}$ is *uncountable.* This means we *cannot* list the real numbers as $\mathbb{R} = \{a_1, a_2, a_3, \ldots\}$; there are simply too many real numbers to list this way. This property also distinguishes $\mathbb{R}$ from $\mathbb{Q}$, since the rational numbers *can* be listed in an infinite sequence (i.e. $\mathbb{Q}$ is *countable*). We will soon elaborate on this point as well.

While $\mathbb{R}$ is an uncountable, totally ordered, complete field, it is not the only field with these properties. How then do we define $\mathbb{R}$? One may be tempted to say that real numbers correspond to points of a physical line; but this definition is unsatisfactory for several reasons. For one, this intuitive description is so imprecise and nebulous that it cannot be used in a proof. Furthermore, it is not even true in the strict physical sense. (You may find this fact hard to swallow, but points on a physical line in space, or in time, are not totally ordered!)

Another popular proposal for defining real numbers appeals to decimal representations, identifying each real number by its decimal expansion. This is better, but still unsatisfactory, for several reasons:

- ◇ In order to be correct, one must first recognize that the decimal expansion of a real number is not necessarily unique; for example, the five expressions

$$1, \qquad 01, \qquad 1.00, \qquad 1.000000\ldots, \qquad \text{and} \qquad 0.999999\ldots$$

  all represent *the same real number.* (The equality $0.9999\ldots = 1$ is in itself is an important fact that we must soon discuss! Without coming to terms with this fact, one cannot understand what the real number system is about.)

- ◇ Even if one resolves the previous technical difficulty of non-unique decimal representations, one is left with the unpleasant fact that despite the practical value of decimal expansions for everyday use, decimal expansions are of practically no value in proving even the most basic facts about $\mathbb{R}$. For example, if one wants to prove the Intermediate Value Theorem, decimal expansions are of no value! so you'd better have a definition of $\mathbb{R}$ that does not refer to decimal expansions. Consider that even if f is continuous, so that small changes in x yield small changes in $f(x)$, this fact is not evident in the decimal digits, where relationship between the decimal digits of x and

$f(x)$ can be very mysterious and hard to predict. (Consider that the decimal digits of x are not related to the decimal digits of $f(x) = x^2 - 2$ in any obvious way. Even in the simplest case where $f(x) = 0$, there is no clear description of the decimal digits of $x = \sqrt{2}$. Nobody even knows if infinitely many decimal digits of $\sqrt{2}$ are equal to 3!)

- ◇ Furthermore, why should the definition of 'real number' make any particular reference to the number 10? Simply because humans have (usually) ten fingers? Is $\mathbb{R}$ as universal as our intuition says, or would we obtain an essentially different number system if we represented numbers in binary (base 2)? or in some other base? or in some other notational system that avoids bases altogether?

One should regard decimal expansions as one possible (and often convenient) way to represent real numbers; but $\mathbb{R}$ exists independently of any particular representation. For example, the expressions

$$3.14159265358979323846264338327950288419 71\ldots \quad \text{(decimal expansion)},$$

$$11.00100100001111110110101010001000100 1011\ldots \quad \text{(binary expansion)},$$

$$3+\cfrac{1}{7+\cfrac{1}{15+\cfrac{1}{1+\cfrac{1}{292+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\ddots}}}}}}} \quad \text{(continued fraction expansion), and}$$

$$\int_{-1}^{1} \sqrt{4-4x^2}\,dx \quad \text{(one of many possible integral representations)}$$

all represent the same real number that we call π.

What then *is* a real number? We cannot answer directly; our definition of the real numbers is founded upon the rational number system $\mathbb{Q}$, just as $\mathbb{Q}$ is in turn founded upon $\mathbb{Z}$, and $\mathbb{Z}$ upon the natural number system $\mathbb{N}$. So a strictly logical approach to constructing the real numbers would start at the bottom with a definition of the natural numbers; upon this foundation one proceeds to build the rational numbers; and from there we shall be able to define the real number system. Unfortunately all of this would take too long and involve several notions from outside basic field theory. Because our first goal is a primarily algebraic introduction to fields, we will unfortunately need to take much of this for granted.

The Rational Numbers

Building the rational number system $\mathbb{Q}$ is considerably easier than building $\mathbb{R}$; yet it is not entirely straightforward. In trying to define $\mathbb{Q}$, we may take it for granted that the integer number system $\mathbb{Z}$ has already been defined, and that its basic properties are known. The naive definition

$$\mathbb{Q} = \left\{ \tfrac{a}{b} \ : \ a, b \in \mathbb{Z}, \ b \neq 0 \right\}$$

is problematic because division of integers has not been defined in $\mathbb{Z}$. We will resolve this difficulty a little later, essentially by taking $\frac{a}{b}$ as a new symbol; but this must be done with care, since just as with $\mathbb{R}$, different symbols do not necessarily represent different rational numbers (e.g. the distinct symbols $\frac{10}{-4}$ and $\frac{-5}{2}$ represent the same element of $\mathbb{Q}$). Let's resolve these issues a little later.

The rational number system is a field; in fact, it is a subfield of $\mathbb{R}$. Like $\mathbb{R}$, it is totally ordered. Unlike $\mathbb{R}$, $\mathbb{Q}$ is countable. And unlike $\mathbb{R}$, $\mathbb{Q}$ has no proper subfields: the only subfield of $\mathbb{Q}$ is $\mathbb{Q}$ itself, whereas $\mathbb{R}$ has *many* subfields.

The Complex Numbers

What is the complex number system $\mathbb{C}$? If one understands the real number system $\mathbb{R}$, then it is just a small step up to the complex number system

$$\mathbb{C} = \{a + bi : a \in \mathbb{R}\}$$

where $i^2 = -1$. (We typically write $i = \sqrt{-1}$, but beware: which of the two square roots of -1 does this refer to? i.e. which one is i, and which one is $-i$? Keep in mind that there is no prior concept, based upon the theory of real numbers, that allows us to answer this question.) We see that $\mathbb{C}$ is a field; the essential feature, namely the ability to divide by any nonzero complex number, is exemplified by

$$\frac{1.2 + 3.1i}{1.1 - 0.7i} = \frac{1.2 + 3.1i}{1.1 - 0.7i} \cdot \frac{1.1 + 0.7i}{1.1 + 0.7i} = \frac{-0.85 + 4.25i}{1.7} = -0.5 + 2.5i.$$

The complex number system is *algebraically closed*: Every polynomial

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where $a_0, a_1, \ldots, a_n \in \mathbb{C}$ factors as

$$f(x) = a_n (x - r_1)(x - r_2) \cdots (x - r_n)$$

for some $r_1, r_2, \ldots, r_n \in \mathbb{C}$. This assertion is known as the Fundamental Theorem of Algebra. Any field F with this property (that polynomials in $F[x]$ factor into linear factors in $F[x]$; equivalently, every polynomial of degree $n \geqslant 1$ has a root in F) is called *algebraically closed*. So the Fundamental Theorem of Algebra is the assertion that $\mathbb{C}$ is algebraically closed. Note that $\mathbb{R}$ is not algebraically closed, since there exist polynomials such as $2 + 3x + 5x^2 \in \mathbb{R}[x]$ which cannot be factored into linear factors in $\mathbb{R}[x]$. Similarly, $\mathbb{Q}$ is not algebraically closed. Although algebraic closure is the most important property of $\mathbb{C}$, there are many other fields that are algebraically closed. In fact there are smaller fields than $\mathbb{C}$ that are algebraically closed. (The field $\mathbb{C}$ is uncountable, in fact $|\mathbb{C}| = |\mathbb{R}|$; and there exist countable fields that are algebraically closed. Indeed, $\mathbb{C}$ has proper subfields with this property.)

The complex number system has an automorphism known as complex conjugation, satisfying $\overline{a + bi} = a - bi$ for all $a, b \in \mathbb{R}$. Neither $\mathbb{R}$ nor $\mathbb{Q}$ has any automorphism other than the identity map $x \mapsto x$ (although this is not obvious). By contrast, the complex number system has at least two automorphisms, the identity and complex conjugation. In fact, the complex number system has *very many* automorphisms, although this is far from obvious.

The term 'complex' derives from the fact that every complex number $a + bi$ has two parts $a, b \in \mathbb{R}$. Here a 'complex' is a combination of two or more parts, just as a sports complex may be a combination of arena, stadium, locker rooms, etc.; and vitamin B-complex is the combination of vitamins B_1, B_2, B_3, B_5, etc. This usage of the term 'complex' has nothing to do with 'complicated'; in fact, complex numbers *simplify* the study of real numbers in many ways. (In particular, the complicated nature of factorization of polynomials in $\mathbb{R}[x]$ is best understood by first understanding the simpler situation in $\mathbb{C}[x]$.)

Just as the 'real numbers' are useful in modeling many physical quantities such as distance, time, mass, temperature, etc., so also the 'complex numbers' are the best device available for representing many other physical quantities. A classical example arises in the modelling of waves and vibrations, where a single complex-valued function $A(t)$ may be used to represent two aspects of an oscillating quantity by way of its real and imaginary parts (e.g. voltage and current in an LC circuit; or kinetic and potential energy in a swinging pendulum). This is a weak example, since the role of complex numbers serves here merely as a bookkeeping device, achieving a slightly more concise description of the quantities involved—the formulas are shortened by at most 50%. A better example is found in the theory of electromagnetic fields: the field of electrostatic potential in a charge-free two-dimensional region is conveniently represented by a complex analytic function whose

real and imaginary parts u and v describe two features of the field, including its static features and the dynamics of a small test charge released in the field. Here one can describe the same features using real analysis: the functions u and v constitute a conjugate pair of harmonic real-valued functions. However, the single complex analytic function $u + iv$ is *much* easier to describe than either of the individual functions u or v, and the complex viewpoint offers a savings of much more than 50% in terms of analytic skill and shear paper required for the relevant calculations.

However, a much more convincing example of the physical relevance of complex numbers arises in the theory of quantum mechanics, where the role of complex numbers is practically inescapable. The wave function of a quantum mechanical system (such as a particle or bundle of particles) is a single complex-valued function Ψ. While it is in principle possible to work with the real and imaginary parts of Ψ separately, the technical requirements of doing so would be so formidable; and such loss of clarity is suffered in the process, that *no* serious treatment of quantum mechanics is possible unless one works over the complex numbers. In short, from a practical viewpoint, quantum mechanics *requires* the complex numbers. This point becomes even more forceful when considering the deeper physical theories arising from unified field theories. The moral: We cannot avoid $\mathbb{C}$ if we want to do physics.

However, $\mathbb{C}$ is *not everything*. One can (and should, if one wants to properly understand physics) extend $\mathbb{C}$ to the 4-dimensional skewfield of *real quaternions*

$$\mathbb{H} = \{a + bi + cj + dk \; a, b, c, d \in \mathbb{R}\}$$

where $i^2 = j^2 = k^2 = ijk = -1$, and to the still larger 8-dimensional algebra of *octonions* $\mathbb{O}$. We won't be spending time on $\mathbb{H}$ or $\mathbb{O}$ beyond simply mentioning them, because they are not commutative and hence not strictly fields (in fact $\mathbb{O}$ is not even associative). Modern physics requires even larger systems, including Clifford algebras of much larger dimension, not to mention infinite-dimensional algebras and vector spaces.

Finite Fields

For every prime power $q = p^r$ (where p is prime and r is a positive integer) there is a field of *order* q, i.e. having exactly q elements, denoted by $\mathbb{F}_q$, or sometimes $GF(q)$ (where GF stands for 'Galois field'). The order of any finite field is necessarily a prime power (and we will soon see an easy proof of this fact). More subtle is the fact that for every prime power q, there exists a field of order q; and it is unique up to isomorphism. So there exists a unique field of each order 2, 3, 4, 5, 7, 8, 9, 11, etc.; but there do not exist fields of order

6, 10, 12, etc. In the special case of prime order p (i.e. for $r = 1$), the field[1] of order p is given simply by

$$\mathbb{F}_p = \{0, 1, 2, \ldots, p-1\}$$

with addition and multiplication defined 'mod p'.

As an example of a computation in $\mathbb{F}_{11}$, we ask for the solution of the linear system in two unknowns:

$$7x + 4y = 2;$$

$$2x + 10y = 8.$$

Multiplying the first equation by $\frac{1}{7} = 8$ and the second equation by $\frac{1}{2} = 6$ yields the equivalent linear system

$$x + 10y = 5;$$

$$x + 5y = 4.$$

Subtracting yields $5y = 5 - 4 = 1$ so that $y = \frac{1}{5} = 9$. Finally, $x = 5 - 10y = 5 - 90 = -85 = 3$.

The most difficult steps required in the latter example are divisions. In order to perform division in $\mathbb{F}_p$, one can use a quick search for small p; but for large p, the extended Euclidean algorithm works very well. This should be familiar from previous courses, including Math 3500; but I will provide a review of this and related background material (check the Masth 4520 course website). Please note that all computations in the latter example take place in $\mathbb{F}_{11}$. One can alternatively work in $\mathbb{Z}$ if one is careful to replace equalities by congruences mod 11, e.g. the last step is $-85 \equiv 3 \mod 11$.

The field of order 4 is given by $\mathbb{F}_4 = \{0, 1, \alpha, \beta\}$ where addition and multiplication are given by the tables

$+$	0	1	α	β
0	0	1	α	β
1	1	0	β	α
α	α	β	0	1
β	β	α	1	0

$\cdot$	0	1	α	β
0	0	0	0	0
1	0	1	α	β
α	0	α	β	1
β	0	β	1	α

The field of order 9 may be expressed as

$$\mathbb{F}_9 = \{a + bi \,:\, a, b \in \mathbb{F}_3\}$$

[1] Many undergraduate textbooks denote the field of order p by $\mathbb{Z}_p$. This unfortunately conflicts with the standard notation $\mathbb{Z}_p$ for the ring of p-adic integers, described below. Moreover since $\mathbb{F}_q$ is *not* the integers mod q unless q is prime, it is better not to write $\mathbb{Z}_q$ if one intends to refer to a finite field.

where $i = \sqrt{-1}$. Let's show some typical computations in this field. If $u = 1 + 2i$ and $v = 1 + i$ then

$$u + v = 2 + 3i = 2;$$
$$u - v = i;$$
$$uv = 1 + 2i + i - 2 = 1;$$
$$\frac{u}{v} = \frac{1+2i}{1+i} \cdot \frac{1-i}{1-i} = \frac{i}{2} = 2i.$$

This example gives a good idea how more general finite fields may be constructed, of arbitrary prime power order.

Finite fields are not ordered. Consider $\mathbb{F}_2 = \{0, 1\}$ as an example: if $0 < 1$ then adding 1 to both sides yields $1 < 2$, i.e. $1 < 0$, a contradiction. There is no way to put an ordering on the elements of a finite field $\mathbb{F}_q$, while satisfying the usual requirements of an ordering.

Finite fields have numerous applications, for example in coding theory and cryptography. (By 'coding theory' we mean the theory of error-correcting codes, used to protect information from accidental environmental contamination due to environmental factors such as static and noise.) Most current literature using finite fields is published not by professional mathematicians, but rather by electrical engineers. This may come as no surprise, considering the the discrete nature of most information (which is typically digitized, if not already in digital form). In fact given the possibility of representing information as streams of 0's and 1's, you may suspect that it is really only $\mathbb{F}_2$ that we require; but this would hardly justify the pretense of invoking finite fields. In fact it is more general fields of the form $\mathbb{F}_q$ and $\mathbb{F}_q((x))$ that we require, often (but not exclusively) with $q = 2^r$ and r large. (Fields of the form $F((x))$ are introduced below.) Unfortunately we do not have time to dwell on these applications.

For anyone who has struggled with roundoff error arising in practical computations in $\mathbb{R}$ or $\mathbb{C}$, it is a tremendous relief to work instead over finite fields where roundoff error is not an issue at all: computations are exact. Even if large finite fields are involved, overflow is not a concern because modern computers easily implement arbitrary precision arithmetic. Over $\mathbb{R}$ or $\mathbb{C}$, computations are typically approximate; but over $\mathbb{F}_q$ exact computations are not only required, but readily achievable.

Other Fields

Many other examples of fields arise in both theory and practice. Here we mention just a few of the more important ones:

◇ There are many fields F between $\mathbb{Q}$ and $\mathbb{C}$, i.e. $\mathbb{Q} \subseteq F \subseteq \mathbb{C}$. These include fields that arise by throwing in roots of certain polynomials with rational coefficients. Here we refer to the *algebraic number fields* which are described more fully in courses on number theory. An example is

$$\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \,:\, a, b \in \mathbb{Q}\}.$$

A typical division exercise in $\mathbb{Q}[\sqrt{2}]$ illustrates why this is a field:

$$\frac{3+5\sqrt{2}}{7+3\sqrt{2}} = \frac{3+5\sqrt{2}}{7+3\sqrt{2}} \cdot \frac{7-3\sqrt{2}}{7-3\sqrt{2}} = \frac{51+44\sqrt{2}}{31} = \frac{51}{31} + \frac{44}{31}\sqrt{2}\,.$$

◇ If F is any field, then the set of all polynomials in x with coefficients in F forms a ring $F[x]$. This is not a field; but by taking quotients we obtain the *field of rational functions* with coefficients in F, namely

$$F(x) = \left\{ \frac{f(x)}{g(x)} \,:\, f(x), g(x) \in F[x],\ g(x) \neq 0 \right\}.$$

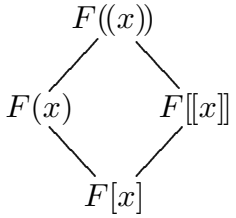
◇ Again, if F is any field, then the set of power series in x with coefficients in F forms a ring

$$F[[x]] = \{a_0 + a_1 x + a_2 x^2 + \cdots \,:\, a_0, a_1, a_2, \ldots \in F\}.$$

This is not a field; but by allowing finitely many terms with negative exponents, we obtain the *field of Laurent series*

$$F((x)) = \{a_n x^n + a_{n+1} x^{n+1} + a_{n+2} x^{n+2} + \cdots \,:\, a_n, a_{n+1}, a_{n+2}, \ldots \in F;\ n \in \mathbb{Z}\}$$

(these being in fact particular examples of *Puiseux series*). The containments between the rings $F[x]$ and $F[[x]]$, and their respective quotient fields $F(x)$ and $F((x))$, are illustrated by the diagram



◇ Although the field $\mathbb{R}$ does not contain any infinite or infinitesimal elements, there is a larger field of *hyperreal numbers* which does contain infinite and infinitesimal elements, in addition to the standard real numbers (which are finite). The field of hyperreal numbers has many uses, both theoretical and practical. In the development

of calculus, in particular, hyperreal numbers have played a much more prominent role historically than is recognized today, where the traditional presentation of calculus has been revised to allow real numbers only. Some experts have argued that a return to the hyperreal formulation of calculus, in which infinite and infinitesimal quantities are legitimized, offers pedagogical advantages over the current traditional approach. The force of tradition is so strong, in mathematics education as in many other disciplines, that in fact many professional mathematicians are not conversent with the hyperreal number system. More about the hyperreal numbers is covered in Math 5090 (Set Theory and Mathematical Logic), which we teach once every 3–4 years.

◇ For every prime p, we have a *field of p-adic numbers* $\mathbb{Q}_p$ constructed by a process of completing $\mathbb{Q}$ in much the same way as $\mathbb{R}$ is obtained from $\mathbb{Q}$; the key difference here is to use a different notion of distance. For example consider the real number

$$\pi = 3 + \frac{1}{10} + \frac{4}{10^2} + \frac{1}{10^3} + \frac{5}{10^4} + \frac{9}{10^5} + \cdots \in \mathbb{R}.$$

The partial sums of this infinite series are the rational numbers 3, 3.1, 3.14, 3.141, etc. which are successively closer approximations to π; and the actual value of π is the limit of this sequence of approximations. This works because the sequence of partial sums forms a *Cauchy sequence*, which (postponing the technical details) means that the terms of the sequence get closer to one another. In a similar way, a typical 11-adic number is

$$a = 3 + 9{\cdot}11 + 4{\cdot}11^2 + 1{\cdot}11^3 + 4{\cdot}11^4 + \cdots \in \mathbb{Q}_{11}\,.$$

The partial sums of this series form a sequence of rational numbers 3, 102, 586, 1917, 60481, ... which converges (not in $\mathbb{R}$, but in $\mathbb{Q}_{11}$) to a number $a \in \mathbb{Q}_{11}$ which satisfies $a^2 = -2$, i.e. $a = \sqrt{-2}$. This is because the approximations solve the equation $x^2 = -2$ modulo increasing powers of 11:

$$\begin{aligned}
3^2 = 9 &\equiv -2 \mod 11;\\
102^2 = 10404 &\equiv -2 \mod 11^2;\\
586^2 = 343396 &\equiv -2 \mod 11^3;\\
1917^2 = 3674889 &\equiv -2 \mod 11^4;\\
60481^2 = 3657951361 &\equiv -2 \mod 11^5;
\end{aligned}$$

etc. Depending on how the semester proceeds, we may or may not have time to spend on p-adic fields. But for now, let us mention that p-adic fields are uncountable fields similar to $\mathbb{R}$, but in some ways simpler. For example, in Math 2205 (Calculus II) we

are forced to recognize that for an infinite series to converge, it is necessary, but *not sufficient*, that its terms approach zero in the limit. Over p-adic fields, an infinite series converges *if and only if* its terms approach zero in the limit. Many questions about $\mathbb{Z}$, or about $\mathbb{Q}$, are best answered by appealing to p-adic fields; this accounts for their importance in number theory. In passing, let me mention that the field $\mathbb{Q}_p$ is the field of quotients of *ring of p-adic integers* $\mathbb{Z}_p$. Every p-adic number $a \in \mathbb{Q}_p$ has a p-adic expansion

$$a = a_n p^n + a_{n+1} p^{n+1} + a_{n+2} p^{n+2} + \cdots$$

where $a_n, a_{n+1}, a_{n+2}, \ldots \in \{0, 1, 2, \ldots, p-1\}$ and $n \in \mathbb{Z}$; if we restrict to $n \geqslant 0$ then we obtain a p-adic integer. (So the subring $\mathbb{Z}_p \subset \mathbb{Q}_p$ is similar to the subring $F[[x]] \subset F((x))$.) For example the 11-adic number $\sqrt{-2}$ shown above, is actually an 11-adic integer.

Caution: p-adic numbers are *not* the same thing as real numbers expressed in base p. Nor are they the same thing as real numbers 'mod p' (which, in fact, doesn't give anything meaningful).

The Axioms of Field Theory

By definition, a *field* is a commutative ring with identity (also called a 'unity' element 1) such that every nonzero element is a unit (i.e. has a multiplicative inverse). This definition builds upon prior definitions, so let's give a self-contained definition which recaps all the prerequisite notions.

Let F be a set with at least two distinct (and distinguished) elements $0, 1 \in F$, together with two binary operations (addition and multiplication). Implicit in the definition of 'binary operation' is the requirement that for all $a, b \in F$, we have well-defined elements $a+b \in F$ and $ab \in F$. (This, and other considerations arising in the task of providing axioms for a field, should be familiar to you from previous experience with axioms for vector spaces, groups and rings.) We say that F is a *field* if the following conditions are satisfied:

(F1) $a + b = b + a$ for all $a, b \in F$;
(F2) $a + 0 = a$ for all $a \in F$;
(F3) $a + (b + c) = (a + b) + c$ for all $a, b, c \in F$;
(F4) For all $a \in F$ there exists an element in F, which we denote by $-a$, such that $a + (-a) = 0$;
(F5) $ab = ba$ for all $a, b \in F$;

(F6) $(ab)c = a(bc)$ for all $a, b, c \in F$;

(F7) $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$ for all $a, b \in F$. Here multiplication takes precedence over addition, e.g. $ab + ac$ means $(ab) + (ac)$;

(F8) $1a = a$ for all $a \in F$; and

(F9) For all *nonzero* $a \in F$, there exists an element in F, which we denote by a^{-1} or $\frac{1}{a}$, such that $a^{-1}a = aa^{-1} = 1$.

Here

- the axioms (F1)–(F4) assert that F is an additive abelian group with identity element 0;
- (F1)–(F4), (F6) and (F7) together assert that F is a ring;
- (F1)–(F8) assert that F is a commutative ring with identity; and
- (F6), (F8) and (F9) assert that the nonzero elements of F form a multiplicative group. By (F5), this group is abelian.

Note that $\mathbb{Z}$ satisfies all these axioms except (F9); it is a commutative ring with identity, whose only units are 1 and -1. Also $\mathbb{H}$ satisfies all these axioms except (F5); it is a noncommutative skewfield (i.e. 'division ring'). We proceed to summarize several properties of an arbitrary field which follow directly from the axioms. Many of these properties were discussed already in Math 3500 (Ring Theory) and so we do not bother reprinting all the proofs here:

For all $a, b \in F$ we have

$$-(-a) = a; \quad a(-b) = (-a)b = -ab; \quad (-a)(-b) = ab.$$

There is a well-defined binary operation of *subtraction* defined by

$$a - b = a + (-b)$$

which satisfies

$$a(b - c) = ab - ac; \quad 0 - a = -a.$$

Also if $a, b \in F$ with $b \neq 0$, we define *division* by

$$\frac{a}{b} = a/b = ab^{-1} = b^{-1}a.$$

Exponentiation is defined by

$$a^0 = 1; \quad a^{m+1} = a^m a \text{ for all } a \in F,\ m \in \mathbb{Z}$$

and satisfies

$$a^{m+n} = a^m a^n; \quad a^{m-n} = \frac{a^m}{a^n}; \quad (ab)^m = a^m b^m.$$

We exclude division by zero in the latter formulas (e.g. 0^n is undefined if $n \leqslant -1$) but we do define $0^0 = 1$. Similarly, integer multiples are defined by

$$na = \underbrace{a + a + \cdots + a}_{n \text{ times}}$$

if $n \geqslant 1$; also $(-na) = n(-a) = -(na)$ if n is a negative integer and $a \in F$; and $0a = 0$ for all $a \in F$. Note that $0a = 0$, $1a = a$ and $(-1)a = -a$ whether we regard the coefficients as elements of F, or integers; fortunately, there is no ambiguity in the meaning of these expressions. Furthermore, integer multiples satisfy the following:

Proposition 1. For all $m, n \in \mathbb{Z}$ and $a, b \in F$ we have

(i) $(m + n)a = ma + na$;

(ii) $m(na) = (mn)a$;

(iii) $m(a + b) = ma + mb$; and

(iv) $m(ab) = (ma)b$.

These identities are easy to prove, but they are *not* directly listed among the field axioms, since we refer not to multiplication in F, but rather to integer multiples of field elements, which are defined in terms of *addition* in F. Because of the importance of these identities in our next theorem, we provide explicit proofs.

Proof. (i)
$$\begin{aligned} ma + na &= \big(\underbrace{a+a+a+\cdots+a}_{m \text{ terms}}\big) + \big(\underbrace{a+a+a+\cdots+a}_{n \text{ terms}}\big) \\ &= \underbrace{a+a+a+a+\cdots+a}_{m+n \text{ terms}} \\ &= (m+n)a; \end{aligned}$$

(ii)
$$\begin{aligned} m(na) &= \underbrace{\big(\underbrace{a+a+\cdots+a}_{n \text{ terms}}\big) + \big(\underbrace{a+a+\cdots+a}_{n \text{ terms}}\big) + \cdots + \big(\underbrace{a+a+\cdots+a}_{n \text{ terms}}\big)}_{m \text{ parenthesized sums}} \\ &= \underbrace{a+a+a+a+\cdots+a}_{mn \text{ terms}} \\ &= (mn)a; \quad \text{and} \end{aligned}$$

(iii)
$$\begin{aligned} m(a+b) &= \underbrace{(a+b)+(a+b)+\cdots+(a+b)}_{m \text{ parenthesized sums}} \\ &= \big(\underbrace{a+a+\cdots+a}_{m \text{ terms}}\big) + \big(\underbrace{b+b+\cdots+b}_{m \text{ terms}}\big) \\ &= ma + mb. \end{aligned}$$

We leave the proof of (iv) as an exercise. □

Field Extensions

Given fields $E \supseteq F$ such that F is a subring of E (i.e. the elements $0, 1 \in F$ agree with the zero and identity elements of E; and the addition and multiplication operations in F are restrictions of the corresponding operations in E) then we say that F is a *subfield* of E, or that E is an *extension field* of F. For example, $\mathbb{C}$ is an extension of $\mathbb{R}$; and both of these are extensions of $\mathbb{Q}$.

Given any extension of fields $E \supseteq F$, the larger field E may always be considered as a vector space over the subfield F. The dimension of this vector space is called the *degree* of the extension, denoted $[E : F]$. For example $[\mathbb{C} : \mathbb{R}] = 2$ since $\{1, i\}$ is a basis for $\mathbb{C}$ over $\mathbb{R}$. (This is because every complex number has a *unique* representation of the form $a1 + bi$ where $a, b \in \mathbb{R}$.) Both $[\mathbb{C} : \mathbb{Q}]$ and $[\mathbb{R} : \mathbb{Q}]$ are infinite. Also $[\mathbb{Q}[\sqrt{2}] : \mathbb{Q}] = 2$ since $\{1, \sqrt{2}\}$ is a basis for $\mathbb{Q}[\sqrt{2}]$ over $\mathbb{Q}$. And $[\mathbb{F}_4 : \mathbb{F}_2] = 2$ since $\{1, \alpha\}$ is a basis for $\mathbb{F}_4$ over $\mathbb{F}_2$. Finally, $[F : F] = 1$ for every field F since $\{1\}$ is a basis for F over itself.

We often consider a *tower* of fields, which is a sequence of extension fields of the form

$$F_n \supseteq F_{n-1} \supseteq \cdots \supseteq F_2 \supseteq F_1 \supseteq F_0.$$

In order to understand the relationship between the various degrees in such a tower, we consider first a simple tower of the form

$$E \supseteq K \supseteq F.$$

We call E, K and F the *extension field*, the *intermediate field*, and the *base field* in this tower. Note that E is an extension of both K and F. The degrees of the extensions are related by

Theorem 2 (Transitivity of Degree for Field Extensions). Given a tower of fields $E \supseteq K \supseteq F$, the degrees satisfy

$$[E : F] = [E : K][K : F].$$

Proof. First suppose all three degrees are finite. Let $\{\alpha_1, \ldots, \alpha_m\}$ be a basis for E over K, and let $\{\beta_1, \ldots, \beta_n\}$ be a basis for K over F. Every element $a \in E$ can be uniquely written as $a = \sum_{i=1}^{m} a_i \alpha_i$ where each $a_i \in K$; and for each i, there exist unique constants $a_{i,j} \in F$ such that $a_i = \sum_{j=1}^{n} a_{i,j} \beta_j$. Now

$$a = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{i,j} \beta_j \right) \alpha_i = \sum_{i,j} a_{i,j} \alpha_i \beta_j$$

where i ranges from 1 to m, and j ranges from 1 to n throughout. So E is spanned by $\{\alpha_i \beta_j : 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n\}$ over K. To see that this is actually a basis, suppose that

$$a = \sum_{i,j} c_{i,j} \alpha_i \beta_j = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} c_{i,j} \beta_j \right) \alpha_i$$

for some $c_{i,j} \in F$. Since the α_i's form a basis for E over K, the uniqueness of coefficients implies that

$$\sum_{j=1}^{n} c_{i,j} \beta_j = \sum_{j=1}^{n} a_{i,j} \beta_j$$

for all $i \in \{1, 2, \ldots, m\}$; and now the uniqueness of coefficients with respect to the β_j's imples that $c_{i,j} = a_{i,j}$ for all i, j. We have shown that every element of E is a *unique* linear combination of the elements $\alpha_i \beta_j$, so

$$[E : F] = mn = [E : K][K : F].$$

We leave it as an exercise to show that $[E : F]$ is infinite iff either $[E : K]$ or $[K : F]$ is infinite. □

One should compare degrees of field extensions, with indices for subgroups in the theory of groups. Recall that if H is a subgroup of G, denoted by $H \leqslant G$, then the *index* of H in G. denoted by $[G : H]$, is the number of left cosets of H in G. Given a tower of groups $K \leqslant H \leqslant G$, the indices are related by

$$[G : K] = [G : H][H : K].$$

The analogy between this result and Theorem 2 exhibits a strong link between group theory and field theory, which is no coincidence. This connection will recur throughout the course.

The Characteristic of a Field

Given a field F, if

$$n1 = \underbrace{1+1+1+\cdots+1}_{n \text{ terms}} = 0$$

for some positive integer n, then the smallest such n is called the *characteristic of* F, and is denoted by $\operatorname{char} F$. Otherwise, we say that F has characteristic zero, denoted $\operatorname{char} F = 0$. (Why not $\operatorname{char} F = \infty$, one might ask? The reason for this definition will present itself in the proof of Theorem 3.) Thus for example the fields $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$ and $\mathbb{Q}_p$ all have characteristic zero; and every prime order field $\mathbb{F}_p$ has characteristic p. An example of an infinite field with positive characteristic p is $\mathbb{F}_p((x))$, where p is prime.

Theorem 3. For every field F, the characteristic of F is either zero or a prime p. Thus F is an extension of either $\mathbb{Q}$ or $\mathbb{F}_p$ respectively.

The corresponding subfield $\mathbb{Q}$ or $\mathbb{F}_p$ is the (unique) smallest subfield of F; it is called the *prime subfield* of F.

Proof. Define $\phi : \mathbb{Z} \to F$ by $\phi(j) = j1$. By Proposition 1, ϕ is a ring homomorphism. Consider its kernel

$$J = \ker \phi = \{j \in \mathbb{Z} \,:\, j1 = 0\}$$

which is an ideal of $\mathbb{Z}$. Since every ideal of $\mathbb{Z}$ is principal, we have $J = n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$ for some integer n. By the First Isomorphism Theorem for Rings, we have a subring of F given by

$$(*) \qquad \mathbb{Z}/n\mathbb{Z} = \mathbb{Z}/J \cong \phi(\mathbb{Z}) = \{k1 : k \in \mathbb{Z}\} \subseteq F.$$

We have two cases:

Case I: $J = \{0\}$, $n = 0$. By definition, $\operatorname{char} F = 0$. In this case ϕ is one-to-one; so by (*), F has a subring $\phi(\mathbb{Z}) \cong \mathbb{Z}$. Since F also contains multiplicative inverses of these elements, F contains a copy of $\mathbb{Q}$.

Case II: $J = n\mathbb{Z}$ where $n \neq 0$; and we may assume $n > 0$ (otherwise replace n by $-n$ without changing the ideal $n\mathbb{Z} \subseteq \mathbb{Z}$). By definition, $\operatorname{char} F = n$. We must have $n \geqslant 2$ (for if $n = 1$ then $1 = 0$ in F, which is not permitted in fields). If n is composite then $n = ab$ for some integers $a, b \in \{2, 3, \ldots, n-1\}$. But then $\phi(a)\phi(b) = \phi(n) = 0$. Since a field has no zero divisors, either $\phi(a) = a1 = 0$ or $\phi(b) = b1 = 0$. Either way, this violates the fact that n is the *smallest* positive integer satisfying $n1 = 0$. Since $n \geqslant 2$ is not composite, $n = p$ for some prime p. Now by (*), F has a subring $\phi(\mathbb{Z}) \cong \mathbb{Z}/p\mathbb{Z} \cong \mathbb{F}_p$. □

In the latter proof, note that the kernel of the homomorphism $\phi : \mathbb{Z} \to F$, $j \mapsto j1$ is generated by an integer n which is either 0 or a prime p. This is the explanation, promised earlier, for saying the characteristic of F is 0 (and *not* ∞), or p.

Corollary 4. Every finite field has order $q = p^r$ for some prime p and integer $r \geqslant 1$.

Proof. Let F be a finite field. Since $|F| < \infty$, F does not have $\mathbb{Q}$ as a subfield; so the prime subfield of F is $\mathbb{F}_p$ for some prime p. Also since $|F| < \infty$, it is clear that the degree $r = [F : \mathbb{F}_p]$ is itself finite. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ be a basis for F over $\mathbb{F}_p$. Then the elements of F are uniquely expressible as

$$a_1\alpha_1 + a_2\alpha_2 + \cdots + a_r\alpha_r \quad \text{where } a_1, a_2, \ldots, a_r \in \mathbb{F}_p.$$

Since there are p choices for each coefficient $a_i \in \mathbb{F}_p$, the number of distinct elements in F is $|F| = p^r$. □